



TATIANE CARVALHO ALVARENGA

**REDES BAYESIANAS NA PREDIÇÃO DE VALORES
ENERGÉTICOS DE ALIMENTOS PARA AVES**

LAVRAS – MG

2019

TATIANE CARVALHO ALVARENGA

**REDES BAYESIANAS NA PREDIÇÃO DE VALORES ENERGÉTICOS DE
ALIMENTOS PARA AVES**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Prof. Dr. Renato Ribeiro de Lima

Orientador

LAVRAS – MG

2019

**Ficha catalográfica elaborada pela Coordenadoria de Processos Técnicos
da Biblioteca Universitária da UFLA**

Alvarenga, Tatiane Carvalho

Redes bayesianas na predição de valores energéticos de alimentos para aves / Tatiane Carvalho Alvarenga. – Lavras : UFLA, 2019.

101 p. : il.

Tese (doutorado)–Universidade Federal de Lavras, 2019.

Orientador: Prof. Dr. Renato Ribeiro de Lima.

Bibliografia.

1. Algoritmo híbrido MMHC. 2. Distribuição empírica. 3. Energia Metabolizável. 4. Equações de predição. 5. Nutrição de aves. I. Universidade Federal de Lavras. II. Título.

TATIANE CARVALHO ALVARENGA

**REDES BAYESIANAS NA PREDIÇÃO DE VALORES ENERGÉTICOS DE
ALIMENTOS PARA AVES**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 20 de Fevereiro de 2019.

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho	DES/ UFLA
Prof. Dr. Joel Augusto Muniz	DES/ UFLA
Prof. Dr. Paulo Borges Rodrigues	DZO/ UFLA
Profa. Dra. Flávia Cristina Martins Queiroz Mariano	UNIFESP

Prof. Dr. Renato Ribeiro de Lima
Orientador

**LAVRAS – MG
2019**

“O temor do Senhor é o princípio da sabedoria, e a ciência do Santo, a prudência.”

Provérbios 9:10.

À Deus... Não a nós, Senhor, nenhuma glória para nós, mas sim ao Teu nome, por Teu amor e por Tua fidelidade! Salmos 115:1.

AGRADECIMENTOS

TATIANE CARVALHO ALVARENGA, filha de Vander Mendes de Alvarenga e Sônia Maria Carvalho Alvarenga, nasceu na cidade de Perdões (MG), em 26 de janeiro de 1991. Em 2009, ingressou na Universidade Federal de Lavras, na qual se graduou em Matemática em 07 de junho de 2013. Ingressou no Programa de Pós-Graduação do Departamento de Estatística na Universidade Federal de Lavras em 2013 obtendo o título de Mestre em Estatística e Experimentação Agropecuária, em 13 de fevereiro de 2015. Em março de 2015 iniciou o curso de Doutorado em Estatística e Experimentação Agropecuária, também na Universidade Federal de Lavras. Em 20 de fevereiro de 2019, submeteu-se à defesa de tese para obtenção do título de Doutor, sendo aprovada.

Primeiramente agradeço a Deus pela oportunidade, saúde e sabedoria no decorrer dessa longa jornada de estudos. Em momentos de tantas incertezas, tive fé que pela graça do amor de Deus seria bem sucedida ao pesquisar sobre o tema da tese, e que por meio dessa conquista iria glorificar o Nome de Deus. A Ele toda honra e glória, OBRIGADA SENHOR, se cheguei até aqui foi porque o Senhor me sustentou. Aos meus pais, Vander Mendes de Alvarenga e Sônia Maria Carvalho Alvarenga pelo apoio e amor incondicional. E às minhas irmãs, Taís Aparecida Alvarenga e Talita Aparecida Carvalho Alvarenga pela amizade e companheirismo em todos os momentos.

À Universidade Federal de Lavras, especialmente ao Departamento de Estatística, pela oportunidade. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, pela concessão da bolsa de doutorado. Ao meu orientador, Prof. Dr. Renato Ribeiro de Lima pelos valiosos ensinamentos, pela paciência, compreensão e confiança deste a orientação que se iniciou na graduação.

Aos professores colaboradores, Flávia Cristina Martins Queiroz Mariano (UNIFESP), Paulo Borges Rodrigues (DZO/UFLA) e Renata Ribeiro Alvarenga (DZO/UFLA) por toda a ajuda, seja com os dados experimentais e nas dúvidas ao longo da elaboração da tese. Aos professores Daniel Furtado Ferreira (DES/UFLA), Renata Ribeiro Alvarenga (DZO/UFLA) e Thelma Sáfadi (DES/UFLA) pela participação na banca de qualificação. Aos professores participantes da banca de defesa.

Aos funcionários pelos ótimos serviços prestados e aos professores do Departamento de Ciências Exatas e Departamento de Estatística da Universidade Federal de Lavras, pelo conhecimento e amizade no decorrer da minha trajetória acadêmica. Em especial ao Prof. Dr. Júlio Sílvio de Sousa Bueno Filho (DES/UFLA), pela orientação e ensinamentos durante o Estágio Docência e disposição em sempre colaborar.

A todos os colegas e amigos da pós-graduação em Estatística e Experimentação Agropecuária da Universidade Federal de Lavras, pelo coleguismo em todos os momentos. Aos colegas e amigos da pós-graduação em Zootecnia da Universidade Federal de Lavras, em especial, Sérgio Domingos Simão, por ser um amigo de todas as horas e não medir esforços para inúmeras ajudas no decorrer do doutorado. Em especial também agradeço ao Luiz Carlos Brandão, colega do Departamento de Engenharia da Universidade Federal de Lavras, pela parceria na elaboração das calculadoras desenvolvidas com os resultados desta tese. Muito obrigada a todos que ajudaram direta ou indiretamente para a realização deste trabalho!

RESUMO

Rações balanceadas para a nutrição de aves dependem do conhecimento da composição química dos alimentos, em especial dos valores de energia metabolizável aparente corrigida pelo balanço de nitrogênio (EMAn). Os valores de EMAn podem ser obtidos em ensaios biológicos, em que a execução é demorada e de custo elevado, assim como pelas tabelas de composição de alimentos. Outro meio de se obter os valores de EMAn são as equações de predição estabelecidas em função da composição química dos alimentos, normalmente de fácil e rápida obtenção. Na literatura existem trabalhos que obtiveram as equações de predição por meio de regressão múltipla, meta-análise e redes neurais. Com o objetivo de encontrar resultados mais acurados, emprega-se as redes bayesianas para prever a EMAn em função da composição química dos alimentos. As redes bayesianas são modelos gráficos (*graphical models*), os quais consistem na representação gráfica (grafo) e probabilística (distribuições de probabilidade condicionais e conjunta) das variáveis. Redes bayesianas foram propostas por Judea Pearl, então conhecido por defender o conhecimento probabilístico no campo da inteligência artificial. Para uma ampla compreensão sobre esta área de pesquisa, utilizou-se a base *Web of Science da Thomson Reuters* para identificar os padrões e tendências das publicações científicas sobre as redes bayesianas, permitindo assim, verificar que a maioria das publicações estão relacionadas à área de Ciências da Computação. Nas áreas aplicadas, principalmente agropecuária, ainda se têm pouquíssimas publicações, no entanto, redes bayesianas é uma linha de pesquisa inédita na nutrição de aves e que pode ser estudada por pesquisadores que tem o interesse na predição dos valores de energia metabolizável. Equações foram propostas por meio das redes bayesianas sendo obtidas pelo algoritmo *Max-Min Hill Climbing* (MMHC) e as mesmas podem ser utilizadas pela indústria de frangos de corte na elaboração de rações, pois apresentaram acurácia na predição de EMAn. Além do mais, as mesmas foram validadas com dados provenientes de ensaios metabólicos e apresentaram precisão e acurácia na predição dos valores energéticos. Essas equações estão disponíveis em uma calculadora que pode ser instalada em celulares, tablets e computadores. Nesta tese também foi proposta uma nova abordagem metodológica na qual considerou incertezas na obtenção de equações a partir de resultados de redes bayesianas híbridas. As estimativas provenientes de médias e modas dos coeficientes das composições químicas dos alimentos advindos de distribuições empíricas construídas com 10000 redes bayesianas híbridas tiveram melhores desempenhos. As equações propostas mostraram resultados precisos à medida que foram avaliadas com os dados de ensaios metabólicos. Em síntese, contribuiu-se tanto em termos metodológicos quanto em termos práticos e produção de produtos tecnológicos inovadores na experimentação agropecuária, mais especificamente na Nutrição de aves.

Palavras-chave: Algoritmo híbrido *mmhc*. Distribuição empírica. Energia Metabolizável. Equações de predição. Nutrição de aves.

ABSTRACT

Balanced diets for poultry nutrition depend on the knowledge of the chemical composition of the feedstuffs, especially the values of apparent metabolizable energy corrected by the nitrogen balance (EMAn). The values of EMAn can be obtained in biological assays, in which the execution is time-consuming and expensive, as well as by feedstuff composition tables. Another way of obtaining the values of EMAn are the prediction equations established according to the chemical composition of the feedstuffs, usually of easy and fast obtaining. In the literature there are studies that obtained the prediction equations through multiple regression, meta-analysis and neural networks. In order to find more accurate results, the Bayesian networks are used to predict EMAn according to the chemical composition of the feedstuffs. Bayesian networks are graphical models (graphical models), which consist of graphical (graph) and probabilistic representation (conditional and joint probability distributions) of the variables. Bayesian networks were proposed by Judea Pearl, then known for defending probabilistic knowledge in the field of artificial intelligence. For a broad understanding of this area of research, Thompson Reuters' Web of Science database was used to identify the patterns and trends of scientific publications on Bayesian networks, thus making it possible to check that most publications are related to the area of Computer Science. In the applied areas, mainly agriculture and livestock, there are still very few publications, however, Bayesian networks is an unprecedented research line in poultry nutrition and can be studied by researchers who are interested in predicting the values of metabolizable energy. Equations have been proposed through of the Bayesian networks, their being obtained by the Max-Min Hill Climbing algorithm (MMHC) and they can be used by the broiler industry in the making of diets, since they presented accuracy in the prediction of EMAn. Moreover, they were validated with data from metabolic assays and showed both precision and accuracy in the prediction of energy values. These equations are available in a calculator that can be installed on phones, tablets and computers. In this thesis a new methodological approach was also proposed in which it considered uncertainties in obtaining equations from the results of hybrid Bayesian networks. The estimates from means and mode of the coefficients of the chemical composition of feedstuffs derived from empirical distributions constructed with 10000 hybrid Bayesian networks performed better. The proposed equations showed accurate results as they were evaluated with metabolic assay data. In short, it has contributed both in methodological terms and in practical terms and the production of innovative technological products in agricultural experimentation, more specifically in poultry nutrition.

Keywords: Empirical distribution. Hybrid algorithm mmhc. Metabolizable Energy. Nutrition of Monogastric. Prediction equations.

LISTA DE FIGURAS

Figura 2.1 – Exemplos ilustrativos das possibilidades de Grafo	13
Figura 2.2 – Exemplo ilustrativo de DAG (adaptado de Nagarajan; Scutari e Lébre, 2013)	14
Figura 2.3 – DAG, separação gráfica e independência probabilística	15
Figura 2.4 – Exemplo de DAG para ilustrar distribuição de probabilidade conjunta . . .	16
Figura 2.5 – Estrutura de rede com nós discretos (adaptado de Scutari e Denis, 2015) . .	17
Figura 2.6 – Exemplo de rede com nós contínuos (adaptado de Scutari e Denis, 2015) .	19

SUMÁRIO

1	INTRODUÇÃO	9
2	REFERENCIAL TEÓRICO	11
2.1	Métodos para obtenção da energia metabolizável	11
2.2	Redes bayesianas	12
2.2.1	Representação gráfica	13
2.2.2	Representação probabilística	16
2.2.3	Redes bayesianas: o caso discreto e contínuo	17
2.3	Redes bayesianas híbridas	19
2.4	Aprendizagem das redes bayesianas	21
2.5	Redes bayesianas e suas aplicações	23
2.5.1	Validação das equações de predição	25
2.6	Softwares para redes bayesianas	26
	REFERÊNCIAS	30
3	ARTIGOS	36
4	CONSIDERAÇÕES	98

PRIMEIRA PARTE

1 INTRODUÇÃO

A indústria ligada à produção de frangos de corte ocupa uma posição de destaque no cenário econômico nacional (MAPA, 2017) e internacional (UNITED STATES DEPARTMENT OF AGRICULTURE, 2017). No Brasil, os principais produtores são os Estados do Paraná, Santa Catarina e Rio Grande do Sul (EMBRAPA, 2018), sendo a produção avícola favorecida pelo baixo custo com investimento em instalações e alta produção de grãos de milho e soja. Fatores genéticos, nutricionais, de ambiência, sanidade e manejo constituem os elementos para a efetivação desse cenário. Contudo, há necessidade de formulação adequada das dietas dos frangos, com a finalidade de se obter o máximo desempenho, conforme as exigências nutricionais.

Conhecer a composição química e os valores de energia metabolizável aparente corrigida pelo balanço de nitrogênio (EMAn) dos alimentos que compõem as rações é essencial para uma formulação adequada. A EMAn pode ser analisada por meio de ensaios metabólicos, mas tem altos custos e requer tempo elevado. Uma das soluções para determinar a EMAn dos alimentos são as equações de predição em função da composição química dos alimentos.

Trabalhos como os de Rodrigues et al. (2001, 2002); Borges et al. (2003); Zhao et al. (2008), obtiveram essas equações por meio de regressão múltipla. Porém, Alvarenga et al. (2011) mostraram que os resultados não eram os mais adequados. Outros estudos com o objetivo de encontrar as equações de predição da EMAn foram realizados por Nascimento et al. (2009, 2011) e Mariano et al. (2012), por meio da meta-análise. Perai et al. (2010), Ahmadi et al. (2007, 2008) e Mariano et al. (2013) utilizaram redes neurais, sendo que em Mariano et al. (2013) o número de alimentos e dados observados foi bem maior.

Assim, almejando obter mais confiabilidade à formulação de rações para aves, o presente trabalho tem como objetivo geral, prever a EMAn de alimentos usados para frangos de corte por meio de redes bayesianas (RB). As RB foram propostas por Judea Pearl (PEARL, 1988), então conhecido por defender o conhecimento probabilístico no campo da inteligência artificial. As aplicações dessa metodologia tem se estendido nas diferentes áreas do conhecimento (POURRET et al., 2008), sendo indicada para lidar com problemas que tratam incertezas.

As RB são modelos gráficos (*graphical models*), os quais consistem na representação gráfica (grafo) e probabilística (distribuições de probabilidade condicionais e conjunta) das variáveis. Essa nova metodologia para o campo de nutrição animal pode ser utilizada na construção de modelo preditivo pela fatoração da distribuição de probabilidade conjunta, assumindo

um conjunto de independências condicionais (JENSEN,1996; LAURITZEN e SPIEGELHATER, 1988).

Os objetivos específicos desse trabalho são:

- Revisar as principais áreas em que se tem empregado as RB, dentre os artigos mais citados na base *Web of Science*, relacionando-os com a pesquisa da tese;
- Predizer a EMAn de alimentos utilizados na formulação de rações para frangos de corte, em função da composição química desses alimentos, sob diferentes procedimentos de RB, a saber:
 - avaliar configurações da RB, a partir de diferentes aleatorizações dos conjuntos de dados (aprendizagem e teste), por meio de diferentes aprendizagens dos algoritmos híbridos do pacote *bnlearn* (SCUTARI, 2010) do software R, considerando diferentes cenários;
 - combinar os resultados das equações de predição obtidas por RB na obtenção de uma equação para alguns alimentos utilizados na formulação de rações para frangos de corte.
- Avaliar e validar as diferentes formas de predição propostas, utilizando-se os dados obtidos por Alvarenga et al. (2011) por meio de ensaios metabólicos. Além disso, os resultados obtidos serão comparadas com os obtidos por Nascimento et al. (2009, 2011) e Mariano et al. (2012, 2013).
- Desenvolver um produto tecnológico inovador para área de nutrição de frango, ou seja, com base nos resultados obtidos neste trabalho, criar uma ferramenta de cálculo para a predição da EMAn e disponibilizá-la em um aplicativo para ser usado em tablets, celulares e computadores.

2 REFERENCIAL TEÓRICO

2.1 Métodos para obtenção da energia metabolizável

A energia metabolizável (EM) é indispensável para as funções vitais das células e desempenha um papel importante na regulação do consumo, permitindo que o animal se desenvolva bem, ganhe peso e renda lucros. No entanto, de acordo com Nascimento et al. (2009), é comum fazer uma correção no valor da EM, ou seja, corrigi-la para o balanço de nitrogênio (EMAn). Essa correção é aplicada aos valores de EM pelo fato de que o nitrogênio retido no corpo, ao ser catabolizado, é excretado na forma de compostos contendo energia, tal como o ácido úrico (SIBBALD, 1982).

Uma maneira de obter os valores de energia metabolizável aparente corrigida pelo balanço de nitrogênio (EMAn), utilizada na formulação de rações balanceadas, são as tabelas de composição dos alimentos e exigências nutricionais. De acordo com Bertechini (2012), exemplos dessas tabelas são as dos Estados Unidos (SCOTT et al., 1982; NRC, 1994; LEESON e SUMMERS, 2001) e Europa (ARC, 1975; I.N.R.A., 1984; A.E.C., 1996). No Brasil, a primeira tabela de exigências nutricionais para aves disponível na literatura foi em 1983, por meio de pesquisas realizadas na Universidade Federal de Viçosa (UFV) (ROSTAGNO et al, 2011). No entanto, de acordo com Nagata et al. (2004) e Nascimento et al. (2011), as tabelas sugerem heterogeneidade de valores por conta da diversidade de variedades, processamento e armazenamento do grão, condições de plantio, regiões geográficas, entre outros. Em outras palavras, os valores encontrados nas tabelas não podem ser considerados como o método mais seguro a ser incorporado pela indústria (NAGATA et al., 2004; NASCIMENTO et al., 2011).

A determinação da energia metabolizável pode também ser obtida por meio de ensaios metabólicos (ZHAO et al., 2008; WAN et al., 2009), nos quais os alimentos e as excretas dos animais são analisados por um calorímetro. Esse método, além de demandar tempo e alto custo, não é muito viável, pelo fato de calorímetros nem sempre se encontrarem à disposição em laboratórios, indústrias e estações de pesquisa do país. Assim, métodos indiretos são utilizados para se obter a energia metabolizável, dos quais as equações de predição, devido a sua forma fácil, rápida e menos onerosa (ALBINO et al., 1992; NASCIMENTO et al., 2011) são comumente utilizadas.

Trabalhos como de Rodrigues et al. (2001, 2002), de Borges et al. (2003) e de Zhao et al. (2008) obtiveram equações de predição por meio da regressão linear múltipla sem nenhuma

ponderação. Porém, Alvarenga et al. (2011) mostraram que os resultados não eram os mais adequados. Tendo em vista o alcance de resultados mais coerentes, Nascimento et al. (2009, 2011) e Mariano et al. (2012) realizaram estudos para obtenção de equações de predição para a energia metabolizável corrigida pelo balanço de nitrogênio (EMAn) por meio da meta-análise. Nos trabalhos de Perai et al. (2010) e Ahmadi et al. (2007, 2008) foram utilizadas redes neurais para a predição da EMAn. Porém, nestes foram utilizados poucos dados observados, de no máximo dois tipos de alimentos e poucas variáveis de entrada para a predição da energia metabolizável. Mariano et al. (2013) utilizaram redes neurais para a predição da EMAn, porém, trabalhou-se com um maior número de alimentos e dados observados para a obtenção das equações de predição.

Brum et al. (2000) enfatizam o contínuo estudo sobre a nutrição de aves, com o objetivo da atualização dos dados e melhoraria nos valores de energia metabolizável dos alimentos para formulações de rações, sendo que a precisão da mesma pode influenciar de maneira direta o desenvolvimento das aves (DALE e FULLER, 1982). Uma metodologia que pode vir a ser utilizada na obtenção das equações de predição são as redes bayesianas, que consistem em modelos de probabilidade conjunta de variáveis aleatórias. Na Zootecnia, essa metodologia tem sido utilizada no melhoramento animal, mais especificamente em trabalhos que tratam de seleção genômica (MOROTA et al., 2013; TREPODE et al., 2013; GIANOLA et al., 2011; FRIEDMAN et al., 2004; FELIPE et al., 2015). No entanto, as mesmas ainda não foram utilizadas na predição do valor energético dos alimentos.

2.2 Redes bayesianas

As redes bayesianas (RB) foram propostas na década de 80 por Judea Pearl (PEARL, 1988), então conhecido por defender o conhecimento probabilístico no campo da inteligência artificial. As aplicações dessa metodologia tem se estendido em diferentes áreas do conhecimento, tais como: Medicina, Genética, Economia, robótica, análise de crime, Demografia, conservação de espécies, mapeamento do potencial mineral, Educação, indústria, comportamento humano, entre outros (POURRET et al., 2008).

Pearl (1988) e Neapolitan (2003) definem as RB, abordam suas propriedades e designam-na como um campo de estudo. Em síntese, RB são modelos gráficos (*graphical models*), os quais consistem na representação gráfica (grafo) e probabilística (distribuições de probabilidade con-

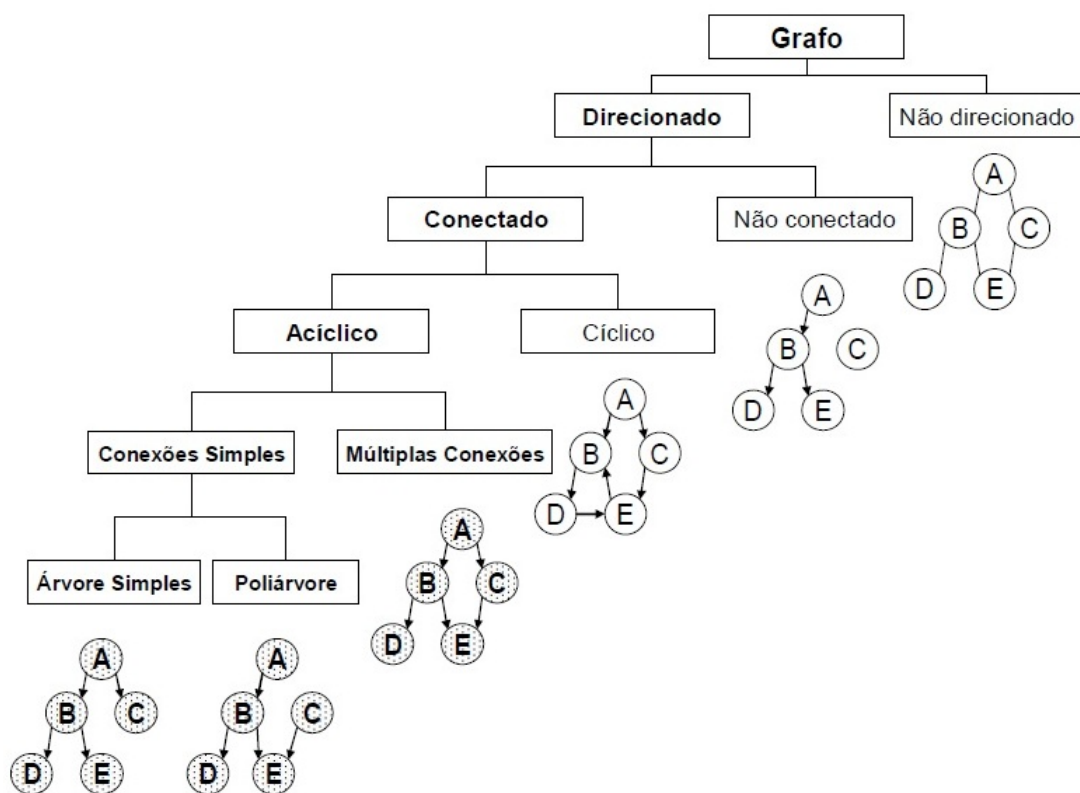
dicionais e conjunta) das variáveis. A seguir, serão apresentadas as definições e propriedades envolvidas nesta área de estudo.

2.2.1 Representação gráfica

Uma RB constitui uma representação gráfica de variáveis e suas relações causais, nomeadas, respectivamente, por nós ou vértices e arestas, arcos ou setas (PEARL, 1988; SPIRITES et al., 2000; KOLLER e FRIEDMAN, 2009; SCUTARI e DENIS, 2015). Essa representação gráfica, denominada de grafo, é estudada na Teoria de Grafos. Mais detalhes dessa teoria podem ser encontrados em Feofiloff et al. (2011).

A representação das relações entre as variáveis (nós), são dadas no grafo por meio de setas, ou seja, $A \rightarrow B$, a qual se lê A causa B ou A tem relação com B ou B depende de A. Nesta, tem-se que o nó A e o nó B estão, respectivamente, na cauda e cabeça da aresta. As estruturas básicas existentes dentro da Teoria dos Grafos são ilustradas na Figura 2.1.

Figura 2.1 – Exemplos ilustrativos das possibilidades de Grafo



Fonte: Ara-Souza, 2010.

Na Figura 2.1, apenas as estruturas gráficas hachuradas correspondem a uma RB e são denotados por DAG (*directed acyclic graph*). Os DAGs apresentam direção nas arestas, todos os

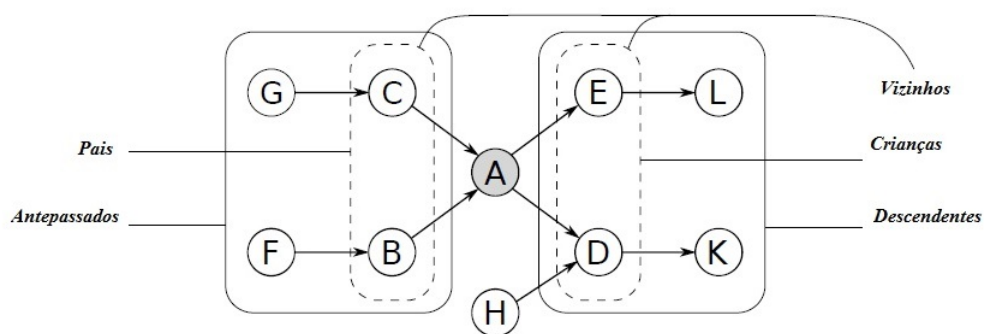
nós são conectados na estrutura gráfica e as direções das arestas não apresentam ciclos, ou seja, após a aresta seguir a direção, ela não retorna ao mesmo nó, classificando-o, respectivamente, por grafo direcionado, conectado e acíclico.

Conforme a Figura 2.1, existe também o grafo não direcionado, ou seja, arestas não indicam causa e efeito entre os nós; o grafo direcionado, isto é, arestas indicam relações de causa e efeito entre os nós; o grafo não conectado, quer dizer, existe pelo menos um nó que não se relaciona com os demais; grafo cíclico, ou melhor, seguindo-se a direção das arestas, há retorno a um mesmo nó. Em síntese, ilustrações de grafos não se enquadram no estudo de RB, apenas o grafo direcionado, conectado e acíclico.

De acordo com o grafo da Figura 2.2, e tomando o nó A como referência, pode-se observar que seus vizinhos (C,B,D,E) são representados pela união de seus pais (C,B) e filhos ou crianças (E,D), chamados também de nós adjacentes (nós pais e filhos). Seus pais também são chamados de antecessores, pois estão ordenados anterior ao A. Os filhos de A são chamados de descendentes, devido a ordem posterior. A ordem das variáveis no grafo chama-se topologia e assim, de acordo com Nagarajan, Scutari e Lébre (2013), um exemplo da ordem topológica da estrutura gráfica apresentada na Figura 2.2 pode ser dada por:

(F,G,H), (C,B), (A), (D,E), (L,K).

Figura 2.2 – Exemplo ilustrativo de DAG (adaptado de Nagarajan; Scutari e Lébre, 2013)



Muitos dos algoritmos utilizados para encontrar a estrutura gráfica das variáveis, requerem ordenação entre os nós. A justificativa da mesma, se deve ao fato de variáveis causais, estarem precedidas de outras variáveis (nós). Alguns desses algoritmos utilizam a cobertura de Markov (*Markov Blanket*) para modelar o fenômeno em estudo. Denota-se cobertura de Markov, o conjunto de nós que separa o nó alvo dos demais nós da estrutura, ou seja, essa cobertura inclui pais, filhos e nós que compartilham os mesmos filhos do nó alvo. A cobertura de Markov

do nó A (Figura 2.2) são os nós: C, B, H, E e D. Estudos cujo objetivo é a predição, apenas as variáveis encontradas nessa cobertura são relevantes ao modelo (TSAMARDINOS et al., 2003).

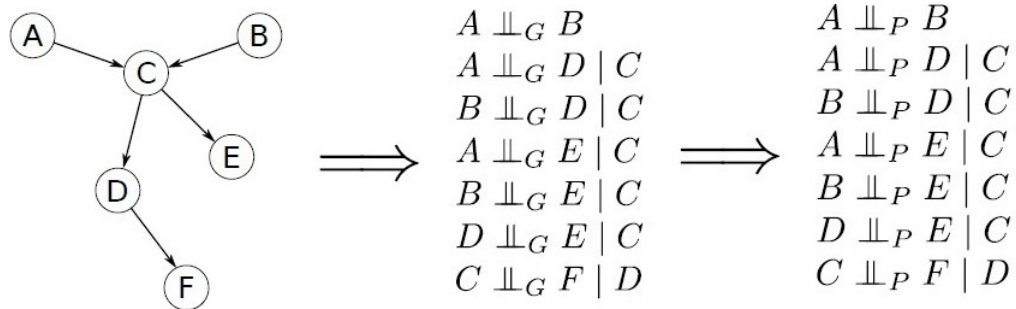
Outras configurações de topologia, podem ser obtidas seguindo o ordenamento dos nós. Robinson (1977 citado por KOSKI e NOBLE, 2009), define o número de possíveis DAGs ($N(d)$) pela Equação 2.1:

$$N(d) = \sum_{i=1}^d (-1)^{i+1} \binom{d}{i} 2^{i(d-1)} N(d-i), \quad (2.1)$$

em que d é o número de nós. $N(d)$ cresce exponencialmente, ou seja, para $d=5$ tem-se $N(d)=29000$ DAGs e para $d=10$, $N(d)$ é aproximadamente $4,2 \times 10^{18}$ DAGs (KOSKI; NOBLE, 2009).

A DAG da Figura 2.3 é apresentada como um mapa de independência da distribuição de probabilidade da variável aleatória. Observa-se que separação gráfica (representada por \perp_G) implica em independência probabilística (representada por \perp_P).

Figura 2.3 – DAG, separação gráfica e independência probabilística



Assim, de acordo com a Figura 2.3, conclui-se que A é independente de B; A é independente de D, dado C; B é independente de D, dado C; A é independente de E, dado C; B é independente de E, dado C; D é independente de E, dado C; C é independente de F, dado D. Essa independência também é chamada de *d-separação*, leia-se: A é *d-separado* de D pelo nó C, ou seja, A e D são separados por C, ou C bloqueia o caminho entre A e D (PEARL, 1988; MARGARITIS, 2003).

Por meio da compreensão de *d-separação*, surge a definição de mapa de independência, mapa de dependência e mapa perfeito. O conjunto das relações de independência condicional que ligam qualquer triple A, B, C de subconjuntos de X , é chamado de estrutura de dependência da distribuição de probabilidade de variáveis aleatórias X , nomeando-a de M . Um grafo G é um mapa de dependência (*D-map*) de M se houver correspondência um-para-um entre as variáveis aleatórias de X e os nós de G , ou seja, $A \perp_P B \mid C \implies A \perp_G B \mid C$. De forma semelhante G

é um mapa de independência (*I-map*) de M se $A \perp\!\!\!\perp_G B \mid C \implies A \perp\!\!\!\perp_P B \mid C$. Caso satisfaça $A \perp\!\!\!\perp_P B \mid C \iff A \perp\!\!\!\perp_G B \mid C$ é chamado de mapa perfeito de M (KOSKI e NOBLE, 2009).

2.2.2 Representação probabilística

Uma RB é uma representação gráfica de uma distribuição de probabilidade conjunta (HECKERMAN, 1995; MARGARITIS, 2003), a qual descreve o domínio por completo. Ela pode ser definida pelo conjunto (ε, θ, X) , em que ε é a estrutura da DAG, θ é o conjunto de parâmetros específicos advindos de distribuições de probabilidades condicionais, envolvendo o conjunto X de variáveis aleatórias.

A fatoração da RB, conforme apresentado na Equação 2.2, consiste no produto das probabilidades condicionais, visto que um nó é condicionalmente independente de seus não-descendentes, dado seus pais (PEARL, 1988; TSAMARDINOS et al., 2003). Essa fatoração resulta na distribuição de probabilidade conjunta, sendo assim, possível inferir e obter respostas a questões de interesse.

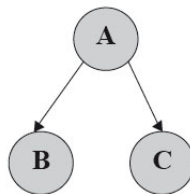
$$P(X_1, X_2, \dots, X_p) = \prod_{i=1}^n P(X_i \mid Pa_i) \quad (2.2)$$

para o caso de nós discretos e contínuos, em que Pa_i são os pais de X_i .

De acordo com a Figura 2.4, a distribuição de probabilidade conjunta obtida pela fatoração das distribuições de probabilidade condicionais é dada por (2.3). Para cada distribuição de probabilidade condicional tem-se o parâmetro, θ_i .

$$P(A, B, C) = P(B \mid A, \theta_1) P(C \mid A, \theta_2) P(A, \theta_3). \quad (2.3)$$

Figura 2.4 – Exemplo de DAG para ilustrar distribuição de probabilidade conjunta



De acordo com Scutari e Denis (2015), o número de possíveis arestas dadas pelos diferentes pares de nós de uma DAG com p nós é obtida por:

$$\frac{1}{2}p(p-1) \quad (2.4)$$

As relações de independências condicionais entre as variáveis são verificadas pelo critério de *d-separação*. Se o modelo de dependência é compatível com uma distribuição de probabilidade, então, pode-se representar de forma exata, por um conjunto de probabilidades condicionais. Assim, a distribuição conjunta é chamada de distribuição global e as distribuições condicionais denotadas por distribuições locais.

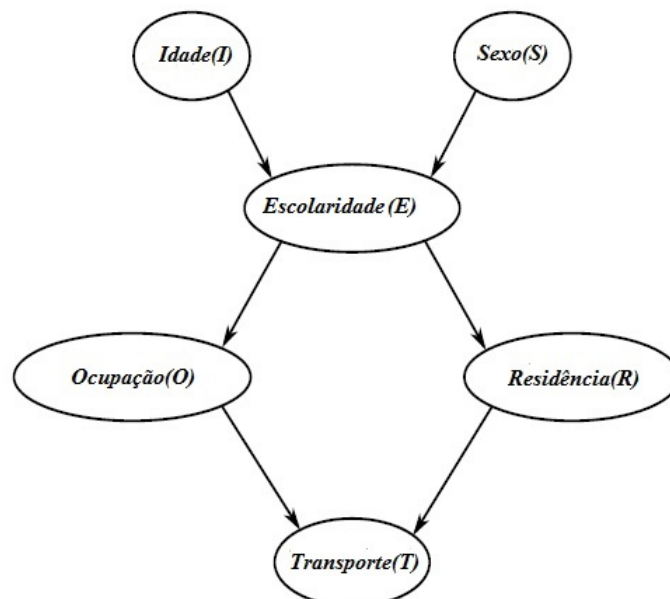
A distribuição de probabilidade conjunta é dada de acordo com a natureza das variáveis. Assim, para o caso discreto, contínuo e ambos, as redes bayesianas são nomeadas, respectivamente, por rede bayesiana multinomial, rede bayesiana gaussiana e rede bayesiana híbrida. Mais detalhes serão apresentados a seguir.

2.2.3 Redes bayesianas: o caso discreto e contínuo

Como exemplo ilustrativo da RB multinomial é apresentada a seguinte situação, adaptado de Scutari e Denis (2015).

Realizou-se uma pesquisa com a finalidade de obter informações de diferentes grupos sociais para fins de planejamento urbano e avaliação das políticas públicas sobre transporte. As variáveis de indicadores demográficos coletadas foram idade (I), sexo (S) e os indicadores socioeconômicos coletados foram escolaridade (E), ocupação (O), residência (R). Na Figura 2.5 ilustra-se a relação dessas variáveis.

Figura 2.5 – Estrutura de rede com nós discretos (adaptado de Scutari e Denis, 2015)



As probabilidades das distribuições locais podem ser calculadas considerando as distribuições condicionais, ou seja,

$$P(A | B) = \frac{P(A \cap B)}{P(B)}. \quad (2.5)$$

Assim, a probabilidade do meio de transporte ser o carro dado que reside na zonal rural (interior) pode ser calculada por

$$P(T = \text{carro} | R = \text{interior}) = \frac{P(T = \text{carro}, R = \text{interior})}{P(R = \text{interior})}. \quad (2.6)$$

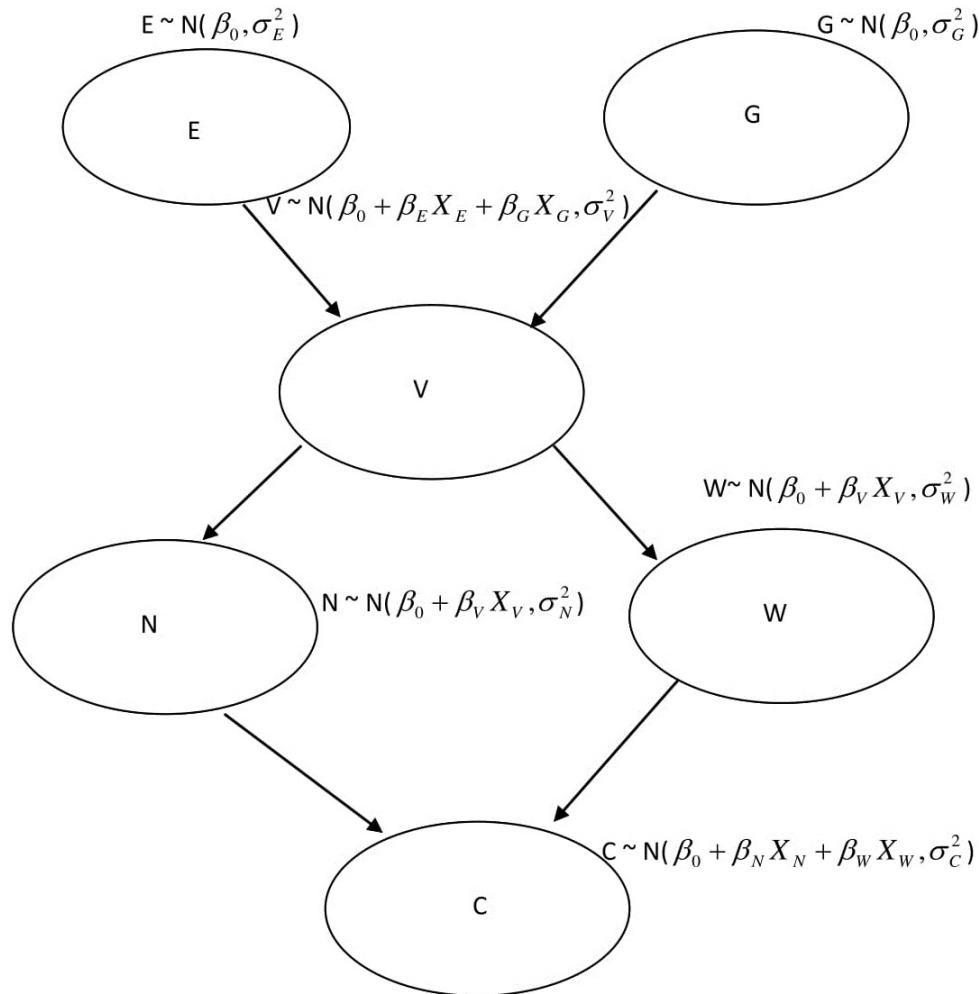
De maneira semelhante, podem ser obtidas as probabilidades $P(I)$, $P(S)$, $P(E | I, S)$, $P(O | E)$, $P(R | E)$ e $P(T | O, R)$, sendo que cada uma delas assume distribuição multinomial e é chamada de distribuição local. A distribuição conjunta é obtida pelo produto de cada probabilidade envolvida dado a independência condicional, a qual é chamada de distribuição global. Neste exemplo, essa é dada por $P(I, S, E, O, R, T) = P(I)P(S)P(E | I, S)P(O | E)P(R | E)P(T | O, R)$.

No entanto, se a natureza da variável é contínua, as distribuições condicionais assumem distribuições normais, caracterizando-se a rede bayesiana Gaussiana (GRB). Na Figura 2.6, tem-se um exemplo de um DAG *fictícia* de variáveis contínuas. Cada nó (G, E, V, N, W, C) segue uma distribuição normal.

Considerando o exemplo da Figura 2.6, observa-se que as distribuições de probabilidade marginais descrevem os nós sem pais. Verifica-se que a distribuição local de cada nó possui média expressa num modelo linear gaussiano, o qual inclui o intercepto e os nós pais (variáveis explanatórias), sem a interação deles. A variância do nó é fixa e não depende dos nós pais, ou seja, o efeito aditivo na média de cada distribuição local, não afeta a variância. A distribuição conjunta de todos os nós é uma normal multivariada.

Conforme mencionado, as variáveis da RB podem ser discretas ou contínuas, assumindo assim distribuição de probabilidade na mesma família, ou seja, distribuição multinomial ou normal multivariada. Quando há razões para se trabalhar com ambas as variáveis ou outras distribuições de probabilidade, aumenta-se a complexidade na aprendizagem e a inferência por parte da programação, sendo chamadas por redes bayesianas híbridas.

Figura 2.6 – Exemplo de rede com nós contínuos (adaptado de Scutari e Denis, 2015)



2.3 Redes bayesianas híbridas

Até o momento foram consideradas apenas as redes bayesianas com variáveis discretas ou contínuas. Sendo que todas as variáveis da RB seguem a mesma distribuição de probabilidade pertencentes à mesma família: multinomial ou normal multivariada. No entanto, existem casos em que as variáveis sob investigação são melhores explicadas por outras distribuições de probabilidade. Assim, surge a ideia de se trabalhar com ambas as variáveis discretas e contínuas em uma mesma rede bayesiana, além da possibilidade de utilizar qualquer tipo de distribuição na modelagem das variáveis.

No entanto, empregar outras distribuições além da multinomial e da normal multivariada, faz com que as RB se tornem difíceis e complicadas de se trabalhar, além de não existir um pacote específico no software R para lidar com elas (SCUTARI e DENIS, 2015). Observa-se que essa metodologia ainda requer pesquisas no campo metodológico estatístico e computaci-

onal para que sua aplicação se torne cada vez mais eficiente. Essa eficiência é esperada desde a obtenção da rede, estimação dos parâmetros, predição das variáveis e otimização do tempo gasto nas análises (*structure learning* e *parameter learning*).

No entanto, existe a possibilidade de se trabalhar com ambas as variáveis, discretas e contínuas, em uma mesma rede bayesiana, empregando a distribuição multinomial para as variáveis discretas e a distribuição normal para as variáveis contínuas. Contudo, existem algumas restrições. Como exemplo tem-se que apenas os nós discretos podem ser pais dos nós contínuos, o contrário não é permitido (KOSKI; NOBLE, 2009).

A rede bayesiana híbrida, assim chamada por contemplar tanto variáveis contínuas quanto discretas, se assemelha a um modelo de regressão linear múltipla, a qual denota-se por rede bayesiana gaussiana condicional linear (KOSKI; NOBLE, 2009). Para tornar claro o conceito apresentado, considere um conjunto de variáveis aleatórias, sendo esse conjunto particionado em variáveis discretas e contínuas, respectivamente X_D e X_C . Assim, a distribuição de probabilidade conjunta $P(X)$ pode ser fatorada da forma apresentada em (2.7).

$$P(X)=P(X_D, X_C) = \prod_{i \in D} P(X_i | Pa_D) \prod_{j \in C} P(X_j | Pa_C, Pa_D) \quad (2.7)$$

em que Pa_D e Pa_C são, respectivamente, os conjuntos de variáveis pais discretos e contínuos.

Para o $\prod_{j \in C} P(X_j | Pa_C, Pa_D)$, a qual contempla as variáveis pais discretas e contínuas, as distribuições locais são regressões lineares gaussianas (CLG) com parâmetros dependentes dos pais discretos, isso significa dizer que:

$$(X_j | Pa_C, Pa_D) \sim N(\mu_j, \sigma_{X_j|Pa_D}^2) \quad (2.8)$$

em que,

$$\mu_j = \beta_{0, X_j|Pa_D} + \beta_{i, X_j|Pa_D} X_{i, Pa_C}. \quad (2.9)$$

Assim, no contexto da predição da EMAn, μ_j refere-se ao valor da EMAn em cada categoria, alimento e animal; $\beta_{0, X_j|Pa_D}$ é média da variável X_j na configuração de seus pais discretos, ou seja, o intercepto da regressão linear multivariada; $\beta_{i, X_j|Pa_D}$ são os coeficientes da regressão, ou seja, a contribuição de cada pai contínuo em cada nível da variável discreta; X_{i, Pa_C} representam as variáveis de composição química dos alimentos, sendo elas neste trabalho, PB, MM, EE e FB. Para maiores detalhes sobre a CLG consultar Lauritzen e Wermuth (1989).

Neste trabalho são propostos modelos de redes bayesianas híbridas, que contemplem inicialmente as variáveis de composição química (PB, MM, EE e FB), além das categorias dos alimentos (energéticos e proteicos), ingredientes (milho, soja, seus subprodutos e outros) e animais (pinto, galo) como variáveis de entrada na rede. A energia metabolizável aparente corrigida pelo balanço de nitrogênio será considerada como variável de saída no modelo. Porém, para elucidar a estrutura de relações entre as variáveis de composição química e de valores energéticos, de cada alimento, serão realizadas implementações nas RB, tal como no processo de seleção das diferentes estruturas obtidas por meio dos algoritmos.

Observa-se que qualquer outra variável, além das variáveis de composição química, que sejam fornecidas nos trabalhos catalogados, pode ser considerada na análise, mesmo que essa variável seja qualitativa. Essa é uma vantagem do uso da rede bayesiana.

2.4 Aprendizagem das redes bayesianas

O aprendizado do modelo é conhecido no campo da RB como *learning*, nome emprestado da inteligência artificial. A princípio ocorre a aprendizagem da estrutura da DAG, ou seja, *structure learning* (SCUTARI, 2010). Em seguida tem-se *parameter learning*, ou seja, aprender sobre as distribuições locais implícitas da estrutura da DAG (SCUTARI et al., 2014). A aprendizagem pode ser sem supervisão, ou seja, apenas com as informações contidas nos dados, ou com aprendizagem supervisionada, na qual conta com um conhecimento *a priori* sobre o fenômeno.

Três algoritmos podem ser utilizados na aprendizagem da RB, eles são: *constraint-based*, *score-based* e *hybrid*. Todos esses algoritmos realizam a aprendizagem assumindo:

- haver uma correspondência de um-para-um entre os nós no DAG e entre as variáveis aleatórias;
- todas as relações entre as variáveis devem ser condicionalmente independentes. Por definição esse é o único tipo de relação que pode existir na RB;
- a distribuição global tem que ser estritamente positiva;
- as observações são tratadas como realizações independentes do conjunto de nós, mesmo que alguma dependência temporal ou espacial estiver presente. Isso deve ser especificado na definição da rede.

Os algoritmos *constraint-based* consistem em vários testes de independência condicional e marginal para inferir as direções das arestas entre os nós. Esse algoritmo constrói, a princípio, uma estrutura de variáveis não direcionadas, apenas relacionadas. Depois são projetadas as *v-estruturas* (isto é, arestas convergindo para o mesmo nó, como $X \rightarrow Y \leftarrow Z$) e por fim os arcos remanescentes são orientados de forma que não ocorra novas *v-estruturas* e ciclos. Os testes de independência condicional utilizados para RB discretas encontrados na literatura são: G^2 , *shrinkage* e qui-quadrado. Para as RB contínuas os testes utilizados são: teste t , teste Z e *shrinkage* (SCUTARI, 2010).

Já o algoritmo *score-based* seleciona entre todas as possíveis estruturas de RB, a DAG com melhor qualidade, ou seja, com maior pontuação (SCUTARI, 2010; SU et al., 2013). Essa seleção pode ser por meio do critério de informação Bayesiano (BIC), critério de informação de Akaike (AIC) e intermediário aos dois quanto à penalização, o $K2$ score. Não foi encontrado na literatura, a seleção da DAG por meio da medida L . A medida L seleciona modelos baseada no preditivismo (IBRAHIM e LAUD, 1994). Portanto, verifica-se a possibilidade de desenvolver pesquisa inédita no campo metodológico de RB, a partir da medida L na seleção da DAG.

O algoritmo *hybrid* é uma combinação dos dois algoritmos apresentados (SCUTARI e DENIS, 2015). Geralmente esse algoritmo utiliza o método *constraint-based* para encontrar o esqueleto da rede bayesiana e posteriormente o método *score-based* é utilizado para estabelecer as relações de causa e efeito na rede.

Conforme abordado, a princípio deve-se definir a estrutura gráfica. Nessa etapa, têm-se os algoritmos de *constraint-based* implementados no *bnlearn* (SCUTARI, 2010) do Software *R* (R CORE TEAM, 2018): *Grow-Shrink* (GS)(MARGARITIS e THRUN, 2000); *Incremental Association Markov Blanket* (IAMB) (TSAMARDINOS et al., 2003); *Fast Incremental Association* (Fast-IAMB) (YAMARAKALA e MARGARITIS, 2005); *Interleaved Incremental Association* (Inter-IAMB); *Max-Min Parents e Children* (MMPC); *Semi-Interleaved Hiton-PC* (SI-HITON-PC). Já os algoritmos de estruturas de aprendizagem *score-based* também implementados no *bnlearn* são: *Hill Climbing* (HC)(MARGARITIS, 2003); *Tabu Search* (Tabu). Já os algoritmos de aprendizagem *hybrid* são: *Max-Min Hill Climbing* (MMHC) (TSAMARDINOS et al., 2006) e *General 2-Phase Restricted Maximization* (RSMAX2).

A aprendizagem da RB pode ser dada por:

$$\underbrace{P(B | D)}_{\text{aprendizagem}} = \underbrace{P(G, \Theta | D)}_{\text{aprendizagem}} = \underbrace{P(G | D)}_{\text{DAG}} \underbrace{P(\Theta | G, D)}_{\text{parametros}}, \quad (2.10)$$

em que D é o conjunto de dados e G representa a estrutura gráfica. A RB é dada por $B = (G, \Theta)$ e Θ são os parâmetros da distribuição global. A partir de modelar D escreve-se $B = (G, \Theta)$. A aprendizagem da estrutura encontra-se em 2.11.

$$P(G | D) \propto P(G)P(D | G) = P(G) \int P(D | G, \Theta)P(\Theta | G)d\Theta. \quad (2.11)$$

Uma vez que a aprendizagem da DAG ocorre a partir dos dados, a estimação dos parâmetros pode ser obtida pela fatoração da distribuição global, nas distribuições locais.

2.5 Redes bayesianas e suas aplicações

Uma concorrente das redes bayesianas são as redes neurais. De acordo com Bishop (2006), as redes neurais artificiais (RNA) são indicadas para estudos cujo objetivo é predição. Felipe et al. (2015) mencionam a utilização dessa metodologia nas ciências animais, cujas aplicações estão na predição da produção de ovos (AHMAD e GOLIAN, 2008; SAVEIDAGO et al., 2011; FARIDI e GOLIAN, 2011); eclodibilidade (MEHRI, 2013); peso e número de ovos (SEMSARIAN et al., 2013); curvas de crescimento (AHMAD, 2009) e nutrição de aves poedeiras (MEHRI, 2012).

O modelo preditivo via rede bayesiana nas ciências animais estão concentrados no melhoramento animal, mais especificamente em trabalhos que abordam seleção genômica (MOROTA et al., 2013; TREPODE et al., 2013; GIANOLA et al., 2011; FRIEDMAN et al., 2004). Porém, RNA e RB têm sido utilizadas para predição, devido à superioridade em capturar (e expressar) relações mais complexas entre as covariáveis e a variável resposta (GIANOLA et al.; 2011 citado por Felipe et al., 2015).

As aplicações de redes bayesianas se estendem nas diferentes áreas do conhecimento (POURRET et al., 2008). Porém, essa metodologia ainda requer pesquisas no campo metodológico estatístico e computacional para que sua aplicação se torne cada vez mais eficiente. Essa eficiência é esperada na obtenção da rede, na estimação dos parâmetros, na predição das variáveis e na otimização do tempo gasto nas análises (*structure learning* e *parameter learning*).

Carvalho e Chiann (2013), em busca de uma maneira distinta de abordagem ao tema, utilizaram redes bayesianas para modelar a estrutura de interdependência entre mercados financeiros. A partir da metodologia empregada com respaldo a teoria de inferência bayesiana, os resultados encontrados indicaram evidências de contágio em períodos de crise financeira. Fo-

ram consideradas as realidades que precediam períodos de crise como situações a priori para os eventos (verossimilhanças) e a nova realidade (a posteriori), servindo assim como priori, para o período subsequente.

Simões et al. (2015) utilizaram as redes bayesianas no apoio ao diagnóstico de câncer de mama, juntamente com a meta-análise, na qual foram considerados artigos e relatórios publicados entre janeiro de 1990 e março de 2013. Por meio das redes bayesianas, o diagnóstico de lesões malignas aumentou significativamente a probabilidade pré-teste para um verdadeiro positivo e diminuiu a probabilidade de um falso negativo. Por fim, chegou-se a conclusão que a metodologia empregada forneceu resultados acurados e não invasivos no apoio ao diagnóstico de câncer de mama.

Vieira et al. (2017) analisaram o risco ocupacional por meio das redes bayesianas. Para tal análise, realizaram uma revisão sistemática de maneira semelhante ao protocolo *Statement for Reporting Systematic Reviews* (PRISMA), ou seja, uma meta-análise. Por meio dessa revisão, concluíram que a aplicação das redes bayesianas na análise de risco ocupacional é essencial no setor industrial e de transporte. Em conclusão ao trabalho, Vieira et al. (2017) relataram a importância das redes bayesianas nas análises de riscos ocupacionais, justamente pelo fato do risco ser uma variável aleatória e, este raciocínio se aplica na modelagem das redes bayesianas, ou seja, simula processos estocásticos.

Felipe et al. (2015) compararam diferentes metodologias para prever a produção total de ovos em codornas. As variáveis analisadas foram: peso, produção parcial de ovos e medidas de qualidade. As metodologias empregadas contemplaram os modelos regressão linear múltipla e rede neural artificial (RNA), além da rede bayesiana para a seleção das variáveis. O modelo com a combinação de rede bayesiana (para a seleção das variáveis) e RNA resultou em melhor desempenho para prever a produção total de ovos.

Töpner et al. (2017) utilizaram as redes bayesianas em experimento de milho para análise das relações entre as características a níveis genômico e residual. Foram utilizados dois algoritmos de aprendizagem, sendo eles *Tabu Search (score-based)* e *Grow-Shrink (constraint-based)*. O algoritmo *Tabu Search* foi combinado com dois critérios: critério de informação bayesiano e pontuação equivalente a gaussiana bayesiana. Já o algoritmo *Grow-Shrink* foi combinado com quatro diferentes testes para independência marginal e condicional: coeficiente de correlação com o teste t, coeficiente de correlação com o teste de permutação de Monte Carlo, informação mútua e teste qui-quadrado, e informação mútua e teste de permutação de Monte

Carlo. As redes obtidas foram classificadas em termos de capacidade de ajuste e habilidade preditiva por meio das equações estruturais.

Por meio desta pesquisa espera-se que a aplicação de RB forneça resultados mais precisos e acurados na predição de EMAn. Além disso, as redes bayesianas compreendem uma nova metodologia para o campo de nutrição animal e pode ser utilizada na construção de modelo preditivo pela fatoração da distribuição de probabilidade conjunta, assumindo um conjunto de independências condicionais (JENSEN,1996). Um passo muito importante após o emprego dessa metodologia inovadora na nutrição de aves é a validação das equações obtidas, sendo as mesmas aplicadas a dados reais, ou seja, o funcionamento da mesma em dados provenientes de ensaios metabólicos.

2.5.1 Validação das equações de predição

Alvarenga et al. (2011) e Mariano et al. (2012) relataram a importância da validação das equações de predição. Os resultados preditos por alguma metodologia, devem ser validados pelos resultados de ensaios metabólicos e de desempenho. A validação tem como objetivo indicar se o método utilizado para prever a energia metabolizável é ou não preciso com relação aos empregados até o momento.

Um modelo de regressão linear simples ($Y = \beta_0 + \beta_1 X_1 + \varepsilon$) pode ser utilizado para validação das equações de predição, sendo os resultados dos ensaios metabólicos e valores preditos, as variáveis dependente e independente, respectivamente (ALVARENGA et al., 2011). Outra maneira de validar as equações de predição pode ser por meio do coeficiente de correlação de concordância entre os valores observados e os valores preditos de EMAn (MARIANO et al., 2012).

As estatísticas utilizadas no processo de validação das equações obtidas nesta pesquisa, foram as mesmas empregadas em Mariano et al. (2013). Na pesquisa de Mariano et al. (2013) utilizou as redes neurais para prever a EMAn dos alimentos utilizados na elaboração de rações para frangos de corte. As estatísticas mencionadas se referem ao coeficiente de determinação (R^2), o desvio médio absoluto (MAD), o erro médio percentual absoluto (MAPE), o viés (bias) e o erro quadrático médio (m.s.e). Modelos mais acurados, são esperados valor de (R^2) próximo de 1, e bias e os erros próximos de zero. Assim, tem-se:

$$R^2 = 1 - \frac{\sum(d_i - \hat{y}_i)^2}{\sum(d_i - \bar{d}_i)^2}; \quad (2.12)$$

$$\text{MAD} = \frac{\sum |d_i - \hat{y}_i|}{n}; \quad (2.13)$$

$$\text{MAPE} = \frac{\sum \frac{|d_i - \hat{y}_i|}{d_i}}{n} * 100, d_i \neq 0; \quad (2.14)$$

$$\text{bias} = \frac{\sum (d_i - \hat{y}_i)}{n}; \quad (2.15)$$

$$\text{m.s.e} = \frac{\sum e^2}{n} = \frac{\sum (d_i - \hat{y}_i)^2}{n} \quad (2.16)$$

em que d_i é o i -ésimo valor observado, \hat{y}_i é o i -ésimo valor predito, n é o número de valores observados e \bar{d}_i é a média dos n valores observados.

2.6 Softwares para redes bayesianas

Alguns recursos computacionais para RB (SCUTARI e DENIS, 2015) são:

- Software R, pacotes: *bnlearn*, *catnet*, *deal*, *pcalg*, *gRbase*, *gRain* e *rbmn*;
- BayesiaLab (<http://library.bayesia.com/display/BlabC/Evaluation+License>);
- HUGIN (<http://hugin.sourceforge.net/download/>);
- GeNIe (<https://download.bayesfusion.com/files.html?category=Business>);
- Netica (<https://www.norsys.com/download.html>).

As redes bayesianas foram implementadas no software livre R (R CORE TEAM, 2018), utilizando-se o pacote *bnlearn* (SCUTARI, 2010) e os algoritmos de aprendizagem *hybrid: Max-Min Hill Climbing* (MMHC) (TSAMARDINOS et al., 2006) e *General 2-Phase Restricted Maximization* (RSMAX2). O pacote *bnlearn* possibilita a aprendizagem da estrutura gráfica, bem como a estimação dos parâmetros, trabalhando com qualquer tipo de variável (quantitativa e qualitativa). Outros pacotes implementados no software R que trabalham com as redes bayesianas encontram-se na Tabela 2.1. Observa-se que o *bnlearn* é bem superior aos demais em se tratando de suas funcionalidades.

Os algoritmos implementados no pacote *bnlearn* são:

- *Constraint-based: Grow-Shrink* (GS); *Incremental Association Markov Blanket* (IAMB); *Fast Incremental Association* (Fast-IAMB); *Interleaved Incremental Association* (Inter-IAMB); *Semi-Interleaved Hiton-PC* (SI-HITON-PC) e *Max-Min Parents e Children* (MMPC);

Tabela 2.1 – Pacotes do R e suas funcionalidades

	<i>bnlearn</i>	<i>catnet</i>	<i>deal</i>	<i>pcalg</i>	<i>gRbase</i>	<i>gRain</i>	<i>rbmn</i>
Dados discretos	X	X	X	X	X	X	
Dados contínuos	X		X	X	X		X
Dados mistos			X				
aprendizagem <i>constraint-based</i>	X			X			
aprendizagem <i>score-based</i>	X	X	X				
aprendizagem <i>hybrid</i>	X						
Manipulação da estrutura	X	X			X		
Estimação dos parâmetros	X	X	X	X			X
predição	X	X				X	X
Inferência aproximada	X					X	

Fonte: Scutari e Denis (2015).

- *Score-based*: Hill Climbing (HC) e Tabu Search (Tabu);
- *Hybrid*: Max-Min Hill Climbing (MMHC) e General 2-Phase Restricted Maximization (RSMAX2).

Nagarajan et al. (2013) abordam os algoritmos para a aprendizagem da rede: *constraint-based* e *score-based*. Os algoritmos de *constraint-based* consistem em vários testes de independência condicional e marginal para inferir as direções dos arcos entre os nós, já os algoritmos *score-based* selecionam entre todas as possíveis estruturas da RB a com melhor qualidade, ou seja, com maior pontuação.

Conforme abordado, a princípio se aprende a estrutura gráfica, os algoritmos de *constraint-based* implementados no *bnlearn* do Software R são: *Grow-Shrink* (GS); *Incremental Association Markov Blanket* (IAMB); *Fast Incremental Association* (Fast-IAMB); *Interleaved Incremental Association* (Inter-IAMB); *Max-Min Parents e Children* (MMPC); *Semi-Interleaved Hiton-PC* (SI-HITON-PC). Cada um desses algoritmos podem ser utilizados com vários testes de independência condicional, sendo: informação mútua, qui-quadrado (para dados categóricos), correlação linear, transformação de Fisher e informação mútua (para dados contínuos) e informação mútua (para dados mistos).

Para os algoritmos *constraint-based* podem ser usados os seguintes testes:

- Para dados categóricos (distribuição multinomial):
 - Informação mútua (testes paramétricos, semi-paramétricos e de permutação);
 - estimador de encolhimento para a informação mútua;

- qui-quadrado de Pearson (testes paramétricos, semiparamétricos e de permutação).
- Para dados contínuos (distribuição normal multivariada):
 - Correlação linear (testes paramétricos, semiparamétricos e de permutação);
 - transformação de Fisher (testes paramétricos, semiparamétricos e de permutação);
 - informação mútua (testes paramétricos, semi-paramétricos e de permutação);
 - estimador de encolhimento para a informação mútua.
- Dados mistos (distribuição Gaussiana condicional):
 - Informação mútua (paramétrica, semiparamétrica).

Já os algoritmos de estruturas de aprendizagem *score-based* são: *Hill Climbing* (HC) e *Tabu Search* (Tabu). Para os algoritmos *score-based* podem ser usadas nas funções de pontuação: o critério de informação akaike (AIC), o critério de informação bayesiano (BIC) e o log-verossimilhança condicional gaussiana para de rede mista; além da densidade posteriori Dirichlet e escore K2 apenas na existência de dados categóricos e o log-verossimilhança gaussiana multivariada e densidade posteriori gaussiana para os dados contínuos.

Para os algoritmos *score-based* podem ser usadas as funções de pontuação:

- Dados categóricos (distribuição multinomial):
 - A log-verossimilhança da multinomial;
 - o critério de informação akaike (AIC);
 - o critério de informação bayesiano (BIC);
 - a densidade posteriori Dirichlet equivalente (BDe);
 - a densidade posteriori Dirichlet esparsa (BDs);
 - densidade posteriori de Dirichlet baseada na priori de Jeffrey (BDJ);
 - o escore de BDe da média local (BDla);
 - o escore K2.
- Dados contínuos (distribuição normal multivariada):

- A log-verossimilhança gaussiana multivariada;
 - o critério de informação akaike correspondente (AIC);
 - o critério de informação bayesiano correspondente (BIC);
 - a densidade posteriori gaussiana equivalente (BGe).
- Dados mistos (distribuição Gaussiana condicional):
 - A log-verossimilhança gaussiana condicional;
 - o critério de informação akaike (AIC);
 - o critério de informação bayesiano (BIC).

Os algoritmos de aprendizagem *hybrid* são: *Max-Min Hill Climbing* (MMHC) e *General 2-Phase Restricted Maximization* (RSMAX2). O MMHC utiliza o algoritmo *constraint-based Max-Min Parents and Child* (MMPC) para a busca do esqueleto do grafo, em que faz-se a estimação dos pais e filhos da cobertura de Markov para cada variável envolvida na RB. Logo em seguida para determinar as direções dos arcos no grafo, utiliza-se o algoritmo *score-based Hill Climbing* (HC). O algoritmo RSMAX2 é uma implementação mais geral do MMHC, em que pode-se utilizar qualquer combinação dos algoritmos *constraint-based* com o *score-based*.

REFERÊNCIAS

- ALVARENGA, R. R.; RODRIGHUES, P. B.; ZANGERONIMO, M. G.; FREITAS, R. T. F.; LIMA, R. R.; BERTECHINI, A. G.; FASSANI, E. J. Energetic values of feedstuffs for broilers determined with in vivo assays and prediction equations. **Animal Feed Science and Technology**, v. 168, n.1, p.257-266, 2011.
- ALBINO, L. F. T.; ROSTAGNO, H. S.; FONSECA, J. B.; TORRES, R. A. Utilização de diferentes sistemas de avaliação energéticas dos alimentos na formulação de rações para frangos de corte. **Revista da Sociedade Brasileira de Zootecnia**, v.21, n.6, p.1037-1046, 1992.
- AHMAD, R. Functional and adaptive significance of differentially expressed lactate dehydrogenase isoenzymes in tissues of four obligatory air-breathing *Channa* species. **Biologia**, v.64, n.1, p.192-196, 2009.
- AHMADI, H.; MOTTAGHITALAB M.; NARIMAN-ZADEH, N. Group method of data handling-type neural network prediction of broiler performance based on dietary metabolizable energy, methionine, and lysine. **Journal of Applied Poultry Research**, v.16, n.1, p.494-501, 2007.
- AHMADI, H.; GOLIAN, A. Neural network model for egg production curve. **Journal of Animal and Veterinary Advances**. v.7, n.1, p.1168-1170. 2008.
- AHMADI, H.; GOLIAN, A.; MOTTAGHITALAB, M.; NARIMAN-ZADEH, N. Prediction model for true metabolizable energy of feather meal and poultry oal meal using group method of data handling-type neural network. **Poultry Science** v.87, n.1, p.1909-1912, 2008.
- ARA-SOUZA, A.L. Redes Bayesianas: Uma introdução aplicada a credit scoring. In: 19º Simpósio Nacional de Probabilidade e Estatística-SINAPE, 2010, **Anais...**, São Pedro: SINAPE, 2010.
- BERTECHINI, A. G. **Nutrição de monogástricos**. Lavras, MG: Ed. UFLA, 2012. 373p.
- BISHOP, C. M. **Pattern Recognition and Machine learning**. Springer, New York. 2006.
- BRUM, P. A. R.; ZANOTTO, D. L.; LIMA G. J. M. M.; VIOLA, E. S. Composição química e energia metabolizável de ingredientes para aves. **Pesquisa Agropecuária Brasileira**, v.35, n.5, p.995-1002, 2000.
- BORGES, F. M. O.; ROSTAGNO, H. S.; SAAD, C. E. P.; RODRIGUEZ, N. M.; TEIXEIRA, E. A.; LARA, L. B.; MENDES, W. S.; ARAÚJO, V. L. Equações de regressão para estimar valores energéticos de grão de trigo e seus subprodutos para frangos de corte, a partir de análises químicas. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.55, n.1, p.734-746, 2003.

CARVALHO, J. V. F., CHIANN, C. Redes Bayesianas: um método para avaliação de interdependência e contágio em séries temporais multivariadas. **Revista Brasileira de Economia**, v. 67, n.2, p.201-217, 2013.

DALE, N.; FULLER, H. L. Applicability of the metabolizable energy system in practical feed formulation. **Poultry Science**, v.61, n.2, p.351-356, 1982.

EMBRAPA. **Embrapa Suínos e Aves**, 2018. Disponível em:<<http://https://www.embrapa.br/suinos-e-aves/cias/estatisticas/>>. Acesso em: 22 mar. 2018.

FARIDI, A; GOLIAN, A. The use of neural network models to estimate early egg production in broiler breeder hens through dietary nutrient intake. **Poultry Science**, v.90, n.12, p.2897-2903. 2011.

FELIPE, V.P.S.; SILVA, M.A.; VALENTE, B.D.; ROSA, G.J.M. Using multiple regression Bayesian networks and artificial neural networks for prediction of total egg production in European quails based on earlier expressed phenotypes. **Poultry Science**, v.94, n.4, p.772-780, 2015.

FEOFILOFF, P.; KOHAYAKAWA, Y.; WAKABAYASHI, Y. **Uma Introdução Sucinta à Teoria dos Grafos**, 2011. 61p. Disponível em: <http://www.ime.usp.br/pf/teoriadosgrafos/>. Acesso em: 25 abril de 2018.

FRIEDMAN, N.; LINIAL, M.; NACHMAN, I.; PEER, D. Using bayesian networks to analyze expression data. **Journal of Computational Biology**, v.7, n.3, p.601-620, 2004.

GIANOLA, D., OKUT, H., WEIGEL, K. A., ROSA, G. J. M. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. **BMC Genetics**, v.12, n.1, p.1-87, 2011.

HECKERMAN, D. **A tutorial on learning with Bayesian Networks**, Technical Report MSR-TR-95-06, Microsoft Research, 1995.

IBRAHIM, J. G.; LAUD, P. W. A predictive approach to the analysis of designed experiments. **Journal of the American Statistical Association**, v.89, n.425, p.309-319, 1994.

INRA. **Le Cheval. Reproduction, sélection, alimentation, exploitation**. Jarrige R. and Martin-Rosset W. Eds. INRA Editions, Paris, 1984. 689p.

JENSEN, F. V. **An Introduction to Bayesian Networks**. New York: Springer Verlag, 1996.

KOLLER, D.; FRIEDMAN, N. **Probabilistic Graphical Models: Principles and Techniques**. MIT Press, 2009, 1233p.

KOSKI, T.; NOBLE, J.M. **Bayesian networks: an introduction**. Chichester: John Wiley Sons Ltda, 2009. 347p.

LAURITZEN, S. L.; SPIEGELHALTER, D. J. Local computations with probabilities on graphical structures and their applications to expert systems. **Journal Royal Statistics Society**, v.50, n.1, p.157-224, 1988.

LAURITZEN, S.L., WERMUTH, N. Graphical models for associations between outcomes, some of which are qualitative and some quantitative. **The Annals of Statistics**. v.17, n.1, p.31-57, 1989.

LEESON, S.; SUMMERS, J.D. **Nutrition of the chicken**. 4.ed. Ontario: University Books, 2001. 413p.

MAPA. **Ministério da Agricultura, Pecuária e Abastecimento**. 2017. Disponível em:<<http://www.agricultura.gov.br/assuntos/saude-animal-e-vegetal/saude-animal/arquivos-das-publicacoes-de-saude-animal/2-dr-ariel-panorama-da-avicultura-nacional-e-perspectivas-para-o-setor.pdf/view>>. Acesso em: 4 mar. 2019.

MARGARITIS, D. **Learning bayesian network model structure from data**. 2003. 126f. PhD Thesis (Doctor of Philosophy) - Carnegie Mellon University, Pittsburgh, Pennsylvania, 2003.

MARGARITIS, D.; THRUN, S. Bayesian Network Induction via Local Neighborhoods. In: **Advances in Neural Information Processing Systems**, 2000, Denver, CO, United States, **Proceedings...**, Denver, 2000.

MARIANO, F. C. M. Q.; LIMA, R. R.; RODRIGUES P. B.; ALVARENGA, R. R.; NARedes bayesianasIMENTO, G. A. J. Equações de predição de valores energéticos de alimentos obtidas utilizando meta-análise e componentes principais. **Ciência Rural**, v.42, n.9, p.1634-1640, 2012.

MARIANO, F. C. M. Q.; PAIXÃO, C. A.; LIMA, R. R.; ALVARENGA, R. R.; RODRIGUES P. B.; NARedes bayesianasIMENTO, G. A. J. Prediction of the energy values of feedstuffs for broilers using meta-analysis and neural networks. **Animal**, v. 7, n. 9, p. 1440-1445, 2013.

MEHRI, M. A comparison of neural network models, fuzzy logic, and multiple linear regression for prediction of hatchability. **Poultry Science**, v.92, n.4, p.1138-1142, 2013.

MEHRI, M. Development of artificial neural network models based on experimental data of response surface methodology to establish the nutritional requirements of digestible lysine, methionine, and threonine in broiler chicks. **Poultry Science**, v.91, n.12, p.3280-3285, 2012.

MOROTA, G.; MASANORI, K.; ROSA, G.J.M; KENT, A.W.; GIANOLA, D. Predicting complex traits using a di usion kernel on genetic markers with an application to dairy cattle and wheat data. **Genetics Selection Evolution**, v.45, n.17, p.1-15, 2013.

NAGARAJAN, R., Redes bayesianasUTARI, M; LÈBRE, S. **Bayesian Networks in R with Applications in Systems Biology**. New York: Springer, 2013. 157p.

NAGATA, A. K.; RODRIGUES, P. B.; FREITAS, R. T. F.; BERTECHINI, A. G.; FIALHO, E. T. Energia metabolizável de alguns alimentos energéticos para frango de corte, determinada por ensaios metabólicos e por equações de predição. **Ciência e Agrotecnologia**, v.28, n.3, p.668-677, 2004.

NASCIMENTO, G. A. J.; RODRIGUES, P. B.; FREITAS, R. T. F.; REIS NETO, R. V.; LIMA, R. R.; ALLAMANN, I. B. Equações de predição para estimar valores da energia metabolizável de alimentos concentrados energéticos para aves utilizando meta-análise. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.63, n.1, p.222-230, 2011.

NASCIMENTO, G. A. J.; RODRIGUES, P. B.; FREITAS, R. T. F.; BERTECHINI, A. G.; LIMA, R. R.; PUCCI, L. E. A. Equações de predição para estimar os valores energéticos de alimentos concentrados de origem vegetal para aves utilizando a metanálise. **Revista Brasileira de Zootecnia**, v.38, n.1, p.1265-1271, 2009.

NEAPOLITAN, R. E. **Learning Bayesian Networks**, Upper Saddle River, NJ: Prentice Hall, 2003. 693p.

NATIONAL RESEARCH COUNCIL. **Nutrient Requirements of Poultry**: Ninth Revised Edition, Washington, DC: The National Academies Press.1994. 176p.

PERAI, A. H.; MOGHADDAM, H. N.; ASADPOUR, S.; BAHRAMPOUR, J.; MANSOORI, Gh. A comparison of artificial neural networks with other statistical approaches for the prediction of true metabolizable energy of meat and bone meal. **Poultry Science**, v.89, n.1, p.1562-1568, 2010.

PEARL, J. **Probabilistic reasoning in intelligent systems: networks of plausible inference**. San Francisco: Editor Morgan Kaufmann, 1988. 552p.

POURRET, O.; NAIM, P.; MARCOT, B. **Bayesian Networks: a Practical Guide to Applications**. Edited: John Wiley Sons Ltd, The Atrium, Southern Gate, Chichester, 2008. 446p.

R CORE TEAM (2018). **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <http://www.R-project.org/>. Acesso em 25 de março de 2018.

RODRIGUES, P. B.; ROSTAGNO, H. S.; ALBINO, L. F. T.; GOMES, P. C.; BARBOZA, W. A.; SANTANA, R. T. Valores energéticos do milheto, do milho e subprodutos do milho, determinados com frangos de corte e galos adultos. **Revista Brasileira de Zootecnia**, v.30, n.6, p.1767-1778, 2001.

RODRIGUES, P. B.; ROSTAGNO, H. S.; ALBINO, L. F. T.; GOMES, P. C.; NUNES, R. V.; TOLEDO, R. S. Valores energéticos da soja e subprodutos da soja, determinados com frangos de corte e galos adultos. **Revista Brasileira de Zootecnia**, v.31, n.1, p.1771-1782, 2002.

ROSTAGNO, H. S.; ALBINO, L. F. T.; DONZELE, J. L.; GOMES, P. C.; OLIVEIRA, R. F.; LOPES, D. C.; FERREIRA, A. S.; BARRETO, S. L. T.; EUCLIDES, R. F. **Tabelas**

brasileiras para aves e suínos: composição de alimentos e exigências nutricionais de aves e suínos. 3ª edição, Viçosa, MG: UFV, 2011. 252p.

SAVEGNAGO, R. P.; NUNES, B. N.; CAETANO, S. L.; FERRAUDO, A. S.; SCHIMIDT, G. S.; LEDUR, M. C.; MUNARI, D. P. Comparison of logistic and neural network models to fit the egg production curve of White Leghorn hens. **Poultry Science**, v.90, n.1, p.705-711, 2011.

SCOTT, M.L.; NESHEIN, M.C.; YOUNG, R.J. **Nutrition of the chicken**. 3.ed. New York: Scott and Assoc. Publ., 1982. 562p.

SCUTARI, M., DENIS, J.B. **Bayesian Networks with Examples in R**. Taylor & Francis Group. 2015. 221p.

SCUTARI, M. Learning Bayesian Networks with the bnlearn R Package, **Journal of Statistical Software**, v.35, n.3, p.1-21, 2010.

SCUTARI, M.; HOWELL, P.; BALDING, D.J.; MACKAY, I.J. Multiple quantitative trait analysis using bayesian networks. **Genetics**, v.198, n.1, p.129-137, 2014.

SEMSARIAN, S.; NASAB, M. P. E.; ZAREHDARAN, S.; DEHGHANI, A. A. Prediction of the weight and number of eggs in Mazandaran native fowl using artificial neural network. **International Journal of Advanced Biological and Biomedical Research**, v.5, n.1, p.532-537, 2013.

SIBBALD, I. R. Measurement of bioavailable energy in poultry feedingstuffs: a review. **Canadian Journal of Animal Science**, v.62, n.4, p.983-1048, 1982.

SIMÕES, P. W. T. A.; SILVA, G. D.; MORETTI, G. P.; SIMON, C. S.; WINNIKOW, E. P.; NASSAR, S. M.; MEDEIROS, L. R.; ROSA, M. I. Meta analysis of the use of Bayesian networks in breast cancer diagnosis. **Caderno Saúde Pública**, v.31, n.1, p.26-38, 2015.

SPIRITES, P., GLYMOUR, C.; Redes bayesianas HEINES, R. **Causation, Prediction and Search**, Second edition, Cambridge: MIT Press, 2000.

SU, C.; ANDREW, A.; KARAGAS, M. R.; BORSUK, M. E. Using Bayesian networks to discover relations between genes, environment, and disease. **BioData Mining**, v.6, n.6, p.1-21, 2013.

TÖPNER, K.; ROSA, G. J. M.; GIANOLA, D.; Redes bayesianas HÖN, C. C. Bayesian Networks Illustrate Genomic and Residual Trait Connections in Maize (*ZeamaysL.*) **G3: Genes, Genomes, Genetics**, v.7, n.1, p.2779-2789, 2017.

TREPODE, N.W.; FARIAS, C.R.G.; BARRERA, J. A pattern-oriented specification of gene network inference processes. **Computers in Biology and Medicine**, v.43, n.1, p.1415-1427, 2013.

TSAMARDINOS, I.; BROWN, L.E.; ALIFERIS, C.F. Algorithms for large scale markov blanket discovery. In: "Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference". **Proceedings...** 2003, p.376-381. AAAI Press, 2003.

TSAMARDINOS, I.; BROWN, L.E.; ALIFERIS, C.F. The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm. **Machine Learning**, v.65, n.1, p.31-78, 2006.

UNITED STATES DEPARTMENT OF AGRICULTURE. **Livestock and poultr: world markets and trade**. Foreign Agricultural Service, New York, 2017. Disponível em: <http://www.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf>. Acesso em: 11 Set. 2017.

VIEIRA, E. M. A., SILVA, J. M. N., SILVA, L. B. Modeling Bayesian Networks from a conceptual framework for occupational risk analysis. **Production**, v.27, n.1, p.1-12, 2017.

WAN, H. F.; CHEN, W.; QI, Z. L.; PENG, P.; PENG, J. Prediction of true metabolizable energy from chemical composition of wheat milling by-products for ducks. **Poultry Science**, v.88, n.1, p.92-97, 2009.

ZHAO, F.; ZHANG, H. F.; HOU, S. S.; ZHANG, Z. Y. Predicting metabolizable energy of normal corn from its chemical composition in adult pekin ducks. **Poultry Science**, v.87, n.1, p.1603-1608, 2008.

YARAMAKALA, S.; MARGARITIS, D. Speculative Markov Blanket Discovery for Optimal Feature Selection. In: Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM), **Proceedings...** 2005, Houston, Texas, United States, 2005.

3 ARTIGOS

SEGUNDA PARTE
ARTIGOS CIENTÍFICOS

ARTIGO 1: Inovação na predição dos valores energéticos de alimentos para aves: um estudo bibliométrico (1945-2018) da produção científica na base *Web of Science* sobre as redes bayesianas

ALVARENGA, T.C.; LIMA, R.R.; BUENO FILHO, J. S. S.; RODRIGUES, P. B.; ALVARENGA, R. R.; MARIANO, F. C. M. Q.

Artigo a ser submetido na Revista *Worlds Poultry Science Journal*

Resumo

A *Web of Science da Thomson Reuters* é uma base de dados que possibilita identificar padrões e tendências das publicações científicas, permitindo assim, uma ampla compreensão das publicações da área de interesse. Uma área que vem despertando atenção para a comunidade estatística são as redes bayesianas, principalmente pela flexibilidade de modelar os dados, ambos discretos e contínuos, em termos de regressão e alta acurácia nos resultados obtidos, ou seja, são modelos probabilisticamente promissores. O objetivo desta foi de identificar e descrever as principais categorias do *Web of Science* que contemplam pesquisas sobre redes bayesianas, verificar as publicações ao longo dos anos, identificar os tipos de documentos publicados, além das principais agências financiadoras, os principais autores, países e idiomas. Para a realização desta pesquisa foram coletados os dados da base *Web of Science da Thomson Reuters* de 1945 à 2018. Por meio da busca é possível responder a várias questões de interesse, entre elas, se existem publicações de redes bayesianas principalmente nas ciências animais, mais especificamente na formulação de rações para frangos de corte. O interesse pela busca é verificar se a metodologia de redes bayesianas é aplicável na obtenção de equações de predição de energia metabolizável de frangos de corte. Mediante essas questões, foram feitas buscas na base com as palavras relacionadas e palavras de acordo com essa metodologia no que tange a regressão, tais como, comitê e algoritmo *Max-Min Hill Climbing* (MMHC). Os resultados encontrados confirmaram que essa área de conhecimento ainda é bem recente. As primeiras publicações se deram em 1990 e as principais publicações estão concentradas na ciência da computação e não foi encontrado nenhuma pesquisa na predição da energia metabolizável de frangos de corte utilizando essa metodologia.

Palavras-chave: Algoritmo híbrido, avicultura, bibliometria, energia metabolizável.

Abstract

The Thomson Reuters Web of Science is a database that enables to identify patterns and trends in scientific publications, thereby enabling a broad understanding of the publications in the area of interest. An area that has been arising attention to the statistical community is the Bayesian networks, mainly due to the flexibility of modeling the data, both discrete and continuous, in terms of regression and high accuracy in the obtained results, that is, they are probabilistically promising models. The objective of this study was to identify and describe the main categories of the Web of Science that contemplate studies on Bayesian networks, check the publications over the years, identify the types of documents published, as well as the main funding agencies, the main authors, countries and languages. For the accomplishment of this study, the data of the Thomson Reuters Web of Science database were collected from 1945 to 2018. By means of search it is possible to answer several questions of interest, among them, if there are publications of Bayesian networks mainly in the animal sciences, more specifically in the formulation of diets for broilers. The interest in the search is checking if the methodology of Bayesian networks is applicable in obtaining prediction equations of metabolizable energy of broilers. By means of these questions, searches were made on the database of related words and words according to this methodology as far as the regression is concerned such as the committee and Max-Min Hill Climbing algorithm (MMHC). The results confirm that this area of knowledge is still very recent. The first publications took place in 1990 and the main publications are concentrated in computer science and no study was found in the prediction of the metabolizable energy of broilers using this methodology.

Keywords: Hybrid algorithm, poultry, bibliometrics, metabolizable energy.

1 Introdução

O conhecimento da energia metabolizável aparente corrigida pelo balanço de nitrogênio (EMAn) dos alimentos, bem como sua composição química, faz-se necessário na elaboração de rações balanceadas pela indústria de frangos de corte. Os meios da obtenção dos valores da EMAn podem ser obtidos em ensaios biológicos (ZHAO et al., 2008; WAN et al., 2009), bem como pelas tabelas de composição de alimentos (SCOTT et al., 1982; NRC, 1994; LEESON; SUMMERS, 2001; I.N.R.A., 1984; ROSTAGNO et al., 2017). No entanto, a execução dos ensaios biológicos demandam tempo, são de custo

elevado e geralmente as tabelas não são as mais confiáveis pela diversidade de valores que as mesmas sugerem (NAGATA et al., 2004; NASCIMENTO et al., 2011). Outro meio de se obter os valores de EMAn são as equações de predição estabelecidas em função da composição química dos alimentos (ALVARENGA et al., 2011; MARIANO et al., 2013), normalmente de fácil e rápida obtenção. Na literatura existem trabalhos que obtiveram as equações de predição por meio de regressão múltipla (RODRIGUES et al., 2001, 2002; BORGES et al., 2003; ZHAO et al., 2008), meta-análise (NASCIMENTO et al., 2009, 2011; MARIANO et al., 2012) e redes neurais (PERAI et al., 2010; AHMADI et al., 2007, 2008; MARIANO et al., 2013). No entanto, é necessário a inovação em termos de utilização metodológica (ALVARENGA et al., 2015), podendo assim, resultar em melhores estimativas dos valores energéticos.

Uma metodologia que se tem demonstrado promissora, quando o objetivo é predição, são as redes bayesianas (RB) (GIANOLA et al., 2011 citado por Felipe et al., 2015). As RB são modelos gráficos os quais consistem na representação gráfica (grafo) e probabilística (distribuições de probabilidade condicionais e conjunta) das variáveis (SCUTARI; DENIS, 2015). Uma RB constitui uma representação gráfica de variáveis e suas relações causais, nomeadas respectivamente por nós ou vértices e arestas, arcos ou setas (PEARL, 1988; SPIRITES et al., 2000; KOLLER; FRIEDMAN, 2009; SCUTARI; DENIS, 2015). Essa representação gráfica, denominada de grafo, é estudada na Teoria de Grafos. Mais detalhes dessa teoria podem ser encontrados em Feofiloff et al. (2011). A representação das relações entre as variáveis (nós) são dadas no grafo por meio de setas, ou seja, $A \rightarrow B$, a qual se lê A causa B ou A tem relação com B ou B depende de A. Neste tem-se que o nó A e o nó B estão, respectivamente, na cauda e cabeça da aresta. A ilustração gráfica de uma RB é denotada por DAG (*directed acyclic graph*).

As DAGs apresentam direção nas arestas, todos os nós são conectados na estrutura gráfica e as direções das arestas não apresentam ciclos, ou seja, após a aresta seguir uma direção, ela não retorna ao mesmo nó. Essas características classificam um grafo como direcionado, conectado e acíclico. Essa metodologia pode ser utilizada na construção de modelo preditivo pela fatoração da distribuição de probabilidade conjunta, assumindo um conjunto de independências condicionais (JENSEN, 1996; LAURITZEN; SPIEGELHATER, 1988). A fatoração da RB, conforme apresentado na Equação 1, consiste no produto das probabilidades condicionais, visto que um nó é condicionalmente independente de seus não-descendentes, dado seus pais (PEARL, 1988; TSAMARDINOS et al., 2003). Essa fatoração resulta na distribuição de probabilidade conjunta, sendo assim, possível inferir e obter respostas a questões de interesse.

$$P(X_1, X_2, \dots, X_p) = \prod_{i=1}^n P(X_i | Pa_i) \quad (1)$$

para o caso de nós discretos e contínuos, em que Pa_i são os pais de X_i .

Além das RB, outra metodologia utilizada e indicada em estudos cujo objetivo principal é a predição, são as redes neurais artificiais (BISHOP, 2006). Felipe et al. (2015) mencionam a utilização dessa metodologia nas ciências animais, cujas aplicações são a predição da produção de ovos (AHMADI; GOLIAN, 2008; SAVEGNAGO et al., 2011; FARIDI; GOLIAN, 2011); eclobilidade (MEHRI, 2013); do peso e do número de ovos (SEMSARIAN et al., 2013); obtenção de curvas de crescimento (AHMAD, 2009) e estudos sobre nutrição em aves capoeiras (MEHRI, 2012). Assim como as redes neurais artificiais têm sido utilizadas nas ciências animais e os resultados obtidos pelas mesmas tem tido êxito, é de interesse avaliar se as redes bayesianas também têm tido destaque nessa mesma área. Alguns autores têm utilizado as redes bayesianas para predição e afirmam sua superioridade em capturar (e expressar) relações mais complexas entre as covariáveis e a variável resposta (GIANOLA et al. 2011, citado por Felipe et al., 2015).

Com o objetivo de identificar e descrever as principais categorias do *Web of Science* que contemplam pesquisas sobre redes bayesianas acreditando no potencial destas para a área de avicultura, mais especificamente na predição de valores energéticos dos alimentos, verificar as publicações ao longo dos anos, identificar os tipos de documentos publicados, além das principais agências financiadoras, os principais autores, países e idiomas sobre as publicações, foi realizada essa pesquisa bibliométrica. Fez-se buscas na base *Web of Science da Thomson Reuters* de 1945 a 2018 sobre publicações que contemplaram as palavras “*bayesian network*”, “*bayesian network and animal*”, “*bayesian network and poultry*”, “*bayesian network and committee*”, “*bayesian network and MMHC*”, entre outras. Com essa pesquisa buscou-se avaliar o panorama das publicações e assim, ter um direcionamento para pesquisas posteriores sobre as redes bayesianas e as palavras chaves de interesse.

2 Metodologia de pesquisa

O presente estudo se caracteriza como uma análise bibliométrica de publicações científicas. A bibliometria utiliza normalmente de técnicas quantitativas para analisar a produção acadêmica por meio de palavras-chaves que podem ser identificadas na publicação. Para a realização desta pesquisa utilizou-se as seguintes etapas, as quais estão resumidas na Tabela 1.

Tabela 1 – *Framework* de análise bibliométrica.

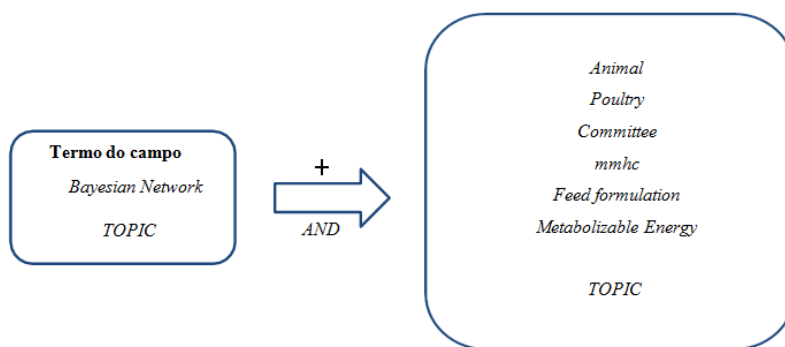
Etapa	Procedimento	Descrição
1	Operacionalização da pesquisa	1.1 Escolha da(s) base(s) científica(s) ou periódicos
		1.2 Delimitação do termo que representa o campo
		1.3 Delimitação de outros termos para apurar os resultados
2	Procedimentos de busca (filtros)	2.1 <i>Topic</i> (termo do campo) <i>AND topic</i> (direcionamento)
		2.2 Utilização de <i>underline</i> : expressão exata
		2.3 Filtro 1: Todos os anos
		2.4 Filtro 2: Todas as áreas
		2.5 Filtro 3: Todos os idiomas
		2.6 Filtro 4: Delimitação em somente artigos
3	Análise da produção científica	3.1 Análise do volume das publicações e tendências temporais
		3.2 Análise das categorias da base
		3.3 Análise dos meios de publicação
		3.4 Análise dos países
		3.5 Análise das universidades
		3.6 Análise das áreas de pesquisa
		3.7 Análise das agências de financiamento
		3.8 Análise dos idiomas
		3.9 Descrição dos principais resultados com as outras buscas

Fonte: Adaptado de Prado et al., 2016.

2.1 Etapa 1: Operacionalização da pesquisa

A base *Web of Science da Thomson Reuters* foi utilizada para identificar as publicações de interesse. A delimitação do termo que representa o campo além de outros termos para apurar os resultados se encontram na Figura 1.

Figura 1 – Procedimentos de busca na *Web of Science*.



Fonte: Elaborado pelos autores (2018).

2.2 Etapa 2: Procedimentos de busca (filtros)

Para a realização das buscas (procedimento 2.1 - Tabela 1) utilizou-se no "*topic*" apenas o termo ("*bayesian network*"). Num segundo momento foi inserido o conectivo "*and*" para ligar o primeiro campo de pesquisa ao segundo, conforme segue ilustrado na Figura 1.

As buscas foram realizadas com as combinações dos termos em “*topic*” totalizando várias pesquisas. Utilizou-se underline para buscar a expressão exata no termo de campo. Foram considerados todas as publicações e posteriormente só artigos científicos que abordavam sobre redes bayesianas em todos os anos disponíveis na base. Posteriormente restringiu-se a busca para o período de 1990 (primeiro trabalho encontrado) a outubro de 2018 e delimitou-se por todas as áreas do conhecimento.

2.3 Etapa 3: Análise da produção científica

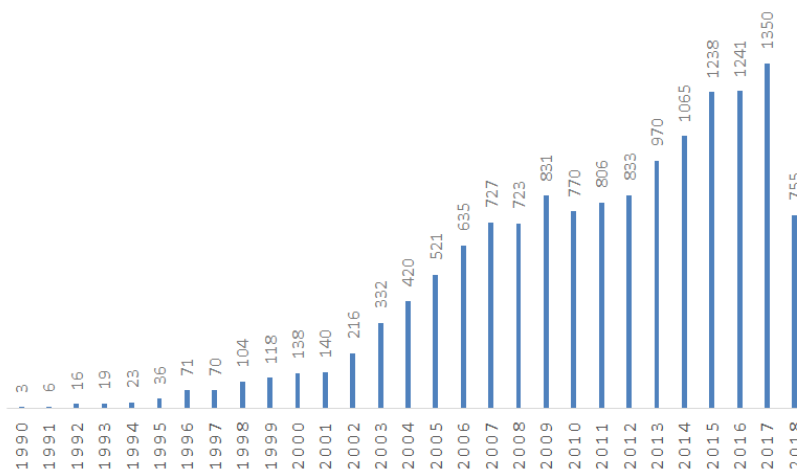
Foram geradas tabelas e gráficos tendo por objetivo apresentar em termos quantitativos as publicações do campo de estudo em foco.

3 Resultados e Discussão

As buscas foram realizadas no “*topic*” totalizando 14174 publicações, com o termo redes bayesianas. Utilizou-se *underline* para buscar a expressão exata. Foram considerados todos os tipos de publicações disponíveis na base. A partir disso se restringiu a busca para o período de 1990 a qual se encontrou a primeira publicação intitulada de *Application of bayesian networks to multitarget tracking* (KOVACICH, 1990) em que não teve nenhuma citação. O segundo trabalho publicado foi *Identifying independence in bayesian networks* (GEIGER et al., 1990), a qual teve 185 citações. O mesmo foi publicado na revista Networks. A terceira publicação foi o artigo da revista *Artificial intelligence* nomeado por *Gibbs sampling in bayesian networks* (HRYCEJ, 1990), o qual teve um total de 24 citações. As publicações mais recentes são *Margins of discrete bayesian networks* (EVANS, 2018), *High-dimensional consistency in score-based and hybrid structure learning* (NANDY et al., 2018) e *Applications of Business Analytics in Predicting Flight On-time Performance in a Complex and Dynamic System* (TRUONG et al., 2018). As duas primeiras publicações são do *Annals of Statistics* e a terceira publicação é da *Transportation Journal*. Para as demais consultas das publicações, as mesmas podem ser consultadas na *Web of Science*.

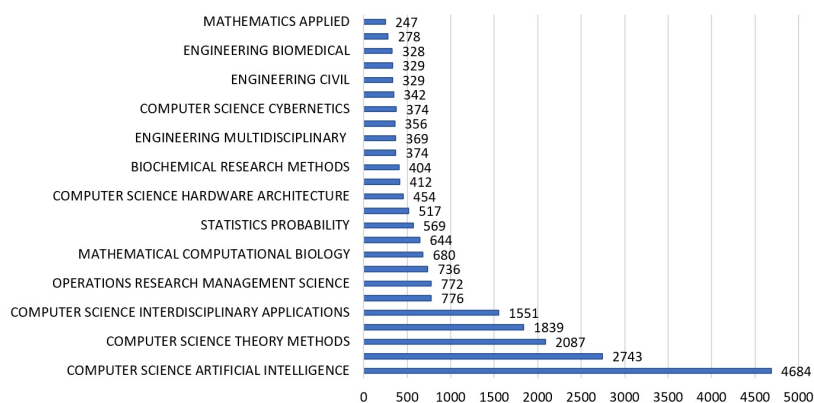
Por meio da Figura 2 observa-se o número de publicações sobre as redes bayesianas ao longo dos anos. O início das publicações foi em 1990, com uma tendência crescente até 2007. Entre os anos 2008 e 2010, houve uma pequena oscilação quanto ao número de publicações e depois novamente uma tendência crescente. Para o ano de 2018 não se tem o seu número total, pois as buscas dos dados se deram apenas até o mês de outubro. O ano de 2017, 2016 e 2015 são os anos que tiveram um maior índice de publicações, até o momento.

Figura 2 – Publicações sobre redes bayesianas (1990 - Outubro de 2018).



Fonte: *Web of Science* (2018).

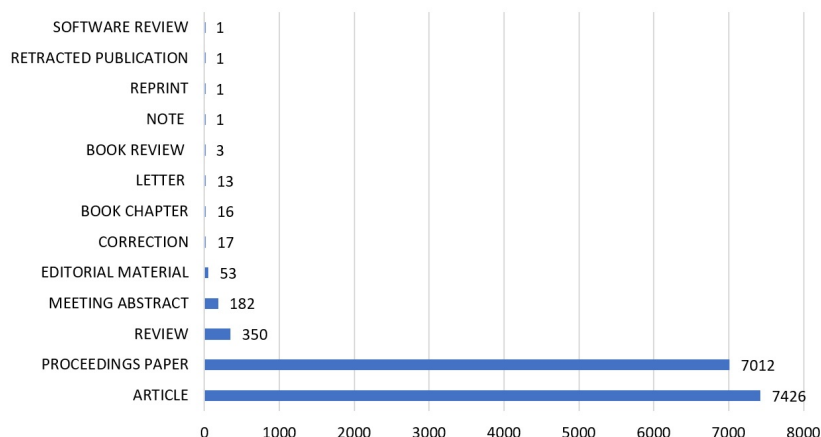
Na Figura 3 pode-se observar as principais categorias do *Web of Science* e seus respectivos números de publicações sobre redes bayesianas. A categoria que apresenta um maior número de publicações é da Inteligência Artificial - Ciência da Computação no total de 4684 publicações. A categoria Probabilidade e Estatística tem um total de 569 publicações. Observa-se que se tratando da categoria relacionada às áreas agrônômica ou zootécnica, não foram encontrados registros nas primeiras 25 categorias de maior número de publicações. Isso mostra que redes bayesianas de alguma forma não é muito conhecida e nem tão utilizada na experimentação agropecuária.

Figura 3 – Categorias do *Web of Science* mais publicadas sobre redes bayesianas.

Fonte: *Web of Science* (2018).

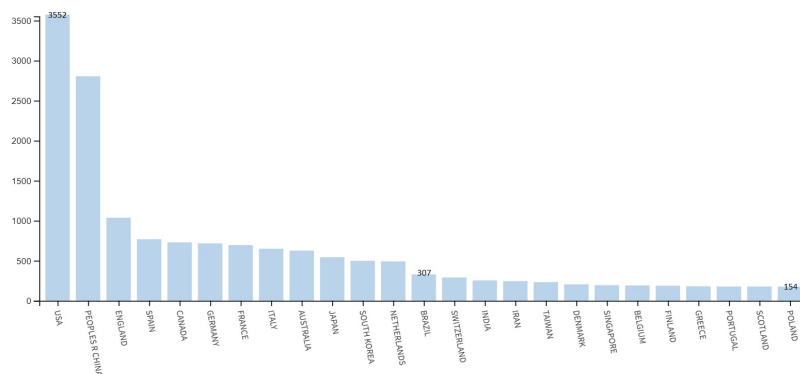
Na Figura 4 encontram-se os principais meios de publicação que abordam sobre as redes bayesianas. Observa-se que a maioria das publicações são em forma de artigo e este número é consideravelmente superior aos demais meios, exceto em relação aos publicados em anais de eventos científicos, que apresenta também um número consideravelmente alto.

Figura 4 – Principais meios de publicações que abordaram sobre redes bayesianas.



Fonte: *Web of Science* (2018).

Figura 5 – Principais países que mais publicaram sobre as redes bayesianas.



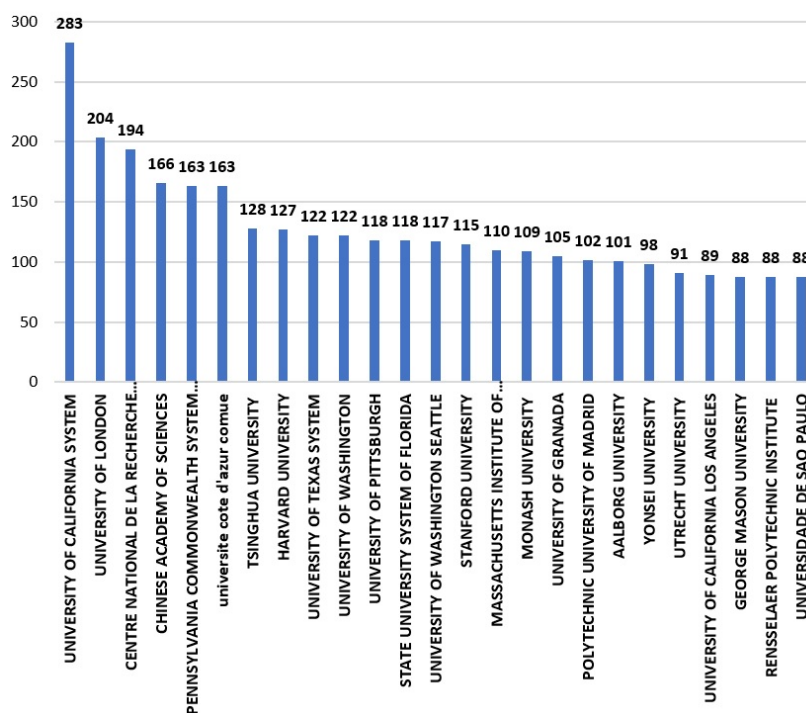
Fonte: *Web of Science* (2018).

Na Figura 5 encontram-se os principais países que publicam sobre as redes bayesianas. Observa-se que os Estados Unidos lidera o ranking com 3552 publicações. O Brasil possui 307 publicações e, das 25 primeiras posições mostradas na Figura 5, o país com menor índice de publicação foi a Polônia, com apenas com 154 publicações. Por meio desse cenário, avaliou-se o número

de publicações com relação às universidades, cujo resultado está apresentado na Figura 6.

Na Figura 6 encontram-se as 25 universidades que mais publicaram sobre as redes bayesianas. Os resultados corroboram a Figura 5, já que os Estados Unidos lideram o ranking de país com maior número de publicações. A universidade com maior publicação pertence a este país, com um total de 283 publicações. A universidade brasileira também tem destaque entre as 25 universidades mencionadas na Figura 6, indicando que a Universidade de São Paulo publicou em 88 meios de divulgação. Para complementar o conhecimento sobre as universidades, investigou-se sobre as principais áreas de pesquisas que trabalharam com as redes bayesianas. Este resultado se encontra na Figura 7.

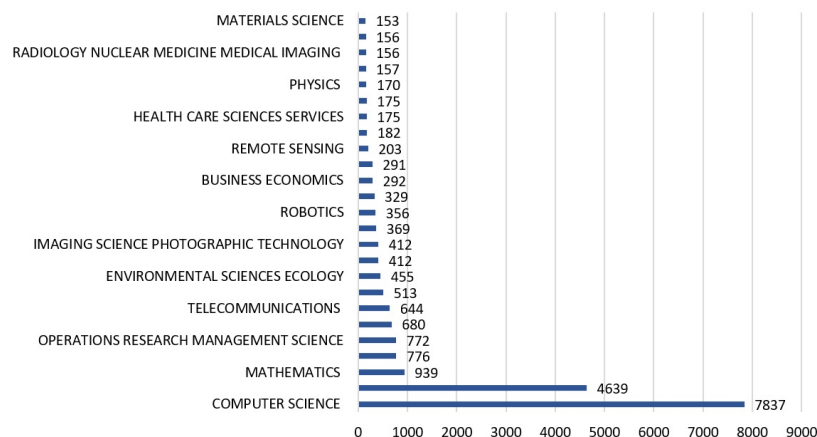
Figura 6 – Principais universidades que mais publicaram sobre as redes bayesianas.



Fonte: *Web of Science* (2018).

Na Figura 7 apresentam as 25 áreas que mais publicaram sobre as redes bayesianas. Os resultados confirmam que, de acordo com a categoria de classificação da *Web of Science*, a área de Ciências da Computação tem um maior número de registros, no total 7837 publicações. Seguindo essa classificação destaca-se a área da Engenharia, com 4639 publicações. As

Figura 7 – Principais áreas de pesquisas que publicaram sobre redes bayesianas.



Fonte: *Web of Science* (2018).

demais áreas possuem um número abaixo de 1000 publicações. Também é de interesse verificar agências que financiam essas publicações.

Por meio da *Web of Science* verifica-se que a agência *National Natural Science Foundation of China* foi a que mais financiou publicações sobre as redes bayesianas. O número de publicações desta agência superou 600 publicações, a qual é um número bem superior de publicações comparado às publicações das demais agências consultadas na base, aos quais as mesmas não superam 150 publicações. Entre as 25 agências que mais financiaram as pesquisas, o CNPq teve destaque também com 31 trabalhos publicados por intermédio de seu auxílio financeiro.

Em consulta a base *Web of Science* é verificado que dentre os 12 idiomas que publicaram sobre as redes bayesianas, 14068 publicações são do idioma em inglês. Nos outros idiomas, alemão, chinês, espanhol, francês, japonês, russo, entre outros, o número de publicações destes é inferior a 50.

Seguindo os passos apresentados na metodologia fez-se um filtro nas buscas apenas selecionando os artigos. Na Tabela 2 apresentam-se os 10 artigos mais citados dentre os 7426 artigos selecionados na área de redes bayesianas. Além dos artigos, na Tabela 2, é apresentado os autores e o número de citações dos mesmos. O artigo com o maior número de citações foi *Bayesian network classifiers* (FRIEDMAN et al.,1997), com 1833 citações.

Na busca pelas palavras “*Bayesian Network AND Animal*”, ambas buscadas no corpo do texto, título, palavras-chaves, foram encontrados 80 registros com estas palavras, das quais as três principais categorias de publicação na base foram na *Computer Science Artificial Intelligence* (11, 25% das publicações), *Mathematical Computational Biology* (10% publicações) e

Computer Science Interdisciplinary Applications (7,5% das publicações). Os Estados Unidos foi o país que liderou as publicações com 47,5% das publicações, seguido pela Austrália e Inglaterra, ambos com 11,25% das publicações. As principais publicações foram da *Pennsylvania Commonwealth System of Higher Education Pcshe*, *Duke University*, *University of Pittsburgh*, *University of Queensland*, *Harvard University*, *Procter Gamble*, *University of Melbourne* e *University of Wisconsin-Madison*.

Tabela 2 – Artigos mais citados - *Web of Science*.

Artigo	Autor	Citações
<i>Bayesian network classifiers</i>	Friedman et al. (1997)	1833
<i>Using Bayesian networks to analyze expression data</i>	Friedman et al. (2000)	1571
<i>Learning bayesian networks - the combination of knowledge and statistical-data</i>	Heckerman et al. (1995)	1341
<i>A Bayesian networks approach for predicting protein-protein interactions from genomic data</i>	Jansen et al. (2003)	797
<i>The max-min hill-climbing Bayesian network structure learning algorithm</i>	Tsamardinos et al. (2006)	372
<i>Advances to Bayesian network inference for generating causal networks from observational biological data</i>	Yu et al. (2004)	384
<i>Advantages and challenges of Bayesian networks in environmental modelling</i>	Uusitalo (2007)	372
<i>Improving the analysis of dependable systems by mapping fault trees into Bayesian networks</i>	Bobbio et al. (2001)	346
<i>Learning Bayesian networks from data: An information-theory based approach</i>	Cheng et al. (2002)	338
<i>Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks</i>	Friedman; Koller (2003)	329

Fonte: Elaborado pelos autores (2018).

Na busca pelas palavras “*Bayesian Network AND poultry*”, ambas com busca no corpo do texto, foram encontrados apenas 5 registros, sendo eles *Using Monte-Carlo simulations and Bayesian Networks to quantify and demonstrate the impact of fertiliser best management practices* (NASH; HANNAH, 2011) com 29 citações e publicado no periódico *Environmental Modelling Software*. O artigo *Detection of poultry egg freshness by dielectric spectroscopy and machine learning techniques* (SOLTANI; OMID, 2015) foi publicado na *Lwt-Food Science and Technology* e tem 9 citações. Empatam em número de citações os artigos, *Using multiple regression, Bayesian networks and artificial neural networks for prediction of total egg production in European quails based on earlier expressed phenotypes* (FELIPE et al., 2015) publicado na *Poultry Science* e o artigo *Bayesian network as an aid for Food Chain Information use for meat inspection* (LUPO et al., 2013), que foi publicado no periódico *Preventive Veterinary Medicine*. Apenas com 2 citações, o artigo *Multivariate Analysis of the Determinants of the End-Product Quality of Manure-Based Composts and Vermicomposts Using Bayesian Network Modelling* (FAVERIAL et al., 2016) foi publicado na *Plos One*. Percebe-se

que as publicações são bem recentes e que o campo ainda pode ser explorado.

Na busca pelas palavras “*Bayesian Network AND committee*”, ambas com busca no corpo do texto, foram encontrados 24 publicações. A primeira publicação e a mais citada foi *Bayesian neural networks for classification: how useful is the evidence framework?* (PENNY; ROBERTS, 1999) publicado no periódico *Neural Networks*, este artigo teve 62 citações. A publicação mais recente com a busca de palavras mencionadas foi o artigo *Identifying opportunities to improve piped water continuity and water system monitoring in Honduras, Nicaragua, and Panama: Evidence from Bayesian networks and regression analysis* (CRONK; BARTRAM, 2018), publicado no *Journal of Cleaner Production*. No entanto, comitês a qual combinam diferentes redes bayesianas híbridas não foram encontradas publicações.

Na busca pelas palavras “*Bayesian network AND MMHC*”, ambas com busca no corpo do texto, foram encontradas apenas 16 publicações. A justificativa de busca pela palavra “*MMHC*” se baseia na literatura, já que a mesma aborda que é um algoritmo computacionalmente eficiente e utilizado quando o estudo é predição (FELIPE et al., 2015) e por meio dele pode-se obter regressões multivariadas. Por meio dessa busca verificou-se que o primeiro artigo e mais citado foi *The max-min hill-climbing Bayesian network structure learning algorithm* (TSAMARDINOS et al., 2006) publicado na *Machine Learning*, foram 394 citações, no qual é proposto o método. Observa-se que a maioria das publicações se encontram nas categorias da *Web of Science* relacionadas à computação.

Para conhecimento de publicações na produção avícola foram feitas buscas pelas palavras “*Bayesian network AND Feed formulation*” e “*Bayesian network AND Metabolizable Energy*”, ambas palavras com busca no corpo do texto. Por meio dessas buscas não foram encontrados nenhum registro de publicações.

4 Considerações Finais

Por meio de todas as buscas na base *Web of Science* observa-se que a maioria das publicações estão relacionadas à área de Ciências da Computação. Nas áreas aplicadas, principalmente agropecuária, ainda se têm pouquíssimas publicações ou nenhuma. Conclui-se que redes bayesianas é uma linha de pesquisa inédita na nutrição de aves e que pode ser estudada por pesquisadores que tem o interesse na predição dos valores de energia metabolizável.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo auxílio financeiro.

Referências

ALVARENGA, R. R.; RODRIGUES, P. B.; ZANGERONIMO, M. G.; FREITAS, R. T. F.; LIMA, R. R.; BERTECHINI, A. G.; FASSANI, E. J. Energetic values of feedstuffs for broilers determined with in vivo assays and prediction equations. **Animal Feed Science and Technology**, v.168, n.1, p.257-266, 2011.

ALVARENGA, R. R. ; RODRIGUES, P. B. ; ZANGERONIMO, M. G. ; OLIVEIRA, E. C. ; MARIANO, F. C. M. Q. ; LIMA, E. M. C. ; GARCIA JR, A. A. P. ; NAVES, L. P. ; NARDELLI, N. B. S. . Validation of Prediction Equations of Energy Values of a Single Ingredient or Their Combinations in Male Broilers. **Asian-Australians Journal of Animal Science**, v.28, n.1, p.1335-1344, 2015.

AHMAD, R. Functional and adaptive significance of differentially expressed lactate dehydrogenase isoenzymes in tissues of four obligatory air-breathing Channa species. **Biologia**, v.64, n.1, p.192-196, 2009.

AHMADI, H.; GOLIAN, A. Neural network model for egg production curve. **Journal of Animal and Veterinary Advances**, v.7, n.1, p. 1168-1170. 2008.

AHMADI, H.; MOTTAGHITALAB M.; NARIMAN-ZADEH, N. Group method of data handling-type neural network prediction of broiler performance based on dietary metabolizable energy, methionine, and lysine. **Journal of Applied Poultry Research**, v.16, n.1, p.494-501, 2007.

AHMADI, H.; GOLIAN, A.; MOTTAGHITALAB, M.; NARIMAN-ZADEH, N. Prediction model for true metabolizable energy of feather meal and poultry oal meal using group method of data handling-type neural network. **Poultry Science**, v.87, n.1, p.1909-1912, 2008.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. New York: Springer, 2006.

BOBBIO, A.; PORTINALE, L.; MINICHINO, M.; CIANCAMERLA, E. Improving the analysis of dependable systems by mapping fault trees into

Bayesian networks. **Reliability Engineering and System Safety**, v.71, n.3, p.249-260, 2001.

BORGES, F. M. O.; ROSTAGNO, H. S.; SAAD, C. E. P.; RODRIGUEZ, N. M.; TEIXEIRA, E. A.; LARA, L. B.; MENDES, W. S.; ARAÚJO, V. L. Equações de regressão para estimar valores energéticos de grão de trigo e seus subprodutos para frangos de corte, a partir de análises químicas. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.55, n.1, p.734-746, 2003.

CHENG, J.; GREINER, R.; KELLY, J.; BELL, D.; LIU, W. Learning Bayesian networks from data: An information-theory based approach. **Artificial Intelligence**, v.137, n.1, p.43-90, 2002.

CRONK, R; BARTRAM, J. Identifying opportunities to improve piped water continuity and water system monitoring in Honduras, Nicaragua, and Panama: Evidence from Bayesian networks and regression analysis. **Journal of Cleaner Production**, v.196, n.1, p.1-10, 2018.

EVANS, R.J. Margins of discrete Bayesian networks. **The Annals of Statistics**, v.46, n.6, p.2623-2656, 2018.

FARIDI, A; GOLIAN, A. The use of neural network models to estimate early egg production in broiler breeder hens through dietary nutrient intake. **Poultry Science**, v.90, n.12, p.2897-2903. 2011.

FAVERIAL, J.; CORNET, D.; PAUL, J.; SIERRA, J. Multivariate Analysis of the Determinants of the End-Product Quality of Manure-Based Composts and Vermicomposts Using Bayesian Network Modelling. **PLoS ONE**, v.11, n.6, p.1-16, 2016.

FELIPE, V.P.S.; SILVA, M.A.; VALENTE, B.D.; ROSA, G.J.M. Using multiple regression Bayesian networks and artificial neural networks for prediction of total egg production in European quails based on earlier expressed phenotypes. **Poultry Science**, v.94, n.4, p.772-780, 2015.

FEOFILOFF, P.; KOHAYAKAWA, Y.; WAKABAYASHI, Y. **Uma Introdução Sucinta à Teoria dos Grafos**. 2011, 61p. Disponível em: <http://www.ime.usp.br/pf/teoriadosgrafos/>. Acesso em 25 de agosto de 2018.

- FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. Bayesian network classifiers. **Machine Learning**, v.29, n.2-3, p.131-163, 1997.
- FRIEDMAN, N.; KOLLER, D. Being Bayesian about network structure: A Bayesian approach to structure discovery in Bayesian networks. **Machine Learning**, v.50, n.1, p.95-126, 2003.
- FRIEDMAN, N.; LINIAL, M.; NACHMAN, I.; PE'ER, D. Using Bayesian networks to analyze expression data. **Journal of Computational Biology**, v.7, n.3-4, p.601-620, 2000.
- GEIGER, D.; VERMA, T.S.; PEARL, J. Identifying independence in Bayesian networks. **Networks**, v.20, n.5, p.507-534, 1990.
- HECKERMAN, D.; GEIGER, D.; CHICKERING, D. M. Learning Bayesian networks: the combination of knowledge and statistical data. **Machine Learning**, v.20, n.3, p.197-243, 1995.
- HRYCEJ, T. Gibbs Sampling in Bayesian Networks. **Artificial Intelligence**, v.46, n.3, p.351-363, 1990.
- INRA. **Le Cheval. Reproduction, sélection, alimentation, exploitation.** Jarrige R. and Martin-Rosset W. Eds. INRA Editions, Paris, 1984. 689p.
- JANSEN, R.; YU, H.; GREENBAUM, D.; KLUGER, Y.; KROGAN, N.J.; CHUNG, S.; EMILI, A.; SNYDER, M.; GREENBLATT, J.F.; GERSTEIN, M. A Bayesian networks approach for predicting protein-protein interactions from genomic data. **Science**, v.302, n.5644, p.449-453, 2003.
- JENSEN, F. V. **An Introduction to Bayesian Networks.** New York: Springer Verlag, 1996.
- KOLLER, D.; FRIEDMAN, N. **Probabilistic Graphical Models: Principles and Techniques.** MIT Press, 2009, 1233p.
- KOVACICH, M. Application of Bayesian networks to multitarget tracking. **SPIE Proceedings**, v.1305, n.1, 1990. doi: 10.1117/12.2321776.
- LAURITZEN, S. L.; SPIEGELHALTER, D. J. Local computations with probabilities on graphical structures and their applications to expert

- systems. **Journal Royal Statistics Society**, v.50, n.1, p.157-224, 1988.
- LEESON, S.; SUMMERS, J.D. **Nutrition of the chicken**. 4.ed. Ontario: University Books, 2001. 413p.
- LUPO,C.; LE BOUQUIN, S.; BALAINE, L.; MICHEL, V.; PÉRASTE, J.; PETETIN, I.; COLIN, P.; JOUFFE, L.; CHAUVIN, C. Bayesian network as an aid for food chain information use for meat inspection. **Preventive Veterinary Medicine**, v.109, n.1, p.25-36, 2013.
- MARIANO, F.C.M.Q. et al. Prediction of the energy values of feedstuffs for broilers using meta-analysis and neural networks. **Animal**, v.7, n.9, p.1440-1445, 2013.
- MARIANO, F. C. M. Q.; LIMA, R. R.; RODRIGUES P. B.; ALVARENGA, R. R.; NASCIMENTO, G. A. J. Equações de predição de valores energéticos de alimentos obtidas utilizando meta-análise e componentes principais. **Ciência Rural**, v.42, n.9, p.1634-1640, 2012.
- MEHRI,M. A comparison of neural network models, fuzzy logic, and multiple linear regression for prediction of hatchability. **Poultry Science**, v.92, n.4, p.1138-1142, 2013.
- MEHRI, M. Development of artificial neural network models based on experimental data of response surface methodology to establish the nutritional requirements of digestible lysine, methionine, and threonine in broiler chicks. **Poultry Science**, v.91, n.12, p.3280-3285, 2012.
- NAGATA, T.; ALMEIDA, A.; RESENDE, R.; DE ÁVILA, A. The competence of four thrips species to transmit and replicate our Tospoviruses. **Plant Pathology**, v.53, n.1, p.136-140, 2004.
- NANDY, T.; HAUSER, A.; MAATHUIS, M.H. High-dimensional consistency in score-based and hybrid structure learning. **The Annals of Statistics**, v.46, n.6A, p. 3151-3183, 2018.
- NASH, D.; HANNAH, M. Using Monte-Carlo simulations and Bayesian networks to quantify and demonstrate the impact of fertiliser best management practices. **Elsevier, Environmental Modelling Software**, v.26, n.1, p.1079-1088, 2011.

NASCIMENTO, G.A.J.; RODRIGUES, P.B.; FREITAS, R.T.F.; REIS NETO, R.V.; LIMA, R.R.; ALLAMANN, I.B. Equações de predição para estimar valores da energia metabolizável de alimentos concentrados energéticos para aves utilizando meta-análise. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.63, n.1, p.222-230, 2011.

NASCIMENTO, G.A.J.; RODRIGUES, P.B.; FREITAS, R.T.F.; BERTECHINI, A.G.; LIMA, R.R.; PUCCI, L.E.A. Equações de predição para estimar os valores energéticos de alimentos concentrados de origem vegetal para aves utilizando a metanálise. **Revista Brasileira de Zootecnia**, v.38, n.1, p.1265-1271, 2009.

NATIONAL RESEARCH COUNCIL. **Nutrient Requirements of Poultry**. Ninth Revised Edition, Washington, DC: The National Academies Press.1994. 176p.

PEARL, J. **Probabilistic reasoning in intelligent systems: networks of plausible inference**. San Francisco: Editor Morgan Kaufmann, 1988. 552p.

PENNY, W.D.; ROBERTS, S.J. Bayesian neural networks for classification: how useful is the evidence framework?. **Neural Networks**, v.12, n.6, p.877-892, 1999.

PERAI, A. H.; MOGHADDAM, H. N.; ASADPOUR, S.; BAHRAMPOUT, J.; MANSOORI, Gh. A comparison of artificial neural networks with other statistical approaches for the prediction of true metabolizable energy of meat and bone meal. **Poultry Science**, v.89,n.1, p.1562-1568, 2010.

PRADO, J. W.; ALCÂNTARA, V.C.; CARVALHO, F.M.; VIEIRA, K. C.; MACHADO, L. K. C.; TONELLI, D. F. Multivariate analysis of credit risk and bankruptcy research data: a bibliometric study involving different knowledge fields (1968-2014). **Scientometrics**, v.106, n.3, p.1007-1029, 2016.

RODRIGUES, P. B.; ROSTAGNO, H. S.; ALBINO, L. F. T.; GOMES, P. C.; BARBOZA, W. A.; SANTANA, R. T. Valores energéticos do milheto, do milho e subprodutos do milho, determinados com frangos de corte e galos adultos. **Revista Brasileira de Zootecnia**, v.30, n.6, p.1767-1778, 2001.

RODRIGUES, P. B.; ROSTAGNO, H. S.; ALBINO, L. F. T.; GOMES, P.

C.; NUNES, R. V.; TOLEDO, R. S. Valores energéticos da soja e subprodutos da soja, determinados com frangos de corte e galos adultos. **Revista Brasileira de Zootecnia**, v.31, p.1771-1782, 2002.

ROSTAGNO, H. S. et al. **Tabelas brasileiras para aves e suínos: composição de alimentos e exigências nutricionais de aves e suínos**. 4ª edição, Viçosa, MG: UFV, 488 p., 2017.

SAVEGNAGO, R. P.; NUNES, B. N.; CAETANO, S. L.; FERRAUDO, A. S.; SCHIMIDT, G. S.; LEDUR, M. C.; MUNARI, D. P. Comparison of logistic and neural network models to fit the egg production curve of White Leghorn hens. **Poultry Science**, v.90, p.705-711, 2011.

SCOTT, M.L.; NESHEIN, M.C.; YOUNG, R.J. **Nutrition of the chicken**. 3.ed. New York: Scott and Assoc. Publ., 1982. 562p.

SCUTARI, M., DENIS, J.B. **Bayesian Networks with Examples in R**. Taylor Francis Group. 2015. 221p.

SEMSARIAN, S.; NASAB, M. P. E.; ZAREHDARAN, S.; DEHGHANI, A. A. Prediction of the weight and number of eggs in Mazandaran native fowl using artificial neural network. **International Journal of Advanced Biological and Biomedical Research**, v.5, n.1, p.532-537, 2013.

SOLTANI, M.; OMID, M. Detection of poultry egg freshness by dielectric spectroscopy and machine learning techniques LWT. **Food Science and Technology**, v.62, n.1, p.1034-1042, 2015.

SPIRITES, P.; GLYMOUR, C.; SCHEINES, R. **Causation, Prediction, and Search**. Second edition, Cambridge: MIT Press, 2000.

TRUONG, D.; FRIEND, M.A.; CHEN, H. Applications of business analytics in predicting flight on-time performance in a complex and dynamic system. **Transportation Journal**, v.57, n.1, p.24-52, 2018.

TSAMARDINOS, I.; BROWN, L.E.; ALIFERIS, C.F. The max-min hill-climbing Bayesian network structure learning algorithm. **Machine Learning**, v.65, n.1, p.31-78, 2006.

TSAMARDINOS, I.; BROWN, L.E.; ALIFERIS, C.F. Algorithms for large scale markov blanket discovery. In: "Proceedings of the Sixteenth

International Florida Artificial Intelligence Research Society Conference".
Proceeding... 2003, p.376-381. AAAI Press. 2003.

UUSITALO, L. Advantages and challenges of Bayesian networks in environmental modeling. **Ecological Modelling**, v.203, n.3-4, p.312-318, 2007.

WAN, X.; O'QUINN, R.P.; PIERCE, H.L.; JOGLEKAR, A.P.; GALL, W.E.; DELUCA, J.G.; CARROLL, C.W.; LIU, S.T.; YEN, T.J.; MCEWEN, B.F.; STUKENBERG, P.T.; DESAI, A.; SALMON, E.D. Protein architecture of the human kinetochore microtubule attachment site. **Cell Press**, v.137, n.1, p.672-684, 2009.

YU, J.; SMITH, V.A.; WANG, P.P.; HARTEMINK, A.J.; JARVIS, E.D. Advances to bayesian network inference for generating causal networks from observational biological data. **Bioinformatics**, v.20, n.1, p.3594-3603, 2004.

ZHAO, F.; ZHANG, H. F.; HOU, S. S.; ZHANG, Z. Y. Predicting metabolizable energy of normal corn from its chemical composition in adult pekin ducks. **Poultry Science**, v.87, n.1, p.1603-1608, 2008.

ARTIGO 2: Redes Bayesianas na predição de valores energéticos de alimentos para aves

ALVARENGA, T.C.; LIMA, R.R.; BRANDAO JUNIOR, L. C.; BUENO FILHO, J. S. S.; MARIANO, F. C. M. Q.; RODRIGUES, P. B.; ALVARENGA, R. R.

Artigo a ser submetido na Revista *Animal (Cambridge)*

Resumo

Elaboração de rações balanceadas consistem no conhecimento preciso da energia metabolizável aparente corrigida pelo balanço de nitrogênio (EMAn), em especial da composição química dos alimentos. Equações de predição estabelecidas em função da composição química dos alimentos foram propostas por vários pesquisadores. Na literatura existem trabalhos que obtiveram as equações de predição por meio de regressão múltipla, meta-análise e redes neurais. Porém, outras metodologias estatísticas cujo potencial seja promissor podem ser utilizadas na obtenção de melhores estimativas dos valores energéticos. O objetivo do presente estudo foi propor equações de predição para estimar os valores de EMAn de alimentos energéticos e proteicos, de origem vegetal, utilizados na formulação de rações de frangos de corte a partir de redes bayesianas. Redes bayesianas (RB) são modelos gráficos os quais consistem na representação gráfica (grafo) e probabilística (distribuições de probabilidade condicionais e conjunta) das variáveis, sendo esta uma metodologia promissora quando o objetivo é predição, de modo a empregar o conhecimento probabilístico no campo da Inteligência Artificial. Utilizou-se de várias aprendizagens da rede bayesiana, considerando diferentes cenários, com a finalidade de selecionar o modelo de melhor predição. O pacote *bnlearn* do software R foi utilizado para prever a EMAn a partir das covariáveis: proteína bruta, fibra bruta, extrato etéreo, matéria mineral, além da categoria de alimentos, ou seja, energéticos (milho, subprodutos de milho e outros) e proteicos (soja, subprodutos de soja e outros) e o tipo de animal (pinto ou galo). Os dados são todos de experimentos realizados no Brasil, no total de 568 experimentos. Além desses, foram utilizados dados de ensaios metabólicos conduzidos no Setor de Avicultura do Departamento de Zootecnia da Universidade Federal de Lavras (UFLA), em Lavras, Minas Gerais, Brasil. O modelo que apresentou maior acurácia foi obtido com a aprendizagem do algoritmo *Max-Min Hill Climbing* (MMHC) e o conjunto de treinamento e teste, respectivamente em 80% e 20% dos dados. As equações propostas foram validadas com os dados de ensaios metabólicos e obtiveram resultados precisos e acurados. Conclui-se que as equações propostas por RB, sendo

a mesma inédita na Nutrição de aves, podem ser utilizadas pela indústria de frangos de corte na elaboração de rações. Estas equações estão disponíveis em calculadoras no sistema Android e Java para o cálculo dos valores energéticos de alimentos para frangos de corte.

Palavras-chave: Algoritmo *mmhc*. Frangos de corte. Energia metabolizável. Nutrição de aves.

Abstract

The making of balanced diets consists of the precise knowledge of the apparent metabolizable energy corrected by the nitrogen balance (EMAn), in especial of the chemical composition of the feedstuffs. Prediction equations established according to the chemical composition of feedstuffs were proposed by several researchers. In the literature there are studies that obtained the prediction equations through multiple regression, meta-analysis and neural networks. However, other statistical methodologies with promising potential can be used to obtain better estimates of energy values. The objective of the present study was to propose prediction equations to estimate the AMEn values of energy and protein feeds of plant origin, used in the formulation of poultry diets by means of Bayesian networks. Bayesian networks (BN) are graphical models which consist of graphical (graph) and probabilistic representation (conditional and joint probability distributions) of the variables, being this a promising methodology when the objective is prediction, so as to employ probabilistic knowledge in the field of Artificial Intelligence. A great deal of learning of the Bayesian network was used, considering different scenarios with the purpose of selecting the best prediction model. The *bnlearn* package of R software was used to predict EMAn from covariates: crude protein, crude fiber, ether extract, mineral matter, as well as feedstuff category, ie energy-yielding (corn, corn byproducts and others) and protein (soybean meal, soybean byproducts and others) and the type of animal (chick or rooster). The data are all of experiments carried out in Brazil, in a total of 568 experiments. In addition to these, data from metabolic tests conducted in the Poultry production Sector of the Department of Animal Science of the Federal University of Lavras (UFLA), in Lavras, Minas Gerais, Brazil, were used. The model that presented the highest accuracy was obtained with the learning of the Max-Min Hill Climbing algorithm (MMHC) and the training and test set, respectively, in 80% and 20% of the data. The proposed equations were validated with data from metabolic assays and obtained accurate

and precise results. It follows that the equations proposed by RB, they being unprecedented in poultry production, can be used by the broiler industry in the making of diets. These equations are available in calculators in the Android and Java system for the calculation of the energy values of feedstuffs for broilers.

Keywords: Algorithm *MMHC*. Broilers. Metabolizable energy. Nutrition of poultry.

Implicações

Executar ensaios biológicos para determinar valores de EMAn são demorados e de custo elevado, além de que, existe diversidade nos valores de EMAn fornecidos pelas tabelas de composição de alimentos. No entanto, têm-se utilizado metodologias de *Machine Learning* na estimação da EMAn, tal como as redes neurais, o que pode ser feito de maneira rápida, econômica e eficiente. Neste empregou-se as redes bayesianas, uma metodologia inovadora para área de nutrição animal. Redes bayesianas empregam incertezas na modelagem, além de capturar dependências condicionais entre as variáveis envolvidas, o que não é possível com as metodologias tradicionais.

1 INTRODUÇÃO

A indústria ligada à produção de frangos de corte ocupa uma posição de destaque no cenário econômico nacional e internacional (UNITED STATES DEPARTMENT OF AGRICULTURE, 2017). Conhecer a composição química e os valores de energia metabolizável aparente corrigida pelo balanço de nitrogênio (EMAn) dos alimentos que compõem as rações, possibilita uma formulação adequada. A EMAn pode ser analisada utilizando ensaios metabólicos, porém, estes apresentam altos custos e tempo elevado. Uma das soluções para determinar a EMAn dos alimentos é o uso das equações de predição em função da composição química dos alimentos. Trabalhos como de Rodrigues et al. (2001, 2002); Borges et al. (2003); Zhao et al. (2008), obtiveram as equações de predição por meio da regressão múltipla sem nenhuma restrição. Porém, Alvarenga et al. (2011) mostraram que os resultados não eram os mais adequados. Estudos com o objetivo de encontrar as equações de predição para a energia metabolizável corrigida pelo balanço de nitrogênio (EMAn) de forma coerente foram realizados por Nascimento et al. (2009, 2011) e Mariano et al. (2012), por meio da meta-análise. Em Perai et al. (2010), Ahmadi et al. (2007, 2008) e Mariano et al. (2013) foram utilizadas

redes neurais, sendo que em Mariano et al. (2013), foram utilizados maior número de alimentos e dados observados.

De acordo com Bishop (2006), as redes neurais (RN) são indicadas para estudos cujo objetivo é predição. Assim, Mariano et al. (2013) utilizaram RN na predição de valores energéticos de alimentos para aves. Felipe et al. (2015) mencionam a utilização de RN em Zootecnia e apontam o uso de redes bayesianas também. Porém, o modelo preditivo via redes bayesianas (RB) na Zootecnia estão concentrados no melhoramento animal, mais especificamente em trabalhos que abordam a seleção genômica (MOROTA et al., 2013; TREPODE et al., 2013; GIANOLA et al., 2011; FRIEDMAN et al., 2000; FELIPE et al., 2015), ou seja, na nutrição animal ainda é novidade aplicações dessa metodologia. Logo, RN e RB têm sido utilizadas na predição, devido superioridade em capturar (e expressar) relações mais complexas entre as covariáveis e a variável resposta (GIANOLA et al., 2011 citado por Felipe et al., 2015). RB pode ser utilizada na construção de um modelo preditivo pela fatoração da distribuição de probabilidade conjunta, assumindo um conjunto de independências condicionais (JENSEN,1996; LAURITZEN e SPIEGELHATER, 1988). As aplicações de RB tem se estendido em diferentes áreas do conhecimento, tais como: Medicina, Genética, Economia, robótica, análise de crime, Demografia, conservação de espécie, mapeamento do potencial mineral, Educação, indústria, comportamento humano, entre outros (POURRET et al., 2008).

Carvalho e Chiann (2013), em busca de uma maneira distinta de abordagem ao tema, utilizaram redes bayesianas para modelar a estrutura de interdependência entre mercados financeiros. A partir da metodologia empregada com respaldo da teoria de inferência bayesiana, os resultados encontrados indicaram evidências de contágio em períodos de crise financeira. Simões et al. (2015) utilizaram as redes bayesianas no apoio ao diagnóstico de câncer de mama, juntamente com a meta-análise. Por fim, chegaram-se a conclusão que a metodologia empregada forneceu resultados acurados e não invasivos no apoio ao diagnóstico de câncer de mama. Vieira et al. (2017) analisaram o risco ocupacional por meio das redes bayesianas. Para tal análise, realizaram uma revisão sistemática de maneira semelhante ao protocolo *Statement for Reporting Systematic Reviews* (PRISMA), ou seja, uma meta-análise. Em conclusão ao trabalho, Vieira et al. (2017) relataram a importância das redes bayesianas nas análises de riscos ocupacionais, justamente pelo fato do risco ser uma variável aleatória e, este raciocínio se aplica na modelagem das redes bayesianas, ou seja, simula processos estocásticos. Felipe et al. (2015) compararam diferentes metodologias para prever a produção total de ovos em codornas de linhagens diferentes. As variáveis analisadas foram: peso, produção parcial de ovos e medidas de qualidade. As metodologias emprega-

das contemplaram os modelos regressão linear múltipla e rede neural artificial (RNA), além da rede bayesiana para a seleção das variáveis. O modelo com a combinação de rede bayesiana (para a seleção das variáveis) e RNA resultou em melhor desempenho para prever a produção total de ovos. Töpner et al. (2017) utilizaram as redes bayesianas em um experimento de milho para análise das relações entre as características a níveis genômico e residual. As redes bayesianas obtidas neste, foram classificadas em termos de capacidade de ajuste e habilidade preditiva por meio das equações estruturais. Concluíram que ao ilustrar as conexões de características em relação à sua natureza genômica e residual, as mesmas se tornam mais claras, o que torna útil para predição dos múltiplos traços e seleção indireta.

Em síntese, redes bayesianas são modelos gráficos (*graphical models*), os quais consistem na representação gráfica (grafo) e probabilística (distribuições de probabilidade condicionais e conjunta) das variáveis. As RB constituem uma representação gráfica de variáveis e suas relações causais, nomeadas respectivamente por nós ou vértices; e arestas, arcos ou setas (PEARL, 1988; SPIRITES et al., 2000; KOLLER e FRIEDMAN, 2009; SCUTARI e DENIS, 2015). A fatoração da RB consiste no produto das probabilidades condicionais, visto que um nó é condicionalmente independente de seus não-descendentes dado seus pais (PEARL, 1988; TSAMARDINOS et al., 2003). Essa fatoração resulta na distribuição de probabilidade conjunta, sendo assim, possível inferir e obter respostas a questões de interesse. As relações de independências condicionais entre as variáveis são verificadas pelo critério de *d-separação*. A distribuição de probabilidade conjunta é dada de acordo com a natureza das variáveis. Para o caso discreto, contínuo e ambos os casos, as redes bayesianas são nomeadas, respectivamente, por rede bayesiana multinomial, rede bayesiana gaussiana e rede bayesiana híbrida (SCUTARI e DENIS, 2015).

O aprendizado do modelo é conhecido no campo da RB como *learning*, nome emprestado da inteligência artificial. A princípio ocorre a aprendizagem da estrutura, ou seja, *structure learning* (SCUTARI, 2010), em seguida é *parameter learning*, que significa aprender sobre as distribuições locais implícitas da estrutura (SCUTARI et al., 2014). Nagarajan et al. (2013) abordam os algoritmos para a aprendizagem da rede: *constraint-based*, *score-based* e *hybrid*. Os algoritmos de *constraint-based* consistem em vários testes de independência condicional e marginal para inferir as direções dos arcos entre os nós. Já os algoritmos *score-based* selecionam entre todas as possíveis estruturas da BN a com melhor qualidade, ou seja, com maior pontuação. Os algoritmos de *constraint-based* implementados no *bnlearn* (SCUTARI, 2010) do Software R (R CORE TEAM, 2018) são: *Grow-Shrink* (GS) (MARGARITIS e THRUN, 2000); *Incremental Association Markov Blanket* (IAMB)

(TSAMARDINOS et al., 2003); *Fast Incremental Association* (Fast-IAMB) (YAMARAKALA e MARGARITIS, 2005); *Interleaved Incremental Association* (Inter-IAMB); *Max-Min Parents e Children* (MMPC); *Semi-Interleaved Hiton-PC* (SI-HITON-PC). Cada um destes algoritmos podem ser utilizados com vários testes de independência condicional.

Já os algoritmos de estruturas de aprendizagem *score-based* são: *Hill Climbing* (HC) (MARGARITIS, 2003) e *Tabu Search* (Tabu). Para os algoritmos *score-based* podem ser usados nas funções de pontuação: o critério de informação Akaike (AIC), o critério de informação Bayesiano (BIC), entre outros. Os algoritmos de aprendizagem *hybrid* são: *Max-Min Hill Climbing* (MMHC) (TSAMARDINOS et al., 2006) e *General 2-Phase Restricted Maximization* (RSMAX2). O MMHC utiliza o algoritmo *constraint-based Max-Min Parents and Child* (MMPC) para a busca do esqueleto do grafo, em que se faz a estimação dos pais e filhos da cobertura de *Markov* para cada variável envolvida na RB. Logo em seguida para determinar as direções dos arcos no grafo, utiliza-se o algoritmo *score-based Hill Climbing* (HC). O algoritmo *RSMAX2* é uma implementação mais geral do MMHC, em que se pode utilizar qualquer combinação dos algoritmos *constraint-based* com o *score-based* (SCUTARI e DENIS, 2015).

O objetivo deste estudo foi propor equações de predição para a EMAn de alimentos energéticos e proteicos utilizados em rações de frangos de corte, considerando redes bayesianas nos estudos de meta-análise. As equações obtidas por meio deste estudo foram implementadas em calculadoras nos sistemas: *Android* e *Java*, com finalidade de dar apoio aos nutricionistas na indústria de alimentos para frangos de corte.

2 MATERIAL E MÉTODOS

Os dados utilizados neste trabalho foram catalogados por Nascimento et al. (2009, 2011) e compreendem valores de EMAn e composição química de alimentos energéticos e proteicos, de origem vegetal, geralmente utilizados na formulação de dietas de frangos de corte. Os dados da meta-análise são referentes a experimentos realizados no Brasil, no período de 1967 a 2007, resultando em 568 experimentos, dentre os quais 370 são de concentrados energéticos e 198 de concentrados proteicos (Tabela 1).

Os valores da variável resposta - EMAn, foram estimados utilizando as covariáveis: proteína bruta (PB), extrato etéreo (EE), matéria mineral (MM), fibra bruta (FB), categoria de alimentos (CAT) (1 - concentrado energético, 2 - concentrado proteico), ingrediente na categoria (ING) 1 - concentrado energético: 1 - milho, 2 - subprodutos de milho, 3 - outros, ingrediente na categoria 2 (ING) - concentrado proteico: 1 - farelo de soja, 2 - subprodutos

Tabela 1 – Número de experimentos considerados neste estudo.

Concentrados ¹	Alimentos	Número de experimentos	Total
Energéticos	Milho	168	370
	Subprodutos de milho	29	
	Outros	173	
Proteicos	Soja	59	198
	Subprodutos da soja	59	
	Outros	80	

¹Concentrados de origem vegetal.

Fonte: Mariano et al. (2013).

de soja, 3 - outros e o tipo de animal (ANIMAL) utilizado no bioensaio: 1 - pinto, 2 - galo, conforme apresentados na Tabela 2.

Tabela 2 – Dados amostrais de meta análise usados na rede bayesiana.

N	PB	EE	MM	FB	CAT	ING	ANIMAL	EMAn
1	10,92	4,03	1,81	1,92	1	1	1	3573
2	10,75	3,53	1,82	2,00	1	1	1	3560
.
.
302	11,805	2,40	3,20	6,50	1	2	1	2700
303	1,76	0,063	3,01	17,67	1	3	2	2340
.
.
568	32,02	10,4	4,95	2,50	2	3	2	2310

Fonte: Mariano et al. (2013).

As estatísticas descritivas dos conjuntos dos dados utilizados para aprendizagem e teste do modelo proposto por meio das redes bayesianas, encontram-se na Tabela 3.

Tabela 3 – Estatísticas descritivas dos conjuntos de dados do treinamento (80%) e de teste (20%).

	Estatísticas	EMAn	PB	EE	MM	FB
Treinamento (80%)						
	Mínimo	1170	1,470	0,030	0,30	0,020
	Mediana	3501	14,130	3,480	2,11	3,020
	Média	3176	23,360	4,872	3,56	4,928
	Máximo	4386	71,440	26,210	12,61	26,500
Teste (20%)						
	Mínimo	1148	1,700	0,030	0,560	0,320
	Mediana	3275	15,270	3,150	3,010	3,985
	Média	3050	23,360	4,135	3,826	5,683
	Máximo	4160	68,810	25,540	11,050	27,630

Os dados utilizados na validação das equações propostas são provenientes de dois ensaios metabólicos para determinação do valor energético de alimentos proteicos (ensaio I) e energéticos (ensaio II), com pintos em crescimento (método tradicional de coleta total de excretas), respectivamente em fevereiro/março e julho de 2008. Os experimentos foram conduzidos no Setor de Avicultura do Departamento de Zootecnia da Universidade Federal de Lavras (UFLA), em Lavras, Minas Gerais, Brasil (ALVARENGA et al., 2011). Algumas estatísticas descritivas do conjunto dos dados utilizados na validação das equações obtidas por meio das redes bayesianas, encontram-se nas Tabelas 4 e 5.

Tabela 4 – Estatísticas descritivas dos ensaios metabólicos

Estatísticas	EMAn	PB	EE	MM	FB
Mínimo	1941	8,790	0,750	0,980	0,380
Mediana	3400	35,290	4,100	1,805	5,090
Média	3198	30,816	7,735	3,768	5,100
Máximo	4022	68,700	25,850	10,430	11,590

Fonte: Alvarenga et al. (2011).

Modelos de predição

Pode-se observar na Figura 1, a ilustração da estrutura gráfica de uma RB, denotadas por DAG (*directed acyclic graph*). As DAGs apresentam direção nas arestas, todos os nós são conectados na estrutura gráfica e as direções das arestas não apresentam ciclos, ou seja, após a aresta seguir direção, ela não retorna ao mesmo nó, classificando-o respectivamente por grafo direcionado, conectado e acíclico.

Tabela 5 – Estatísticas descritivas dos diferentes tipos de alimentos utilizados nos ensaios metabólicos

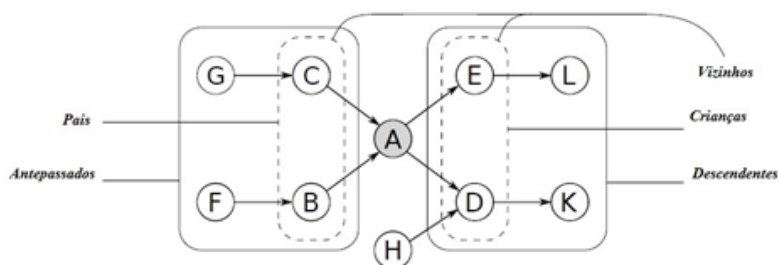
Concentrados ¹	Alimentos	Número de amostras	Média	Desvio	Total
Energéticos	Milho	7	3649	103	22
	Subprodutos de milho	6	2672	758	
	Outros	9	3329	577	
Proteicos	Soja	5	2385	57	26
	Subprodutos da soja	16	3212	612	
	Outros	5	3910	122	

¹Concentrados de origem vegetal.

Fonte: Alvarenga et al. (2011).

De acordo com o grafo da Figura 1, e tomando o nó A como referência, pode-se observar que seus vizinhos (C,B,D,E) são obtidos pela união de seus pais (C,B) e filhos ou crianças (E,D), chamados também de nós adjacentes (nós pais e filhos). Seus pais também são chamados de antecessores, pois estão ordenados anterior ao A. Os filhos de A são chamados de descendentes, devido a ordem posterior. A ordem das variáveis no grafo chama-se topologia e assim, de acordo com Nagarajan, Scutari e Lébrea (2013), um exemplo da ordem topológica da estrutura gráfica apresentada na Figura 1 pode ser dada por: (F,G,H), (C,B), (A), (D,E) e (L,K).

Figura 1 – Exemplo ilustrativo DAG



Fonte: Adaptado de Nagarajan; Scutari e Lébrea, 2013.

Muitos dos algoritmos utilizados para encontrar a estrutura gráfica das variáveis, requerem ordenação entre os nós. A justificativa da mesma, se deve ao fato de variáveis causais, estarem precedidas de outras variáveis (nós). Alguns desses algoritmos utilizam a Cobertura de *Markov* (*Markov Blanket*) para modelar o fenômeno em estudo. Denota-se Cobertura de *Markov*, o conjunto de nós que separa o nó alvo dos demais nós da estrutura, ou seja, essa cobertura inclui pais, filhos e nós que compartilham os mesmos filhos do nó alvo. A Cobertura de Markov do nó A (Figura 1) são os nós: C, B, H,

E e D. Em estudos cujo objetivo é predição, apenas as variáveis encontradas nessa cobertura são relevantes ao modelo (TSAMARDINOS et al., 2003).

Uma RB é uma representação gráfica de uma distribuição de probabilidade conjunta (HECKERMAN, 1995; MARGARITIS, 2003), a qual descreve o domínio por completo. Ela pode ser definida pela estrutura da DAG, o conjunto de parâmetros específicos advindos de distribuições de probabilidades condicionais, envolvendo o conjunto das X variáveis aleatórias. A fatoração da RB, conforme apresentado na Equação 1, consiste no produto das probabilidades condicionais, visto que um nó é condicionalmente independente de seus não-descendentes dado seus pais (PEARL, 1988; TSAMARDINOS et al., 2003). Essa fatoração resulta na distribuição de probabilidade conjunta, sendo assim, possível inferir e obter respostas a questões de interesse.

$$P(X_1, X_2, \dots, X_p) = \prod_{i=1}^n P(X_i | Pa_i) \quad (1)$$

para o caso de nós discretos e contínuos, em que Pa_i são os pais de X_i .

O pacote *bnlearn* (SCUTARI, 2010) do software R (R CORE TEAM, 2018) foi utilizado para implementação das redes bayesianas. Devido à superioridade computacional do algoritmo híbrido MMHC na detecção de relações entre variáveis, o mesmo foi utilizado, além do algoritmo híbrido *RSMAX2*. O algoritmo híbrido geralmente utiliza o *constraint-based* para encontrar o esqueleto rede bayesiana e, posteriormente, o *score-based* é usado para estabelecer relações de causa e efeito na rede (SCUTARI e DENIS, 2015).

Os valores da EMAN foram preditos a partir das equações de predição obtidas utilizando uma rede bayesiana híbrida. A rede bayesiana híbrida, assim chamada por contemplar tanto variáveis contínuas quanto discretas, se assemelha a um modelo de regressão linear múltipla (KOSKI; NOBLE, 2009). Para tornar claro o conceito apresentado, considere um conjunto de variáveis aleatórias, sendo esse conjunto particionado em variáveis discretas e contínuas, respectivamente X_D e X_C . Assim, a distribuição de probabilidade conjunta $P(X)$ pode ser fatorada da forma apresentada em (2).

$$P(X) = P(X_D, X_C) = \prod_{i \in D} P(X_i | Pa_D) \prod_{j \in C} P(X_j | Pa_C, Pa_D) \quad (2)$$

em que Pa_D e Pa_C são, respectivamente, os conjuntos de variáveis pais discretos e contínuos.

Para o $\prod_{j \in C} P(X_j | Pa_C, Pa_D)$, a qual contempla as variáveis pais discretas e contínuas, as distribuições locais são regressões lineares com parâmetros dependentes dos pais discretos, isso significa dizer que:

$$(X_j | Pa_C, Pa_D) \sim N(\mu_j, \sigma_{X_j|Pa_D}^2) \quad (3)$$

em que,

$$\mu_j = \beta_{0, X_j|Pa_D} + \beta_{i, X_j|Pa_D} X_{i, Pa_C}. \quad (4)$$

Para maiores detalhes consultar Lauritzen e Wermuth (1989). Assim, no contexto da predição da EMAn, μ_j refere-se ao valor da EMAn em cada categoria, alimento e animal; $\beta_{0, X_j|Pa_D}$ é média da variável X_j na configuração de seus pais discretos, ou seja, o intercepto da regressão linear multivariada; $\beta_{i, X_j|Pa_D}$ são os coeficientes da regressão, ou seja, a contribuição de cada pai contínuo em cada nível da variável discreta; X_{i, Pa_C} representam as variáveis de composição química dos alimentos, sendo elas neste trabalho, PB, MM, EE e FB.

Neste trabalho são propostos modelos de redes bayesianas híbridas, que contemplem inicialmente as variáveis de composição química (PB, MM, EE e FB), além das categorias dos alimentos (energéticos e proteicos), ingredientes (milho, soja, seus subprodutos e outros) e animais (pinto, galo) como variáveis de entrada na rede. A energia metabolizável aparente corrigida pelo balanço de nitrogênio será considerada como variável de saída no modelo. Porém, para elucidar a estrutura de relações entre as variáveis de composição química e de valores energéticos, de cada alimento, serão realizadas implementações nas RB, tal como no processo de seleção das diferentes estruturas obtidas por meio dos algoritmos.

Os dados advindos de estudos de meta-análise foram divididos aleatoriamente, consistindo em um conjunto de treinamento (70%, 75% e 80% dos dados) e um conjunto de teste do modelo (30%, 25% e 20% dos dados restantes). As equações propostas para obtenção da EMAn foram avaliadas usando os dados de validação (ALVARENGA et al., 2011), que consistem em dados provenientes de ensaios metabólicos. As estatísticas utilizadas no teste e validação das equações foram a correlação (r), o coeficiente de determinação (R^2), erro quadrático médio (m.s.e), desvio médio absoluto (MAD), erro médio de porcentagem absoluta (MAPE), viés (bias) (MARIANO et al., 2014) e erro quadrático médio de predição (PMSE) (FELIPE et al., 2015).

3 RESULTADOS E DISCUSSÃO

Foram avaliadas diferentes redes bayesianas, obtidas de diferentes aleatorizações nos conjuntos de treinamento (80%, 75% e 70% dos dados advindos de estudos da meta-análise) e para teste da rede (20%, 25% e 30% dos dados restantes advindos de estudos da meta-análise). Utilizou-se os algoritmos híbridos para aprender as estruturas gráficas (DAGs). Os modelos obtidos via

redes bayesianas foram avaliados por meio de comparações entre os valores observados (conjunto de teste) e preditos pelo modelo. As estatísticas utilizadas foram correlação (r), coeficiente de determinação (R^2), erro quadrático médio (m.s.e), desvio médio absoluto (MAD), erro médio percentual absoluto (MAPE), bias (VIÉS) e erro quadrático médio de predição (PMSE). Modelos mais acurados, são esperados valor de R^2 próximo de 1, e bias e os erros próximos de zero. O modelo de rede bayesiana que apresentou melhor acurácia e precisão na predição dos valores energéticos foi selecionado. Os valores das estatísticas referentes a esse modelo estão na Tabela 6. Além das estatísticas de avaliação mencionadas, apresenta-se também a correlação entre o observado e predito, nos conjuntos de treinamento e teste, respectivamente expressos por, $r_{treinamento}$ e r_{teste} (Tabela 6).

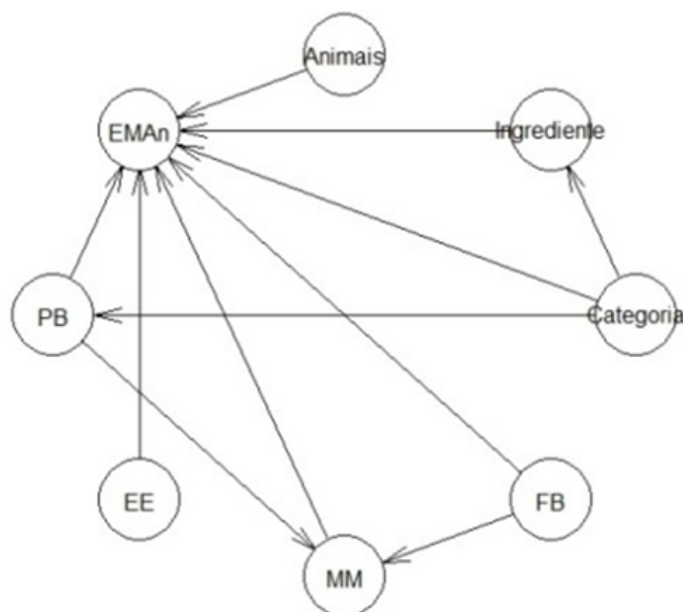
Tabela 6 – Melhor ajuste dos cenários estudados.

Algoritmo	$r_{treinamento}$	r_{teste}	R^2	m.s.e	MAD	MAPE	Bias	PMSE
<i>mmhc</i>	0,92	0,94	0,87	66529,8	191,2	7,52	-43,09	257,93

O melhor resultado obtido foi por meio do algoritmo de aprendizagem MMHC com a aleatorização de 80% dos dados para a aprendizagem e os 20% dos dados restantes para teste. A ilustração gráfica do modelo de rede bayesiana com maior acurácia encontra-se na Figura 2. Este modelo apresentou maior acurácia que os modelos obtidos por Nascimento et al. (2009), Mariano et al. (2012) e Mariano et al. (2013). Esses autores utilizaram regressões tais como, a meta-análise, componentes principais e redes neurais na predição dos valores energéticos de alimentos para frangos de corte, sendo esses valores de R^2 respectivamente aproximados a 0,83, 0,74 e 0,86. Redes bayesianas é uma metodologia inovadora para área de nutrição animal, sendo modelos que permitem a representação de incertezas na modelagem, além de capturar dependências condicionais entre as variáveis envolvidas, o que não é possível com as metodologias tradicionais. O valor do m.s.e do melhor modelo deste trabalho foi menor do que encontrado em Mariano et al. (2013) que contemplou tanto alimentos energéticos quanto proteicos na modelagem das redes neurais. Perai et al. (2010) utilizaram apenas alimentos proteicos na rede neural para predizer os valores energéticos e foram encontrados valores de m.s.e entre o intervalo 2338.1875 e 91335.17, sendo que os valores de m.s.e obtidos empregando as redes bayesianas estão neste intervalo.

Conforme a RB consiste na representação gráfica e probabilística, assim, a distribuição global obtida por meio da estrutura gráfica apresentada na Figura 2 é dada por: $P(\text{EMAn,PB,EE,MM,FB,Categoria,Ingrediente,Animais}) = P(\text{EE})P(\text{FB})P(\text{Categoria})P(\text{Animais})P(\text{PB|Categoria})P(\text{Ingrediente|Categoria})P(\text{MM|PB:FB})P(\text{EMAn|PB:EE:MM:FB:Categoria:Ingrediente:Animais})$, toda

Figura 2 – DAG das variáveis em estudo.



a modelagem é por meio de distribuições de probabilidade, por isso, a representação $P(\cdot)$. Isso significa que em termos de probabilidade EE, FB, Categoria, Animais não estão tendo influência de outras variáveis. No entanto, PB é influenciada pela Categoria de alimentos, além de que, a MM é influenciada pela interação entre PB e FB. Já a variável $EMAn$ é influenciada pela interação entre todas as variáveis em estudo, essa interação entra no modelo de redes bayesianas como um efeito aditivo. As variáveis categóricas assumem distribuição multinomial e as variáveis contínuas assumem distribuição condicional gaussiana, em que a média da distribuição converge para regressões multivariadas. Já que todas as variáveis estão influenciando a $EMAn$, assim, as regressões multivariadas para a variável resposta $EMAn$ são advindas de cada combinação das variáveis categóricas, ou seja, 12 regressões (2 categorias de concentrados sendo cada uma delas com 3 sub-níveis de ingredientes, além de 2 tipos de animais envolvidos no experimento = $2 \times 3 \times 2 = 12$ regressões). As mesmas estão em função das variáveis contínuas, ou seja, as composições químicas dos alimentos (PB, FB, MM, EE). Como o objetivo foi propor equações para prever $EMAn$, destaca-se apenas a distribuição local de interesse que assume distribuição condicional gaussiana $P(EMAn|PB:EE:MM:FB:Categoria:Ingrediente:Animais)$.

A estrutura gráfica da rede bayesiana (Figura 2) utilizaram-se 8 nós no total, 11 arcos direcionados e o tamanho da cadeia de *Markov* foi de 7 nós.

O algoritmo de aprendizagem selecionado foi o *Max-Min Hill-Climbing*, ou seja, o método *constraint-based Max-Min Parent Children* com o teste de independência condicional, a informação mútua. O método *score-based foi o Hill-Climbing*, empregando o BIC como critério de seleção. No total foram realizados 165 testes no processo de aprendizagem otimizado.

As equações propostas por meio de redes bayesianas seguem apresentadas na Tabela 7, por $EMAn_{i,j,k}$. Têm-se que o índice i refere-se a categoria, se 1-energético ou 2-proteico. O índice j refere-se a classificação do alimento, se $i=1$ então j : 1-milho ou 2-subprodutos de milho ou 3-outros subprodutos de milho; se $i=2$ então j é 1-soja ou 2-subprodutos de soja ou 3-outros subprodutos de soja. Para k , os valor de 1 refere-se a pinto e 2 ao galo.

Tabela 7 – Equações obtidas por meio de redes bayesianas.

Equações propostas
$EMAn_{1,1,1} = 3658,16 - 2,41PB - 11,25EE + 83,41MM + 16,76FB$
$EMAn_{2,1,1} = 3684,83 - 19,84PB - 71,15EE + 18,14MM - 8,93FB$
$EMAn_{1,2,1} = 4209,57 - 34,56PB + 32,84EE - 25,15MM - 142,57FB$
$EMAn_{2,2,1} = 2951,05 + 0,09PB + 37,96EE + 5,04MM - 17,60FB$
$EMAn_{1,3,1} = 4335,88 - 50,91PB + 35,40EE - 67,35MM - 87,06FB$
$EMAn_{2,3,1} = 2327,69 + 24,23PB + 77,72EE - 167,06MM - 22,28FB$
$EMAn_{1,1,2} = 3321,82 + 51,31PB + 39,42EE - 377,11MM + 113,92FB$
$EMAn_{2,1,2} = 4143,45 - 3,18PB - 43,45EE - 213,55MM + 6,71FB$
$EMAn_{1,2,2} = 4716,45 - 227,63PB + 144,47EE$
$EMAn_{2,2,2} = 518,54 + 26,25PB + 47,10EE + 184,42MM + 69FB$
$EMAn_{1,3,2} = 4133,39 - 89,45PB + 100,32EE - 5,5MM - 96,37FB$
$EMAn_{2,3,2} = 6033,28 - 15,02PB - 105,81EE - 556,50MM + 91,23F$

Os valores de EMAN observados (dados de teste) e preditos (resultado da RB) estão apresentados na Figura 3. Pode-se observar que estes são bem similares, já que os valores de energia estão próximos da linha da bissetriz. Rodrigues et al. (2001), Moreira et al. (2002) e Brunelli et al. (2006), justificam que as diferenças existentes são devido a variabilidade na composição química dos alimentos que não são comumente utilizados nas rações de frangos.

Foi feita a validação das equações com os dados de ensaios metabólicos (ALVARENGA et al., 2011). A avaliação foi por meio de comparações entre os valores observados e preditos pelas equações, mensurando o m.s.e, MAD, MAPE e bias de cada equação proposta para os alimentos: milho, soja, subprodutos de milho, subprodutos de soja, outros energéticos e outros proteicos. Os valores calculados dessas medidas encontram-se na Tabela 8. Como os ensaios metabólicos foram realizados apenas com pintos, foram avaliadas apenas as equações propostas com esse tipo de animal.

Os valores de EMAN observados (dados de validação) e preditos (resultados das equações) estão apresentados na Figura 4. Os valores de EMAN de

Figura 3 – Valores de EMAn preditos versus observados obtidos dos dados de teste de alimentos concentrados para frangos.

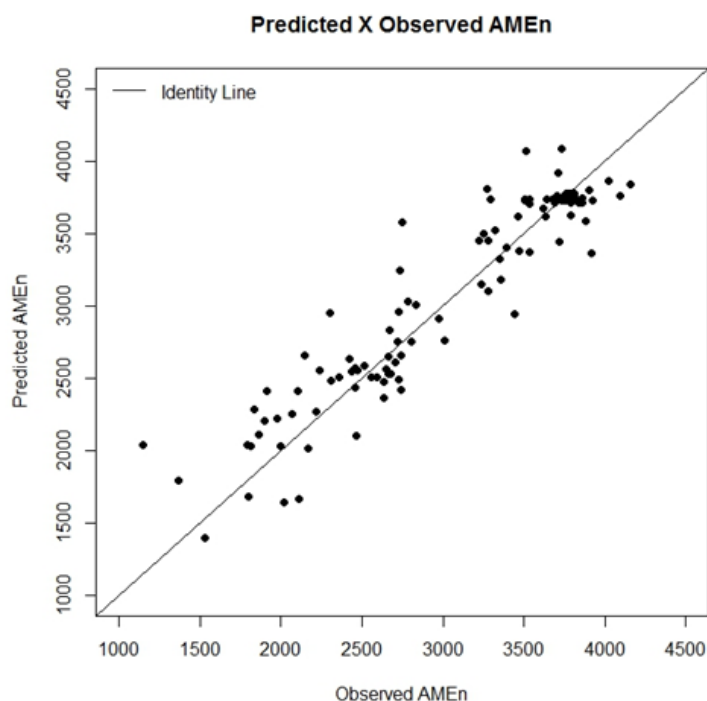


Tabela 8 – Avaliações das equações propostas por dados de ensaios metabólicos.

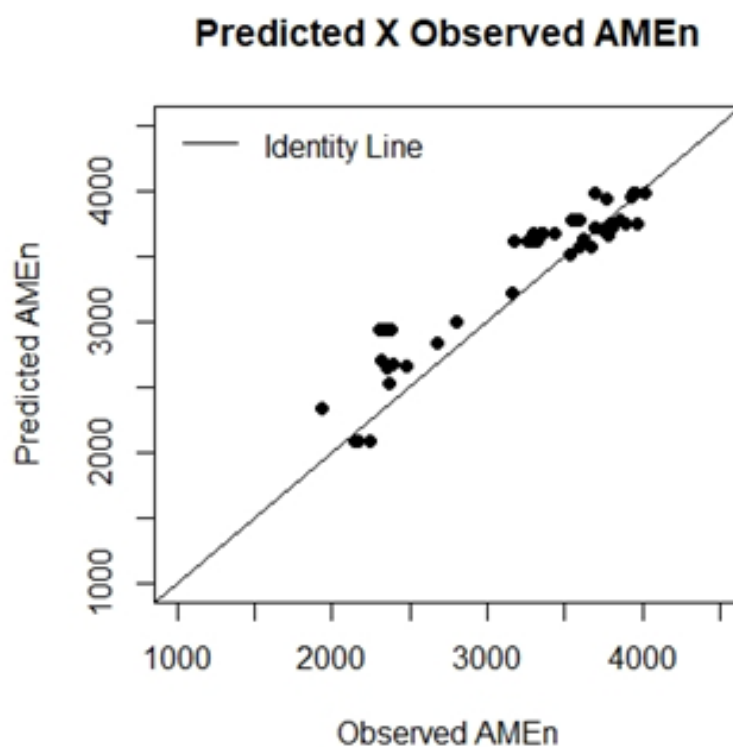
Equações	m.s.e	MAD	MAPE	Bias
0	27314,73	143,28	3,98	-104,56
1	71737,46	254,97	10,74	-254,97
2	9051,84	82,85	3,44	81,26
3	131069,90	299,25	10,84	-227,12
4	64831,12	213,58	7,42	-184,02
5	16473,99	81,66	2,16	-64,51

Equações representam os seguintes alimentos 0: milho; 1: soja; 2: subprodutos de milho; 3: subprodutos de soja; 4: outros energéticos e 5: outros proteicos.

alguma forma estão próximos da linha da bissetriz, indicando alta acurácia nas predições.

Calculadoras nos sistemas operacionais, android e java foram desenvolvidas a partir dos resultados obtidos nesta pesquisa, encontram-se disponíveis para download. As calculadoras fornecem EMAn de forma eficiente e fácil para zootecnistas. O aplicativo desenvolvido pode ser visualizado na Figura 5 (sistema operacional android) e Figura 6 (sistema operacional java). Apenas é necessário especificar os níveis das variáveis categóricas e entrar com os valores das variáveis proteína bruta (CP), extrato étereo (EE), matéria mi-

Figura 4 – Valores de EMAn preditos versus observados obtidos dos dados de ensaio metabólico de alimentos concentrados para frangos.



neral (MM) e fibra bruta (CF) e a calculadora fornece os valores de EMAn e ainda calcula os valores de EMAn em termos da matéria seca.

4 CONCLUSÕES

As equações obtidas por meio das redes bayesianas podem ser utilizadas pela indústria de frangos de corte na elaboração de rações, pois apresentaram alta acurácia na predição de EMAn. As equações estão disponíveis em uma calculadora que pode ser instalada em celulares, tablets e computadores.

Agradecimentos

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo auxílio financeiro.

Figura 5 – Calculadora que pode ser instalada em celulares e tablets



Referências

ALVARENGA, R. R.; RODRIGHUES, P. B.; ZANGERONIMO, M. G.; FREITAS, R. T. F.; LIMA, R. R.; BERTECHINI, A. G.; FASSANI, E. J. Energetic values of feedstuffs for broilers determined with in vivo assays and prediction equations. **Animal Feed Science and Technology**, v.168, n.1, p.257-266, 2011.

AHMADI, H.; GOLIAN, A.; MOTTAGHITALAB, M.;

Figura 6 – Calculadora que pode ser instalada em computadores

AMEn Result

NOTHING

AMEn value expressed on natural matter basis

NOTHING

DM = [] [%]

Chemical Composition Values (CCV)

CP = [] [%]

EE = [] [%]

MM = [] [%]

CF = [] [%]

Concentrate

Concentrate Energetic

Concentrate Protein

Ingredient

Maize or Soybean Meal

Maize or Soybean by-product

Other feedtuff

Animal

Chick

Rooster

Calculate Cleaner

Information

AMEn Calculator

Prediction of AMEn depending on the chemical composition and specifications of the feedstuffs used by the poultry feed industry using baysean network model.

AMEn Definition

AMEn: Nitrogen-corrected Metabolisable Energy, Values expressed on dry matter basis.

Ranges (CCV)

Ranges of the chemical composition values considered in the development of the baysean networks model

CP [%]	[1.47 ; 71.44]
EE [%]	[0.03 ; 26.21]
MM [%]	[0.30 ; 12.61]
CF [%]	[0.02 ; 27.63]

NARIMAN-ZADEH, N. Prediction model for true metabolizable energy of feather meal and poultry oal meal using group method of data handling-type neural network. **Poultry Science**, v.87, n.1, p.1909-1912, 2008.

AHMADI, H.; MOTTAGHITALAB M.; NARIMAN-ZADEH, N. Group method of data handling-type neural network prediction of broiler performance based on dietary metabolizable energy, methionine, and lysine. **Journal of Applied Poultry Research**, v.16, n.1, p.494-501, 2007.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. New York: Springer, 2006.

BORGES, F.M.O.; ROSTAGNO, H.S.; SAAD, C.E.P. et al. Equações de regressão para estimar valores energéticos de grão de trigo e seus subprodutos para frangos de corte, a partir de análises químicas. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.55, n.1, p.734-746, 2003.

BRUNELLI, S.R.; PINHEIRO, J.W.; SILVA, C.A.; FONSECA, N.A.N;

OLIVEIRA, D.D.; CUNHA, G.E.; SOUZA, L.F.A. Feeding increasing defatted corn germ meal levels to broiler chickens. **Brazilian Journal of Animal Science**, v. 35, n.1, p. 1349-1358, 2006.

CARVALHO, J. V. F., CHIANN, C. Redes Bayesianas: um método para avaliação de interdependência e contágio em séries temporais multivariadas. **Revista Brasileira de Economia**, v. 67, n.2, p. 201-217, 2013.

FELIPE, V.P.S.; SILVA, M.A.; VALENTE, B.D.; ROSA, G.J.M. Using multiple regression Bayesian networks and artificial neural networks for prediction of total egg production in European quails based on earlier expressed phenotypes. **Poultry Science**, v.94, n.4, p.772-780, 2015.

FEOFILOFF, P.; KOHAYAKAWA, Y.; WAKABAYASHI, Y. **Uma Introdução Sucinta à Teoria dos Grafos**. 2011, 61p. Disponível em: <http://www.ime.usp.br/pf/teoriadosgrafos/>. Acesso em 25 de agosto de 2018.

FRIEDMAN, N.; LINIAL, M.; NACHMAN, I.; PE'ER, D. Using Bayesian networks to analyze expression data. **Journal of Computational Biology**, v.7, n.3-4, p.601-620, 2000.

GIANOLA, D., OKUT, H., WEIGEL, K. A., ROSA, G. J. M. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. **BMC Genetics**, v.12, n.1, p.1-87, 2011.

HECKERMAN, D.; GEIGER, D.; CHICKERING, D. M. Learning Bayesian networks: the combination of knowledge and statistical data. **Machine Learning**, v.20, n.3, p.197-243, 1995.

JENSEN, F. V. **An Introduction to Bayesian Networks**. New York: Springer Verlag, 1996.

KOLLER, D.; FRIEDMAN, N. **Probabilistic Graphical Models: Principles and Techniques**. MIT Press, 2009, 1233p.

KOSKI, T.; NOBLE, J.M. **Bayesian networks: an introduction**. Chichester: John Wiley Sons Ltd, 2009. 347p.

LAURITZEN, S. L.; SPIEGELHALTER, D. J. Local computations with probabilities on graphical structures and their applications to expert

- systems. **Journal Royal Statistics Society**, v.50, n.1, p.157-224, 1988.
- LAURITZEN, S.L., WERMUTH, N. Graphical models for associations between outcomes, some of which are qualitative and some quantitative. **The Annals of Statistics**. v.17, n.1, p.31-57, 1989.
- MARGARITIS, D. **Learning bayesian network model structure from data**. 126f. PhD Thesis (Doctor of Philosophy) - Carnegie Mellon University, Pittsburgh, Pennsylvania, 2003.
- MARGARITIS, D.; THRUN, S. Bayesian Network Induction via Local Neighborhoods. In: Advances in Neural Information Processing Systems, 2000, Denver, CO, United States, **Proceedings. . .**, Denver, 2000.
- MARIANO, F. C. M. Q.; LIMA, R. R.; ALVARENGA, R. R.; RODRIGUES P. B.; P. B. Neural network committee to predict the AMEn of poultry feedstuffs. **Neural Computing and Applications**, v. 25, n.7-8, p.1903-1911, 2014.
- MARIANO, F. C. M. Q.; LIMA, R. R.; RODRIGUES P. B.; ALVARENGA, R. R.; NASCIMENTO, G. A. J. Equações de predição de valores energéticos de alimentos obtidas utilizando meta-análise e componentes principais. **Ciência Rural**, v.42, n.9, p.1634-1640, 2012.
- MARIANO, F.C.M.Q.; PAIXÃO, C. A.; LIMA, R. R.; ALVARENGA, R. R.; RODRIGUES P. B.; NASCIMENTO, G. A. J. Prediction of the energy values of feedstuffs for broilers using meta-analysis and neural networks. **Animal**, v.7, n.9, p.1440-1445, 2013.
- MOREIRA, I.; RIBEIRO, C.R.; FURLAN, A.C.; SCAPINELLO, C.; KUTSCHENKO, M. Utilization of defatted corn germ meal on growing-finishing pigs feeding digestibility and performance. **Brazilian Journal of Animal Science**, v.31, n.1, p.2238-2246, 2002.
- MOROTA, G.; MASANORI, K.; ROSA, G.J.M; KENT, A.W.; GIANOLA, D. Predicting complex traits using a diffusion kernel on genetic markers with an application to dairy cattle and wheat data. **Genetics Selection Evolution**, v.45, n.17, p.1-15, 2013.
- NAGARAJAN, R., SCUTARI, M; LÈBRE, S. **Bayesian Networks in R with Applications in Systems Biology**. Springer, New York. 2013.

157p.

NASCIMENTO, G.A.J.; RODRIGUES, P.B.; FREITAS, R.T.F.; REIS NETO, R.V.; LIMA, R.R.; ALLAMANN, I.B. Equações de predição para estimar valores da energia metabolizável de alimentos concentrados energéticos para aves utilizando meta-análise. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.63, n.1, p.222-230, 2011.

NASCIMENTO, G.A.J.; RODRIGUES, P.B.; FREITAS, R.T.F.; BERTECHINI, A.G.; LIMA, R.R.; PUCCI, L.E.A. Equações de predição para estimar os valores energéticos de alimentos concentrados de origem vegetal para aves utilizando a metanálise. **Revista Brasileira de Zootecnia**, v.38, n.1, p.1265-1271, 2009.

PEARL, J. **Probabilistic reasoning in intelligent systems: networks of plausible inference**. San Francisco: Editor Morgan Kaufmann, 1988. 552p.

PERAI, A. H.; MOGHADDAM, H. N.; ASADPOUR, S.; BAHRAMPOUT, J.; MANSOORI, Gh. A comparison of artificial neural networks with other statistical approaches for the prediction of true metabolizable energy of meat and bone meal. **Poultry Science**, v.89,n.1, p.1562-1568, 2010.

POURRET, O.; NAIM, P.; MARCOT, B. **Bayesian Networks: a Practical Guide to Applications**. Edited: John Wiley Sons Ltd, The Atrium, Southern Gate, Chichester, 2008. 446p.

R CORE TEAM (2018). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <http://www.R-project.org/>. Acesso em 25 de março de 2018.

RODRIGUES, P. B.; ROSTAGNO, H. S.; ALBINO, L. F. T.; GOMES, P. C.; BARBOZA, W. A.; SANTANA, R. T. Valores energéticos do milheto, do milho e subprodutos do milho, determinados com frangos de corte e galos adultos. **Revista Brasileira de Zootecnia**, v.30, n.6, p.1767-1778, 2001.

RODRIGUES, P. B.; ROSTAGNO, H. S.; ALBINO, L. F. T.; GOMES, P. C.; NUNES, R. V.; TOLEDO, R. S. Valores energéticos da soja e subprodutos da soja, determinados com frangos de corte e galos adultos. **Revista Brasileira de Zootecnia**, v.31, p.1771-1782, 2002.

SCUTARI, M., DENIS, J.B. **Bayesian Networks with Examples in R**. Taylor Francis Group. 2015. 221p.

SCUTARI, M. Learning Bayesian Networks with the bnlearn R Package. **Journal of Statistical Software**, v.35, n.3, p.1-21, 2010.

SCUTARI, M.; HOWELL, P.; BALDING, D.J.; MACKAY, I.J. Multiple quantitative trait analysis using bayesian networks. **Genetics**, v.198, n.1, p.129-137, 2014.

SIMÕES, P. W. T. A.; SILVA, G. D.; MORETTI, G. P.; SIMON, C. S.; WINNIKOW, E. P.; NASSAR, S. M.; MEDEIROS, L. R.; ROSA, M. I. Meta analysis of the use of Bayesian networks in breast cancer diagnosis. **Caderno Saúde Pública**, v.31, n.1, p.26-38, 2015.

SPIRITES, P.; GLYMOUR, C.; SCHEINES, R. **Causation, Prediction, and Search**. Second edition, Cambridge: MIT Press, 2000.

TÖPNER, K.; ROSA, G. J. M.; GIANOLA, D.; SCHÖN, C. C. Bayesian Networks Illustrate Genomic and Residual Trait Connections in Maize (*ZeamaysL.*) **G3: Genes, Genomes, Genetics**, v.7, n.1, p.2779-2789, 2017.

TREPODE, N.W.; FARIAS, C.R.G.; BARRERA, J. A pattern-oriented specification of gene network inference processes. **Computers in Biology and Medicine**, v.43, n.1, p.1415-1427, 2013.

TSAMARDINOS, I.; BROWN, L.E.; ALIFERIS, C.F. Algorithms for large scale markov blanket discovery. In: "Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference". **Proceeding...** 2003, p.376-381. AAAI Press. 2003.

UNITED STATES DEPARTMENT OF AGRICULTURE. **Livestock and poultre: world markets and trade**. Foreign Agricultural Service, New York, 2017. Disponível em:
<http://www.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf>. Acesso em: 11 Set. 2017.

VIEIRA, E. M. A., SILVA, J. M. N., SILVA, L. B. Modeling Bayesian Networks from a conceptual framework for occupational risk analysis. **Production**, v.27, n.1, p.1-12, 2017.

ZHAO, F.; ZHANG, H. F.; HOU, S. S.; ZHANG, Z. Y. Predicting metabolizable energy of normal corn from its chemical composition in adult pekin ducks. **Poultry Science**, v.87, n.1, p.1603-1608, 2008.

YARAMAKALA, S.; MARGARITIS, D. Speculative Markov Blanket Discovery for Optimal Feature Selection. In: Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM), **Proceedings...** 2005, Houston, Texas, United States, 2005.

ARTIGO 3: Equações de predição utilizando distribuições empíricas obtidas a partir de redes bayesianas

LIMA, R.R.; ALVARENGA, T.C.; BUENO FILHO, J. S. S.; MARIANO, F. C. M. Q.; ALVARENGA, R. R.; RODRIGUES, P. B.

Artigo a ser submetido na Revista *Neural Computing And Applications*

Resumo

Judea Pearl, então conhecido por defender o conhecimento probabilístico no campo da inteligência artificial propôs as redes bayesianas na década de 80. As redes bayesianas são modelos gráficos, as quais se constituem pela representação gráfica (grafo) e probabilística (distribuições de probabilidade condicionais e conjunta) das variáveis. Assim, o objetivo desta pesquisa foi propor uma nova abordagem metodológica para obtenção de equações de predição advindas de redes bayesianas híbridas. Nomeia-se redes bayesianas híbridas pelo fato de envolver na modelagem da rede, tanto variáveis de natureza quantitativa quanto de natureza qualitativa. A vantagem de se trabalhar com as redes bayesianas é que as mesmas empregam incertezas na modelagem, além de capturar dependências condicionais entre as variáveis envolvidas, o que não é possível com as metodologias tradicionais. Assim, a proposta desta pesquisa é, além de trabalhar com incertezas na obtenção da rede, empregar a incerteza também nas estimativas das covariáveis da equação utilizando várias partições no conjunto dos dados. Isso se justifica pois verifica-se variabilidade nas estimativas dos coeficientes à medida que se faz a busca das redes no espaço paramétrico das mesmas. Foram obtidas 100, 1000 e 10000 redes bayesianas híbridas com diferentes partições dos dados e, desse modo, foram obtidas distribuições empíricas para as estimativas dos coeficientes das covariáveis. Assim, calculou-se a média, mediana e moda dessas distribuições para construir equações em função desses valores. Neste estudo as covariáveis compreendem a composição química dos alimentos, sendo as mesmas utilizadas nas equações para prever a energia de alimentos para frangos de corte. Observa-se que a maneira de ser obter as estimativas das covariáveis abordada neste artigo é uma proposta inédita na área de pesquisa sendo a mesma aplicada na nutrição de aves. Os dados utilizados na obtenção das equações foram catalogados por Nascimento et al. (2009) e os dados para a validação das equações obtidas neste artigo são de Alvarenga et al. (2011). As equações propostas foram avaliadas por medidas de precisão, tais como, o erro quadrático médio, desvio médio absoluto, erro médio percentual absoluto e viés. As melhores equações para a predição de energia

foram provenientes de estimativas da média e moda para os resultados de 10000 redes bayesianas híbridas. Conclui-se que esta nova abordagem para a obtenção de equações de predição advindas de redes bayesianas híbridas demonstraram acurácia e precisão ao serem aplicadas na obtenção de valores energéticos de alimentos para frangos de corte. Logo, há uma contribuição em termos metodológicos para se obter equações de predição.

Palavras-chave: Algoritmo *MMHC*. *Bnlearn*. Distribuição empírica. Energia metabolizável.

Abstract

Judea Pearl, then known for defending probabilistic knowledge in the field of artificial intelligence proposed the Bayesian networks in the 1980s. Bayesian networks are graphical models, which se consist of the graphical (graph) and probabilistic (conditional and joint probability distribution) representation of the variables. Thus, the objective of this study was to propose a new methodological approach to obtain prediction equations from hybrid Bayesian networks. They are named Hybrid Bayesian networks because they involve both quantitative and qualitative variables in the modeling of the network. The advantage of employing Bayesian networks is that they use uncertainties in modeling, in addition to capturing conditional dependencies among the variables involved, which is not possible with traditional methodologies. Thus, the proposal of this study is, besides working with uncertainties in obtaining the network, using uncertainty also in the estimates of the covariates of the equation employing several partitions in the data set. This is justified because there is variability in the coefficient estimates as the networks are searched in the parametric space of them. 100, 1,000 and 10,000 hybrid Bayesian networks were obtained with different partitions of the data and, thus, empirical distributions were acquired for the estimates of the coefficients of the covariates. Thus, the mean, median and mode of these distributions were calculated to construct equations according to these values. In this study the covariates comprise the chemical composition of the feedstuffs, being the same used in the equations to predict the energy of feedstuffs for broilers. It is found that the way to obtain the estimates of the covariables addressed in this article is an unpublished proposal in the area of research, it being applied in the nutrition of poultry. The data used to obtain the equations were cataloged by Nascimento et al. (2009) and the data for the validation of the equations obtained in this article are from Alvarenga

et al. (2011). The proposed equations were evaluated by precision measurements, such as mean square error, absolute mean deviation, mean absolute error and bias. The best energy prediction equations came from mean and mode estimates for the results of 10,000 hybrid Bayesian networks. It follows that this new approach to obtain prediction equations from hybrid Bayesian networks demonstrated accuracy when applied in obtaining energy values of feedstuffs for broilers. Therefore, there is a contribution in methodological terms to obtain prediction equations.

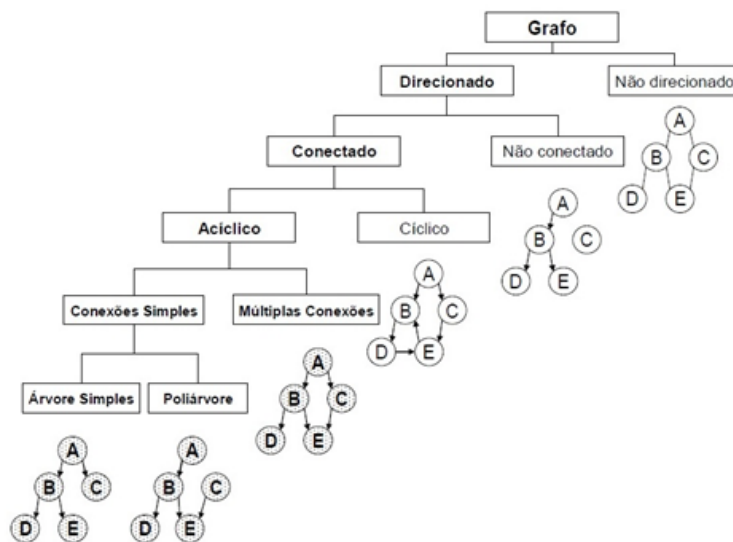
Keywords: Algorithm *MMHC*. Bnlearn. Empirical distribution. Metabolizable energy.

1 INTRODUÇÃO

Pearl (1988) e Neapolitan (2003) definem as redes bayesianas (RB), abordam suas propriedades e designam a como um campo de estudo. Uma RB constitui uma representação gráfica de variáveis e suas relações causais (SPIRITES et al., 2000; KOLLER e FRIEDMAN, 2009; SCUTARI e DENIS, 2015), nomeadas, respectivamente, por nós e arestas. Essa representação gráfica, denominada de grafo, é estudada na Teoria de Grafos. Mais detalhes dessa teoria podem ser encontrados em Feofiloff et al. (2011). A representação das relações entre as variáveis (nós), são dadas no grafo por meio de setas, ou seja, $A \rightarrow B$, a qual se lê A causa B ou A tem relação com B ou B depende de A. Nesta, tem-se que o nó A e o nó B estão, respectivamente, na cauda e cabeça da aresta. As estruturas básicas existentes dentro da Teoria dos Grafos são ilustradas na Figura 1.

Na Figura 1, apenas as estruturas gráficas hachuradas correspondem a RB e são denotados por DAG (*directed acyclic graph*). Os DAGs apresentam direção nas arestas, todos os nós são conectados na estrutura gráfica e as direções das arestas não apresentam ciclos, ou seja, após a aresta seguir a direção, ela não retorna ao mesmo nó, classificando-o, respectivamente, por grafo direcionado, conectado e acíclico. Conforme a Figura 1, existe também o grafo não direcionado, ou seja, arestas não indicam causa e efeito entre os nós; o grafo direcionado, isto é, arestas indicam relações de causa e efeito entre os nós; o grafo não conectado, quer dizer, existe pelo menos um nó que não se relaciona com os demais; grafo cíclico, ou melhor, seguindo-se a direção das arestas, há retorno a um mesmo nó. Em síntese, essas demais ilustrações de grafos não se enquadram no estudo de RB, apenas o grafo direcionado, conectado e acíclico. De acordo com o grafo da Figura 2, e tomando o nó A como referência, pode-se observar que seus vizinhos (C,B,D,E) são obtidos pela união de seus pais (C,B) e filhos ou crianças (E,D), chamados também

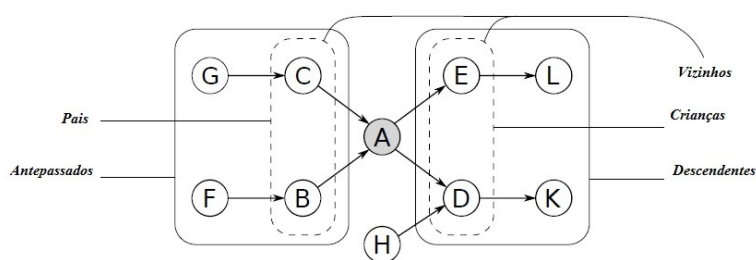
Figura 1 – As estruturas básicas existentes dentro da Teoria dos Grafos.



Fonte: Ara-Souza, 2010.

de nós adjacentes (nós pais e filhos). Seus pais também são chamados de antecessores, pois estão ordenados anterior ao A. Os filhos de A são chamados de descendentes, devido a ordem posterior. A ordem das variáveis no grafo chama-se topologia e assim, de acordo com Nagarajan, Scutari e Lébrea (2013), um exemplo da ordem topológica da estrutura gráfica apresentada na Figura 2 pode ser dada por: (F,G,H), (C,B), (A), (D,E), (L,K).

Figura 2 – Exemplo ilustrativo DAG.

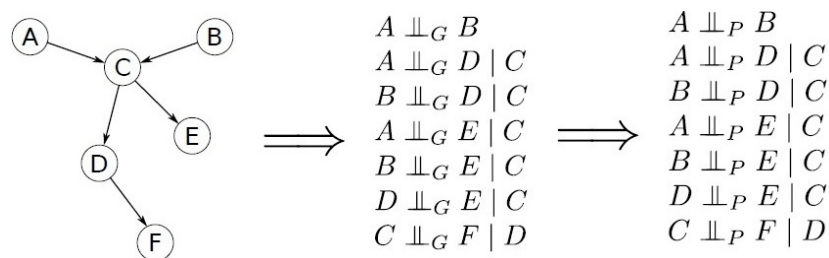


Fonte: Adaptado de Nagarajan; Scutari e Lébrea, 2013.

Muitos dos algoritmos utilizados para encontrar a estrutura gráfica das variáveis requerem ordenação entre os nós. A justificativa da mesma, se deve ao fato de variáveis causais estarem precedidas de outras variáveis (nós). Al-

guns desses algoritmos utilizam a cobertura de *Markov* (*Markov Blanket*) para modelar o fenômeno em estudo. Denota-se cobertura de *Markov*, o conjunto de nós que separa o nó alvo dos demais nós da estrutura, ou seja, essa cobertura inclui pais, filhos e nós que compartilham os mesmos filhos do nó alvo. A cobertura de Markov do nó A (Figura 2) são os nós: C, B, H, E e D. Estudos cujo objetivo é previsão, apenas as variáveis encontradas nessa cobertura são relevantes ao modelo (TSAMARDINOS et al., 2003). Outras configurações de topologia, podem ser obtidas seguindo o ordenamento dos nós. Robinson (1977, citado por KOSKI e NOBLE, 2009), define o número de possíveis DAGs ($N(d)$) no espaço paramétrico, em que d é o número de nós. $N(d)$ cresce exponencialmente, ou seja, para $d = 5$ tem-se $N(d) = 29000$ DAGs e para $d = 10$, $N(d)$ é aproximadamente $4,2 \times 10^{18}$ DAGs (KOSKI; NOBLE, 2009). A DAG da Figura 3 é apresentada como um mapa de independência da distribuição de probabilidade da variável aleatória. Observa-se que separação gráfica (representada por \perp_G) implica em independência probabilística (representada por \perp_P).

Figura 3 – DAG, separação gráfica e independência probabilística.



Assim, de acordo com a Figura 3, conclui-se que A é independente de B; A é independente de D, dado C; B é independente de D, dado C; A é independente de E, dado C; B é independente de E, dado C; D é independente de E, dado C; C é independente de F, dado D. Essa independência também é chamada de *d-separação*, leia-se: A é d-separado de D pelo nó C, ou seja, A e D são separados por C, ou C bloqueia o caminho entre A e D (PEARL, 1988; MARGARITIS, 2003).

Uma RB é uma representação gráfica de uma distribuição de probabilidade conjunta (HECKERMAN, 1995; MARGARITIS, 2003), a qual descreve o domínio por completo. Ela pode ser definida pelo conjunto (ϵ, θ, X) , em que ϵ é a estrutura da DAG, θ é o conjunto de parâmetros específicos advindos de distribuições de probabilidades condicionais, envolvendo o conjunto X de variáveis aleatórias. A fatoração da RB, conforme apresentado na Equação 1, consiste no produto das probabilidades condicionais, visto que um nó é condicionalmente independente de seus não descendente, dado seus pais

(PEARL, 1988; TSAMARDINOS et al., 2003). Essa fatorização resulta na distribuição de probabilidade conjunta, sendo assim, possível inferir e obter respostas a questões de interesse.

$$P(X_1, X_2, \dots, X_p) = \prod_{i=1}^n P(X_i | Pa_i) \quad (1)$$

para o caso de nós discretos e contínuos, em que Pa_i são os pais de X_i .

As relações de independências condicionais entre as variáveis são verificadas pelo critério de *d-separação*. Se o modelo é compatível com uma distribuição de probabilidade, então, pode-se representar de forma exata, por um conjunto de probabilidades condicionais. Assim, a distribuição conjunta é chamada de distribuição global e as distribuições condicionais denotadas por distribuições locais. A distribuição de probabilidade conjunta é dada de acordo com a natureza das variáveis. Assim, para o caso discreto, contínuo e ambos, as redes bayesianas são nomeadas, respectivamente, por rede bayesiana multinomial, rede bayesiana gaussiana e rede bayesiana híbrida (SCUTARI e DENIS, 2015).

A rede bayesiana multinomial e a rede bayesiana gaussiana consideram apenas as variáveis discretas ou contínuas. Sendo que todas as variáveis da RB seguem a mesma distribuição de probabilidade pertencentes à mesma família: multinomial ou normal multivariada. No entanto, existem casos em que as variáveis sob investigação são melhores explicadas por outras distribuições de probabilidade. Assim, surge a ideia de se trabalhar com ambas as variáveis discretas e contínuas em uma mesma rede bayesiana, além da possibilidade de utilizar qualquer tipo de distribuição na modelagem das variáveis. No entanto, empregar outras distribuições além da multinomial e da normal multivariada, faz com que as RB se tornem difíceis e complicadas de se trabalhar, além de que não existe um pacote específico no software R para lidar com elas (SCUTARI e DENIS, 2015). Observa-se que essa metodologia ainda requer pesquisas no campo metodológico estatístico e computacional para que sua aplicação se torne cada vez mais eficiente. Essa eficiência é esperada desde a obtenção da rede, estimação dos parâmetros, predição das variáveis e otimização do tempo gasto nas análises (*structure learning e parameter learning*).

No entanto, existe a possibilidade de se trabalhar com ambas as variáveis discretas e contínuas em uma mesma rede bayesiana, empregando a distribuição multinomial para as variáveis discretas e a distribuição normal para as variáveis contínuas. Contudo, existem algumas restrições, tais como, apenas os nós discretos podem ser pais dos nós contínuos, o contrário não é permitido (KOSKI; NOBLE, 2009). A rede bayesiana híbrida, assim chamada por contemplar tanto variáveis contínuas quanto discretas, se assemelha a um modelo

de regressão linear múltipla, a qual denota-se por rede bayesiana gaussiana condicional linear (KOSKI; NOBLE, 2009). Assim, para cada combinação das variáveis discretas é encontrado uma regressão linear múltipla.

Assim, a proposta desta pesquisa foi utilizar incertezas na obtenção de cada rede bayesiana individual, além de empregar a incerteza nas estimativas das covariáveis das equações obtidas por um conjunto de 100, 1000 e 10000 redes bayesianas. Dessa forma, utilizando-se esse critério foram obtidas equações para prever os valores de energia metabolizável aparente corrigida pelo balanço de nitrogênio (EMAn) para cada alimento utilizado na elaboração de rações de frangos de corte. Os coeficientes das equações foram obtidos a partir da média, moda e mediana de suas distribuições empíricas, sendo estes valores os coeficientes das covariáveis: proteína bruta (PB), fibra bruta (FB), extrato etéreo (EE), matéria mineral (MM). Após aplicar a proposta, para garantir validade às equações, houve a validação das mesmas por dados provenientes de ensaios metabólicos condizidos por Alvarenga et al. (2011).

2 MATERIAL E MÉTODOS

2.1 Dados

Os dados referem-se a valores de EMAn e composição química de alimentos energéticos e proteicos, de origem vegetal, geralmente utilizados na formulação de dietas de frangos de corte. Para obtenção das equações foram utilizados os dados da meta-análise (NASCIMENTO et al., 2009), referentes a experimentos executados no Brasil, no período de 1967 a 2007, resultando em 568 experimentos, dentre os quais 370 são de concentrados energéticos e 198 de concentrados proteicos.

Os dados utilizados na validação das equações propostas foram obtidos por Alvarenga et al. (2011). Esses dados são provenientes de dois ensaios metabólicos para determinação do valor energético de alimentos proteicos (ensaio I) e energéticos (ensaio II), com pintos em crescimento (método tradicional de coleta total de excretas), respectivamente em fevereiro/março e julho de 2008. Os experimentos foram conduzidos no Setor de Avicultura do Departamento de Zootecnia da Universidade Federal de Lavras (UFLA), em Lavras, MG.

Os valores da variável alvo EMAn, foram estimados pelas covariáveis da rede, sendo: proteína bruta (PB), extrato etéreo (EE), matéria mineral (MM), fibra bruta (FB), classificação da categoria de alimentos (concentrado energético, concentrado de proteína), ingrediente na categoria (concentrado energético: milho, subprodutos de milho e outros) e ingrediente na categoria (concentrado de proteína: farelo de soja, subprodutos de soja e outros).

2.2 Modelo de predição

A proposta desse artigo foi combinar diferentes resultados de redes bayesianas. Para encontrar cada uma das redes bayesianas utilizou-se o algoritmo híbrido *Max-Min Hill Climbing* (MMHC) (TSAMARDINOS et al., 2006) que é uma combinação de dois algoritmos, ou seja, o algoritmo *constraint-based Max-Min Parents and Child* (MMPC) e algoritmo *score-based Hill Climbing* (HC). O algoritmo *constraint-based* é utilizado para encontrar o esqueleto da rede bayesiana e posteriormente o algoritmo *score-based* é utilizado para estabelecer as relações de causa e efeito na rede. Os algoritmos *constraint-based* consistem em vários testes de independência condicional e marginal para inferir as direções das arestas entre os nós. Esse algoritmo constrói a princípio uma estrutura de variáveis não direcionadas, apenas relacionadas. Depois são projetadas as *v-estruturas* (isto é, arestas convergindo para o mesmo nó) e por fim os arcos remanescentes são orientados de forma que não ocorra novas *v-estruturas* e ciclos (SCUTARI, 2010). Já o algoritmo *score-based* seleciona entre todas as possíveis estruturas de RB, a DAG com melhor qualidade, ou seja, com maior pontuação (SCUTARI, 2010). Essa seleção pode ser por meio do critério de informação Bayesiano (BIC), critério de informação de Akaike (AIC) e intermediário aos dois quanto à penalização, K2 score. No processo de seleção das redes bayesianas desta pesquisa utilizou-se a informação mútua no algoritmo *constraint-based*, a qual consiste em um teste de independência condicional e marginal. Já no algoritmo *score-based* utilizou-se o critério de informação Bayesiano (BIC) para seleção da rede com melhor ajuste. As redes foram obtidas de maneira aleatória no conjunto de aprendizagem e teste, aos quais consistiam respectivamente, em 80% e 20% dos dados advindos de estudos de meta-análise (NASCIMENTO et al., 2011).

Cada rede bayesiana individual compreende doze equações, as quais seguem apresentadas, por $EMAn_{i,j,k}$. Têm-se que o índice i refere-se a categoria, se 1-energético ou 2-proteico. O índice j refere-se a classificação do alimento, se $i=1$ então j : 1-milho ou 2-subprodutos de milho ou 3-outros; se $i=2$ então j é 1-soja ou 2-subprodutos de soja ou 3-outros. Para k , os valor de 1 refere-se a pinto e 2 ao galo.

$$\text{Equação 1: } EMAn_{1,1,1} = \beta_0 + \beta_1PB + \beta_2EE + \beta_3MM + \beta_4FB$$

$$\text{Equação 2: } EMAn_{1,2,1} = \beta_5 + \beta_6PB + \beta_7EE + \beta_8MM + \beta_9FB$$

$$\text{Equação 3: } EMAn_{1,3,1} = \beta_{10} + \beta_{11}PB + \beta_{12}EE + \beta_{13}MM + \beta_{14}FB$$

$$\text{Equação 4: } EMAn_{2,1,1} = \beta_{15} + \beta_{16}PB + \beta_{17}EE + \beta_{18}MM + \beta_{19}FB$$

$$\text{Equação 5: } EMAn_{2,2,1} = \beta_{20} + \beta_{21}PB + \beta_{22}EE + \beta_{23}MM + \beta_{24}FB$$

$$\text{Equação 6: } EMAn_{2,3,1} = \beta_{25} + \beta_{26}PB + \beta_{27}EE + \beta_{28}MM + \beta_{29}FB$$

$$\text{Equação 7: } EMAn_{1,1,2} = \beta_{30} + \beta_{31}PB + \beta_{32}EE + \beta_{33}MM + \beta_{34}FB$$

$$\text{Equação 8: } EMAn_{1,2,2} = \beta_{35} + \beta_{36}PB + \beta_{37}EE + \beta_{38}MM + \beta_{39}FB$$

Equação 9: $EMAn_{1,3,2} = \beta_{40} + \beta_{41}PB + \beta_{42}EE + \beta_{43}MM + \beta_{44}FB$

Equação 10: $EMAn_{2,1,2} = \beta_{45} + \beta_{46}PB + \beta_{47}EE + \beta_{48}MM + \beta_{49}FB$

Equação 11: $EMAn_{2,2,2} = \beta_{50} + \beta_{51}PB + \beta_{52}EE + \beta_{53}MM + \beta_{54}FB$

Equação 12: $EMAn_{2,3,2} = \beta_{55} + \beta_{56}PB + \beta_{57}EE + \beta_{58}MM + \beta_{59}FB$

Como a proposta foi obter diferentes DAGs, a partir de diferentes partições do conjunto de dados, já que as possibilidades do número de DAGs cresce de maneira exponencial a medida que o número de variáveis aumenta (KOSKI; NOBLE, 2009). Assim, foram obtidas 100, 1000 e 10000 estimativas de cada um dos 60 β 's apresentados anteriormente. Para o conjunto de estimativas de cada um dos β 's foram encontradas distribuições empíricas para cada coeficiente da equação de predição. A partir das distribuições empíricas utilizaram-se do valor da média, da mediana e da moda para definir a equação final de cada uma das doze equações mencionadas anteriormente, pois para cada combinação de categoria, alimento e animal existe uma equação. Assim, os coeficientes das equações representam valores da média ou valores da mediana ou valores da moda das distribuições empíricas, o que permitiu maior generalização e ganho de informação nos resultados, empregando o efeito aleatório tanto na obtenção das redes quanto nas estimativas dos coeficientes das equações.

As estatísticas utilizadas no processo de avaliação (teste e validação) das redes foram o erro quadrático médio (m.s.e.), desvio médio absoluto (MAD), erro médio de porcentagem absoluta (MAPE) e viés (*bias*) (MARIANO et al., 2014). Os programas computacionais utilizados neste trabalho foram desenvolvidos no software livre R (R CORE TEAM, 2018), utilizando-se o pacote *bnlearn* (SCUTARI, 2010) e o algoritmo de aprendizagem *hybrid Max-Min Hill Climbing* (MMHC) (TSAMARDINOS et al., 2006). O pacote *bnlearn* possibilita a aprendizagem da estrutura gráfica de maneira individual, bem como a estimação dos parâmetros, trabalhando com qualquer tipo de variável (quantitativa e qualitativa).

De maneira resumida, a proposta desse artigo consistiu em:

Seja $N = 100$ ou 1000 ou 10000 redes bayesianas.

Para

1: N :

- Faça a partição do conjunto de treinamento (80% dos dados) e teste (20% dos dados restantes) de maneira aleatória;
- Utilize o algoritmo híbrido MMHC para aprendizagem da rede; Dentro deste algoritmo: o teste de independência utilizado será a informação mútua e a seleção da melhor rede será por meio do critério de BIC;
- Repita o processo N vezes;
- Separe os coeficientes (intercepto, PB, EE, MM, FB) de cada rede;
- Obtenha a distribuição empírica dos coeficientes encontrados na etapa anterior;
- Calcule a média, moda e mediana das distribuições encontradas no passo anterior;
- Obtenha as equações para EMAN em função de PB, EE, MM e FB por meio das estatísticas calculadas no passo anterior;
- Validar as equações propostas com os dados de Alvarenga et al. (2011).

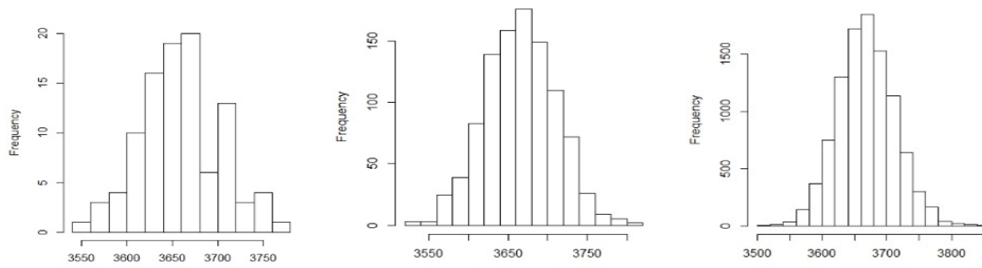
3 RESULTADOS E DISCUSSÃO

As estimativas dos coeficientes (intercepto: β_0 , PB: β_1 , EE: β_2 , MM: β_3 e FB: β_4) advindos de 100, 1000 e 10000 redes bayesianas para Equação 1 (Categoria = energéticos, alimento = milho e animal = pinto): $EMAn = \beta_0 + \beta_1 PB + \beta_2 EE + \beta_3 MM + \beta_4 FB$, estão apresentadas na Figura 4a - 4e. Foram encontradas as estimativas dos coeficientes para as outras equações (2 - 12), contudo, para fins de ilustração optou-se em apresentar apenas as distribuições empíricas das estimativas dos coeficientes a para Equação 1.

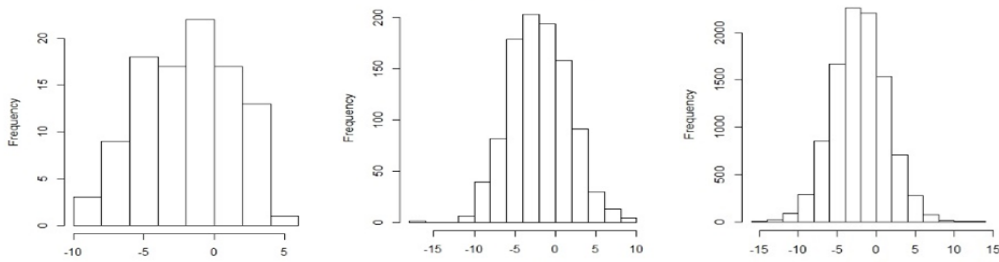
Observa-se que os valores dos β_0 (Figura 4.a) estão em torno de 3550 a 3800 para a maioria das distribuições empíricas de diferentes tamanhos (N). No entanto para $N=1000$ e 10000 redes, a distribuição encontrada é mais simétrica, e assim, os valores da media, moda e mediana são próximos.

Já os valores dos β_1 (Figura 4.b) estão em torno de -10 e 5 para 1000 redes, -15 a 10 para 100 redes e -15 a 15 para 10000 redes. Há maior probabilidade dos valores de β_1 estarem em torno de -5 e 0 para a maioria das distribuições empíricas de tamanhos diferentes (N).

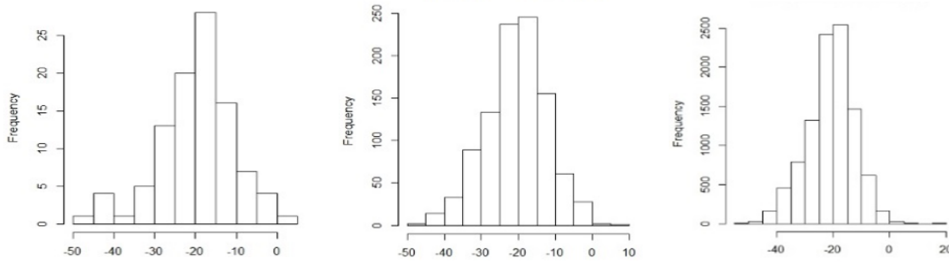
Figura 4 – Estimativas dos β_i , respectivamente em 100, 1000 e 10000 repetições.



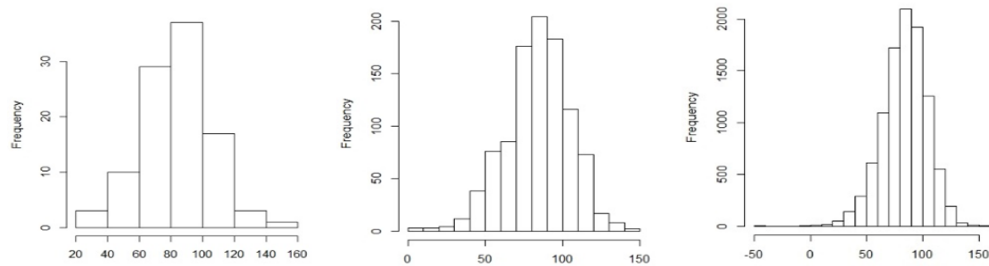
a) Estimativas dos β_0 , respectivamente em 100, 1000 e 10000 repetições



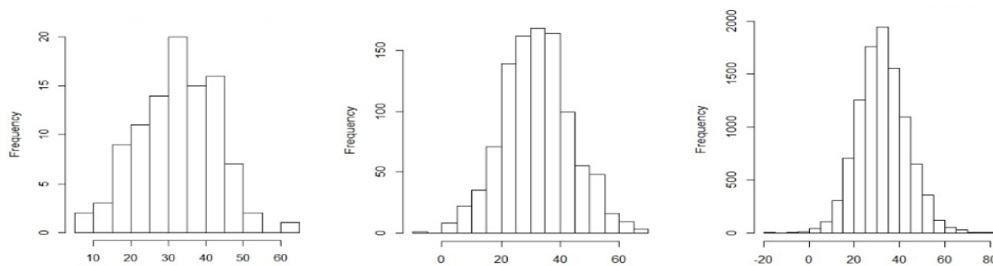
b) Estimativas dos β_1 , respectivamente em 100, 1000 e 10000 repetições



c) Estimativas dos β_2 , respectivamente em 100, 1000 e 10000 repetições



d) Estimativas dos β_3 , respectivamente em 100, 1000 e 10000 repetições



e) Estimativas dos β_4 , respectivamente em 100, 1000 e 10000 repetições

Os valores dos β_2 (Figura 4.c) estão concentrados em torno de -50 e 20, com maior frequência dos β_2 em torno do -20. As distribuições empíricas advindas de 100, 1000 e 10000 redes, respectivamente, não apresenta simetria, é simétrica, é há indícios de assimetria à direita.

Os valores dos β_3 (Figura 4.d) estão concentrados em torno de 20 e 160; 0 e 150; -50 e 150, respectivamente para as distribuições empíricas advindas de 100, 1000 e 10000 redes. Observa-se que principalmente nas caudas das distribuições empíricas de 1000 e 10000 redes os valores de β_3 possuem pouca probabilidade de ocorrência. A maior ocorrência dos valores concentra-se em torno de 80 a 100.

Os valores dos β_4 (Figura 4.e) concentram-se com maior probabilidade em torno de 30 a 40, já os limites inferior e superior para os diferentes tamanhos de redes utilizadas não mantem um padrão, para 10000 redes o intervalo que contém os β_4 é superior aos demais, porém a probabilidade nas caudas da distribuição empírica é relativamente baixa.

No entanto foram encontradas distribuições empíricas das estimativas dos demais β_i ($i= 5$ a 59) coeficientes advindos de 100, 1000 e 10000 redes bayesianas. A partir de cada distribuição empírica calculou-se a média, moda e mediana representando cada β_i das equações. Assim, as equações foram expressas por essas medidas estatísticas, porém, os melhores ajustes foram provenientes de 10000 redes bayesianas híbridas. As equações propostas por meio de 10000 redes bayesianas híbridas foram validadas por dados provenientes de ensaios metabólicos (ALVARENGA et al., 2011), conforme apresentam-se os valores de energia na Tabela 1.

De acordo com a Tabela 1, as primeiras 7 amostras (1-7) se referem aos valores energéticos do milho. As próximas 5 amostras (8-12) se referem a valores energéticos de soja. As próximas 6 amostras (13-18) são de alimentos classificados como subprodutos de milho. As seguintes 16 amostras (19-34) se referem a alimentos classificados como subprodutos de soja. As próximas 9 amostras (35-43) se referem a alimentos classificados em outros energéticos. As seguintes 5 amostras (44-48) se referem a outros proteicos.

As equações propostas foram avaliadas por meio de comparações entre os valores observados na amostra e preditos pelas equações propostas, utilizando-se as medidas de acurácia e precisão: erro quadrático médio (MSE), o desvio médio absoluto (MAD), o erro médio percentual absoluto (MAPE) e o bias. Equações mais acuradas, são esperados valores de bias e dos erros próximos de zero. Os valores referentes a essas medidas estão apresentados na Tabela 2.

Os valores em **negrito** na Tabela 2 são das medidas calculadas para encontrar os coeficientes das equações que melhor predisseram os valores de energia. Sendo assim, as equações selecionadas com a proposta desse artigo

Tabela 1 – Valores energéticos obtidos de equações propostas via coeficientes em termos da média, mediana e moda.

Amostra	Observado*	Média	Mediana	Moda
1	3747	3705	3707	3710
2	3699	3719	3721	3724
3	3813	3715	3717	3721
4	3591	3798	3800	3804
5	3568	3798	3800	3804
6	3581	3798	3800	3804
7	3548	3798	3800	3804
8	2373	2578	2509	2491
9	2326	2642	2560	2530
10	2355	2565	2479	2448
11	2396	2556	2469	2432
12	2478	2562	2482	2448
13	3624	3643	3651	3721
14	3676	3552	3565	3639
15	2248	2463	2498	2538
16	2151	2463	2498	2538
17	2173	2463	2498	2538
18	2162	2463	2498	2538
19	3159	3207	3294	3293
20	3779	3682	3752	3745
21	2809	2890	2978	2973
22	3772	3992	4029	4004
23	2387	2939	3030	3020
24	3971	3775	3835	3827
25	3288	3679	3747	3737
26	2314	2939	3030	3020
27	3818	3775	3835	3827
28	3173	3679	3747	3737
29	2339	2939	3030	3020
30	3793	3775	3835	3827
31	3330	3679	3747	3737
32	2309	2939	3030	3020
33	3890	3775	3835	3827
34	3267	3679	3747	3737
35	3598	3547	3548	3547
36	3529	3495	3496	3495
37	3862	3751	3749	3744
38	2682	2836	2853	2866
39	1941	2336	2346	2350
40	3353	3646	3646	3644
41	3297	3646	3646	3644
42	3436	3646	3646	3644
43	3363	3646	3646	3644
44	3934	3995	3992	3988
45	3700	4014	4011	4007
46	3941	4014	4011	4007
47	3954	4014	4011	4007
48	4022	4014	4011	4007

Observado* corresponde aos valores de EMAn obtidos por Alvarenga et al. (2011).

são: $EMAn = 3668,345 - 2,12PB - 20,96EE + 83,95 MM + 32,41FB$ (para milho), $EMAn = 1969,02 + 3,14PB - 12,04EE + 25,63MM + 34,33FB$ (para soja), $EMAn = 3673,48 + 34,59PB + 22,25EE - 51,31MM - 156,86 FB$ (subprodutos de milho), $EMAn = 1592,81 + 10,99PB + 51,78EE + 118,24MM$

Tabela 2 – Valores das estatísticas de avaliação obtidas ao aplicar as equações propostas via coeficientes da média, moda, mediana para cada alimento.

Equações para Milho			
	Média	Mediana	Moda
MAD	152,3684	153,237	154,8076
MAPE	4,231614	4,25661	4,301777
BIAS	-112,6231	-114,6695	-118,3474
MSE	31189,39	31666,83	32544,43
Equações para Soja			
	Média	Mediana	Moda
MAD	195,6061	114,6843	96,33479
MAPE	8,267283	4,873421	4,08692
BIAS	-195,6061	-114,6843	-84,5425
MSE	43968,47	18932,31	13309,8
Equações para Subprodutos de Milho			
	Média	Mediana	Moda
MAD	210,7405	232,8689	259,3209
MAPE	9,231288	10,26257	11,48916
BIAS	-169,5021	-196,1125	-247,1235
MSE	55957,84	69210,35	86973,84
Equações para Subprodutos de Soja			
	Média	Mediana	Moda
MAD	305,494	346,6845	339,8507
MAPE	11,03021	12,65091	12,41136
BIAS	-247,1080	-319,5921	-310,0323
MSE	142391,5	188363,3	181468
Equações para outros energéticos			
	Média	Mediana	Moda
MAD	209,0507	212,1686	213,8163
MAPE	7,25078	7,38156	7,449738
BIAS	-165,8905	-169,0307	-168,7979
MSE	58823,57	60402,99	60788,12
Equações para outros proteicos			
	Média	Mediana	Moda
MAD	103,687	101,7872	99,26804
MAPE	2,73392	2,684356	2,618636
BIAS	-100,8797	-97,70134	-93,48928
MSE	22446,72	21814,14	21007,24

- 6,96FB (subprodutos de soja), $EMAn = 4296,14 - 49,11PB + 35,21EE - 64,50MM - 88,74FB$ (outros energéticos) e $EMAn = 2295,08 + 25,23 PB + 76,90EE - 168,76MM - 21,90FB$ (outros proteicos). Assim, a média descreveu melhor as estimativas dos coeficientes, exceto para os alimentos soja e outros proteicos, aos quais equações de melhores ajustes foram a partir de estimativas modais para os coeficientes. Os valores de EMAn obtidos por meio destas equações via nova abordagem metodológica se encontram nos intervalos obtidos por comitês de redes neurais propostos por Mariano et al. (2014).

Existem na literatura vários trabalhos que adotaram diferentes metodologias para encontrar equações de predição de energia dos alimentos utilizados na elaboração de rações para frangos, devido a necessidade de equações for-

necerem resultados acurados à medida que são aplicadas na indústria. Trabalhos de Rodrigues et al. (2001, 2002); Borges et al. (2003) e Zhao et al. (2008), obtiveram equações de predição por meio de regressão múltipla, no entanto, não houve nenhuma consideração ou ponderação na análise. Nascimento et al. (2009, 2011) e Mariano et al., (2012) obtiveram as equações por meio da meta-análise, sendo que em Mariano et al. (2012) adotou-se a técnica de componentes principais para obter as equações. As redes neurais também têm sido utilizadas para obter equações de predição, exemplos deste são os estudos de Perai et al. (2010), Ahmadi et al. (2007, 2008) e Mariano et al. (2013). Em Mariano et al. (2014) foi proposto um comitê de redes neurais para prever os valores de energia, sendo uma nova abordagem de se obter os valores energéticos, resultando em estimações precisas e acuradas. Alvarenga et al. (2015) afirmam que é necessário a inovação em termos de utilização metodológica na obtenção das equações de predição, podendo assim, resultar em melhores estimativas dos valores energéticos. Neste trabalho utilizou-se uma nova abordagem para se obter as equações de predição a partir das redes bayesianas e as mesmas mostraram precisão nas estimativas de energia. As equações utilizando os galos nos ensaios não foram validadas, no entanto, o mesmo ainda pode ser realizado com dados provenientes de ensaios metabólicos.

4 CONCLUSÃO

Uma nova abordagem foi proposta neste artigo, a qual empregou incertezas na obtenção de equações a partir de resultados de redes bayesianas. As estimativas provenientes de médias e modas dos coeficientes advindos de distribuições empíricas construídas com 10000 redes bayesianas tiveram melhores desempenhos. As equações propostas mostraram resultados precisos à medida que foram avaliadas com os dados de ensaios metabólicos. Assim, contribui-se tanto em termos metodológicos quanto em termos práticos na área de Nutrição de aves.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo auxílio financeiro.

Referências

ALVARENGA, R. R.; RODRIGUES, P. B.; ZANGERONIMO, M. G.; FREITAS, R. T. F.; LIMA, R. R.; BERTECHINI, A. G.; FASSANI, E. J. Energetic values of feedstuffs for broilers determined with in vivo assays and prediction equations. **Animal Feed Science and Technology**, v.168, n.1, p.257-266, 2011.

ALVARENGA, R. R. ; RODRIGUES, P. B. ; ZANGERONIMO, M. G. ; OLIVEIRA, E. C. ; MARIANO, F. C. M. Q. ; LIMA, E. M. C. ; GARCIA JR, A. A. P. ; NAVES, L. P. ; NARDELLI, N. B. S. . Validation of Prediction Equations of Energy Values of a Single Ingredient or Their Combinations in Male Broilers. **Asian-Australians Journal of Animal Science**, v.28, n.1, p.1335-1344, 2015.

AHMADI, H.; GOLIAN, A.; MOTTAGHITALAB, M.; NARIMAN-ZADEH, N. Prediction model for true metabolizable energy of feather meal and poultry oal meal using group method of data handling-type neural network. **Poultry Science**, v.87, n.1, p.1909-1912, 2008.

AHMADI, H.; MOTTAGHITALAB M.; NARIMAN-ZADEH, N. Group method of data handling-type neural network prediction of broiler performance based on dietary metabolizable energy, methionine, and lysine. **Journal of Applied Poultry Research**, v.16, n.1, p.494-501, 2007.

ARA-SOUZA, A.L. Redes Bayesianas: Uma introdução aplicada a credit scoring. In: 19^o Simpósio Nacional de Probabilidade e Estatística-SINAPE, 2010, **Anais. . .**, São Pedro: SINAPE, 2010.

BORGES, F.M.O.; ROSTAGNO, H.S.; SAAD, C.E.P. et al. Equações de regressão para estimar valores energéticos de grão de trigo e seus subprodutos para frangos de corte, a partir de análises químicas. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.55, n.1, p.734-746, 2003.

FEOFILOFF, P.; KOHAYAKAWA, Y.; WAKABAYASHI, Y. **Uma Introdução Sucinta à Teoria dos Grafos**. 2011, 61p. Disponível em: <http://www.ime.usp.br/pf/teoriadosgrafos/>. Acesso em 25 de agosto de 2018.

HECKERMAN, D. **A tutorial on learning with Bayesian Networks**, Technical Report MSR-TR-95-06, Microsoft Research, 1995.

KOLLER, D.; FRIEDMAN, N. **Probabilistic Graphical Models: Principles and Techniques**. MIT Press, 2009, 1233p.

KOSKI, T.; NOBLE, J.M. **Bayesian networks: an introduction**. Chichester: John Wiley Sons Ltd, 2009. 347p.

MARGARITIS, D. **Learning bayesian network model structure from data**. 126f. PhD Thesis (Doctor of Philosophy) - Carnegie Mellon University, Pittsburgh, Pennsylvania, 2003.

MARIANO, F. C. M. Q.; LIMA, R. R.; ALVARENGA, R. R.; RODRIGUES P. B.; P. B. Neural network committee to predict the AMEn of poultry feedstuffs. **Neural Computing and Applications**, v. 25, n.7-8, p.1903-1911, 2014.

MARIANO, F. C. M. Q.; LIMA, R. R.; RODRIGUES P. B.; ALVARENGA, R. R.; NASCIMENTO, G. A. J. Equações de predição de valores energéticos de alimentos obtidas utilizando meta-análise e componentes principais. **Ciência Rural**, v.42, n.9, p.1634-1640, 2012.

MARIANO, F.C.M.Q.; PAIXÃO, C. A.; LIMA, R. R.; ALVARENGA, R. R.; RODRIGUES P. B.; NASCIMENTO, G. A. J. Prediction of the energy values of feedstuffs for broilers using meta-analysis and neural networks. **Animal**, v.7, n.9, p.1440-1445, 2013.

NAGARAJAN, R., SCUTARI, M; LÈBRE, S. **Bayesian Networks in R with Applications in Systems Biology**. Springer, New York. 2013. 157p.

NASCIMENTO, G.A.J.; RODRIGUES, P.B.; FREITAS, R.T.F.; REIS NETO, R.V.; LIMA, R.R.; ALLAMANN, I.B. Equações de predição para estimar valores da energia metabolizável de alimentos concentrados energéticos para aves utilizando meta-análise. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.63, n.1, p.222-230, 2011.

NASCIMENTO, G.A.J.; RODRIGUES, P.B.; FREITAS, R.T.F.; BERTECHINI, A.G.; LIMA, R.R.; PUCCI, L.E.A. Equações de predição para estimar os valores energéticos de alimentos concentrados de origem vegetal para aves utilizando a metanálise. **Revista Brasileira de Zootecnia**, v.38, n.1, p.1265-1271, 2009.

NEAPOLITAN, R. E. **Learning Bayesian Networks**. Upper Saddle River, NJ: Prentice Hall, 2003. 693p.

PEARL, J. **Probabilistic reasoning in intelligent systems: networks of plausible inference**. San Francisco: Editor Morgan Kaufmann, 1988. 552p.

PERAI, A. H.; MOGHADDAM, H. N.; ASADPOUR, S.; BAHRAMPOUT, J.; MANSOORI, Gh. A comparison of artificial neural networks with other statistical approaches for the prediction of true metabolizable energy of meat and bone meal. **Poultry Science**, v.89,n.1, p.1562-1568, 2010.

R CORE TEAM (2018). **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <http://www.R-project.org/>. Acesso em 25 de marco de 2018.

RODRIGUES, P. B.; ROSTAGNO, H. S.; ALBINO, L. F. T.; GOMES, P. C.; BARBOZA, W. A.; SANTANA, R. T. Valores energéticos do milho, do milho e subprodutos do milho, determinados com frangos de corte e galos adultos. **Revista Brasileira de Zootecnia**, v.30, n.6, p.1767-1778, 2001.

RODRIGUES, P. B.; ROSTAGNO, H. S.; ALBINO, L. F. T.; GOMES, P. C.; NUNES, R. V.; TOLEDO, R. S. Valores energéticos da soja e subprodutos da soja, determinados com frangos de corte e galos adultos. **Revista Brasileira de Zootecnia**, v.31, p.1771-1782, 2002.

SCUTARI, M., DENIS, J.B. **Bayesian Networks with Examples in R**. Taylor Francis Group. 2015. 221p.

SCUTARI, M. Learning Bayesian Networks with the bnlearn R Package. **Journal of Statistical Software**, v.35, n.3, p.1-21, 2010.

SPIRITES, P.; GLYMOUR, C.; SCHEINES, R. **Causation, Prediction, and Search**. Second edition, Cambridge: MIT Press, 2000.

TSAMARDINOS, I.; BROWN, L.E.; ALIFERIS, C.F. The max-min hill-climbing Bayesian network structure learning algorithm. **Machine Learning**, v.65, n.1, p.31-78, 2006.

TSAMARDINOS, I.; BROWN, L.E.; ALIFERIS, C.F. Algorithms for large

scale markov blanket discovery. In: "Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference". **Proceeding...** 2003, p.376-381. AAAI Press. 2003.

ZHAO, F.; ZHANG, H. F.; HOU, S. S.; ZHANG, Z. Y. Predicting metabolizable energy of normal corn from its chemical composition in adult pekin ducks. **Poultry Science**, v.87, n.1, p.1603-1608, 2008.

4 CONSIDERAÇÕES

**CONSIDERAÇÕES FINAIS:
REDES BAYESIANAS NA PREDIÇÃO DE VALORES ENERGÉTICOS
DE ALIMENTOS PARA AVES**

Considerações Gerais

Por meio de todas as buscas na base *Web of Science* observa-se que a maioria das publicações sobre redes bayesianas estão relacionadas à área de Ciência da Computação. Nas áreas aplicadas, principalmente agropecuária, ainda se têm pouquíssimas publicações ou nenhuma. Conclui-se que redes bayesianas é uma linha de pesquisa inédita na nutrição de aves e que pode ser estudada por pesquisadores que tem o interesse na predição dos valores de energia metabolizável.

As equações propostas obtidas ao utilizar redes bayesianas podem ser utilizadas pela indústria de frangos de corte na elaboração de rações, pois apresentaram acurácia na predição de EMAn. As equações propostas por meio de redes bayesianas são apresentadas por $EMAn_{i,j,k}$. Tem-se que o índice i refere-se a categoria, se 1 - energético ou 2 - proteico. O índice j refere-se a classificação do alimento, se $i = 1$ então j : 1 - milho ou 2 - subprodutos de milho ou 3 - outros energéticos; se $i = 2$ então j é 1 - soja ou 2 - subprodutos de soja ou 3 - outros proteicos. Para k , o valor de 1 refere-se a pinto e 2 ao galo.

Tabela 1 – Equações obtidas por meio de redes bayesianas.

Equações propostas
$EMAn_{1,1,1} = 3658,16 - 2,41PB - 11,25EE + 83,41MM + 16,76FB$
$EMAn_{2,1,1} = 3684,83 - 19,84PB - 71,15EE + 18,14MM - 8,93FB$
$EMAn_{1,2,1} = 4209,57 - 34,56PB + 32,84EE - 25,15MM - 142,57FB$
$EMAn_{2,2,1} = 2951,05 + 0,09PB + 37,96EE + 5,04MM - 17,60FB$
$EMAn_{1,3,1} = 4335,88 - 50,91PB + 35,40EE - 67,35MM - 87,06FB$
$EMAn_{2,3,1} = 2327,69 + 24,23PB + 77,72EE - 167,06MM - 22,28FB$
$EMAn_{1,1,2} = 3321,82 + 51,31PB + 39,42EE - 377,11MM + 113,92FB$
$EMAn_{2,1,2} = 4143,45 - 3,18PB - 43,45EE - 213,55MM + 6,71FB$
$EMAn_{1,2,2} = 4716,45 - 227,63PB + 144,47EE$
$EMAn_{2,2,2} = 518,54 + 26,25PB + 47,10EE + 184,42MM + 69FB$
$EMAn_{1,3,2} = 4133,39 - 89,45PB + 100,32EE - 5,5MM - 96,37FB$
$EMAn_{2,3,2} = 6033,28 - 15,02PB - 105,81EE - 556,50MM + 91,23F$

Estas equações (Tabela 1) foram validadas com dados provenientes de ensaios metabólicos e apresentaram precisão e acurácia na predição dos valores energéticos. As equações estão disponíveis em uma calculadora que pode ser instalada em celulares, tablets e computadores.

Uma nova abordagem metodológica para obter equações de predição foi proposta nesta tese, a qual empregou incertezas na obtenção de equações a partir de resultados de um conjunto de redes bayesianas híbridas. As estimativas provenientes de médias e modas dos coeficientes advindos de distribuições empíricas construídas com 10000 redes bayesianas tiveram melhores desempenhos. As equações propostas mostraram resultados precisos à medida que foram avaliadas com os dados de ensaios metabólicos. Assim, contribui-se

tanto em termos metodológicos quanto em termos práticos na área de Nutrição de aves. Sendo assim, as equações selecionadas com a proposta foram: $EMAn = 3668.345 - 2.12 PB - 20.96 EE + 83.95 MM + 32.41 FB$ (para milho), $EMAn = 1969.02 + 3.14 PB - 12.04 EE + 25.63 MM + 34.33 FB$ (para soja), $EMAn = 3673.48 + 34.59 PB + 22.25 EE - 51.31 MM - 156.86 FB$ (subprodutos de milho), $EMAn = 1592.81 + 10.99PB + 51.78EE + 118.24MM - 6.96FB$ (subprodutos de soja), $EMAn = 4296.14 - 49.11PB + 35.21EE - 64.50MM - 88.74FB$ (outros energéticos) e $EMAn = 2295.08 + 25.23 PB + 76.90 EE - 168.76 MM - 21.90 FB$ (outros proteicos). Os valores de EMAn obtidos por meio destas equações via nova abordagem metodológica se encontram nos intervalos obtidos por comitês de redes neurais propostos por Mariano et al. (2014).

Para trabalhos futuros pretende-se:

- Atualizar a base de dados utilizada por Nascimento et al. (2009, 2011) e Mariano et al. (2012, 2013), buscando aumentar a representatividade dos dados utilizados na predição da EMAn, bem como avaliar a acurácia das predições por meio das redes bayesianas;
- Avaliar o efeito da imputação dos valores de fibra detergente ácido e fibra detergente neutro (*missing data*), pelo fato desses valores não estarem disponíveis na maioria dos experimentos;
- Predizer a EMAn de alimentos utilizados na formulação de rações para frangos de corte, como função da composição química desses alimentos, sob diferentes procedimentos de RB, a saber:
 - Utilizar RB para selecionar as variáveis e obter o ajuste a partir de redes neurais artificiais (FELIPE et al., 2015);
 - Utilizar RB para obter as redes e classificá-las em termos de capacidade de ajuste e habilidade preditiva por meio das equações estruturais (TÖPNER et al., 2017);
 - Utilizar diferentes combinações de testes de independência e de pontuação nos algoritmos de aprendizagem conforme Töpner et al. (2017);
 - Implementar a medida L (IBRAHIM e LAUD, 1994) como ferramenta de seleção dos modelos de RB.
 - Trabalhar com diferentes distribuições de probabilidades para as variáveis, com a finalidade de obter equações de predição, sendo que neste caso têm-se pesquisa em termos teóricos sofisticada para

obter as distribuições conjuntas. Além de implementar os resultados em um pacote específico no software R para lidar com elas, pois observa-se que essa metodologia ainda requer pesquisas no campo metodológico estatístico e computacional para que sua aplicação se torne cada vez mais eficiente. Essa eficiência é esperada desde a obtenção da rede, estimação dos parâmetros, predição das variáveis e otimização do tempo gasto nas análises.

- Desenvolver produto tecnológico inovador a partir da proposta metodológica apresentada no Artigo 3 desta tese, com o objetivo de obter equações de predição para qualquer Área do Conhecimento que se tenha interesse, seja nas Ciências Exatas e da Terra, Ciências Biológicas, Engenharias, Ciências da Saúde, Ciências Agrárias, Ciências Sociais Aplicadas, Ciências Humanas e Linguística, Letras e Artes. Pois a partir dessa proposta já se obtêm as equações de melhores ajustes, não havendo a necessidade da seleção de vários modelos, conforme feito no Artigo 2 desta tese.