UNIVERSIDADE FEDERAL DE LAVRAS

CRISTIANO MESQUITA GARCIA

INCREMENTAL MISSING DATA IMPUTATION VIA MODIFIED GRANULAR EVOLVING FUZZY MODEL

LAVRAS – MG

2018

CRISTIANO MESQUITA GARCIA

INCREMENTAL MISSING DATA IMPUTATION VIA MODIFIED GRANULAR EVOLVING FUZZY MODEL

Dissertation submitted in partial fulfillment of the requirements of the Graduate Program in Systems and Automation Engineering for the Degree of Master of Science at the Federal University of Lavras.

Daniel Furtado Leite
Supervisor

Ahmed Ali Abdalla Esmin
Co-Supervisor

LAVRAS – MG
2018

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Garcia, Cristiano Mesquita.
Incremental Missing Data Imputation via Modified Granular Evolving Fuzzy Model / Cristiano Mesquita Garcia. - 2018.
71 p. : il.

Orientador(a): Daniel Furtado Leite.
Coorientador(a): Ahmed Ali Abdalla Esmin.
Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2018.
Bibliografia.

1. Evolving Intelligence. 2. Fuzzy Systems. 3. Missing Data Imputation. I. Leite, Daniel Furtado. II. Esmin, Ahmed Ali Abdalla. III. Título.

CRISTIANO MESQUITA GARCIA

INCREMENTAL MISSING DATA IMPUTATION VIA MODIFIED GRANULAR EVOLVING FUZZY MODEL

Dissertation submitted in partial fulfillment of the requirements of the Graduate Program in Systems and Automation Engineering for the Degree of Master of Science at the Federal University of Lavras.

APPROVED on July 17th, 2018.

Prof. Heloisa de Arruda Camargo UFSCAR, Brazil
Prof. Marcos Evandro Cintra UFERSA/UNIFESP, Brazil

Daniel Furtado Leite
Supervisor

Ahmed Ali Abdalla Esmin
Co-Supervisor

LAVRAS – MG
2018

À minha família e à glória de Deus.
A todos mencionados nos agradecimentos,
Dedico.

AGRADECIMENTOS

Agradeço a Deus, pelo dom da vida;
Aos meus professores orientadores Daniel e Ahmed, por compartilhar ensinamentos, pela sabedoria, incentivos e diálogos;
Aos meus pais Rudimar e Luisa; e minha irmã Marcela, pelo apoio incondicional em todos os momentos;
À Thati, pelo amor, companheirismo e conselhos;
Às minhas avós Diva e Nenzinha, pelas orações diárias; aos meus tios e tias, primos e primas;
Aos meus avôs Nelo e Lourival, que sempre acreditaram em mim, e em algum lugar, também comemoram este momento comigo;
Aos demais professores do PPGESISA, em especial Sílvia e Bruno, com quem tive mais contato durante o mestrado;
Aos colegas e irmãos acadêmicos do programa de Mestrado em Engenharia de Sistemas e Automação, turma 2017/1; e Eduardo, Diego, Ricardo e Stella, por tornarem mais leve este potencial início de caminhada acadêmica;
Aos colegas de trabalho da DGTI, pelo incentivo, compreensão e amizade;
A todos os meus amigos, em especial os Produceiros, Guilherme e Felipe, da república onde moro, e os amigos do Emaús de Lavras;
À Universidade Federal de Lavras pela estrutura física e à sociedade brasileira, de onde saem os impostos que pagam a educação pública de qualidade.

A todos vocês, o meu MUITO OBRIGADO!

In the middle of the journey, I found out that incremental learning systems replicate the human behavior: they learn from mistakes, adapt to new situations, and forget memories that are not useful anymore.
(The author)

"Whether you think you can, or you think you can't – you're right."
(Henry Ford)

ABSTRACT

Large amounts of data have been produced daily. Extracting information and knowledge from data is meaningful for many purposes and endeavors, such as prediction of future values of time series, classification, semi-supervised learning and control. Computational intelligence and machine learning methods, such as neural networks and fuzzy systems, usually require complete datasets to work properly. Real-world datasets may contain missing values due to, e.g., malfunctioning of sensors or data transfer problems. In online environments, the properties of the data may change over time so that offline model training based on multiple passes over data is prohibited due to its inherent time and memory constraints. This study proposes a method for incremental missing data imputation using a modified granular evolving fuzzy model, namely evolving Fuzzy Granular Predictor (eFGP). eFGP is equipped with an incremental learning algorithm that simultaneously impute missing data and adapt model parameters and structure. eFGP is able to handle single and multiple missing values on data samples by developing reduced-term consequent polynomials and relying on information of time-varying granules. The method is evaluated in prediction and function approximation problems considering the constraints of online data stream. Particularly, the underlying data streams may be subject to missing at random (MAR) and missing completely at random (MCAR) types of missing values. Predictions given by the model evolved after data imputation are compared to those provided by state-of-the-art fuzzy and neuro-fuzzy evolving modeling methods in the sense of accuracy. Results and statistical comparisons with other approaches corroborate to conclude that eFGP is competitive as a general evolving intelligent method and overcomes its counterparts in MAR and MCAR scenarios according to an ANOVA-Tukey statistical hypothesis test.

Keywords: Evolving Intelligence. Fuzzy Systems. Data Stream. Incremental Learning. Missing Data Imputation

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Information uncovered from data can be of a great value to researchers and to a variety of human endeavors. Such information may be useful for detecting trends, understanding customers, predicting sales or the weather, identifying patterns in dynamic systems, to mention a few. Knowledge discovery from datasets requires data analysis tools, statistical methods, data mining and/or machine learning techniques.

Data mining and machine learning techniques usually require the complete set of data to be available to work properly, that is, to perform prediction, classification or data clustering. However, missing values are common in real-world problems. Missing values occur due to incomplete observations, data transfer problems, malfunction in sensors and devices, incomplete information given in surveys (BOYLES, 2011)(PIGOTT, 2001).

The presence of missing data in a dataset may be highly undesirable (NELWAMONDO; GOLDING; MARWALA, 2013). It may put in risk a whole project if they are not handled correctly (AYDILEK; ARSLAN, 2012)(NUOVO, 2011). There are simple approaches to deal with missing data. Although they are not the best in the sense of accuracy, they can be useful in many situations. Completing the missing values with zeros, with the mean value for the attribute, or even ignoring the incomplete observation are some of the simplest approaches to the problem. A critic for the latter is that incomplete observations can still carry important information about a process or a phenomenon, and should not be discarded.

There exist several approaches from machine learning and statistical theories to tackle the missing data problem by estimating and imputing values to complete the observations in a dataset (FARHANGFAR; KURGAN; PEDRYCZ, 2007)(PIGOTT, 2001)(LUENGO; GARCÍA; HERRERA, 2012). Then, complete observations can be delivered to traditional machine learning methods. Among the approaches to impute data, one may find methods such as Linear and Polynomial Interpolation, Gaussian Process Regression, Correlation-based imputation, Nearest-Neighbors, Neural Networks, Genetic Algorithms, and Fuzzy systems (AYDILEK; ARSLAN, 2013)(AZIM; AGGARWAL, 2016)(LI; GU; ZHANG, 2013)(SARAVANAN; SAILAKSHMI, 2015)(KUPPUSAMY; PARAMASIVAM, 2017)(ZHANG et al., 2016)(SAMAT; SALLEH, 2016).

There are some studies on the use of fuzzy methods to perform missing data imputation. A key point of the present study is to develop a fuzzy rule-based prediction and imputation method. Fuzzy imputation methods are discussed in Section 3.3. The aforementioned statistical and intelligent methods handle missing data in closed datasets in an offline way. None of them deals with online data streams in an incremental fashion, which can impose several additional constraints to the imputation problem.

An evolving fuzzy granular rule-based model with modified structure equipped with an incremental learning algorithm for missing data imputation and model adaptation is presented in this study. The evolving Fuzzy Granular Predictor (eFGP) is useful for function approximation and prediction problems. Different from the aforementioned methods, eFGP imputation is suited for time-varying environment and therefore manages the restrictions it imposes: data stream's potentially infinite size, which brings the need to process and immediately discard data samples; data often arrive in a high sampling frequency; data may be unsorted and may have its distribution changing over time (SILVA et al., 2013)(GAMA, 2010). The eFGP rule base is constructed from scratch based on nonstationary data that can be missing at random (MAR) and missing completely at random (MCAR). eFGP handles single missing values on data samples by developing reduced-term consequent polynomials. Additionally, multiple missing values are dealt with using the essence of time-varying granules evolved in the data space.

1.1 Motivations

Missing data is a common phenomenon in real-world engineering and computer science applications. Several factors may yield missing data, such as sensors failure, communication problems, incomplete answered surveys, among others. As most of the algorithms to construct prediction, classification and regression models need complete data, missing data is highly undesirable.

Generally, missing data imputation is a complicated problem even when dealing with static and closed datasets. Sometimes the data is generated by highly nonlinear processes or by time-varying probability distributions. There are some approaches that use machine learning techniques to impute values when some are missing. However, most of them applies multiple passes over the dataset (batch training), which is not allowed in online environments. Dealing with missing data in online streams is a complex task since constraints such as single-pass learning on a per-sample basis, potential existence of an infinite number of observations, and limited processing time and memory, are very strict.

The main motivation for considering the problem of missing data in online data streams is that it is fundamental in a diversity of real-world problems whenever models and real-time processing are needed. Big data challenges (WALKER, 2014)(LV et al., 2015) have stimulated the evolution of models as the data occur.

1.2 Objectives

The objective of this study is to develop a fuzzy rule-based method for missing data imputation in online data streams. The purpose of the fuzzy model is to capture the essence of the information on the data to give predictions of future values and an associated linguistic description.

1.3 Contributions

The main contribution of this study is a new method for missing data imputation in online environments, able to cope with nonstationary data streams which may contain single and multiple missing at random and missing completely at random data. The method is able to construct hybrid (functional and linguistic) fuzzy rules with a different type of rule consequent. It handles the intrinsic characteristics of nonstationary real-time environment. In other words, the proposed fuzzy model and incremental algorithm to deal with missing data supports real-time modeling as it does not scale with the size of the data stream, handles concept changes while at the same time keeps a reasonable and competitive prediction accuracy compared to other evolving intelligent models and algorithms.

1.4 Organization of the manuscript

The remaining of this manuscript is organized as follows:

- **Chapter 2** discusses the main concepts and characteristics of incremental machine learning methods, including nonstationary data streams and fundamentals of evolving fuzzy systems;
- In **Chapter 3**, the types of missing data are discussed. A review on fuzzy approaches to tackle the problem of missing data and adaptation mechanisms for fuzzy set-based evolving modeling are given;
- In **Chapter 4**, the proposed method for missing data imputation and development of prediction models is addressed;

- **Chapter 5** presents the methodology of the experimental study;
- **Chapter 6** brings the results, comparisons with other intelligent evolving systems, and a statistical hypothesis test regarding MAR and MCAR scenarios;
- In **Chapter 7**, the conclusion and an outlook on further research are given.

2 INCREMENTAL MACHINE LEARNING

In this chapter, concepts and definitions of incremental machine learning are given. Moreover, characteristics of nonstationary data streams are leveraged in order to emphasize the constraints of online environment and how incremental learning and time-varying models are able to handle such constraints.

In the last years, the amount of information has increased dramatically. Every day, millions of transactions occur; new data from a diverse range of areas are made available on the Internet and from many devices, sensors; industrial/commercial and virtual settings; and so on. Humanity has witnessed an exponential growth of the amount and rate of appearance of new information (LEITE, 2012)(ANGELOV, 2012)(TAN et al., 2013). Nevertheless, information is not knowledge.

Intelligence from the Computational Intelligence perspective mainly refers to continuous adaptation of models to better perform a specific task (KASABOV, 2007)(BOUCHACHIA; GABRYS; SAHEL, 2007). An intelligent system is a system that is provisioned with an adaptable learning algorithm (BOUCHACHIA; GABRYS; SAHEL, 2007)(ANTSAKLIS; PASSINO, 1993). Adaptable learning systems may have a number of names in the literature, such as Incremental Learning Systems (BOUCHACHIA; GABRYS; SAHEL, 2007)(TSAI; AGGARWAL; HUANG, 2015)(AL-KHATEEB et al., 2016)(GUNN; ARNAIZ-GONZÁLEZ; KUNCHEVA, 2018), Evolving Intelligent Systems (EIS) (KASABOV, 2007)(ANGELOV; LUGHOFER; ZHOU, 2008)(ANDONOVSKI et al., 2018)(LEITE et al., 2015)(RUBIO, 2017) and Evolvable Systems (PEDRYCZ, 2010)(MOSHTAGHI et al., 2015).

The computational intelligence systems developed in the last decades generally depend on gradient or Newtonian-based techniques, least squares and dataset partitioning by clustering to perform classification, clustering, regression and related tasks. These techniques assume that some data are known in advance (ANGELOV, 2012), what may not occur in online environment, where the data may change its characteristics over time. In other words, the aforementioned intelligent approaches are not directly applicable to data streams (ANGELOV, 2012). New challenges that data streams bring to intelligent systems include:

- Handling large amounts of data;
- Data modeling in an online and real-time fashion;
- Ability to adapt models to new situations;

- Efficiency of algorithms to attend time and memory constraints;
- Preservation of the interpretability and transparency of models.

Incremental learning refers to the ability of an algorithm to adapt models to new situations (BOUCHACHIA; GABRYS; SAHEL, 2007). We will call these models Evolving Models henceforth when they:

- Learn online on a sample-per-sample basis;
- Do not need previous data to operate;
- Do not require prior knowledge about the structure of the model (a rule base or neural network);
- Are able to self-adapt their parameters and structure;
- Dispense prior knowledge about the statistical properties of the data;
- Start learning from scratch.

According to Kasabov (2007), Leite et al. (2012), Ditzler & Polikar (2013), an evolving intelligent model updates its own functionality, knowledge and structure continuously.

2.1 Non-stationary Data Stream

In recent years, technological advances and the popularization of electronic devices have increased the generation and sharing of data. Millions of simple transactions occur everyday. Terabytes of data are produced from social networks, search engines, electronic devices, satellites, industrial sensors, health care systems, to mention a few. Large data streams (DS), known by their characteristic of being potentially infinite, have become common.

According to Subbian & Aggarwal (2017), large data stream processing leads to computational and mining issues, such as efficiency, where multiple passes over the data are quite often unfeasible and a single pass is the ideal. The characteristic of sequential data may change over time. Gama et al. (2014) stated that adaptive learning refers to updating predictive models online during their operation to react to concept changes.

Silva et al. (2013) and Gama (2010) state that DS mining is challenging due to its constraints and characteristics, such as:

- Data arrives continuously;
- Data may arrive unordered, e.g. if they come from different sources, are influenced by a random factor, or result from a switched dynamical system;
- DS is potentially infinite;
- Data should be processed and discarded after a short period of time;
- Data and its distribution may change over time, indicating nonstationary behavior.

Incremental learning models should have mechanisms to incorporate concept change, forget outdated data, and adapt to the most recent state of a system (GAMA, 2010)(DOVŽAN; LOGAR; ŠKRJANC, 2015)(SOARES et al., 2017)(LOPES; CAMARGO, 2017)(MACIEL; BALLINI; GOMIDE, 2018). In other words, incremental machine learning algorithms should manage process changes over time and, as processes changes, its model should be promptly updated.

Domingos & Hulten (2001) also reported some properties that are desirable in incremental machine learning systems for data streams. They are: online learning; polynomial scalability to samples, attributes and number of local models (rules, neurons or prototypes); single-pass on the training set; and concept drifts should be taken into account. It is important to take some aspects of the environment into consideration, specially the nonstationarity of online environment, which means that the distribution of the data may change over time and the current incoming data may have low correlation with previous data.

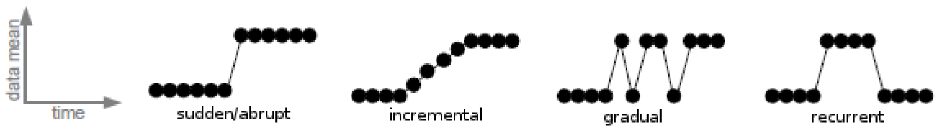
2.1.1 Concept Changes

In dynamic environment, two main types of concept changes can be mentioned: concept drift and concept shift. *Concept drift* refers to the gradual change of the data distribution over time. New concepts or concepts previously known that disappeared and re-appeared after a period of time or even concepts that switch between/among concepts (GAMA et al., 2014) are included in concept drift cases. *Concept shift* refers to the abrupt change of the data distribution (GAMA, 2010). For example, the mean value of a class suddenly changes, or a new class or behavior arises in the data stream.

In Gama et al. (2014), four types of concept change that may occur in an online environment are listed. They are illustrated in Figure 2.1 and described below.

- **Sudden/Abrupt:** occurs when a concept shifts to another concept;
- **Incremental:** occurs when a concept softly switches to another, introducing intermediate points between the concepts;
- **Gradual:** occurs when a concept is toggled to another until becoming the last concept;
- **Recurrent:** occurs when a concept temporarily changes to another.

Figure 2.1 – Types of concept changes.
Source: adapted from Gama et al. (2014)



A challenge is to not mix a real concept change with either *outlier* or noise, as in these cases there is no need for adaptation. In case outlier or noise is taken into account, the updated model becomes less accurate at least temporarily. In real-world problems, a mix of different types of concept changes are observed, therefore additional difficulties are imposed to incremental learning algorithms and evolving modeling.

2.1.2 Temporal Spans

Temporal span is a technique strictly related to the idea of "forgetting" old data in order to maintain a model updated. In a nonstationary context, temporal span is a sort of partial memory, where only the most recent data, in a *first-in first-out* structure, define a time window over the data stream.

Temporal spans can help keeping a model updated. A main point is how to choose an adequate window size. A small window size can help a fast reaction to concept changes. On the other hand, a larger window size cannot react fast to concept changes, but can offer stability and robustness. Therefore, a tradeoff has to be considered.

Approaches to choose the window length may consider fixed size or adaptive window size. Fixed size defines a fixed number of the most recent examples. When a new example arrives, the oldest in the window is discarded. Windows with adaptive size may be relevant to deal with concept changes, expanding if a change is recognized into the window, or shrinking otherwise (GAMA, 2010).

Granular computing theory provides methods for temporal granulation. Temporal granulation may slow down the data flow as several instances may be encapsulated by a granular object. From this point, operations are done over the granules instead of data. Among the benefits of the granular approach are a smaller amount of data "generated" for a posterior space granulation; and synchronization of data streams with different sampling intervals is granted (LEITE; COSTA; GOMIDE, 2012b).

2.2 Recent Proposals on Evolving Fuzzy Modeling

Fuzzy sets refer to a class of objects that can assume partial membership in a set and can be used to deal with uncertainty (ZADEH, 1965). Various models and methods use fuzzy concepts and ideas. For example, Fuzzy C-Means (FCM) is a clustering method where the data, in oppose to nonfuzzy (crisp) clustering methods, may belong in some degree to all existing clusters (BEZDEK; EHRLICH; FULL, 1984).

Fuzzy models consist in a set of rules that can generalize and summarize knowledge. Fuzzy models can be interpreted in a linguistic way. Thus, they are intuitive and convenient to formalize and aggregate knowledge extracted from data and from experts.

A fuzzy rule R is a linguistic statement of the form

$$R_i : IF\ (antecedent)\ THEN\ (consequent),$$

where $i = 1, ..., c$; and c is the number of rules in a rule base. The *antecedent* and *consequent* terms of the rules are different for different types of fuzzy models. They can be linguistic or functional expressions. The most common types of fuzzy rules are known as Mamdani and Takagi-Sugeno types (ANGELOV, 2012). Many modeling approaches for prediction take advantage of fuzzy rules.

Lughofer (2008) presented FLEXFIS, a data-driven incremental learning approach for evolving Takagi-Sugeno fuzzy models. The focus was on providing a connection between the adaptation of antecedent and consequent nonlinear parameters. FLEXFIS has performed better in terms of accuracy than well-known methods such as eTS, Dynamic Evolving Neural-Fuzzy Inference System (DENFIS) (KASABOV; SONG, 2002) and Generalized Dynamic Fuzzy Neural Network (GDFNN) (WU; ER; GAO, 2001).

Leite, Costa & Gomide (2010) introduced a framework that consists of an evolving granular neural network (eGNN). An eGNN is capable of generating and updating a neural model by processing nonstationary data stream using a one-pass incremental algorithm. eGNN has obtained good results for classification in experiments simulating nonstationary environments, combining labeled and unlabeled data, and predicting future values of time series (LEITE; COSTA; GOMIDE, 2012a).

Dovžan & Škrjanc (2010) proposed a Takagi-Sugeno evolving fuzzy method (eFuMo) for modeling physical processes. eFuMo has three major components: a central decision logic, responsible for deciding between either evolving new rules or adapting parameters of local models; procedures for changing the rule base; and procedures for drifting coefficients. eFuMo has achieved competitive performance compared to those of other existing online learning methods, such as DENFIS and eTS considering benchmark problems. eFuMo was also implemented for monitoring and fault detection in water-waste treatment plants, achieving good results using data from a real process (DOVŽAN; LOGAR; ŠKRJANC, 2015)

Pratama et al. (2014) presented PANFIS, a neuro-fuzzy inference system capable of dealing with nonstationary data streams. PANFIS starts learning with an empty rule base. The rules are generated from scratch, as the data are input. The rules are created and removed from a rule base when a new incoming pattern arises. According to the authors, a notable characteristic of PANFIS is the ability of developing ellipsoids in arbitrary positions connected with a projection concept to form the antecedent, linguistic part of the rules. PANFIS has obtained good results compared to other fuzzy modeling methods, such as evolving Takagi-Sugeno (eTS) (ANGELOV et al., 2004) and Adaptive Neuro Fuzzy Inference System (ANFIS) (JANG; SUN; MIZUTANI, 1997).

Angelov (2014) proposed a framework for data analytics called TEDA, which stands for Typicality and Eccentricity based Data Analytics. In TEDA, typicality refers to good examples of a concept; whereas eccentricity gives the idea of an outlying, abnormal observation. TEDA is based on data sequences and their distribution in the data space; it does not depend on predefined parameters or prior knowledge. TEDA can perform tasks such as classification, clustering, prediction and function approximation.

Rubio (2017) described a stable neuro-fuzzy inference system based on multilayer neural network and fuzzy inference system, namely USNFIS. The approach can cope with big data. As a result, USNFIS achieved better results in terms of accuracy for learning from big data provided by nonlinear systems. The method overcame gradient-based and fuzzy inference system methods.

Angelov, Gu & Príncipe (2017) proposed an autonomous learning system for data streams called Autonomous learning of multi-model systems (ALMMo). ALMMo is a non-parametric and data-driven method that develops data cloud-type of fuzzy rules. This system has specific characteristics: (i) it uses density and typicality (ANGELOV; GU; KANGIN, 2017) to reveal data pattern; (ii) it structure is formed by data clouds; (iii) unimodal density based membership functions are identified; and (iv) categorical and real data are supported. The authors claim that ALMMo can be applied to online data analytics, classification and prediction. Using three benchmark datasets, ALMMo has provided the best performance compared to other classification methods. For prediction, ALMMo overcame other methods in terms of an error metric.

3 MISSING DATA

In this chapter, the missing data problem is discussed. State-of-the-art methods to handle missing values in data samples are reviewed with emphasis on intelligent fuzzy and neural methods.

3.1 Preliminaries

Missing data or missing values occur due to various reasons such as malfunction in sensors, communication or transfer problems, informants that give incomplete information while filling in forms, to mention some examples (BOYLES, 2011)(PIGOTT, 2001). A simple example of dataset with missing values is given in Table 3.1. 9 out of 150 samples of the well-known 3-class Iris benchmark dataset are shown for short. Each column refers to an attribute or feature. While the second sample is a complete observation, the forth sample contains missing values for the features sepal lenght and petal width.

Table 3.1 – Sample of the dataset Iris with missing values

Sample	Sepal length	Sepal width	Petal length	Petal width	Class
1	-	3.2	1.4	0.2	Iris-setosa
2	5.3	3.7	1.5	0.2	Iris-setosa
3	5.0	3.3	1.4	0.2	Iris-setosa
4	-	3.2	4.7	-	Iris-versicolor
5	6.4	3.2	4.5	1.5	Iris-versicolor
6	6.9	3.1	4.9	-	Iris-versicolor
7	5.7	2.5	5.0	2.0	Iris-virginica
8	5.8	2.8	-	2.4	Iris-virginica
9	6.4	-	5.3	2.3	Iris-virginica

Slightly different definitions and interpretations about missing data mechanisms may be found across research communities. Based on the nature of the data and the way it is produced, there are three types of missing data:

- Missing completely at random (MCAR), where the propensity for a data point to be missing is completely random, i.e., the probability of a given value to be missing is equal to the remaining values;

- Missing at random (MAR), where the propensity of the values of a specific attribute to be missing is higher than those of other attributes, e.g. a sensor is not working properly or a question in a survey is harder to be answered than others;

- Not missing at random (NMAR), where the missing values are dependent on other missing values, e.g., a sensor is not working properly in a specific range of values or people tend not to answer a question in a survey if their depression level is high.

In the majority of the cases, missing data are of the MAR type (AMIRI; JENSEN, 2016). Methods to tackle the problem of MAR and MCAR data are summarized in the next sections.

3.2 General Aspects of Missing Data Imputation

When searching a good strategy to perform imputation, the nature of the data should be taken into account. For nominal attributes (words), for example, there are approaches such as global most common (GMC), where the missing value is fulfilled with the most common value for the attribute (GRZYMALA-BUSSE et al., 2005).

When dealing with numerical data, there is a variety of approaches to tackle the problem of missing data. Some are simpler and computationally faster whereas others are more complex. For example, the simplest approaches include: replacing missing values with 0 (zero) (TROYANSKAYA et al., 2001); or replacing missing values with the mean for the attribute. In some cases, the incomplete observation is removed from the dataset (ZHU et al., 2001).

The consequences of using the aforementioned techniques is that compelling the mean or zero to the missing value may "generate" an inaccurate dataset with distorted covariances and correlations. If this dataset is applied to a machine learning system, an inaccurate model can be obtained, which may jeopardize a research or experiment, leading to erroneous conclusions. By removing the incomplete observations from the dataset, substantial information may be lost, and again, the dataset may not represent the dynamic behavior of a system, process or phenomenon in a reliable way.

On the other hand, model-based approaches grounded on statistical and machine learning fundamentals for missing data imputation have emerged since 1986 (LITTLE; RUBIN, 1986). Simple statistical and interpolation methods may make the dataset biased. Nonlinear model-based imputation methods focus on replacing the missing value possibly with the most accurate value regarding the essence of the information obtained from previous data are complex (AMIRI; JENSEN, 2016).

Several model-based methods supported by machine learning, computational intelligence and data mining fundamentals have been reported, e.g., nearest neighbors (BATISTA; MONARD, 2003)(TROYANSKAYA et al., 2001), feedforward neural networks (GARCÍA-LAENCINA et al., 2007)(BENGIO; GINGRAS, 1996), fuzzy inference systems (JENSEN; CORNELIS, 2011), meta-heuristics (AYDILEK; ARSLAN, 2013), among others.

An important aspect of missing data problem to be emphasized in the present study is the nonstationarity of the dataset. Is the dataset closed and completely available beforehand or is it time-varying and gradually available by means of a data stream? The answer is decisive to choose a method for the underlying problem. The vast majority of the model-based missing data imputation methods, including those mentioned in the previous paragraph, are not able to cope with the constraints imposed by online and nonstationary data streams. The present study considers Fuzzy Set-based Evolving Modeling (FBeM) (LEITE et al., 2012) as a basic framework. In this manuscript, the FBeM framework was modified structurally to consider Takagi-Sugeno consequent polynomials with a reduced term. Additionally, the FBeM learning algorithm was modified to be capable of handling multiple missing values on data samples.

Differently from any other study on missing data imputation, the proposed evolving framework does not assume prior knowledge about the properties of the data. A linguistic-functional fuzzy model is developed from scratch to give predictions of nonstationary time series while at the same time imputates values (if needed) and provides predictions using incomplete samples.

3.3 A Review on Fuzzy Methods for Missing Data Imputation

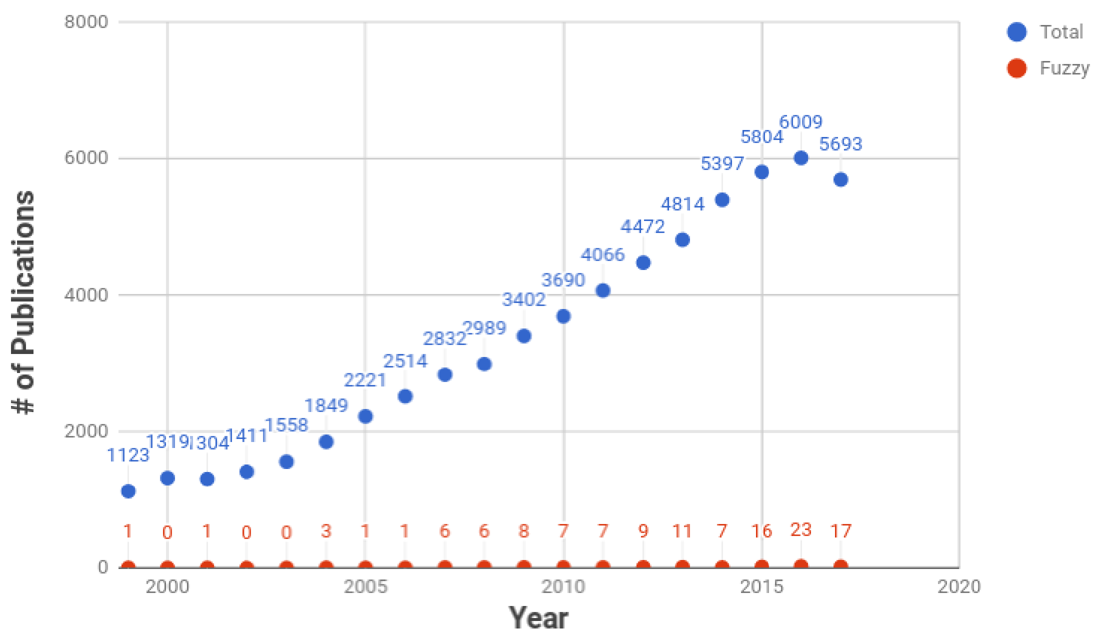
First, a search on the Scopus[1] and IEEExplore[2] databases, using the keywords: (fuzzy) AND (missing) AND (data OR values) AND (imputation).

For the search on Scopus, 115 documents[3] were retrieved, of which 55 were journal papers, 53 conference papers, 3 conference reviews, 1 book chapter, 2 articles in press and 1 review. Figure 3.1 shows the overall amount of publications related to missing data imputation in blue whereas in red are those also related to fuzzy sets.. Notice that the total number of

[1] https://www.elsevier.com/solutions/scopus
[2] http://ieeexplore.ieee.org
[3] Search performed on 2017-Aug-3

publications about missing data imputation has considerably increased over the years. The number of publications about missing data imputation using fuzzy methods, contrariwise, has not proportionally increased. Taking into account only the year of 2017, fuzzy methods applied for the problem represent only 0.003% of the cases.

Figure 3.1 – Publications per year about missing data and fuzzy, according to Scopus database.
Source: the author

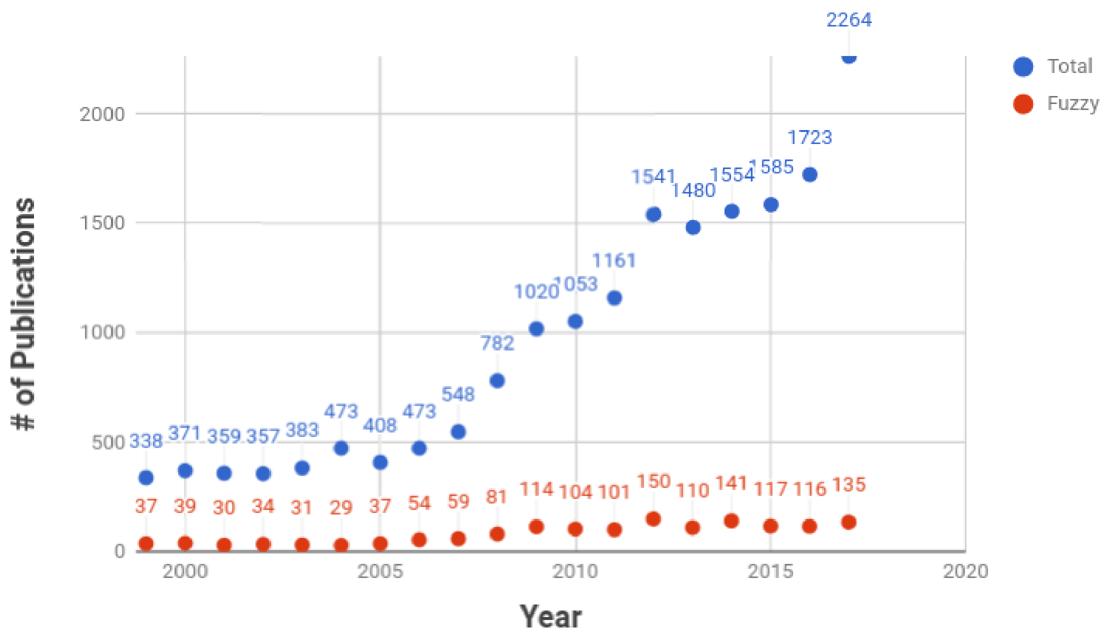


For the IEEExplore database, 32 articles were found[4], of which 30 articles from conference publications and only 2 were published in Journals. The amount of publications per year can be seen in Figure 3.2. Similar to Fig. 3.1, the number of studies on missing data imputation considerably increased over the years. However, fuzzy methods have contributed to that increasing only marginally. Taking these observations into account, there is room for fuzzy approaches to the problem of missing data imputation.

Recent methods closely related to that proposed in the present study are discussed next. It is important to highlight that these studies do not handle online data streams and do require datasets completely available to perform imputation. Computational intelligence methods, such as the Fuzzy C-Means (FCM) (RAZAVI-FAR; SAIF, 2016), sometimes used together with other methods, such as Multi-Layer Perceptron neural network (MLP) (AZIM; AGGARWAL, 2016), Support Vector Regression (SVR) and Genetic Algorithms (GA) (AYDILEK; ARSLAN, 2013) were considered to impute missing values in classification problems.

[4] Search performed on 2017-Aug-3

Figure 3.2 – Publications per year about missing data and fuzzy, according to the IEEExplore database. Source: the author



Nanni, Lumini & Brahnam (2012) described an approach based on random subspace that uses ensembles of models to handle classification with missing values. Random subspaces were created based on Input Decimated Ensembles (IDE) and Support Vector Machines (SVM). Subspaces were then combined using a sum rule. Model development consists of the following steps: normalization, clustering (using the Gustafson-Kessel clustering algorithm with a fixed number of clusters), mean imputation and classification. The approach, called Multiple Imputation Ensemble (MIE), was tested using 12 different datasets for classification. MIE was compared to 30 different methods and combinations of methods, varying the percentage of missing data from 5% to 50%. An extensive study was carried out, and the proposed method was considered as a good approach across the several datasets tested. MIE requires the dataset to be available, and subspaces are stationary after the training process.

Castillo & Cardeñosa (2012) proposed a new method that extends the fuzzy min-max neural network, in order to make this type of network able to deal with categorical data, not only numerical data. The authors applied the proposed method on opinion polls. To evaluate the method, two polls from Sociological Research Center's catalog were used, containing 16345 and 13280 interviews. The method was compared with logistic regression and with the same method using different distance measures: Euclidean, Logarithmic and Kronecker. The results leaded to the conclusion that Euclidean and Logarithmic distance are better than Kronecker distance, whereas between Euclidean and Logarithmic distances there was no significant difference. The authors also observed that the proposed method does not appear to be suitable for cases when the variables are only of categorical type.

Luengo, Sáez & Herrera (2012) presented an extensive analysis on fuzzy rule-based classification systems (FRBCS) for missing data imputation. The authors used three variants in their tests: a fuzzy hybrid-genetic-based machine learning method, a Mamdani-based FRBCS; a fuzzy rule learning model, another Mamdani-based FRBCS; the positive definite fuzzy classifier (PDFC), which is a Takagi-Sugeno-based FRBCS; all of them proposed in other studies. The authors combined 9 methods for the imputation step and compared all of them at the end of the process and tested them with 21 datasets with missing data by default. The authors concluded that the imputation method had positive effect on the behavior of the chosen FRBCS. The analysis was made using the measures: Wilson's noise ratio and the average mutual information difference.

Li, Gu & Zhang (2013) proposed a method called IGA-FCM for incomplete data clustering, acting as an imputation method. They used Nearest Neighbors (NN) with partial distance calculation, GA using real-valued chromosomes, and FCM to cluster the data. The authors tested the IGA-FCM, varying from 0% to 20% the percentage of missing data, using three well-known datasets: Iris, Wine and New-Thyroid, comparing results with other 4 methods. IGA-FCM achieved the best results among them.

Aydilek & Arslan (2013) presented a hybrid method that imputes missing values by using Fuzzy C-Means optimized by Support Vector Regression (SVR) and GA. The authors tested the proposed method using six different datasets: Iris; Wine; Glass, with 214 instances and 11 attributes; Haberman, with 306 instances and 3 attributes; Musk1 with 476 instances and 167 attributes and Yeast, with 1489 instances and 9 attributes. The tests were performed varying the percentage of missing data from 1% to 25%. The results were consistent when compared with FCM together with only GA, SVR together only with GA and imputation of zeros. The metric used to compare was Wilcoxon Sum Rank Test.

Tang et al. (2015) presented a hybrid approach for missing data imputation using FCM optimized by a genetic algorithm (GA). The authors argued that traditional FCM may have its effectiveness degraded by the selection of initial values. Therefore, GA is used to optimize the parameters k (number of clusters) and m (fuzzy partition matrix exponent). Experiments on single loop detection station on freeway networks in Harbin, China, showed that the proposed approach FCM-GA achieved good results, encouraging further research using prediction problems.

Similarly to the method proposed in Aydilek & Arslan (2013), Saravanan & Sailakshmi (2015) proposed a hybrid method that imputes missing values by using Fuzzy Possibilistic C-Means (FPCM) optimized by Support Vector Regression (SVR) and GA. FPCM is a method that combines traditional FCM and Possibilistic C-Means (PCM), taking advantage of the positive aspects of both of them: the possibility of a data to belong to more than one cluster and the ability of PCM to handle noisy data in an effective way. The datasets Iris and MarineDB (700 instances and 7 attributes) were used to test the method, with missing data artificially generated varying the percentage of missing data from 5% to 20%. The method was compared with FCM-SVRGA, proposed in Aydilek & Arslan (2013). Evaluated using RMSE, FPCM-SVRGA obtained the best results compared to FCM-SVRGA.

Azim & Aggarwal (2016) used the FCM method together with MLP network, in a 2-stage process, for missing data imputation. FCM was argued to be more efficient than the traditional K-Means clustering method in this task. Imputation results on benchmark datasets were given. The dataset Seed, with 210 observations, 7 attributes and 3 classes; and the dataset Sonar, which has 208 observations, 60 attributes and 2 classes, were evaluated. The percentage of missing values varied from 5% to 45%, with gap of 10%. In the results, the authors compared mean-based imputation, K-Means, K-Means+MLP, FCM and the proposed model using FCM+MLP. For the Seed dataset, the best result was obtained with 5% of missing values. In addition, difficulties using FCM for higher percentages of missing values were reported. For the Sonar dataset, FCM worked properly and helped the proposed model to reach the best classification results for all the aforementioned situations.

Zhang et al. (2016) proposed a method called HPF to reconstruct intervals of missing attributes. An interval is defined by the nearest neighbors. It is calculated using the concept of partial distance in the first step of the method. As intervals can often be an appropriate representation of data uncertainty, the proposed method tends to yield accurate results in these cases. In the second step, a FCM-Particle Swarm Optimization (PSO) combination was taken into consideration, where PSO particles provide the cluster prototypes and values for the missing attributes. FCM calculates the membership degree on the recovered incomplete dataset and provide a fitness value to guide the PSO. The authors tested their approach using three datasets for classification: Iris, Bupa and Breast. The proposed method was compared with four other methods. Classification results using the HPF method and varying the percent of missing values from 0% to 20% were the most accurate for the Iris and Breast datasets. On the other hand, other methods overcame HPF for the Bupa dataset. The other methods rely on a set of characteristics whereas HPF rely only on nearest-neighbor intervals.

Somehow similar to Zhang et al. (2016) and Tang et al. (2015), Zhang, Bing & Zhang (2015) presented a hybrid clustering algorithm based on missing attribute interval estimation for incomplete data. Their approach, called NIR, is very similar to that presented by Zhang et al. (2016), but instead of GA, the authors used PSO, having the particles in the PSO as the clusters prototypes. To evaluate the approach, the authors used five datasets for classification from UCI Repository: Iris, Bupa, Wine, Haberman and Breast Cancer, varying the percentage of missing data from 0% to 20%. NIR had the best results comparing with 5 other methods.

Saha et al. (2016) proposed a method called LRFDVImpute to complete multiple missing values in sequences of gene expression data. A GA-based method was used to validate their method. The authors argued that they were the first to apply LRFDVimpute on gene expression data. The original LRFDVImpute was compared with its modified version. The following datasets were used: Colorectal Cancer tumours, Breast Cancer, Prostate Cancer and DLBCL-FL. LRFDVImpute obtained perfect accuracy for these datasets.

Samat & Salleh (2016) discussed an approach based on FCM and PSO for missing data imputation. Three datasets were used for tests: Cleveland Heart Disease (297 instances and 13 attributes), Iris and Breast Cancer datasets, varying the percentage of missing data from 1% to 30%. The results showed that FCM and PSO together for missing data imputation can be a good alternative compared to FCM and Nearest Neighbors.

Kuppusamy & Paramasivam (2017) presented a hybrid method for missing data imputation that uses Grey Wolf Optimizer (GWO) together with Adaptative Neural Fuzzy Inference System (ANFIS), called GFNN. The authors used a medical dataset and were challenged by the fact that this aforementioned dataset has numerical and categorical fields. The authors applied the dataset with missing data to the WLI fuzzy clustering, and the centroids of the clusters are obtained, and in the process, the missing data are filled. Thus, GWO algorithm was used to train the ANFIS and the complete data was used as input of the ANFIS. To evaluate the GFNN, two datasets were used: Heart disease and Pima Indian Dataset, varying the percentage of missing data from 10% to 50%. The method was tested with 9 methods. GFNN obtained the best results according to MSE[5] and RMSE[6] measures.

Morshedizadeh et al. (2017) proposed a method for missing data imputation in order to estimate the performance of wind turbines. The method was used for missing data imputation and replacement of outlier values. The data were collected from 21 turbines from a wind farm in Ontario, Canada, over 20 months. Incomplete data were pre-processed using a decision tree. Thus, to estimate the performance of the wind turbines, an Adaptive Neural Fuzzy Inference System (ANFIS) was designed. Their method was compared with SVM, MLP, Deep Neural Network based Meta Regression and Transfer Learning (DNN-MRT), Genetic Programming ensemble of Artificial Neural Networks and Autoregressive Integrated Moving Average. The best results were obtained by the proposed method.

Table 3.2 briefly summarizes the studies described in this Section, emphasizing the sort of task and the datasets used in the experiments.

Table 3.2 – Summary of the aforementioned described approaches.

Study	Tasks	Employed datasets
Nanni, Lumini & Brahnam (2012)	Classification	12 datasets
Castillo & Cardeñosa (2012)	Classification	2 datasets (Polls 2555 and 2750)
Luengo, Sáez & Herrera (2012)	Classification	21 datasets
Li, Gu & Zhang (2013)	Classification	3 (Iris, Wine, New-Thyroid)
Aydilek & Arslan (2013)	Classification	6 datasets
Tang et al. (2015)	Prediction	1 (Loop detection station)
Saravanan & Sailakshmi (2015)	Classification	2 (Iris and MarineDB)
Azim & Aggarwal (2016)	Classification	2 (Seed and Sonar)
Zhang, Bing & Zhang (2015)	Classification	5 datasets
Zhang et al. (2016)	Classification	3 (Iris, Bupa, Breast Cancer)
Saha et al. (2016)	Classification	4 datasets
Samat & Salleh (2016)	Classification	3 (Heart disease, Iris, Breast Cancer)
Kuppusamy & Paramasivam (2017)	Classification	2 (Heart disease and Pima Indian)
Morshedizadeh et al. (2017)	Prediction	1 (Data from a wind farm in Canada)

To summarize, the methods described in this section combine computational intelligence and machine learning methods and models to perform classification or prediction even if some of the samples contain missing values. Nevertheless, all of them require a dataset to be completely available to perform imputation and to set/train the parameters of classifiers or predictors. None of them is suitable for online nonstationary environment. When a concept drift or shift

[5] Mean Squared Error
[6] Root-Mean-Squared Error

is perceived from a data stream, models should be reviewed parametrically and structurally to keep their accuracy acceptable. Moreover, many of the methods described in this section deal with a single missing value per sample. In the next Section, the Fuzzy Set-based Evolving Modeling (FBeM) is presented. Then, a modification of FBeM for handling multiple missing values on data streams is proposed.

4 PROPOSED METHOD

A modified Fuzzy Set-based evolving Model (FBeM), namely evolving Fuzzy Granular Predictor (eFGP), and a new learning algorithm are proposed to simultaneously deal with nonstationary data streams, provide predictions of future values, and to cope with missing data. In this Section, FBeM, eFGP and its learning algorithm are addressed.

4.1 Fuzzy Set-based Evolving Modeling (FBeM)

Fuzzy Set-based Evolving Modeling (FBeM) is a data-driven evolving approach that provides simultaneous singular and granular function approximation and linguistic description of the behavior of a system (LEITE et al., 2012). Granule, in this context, consists of a set of attributes and values that have common characteristics as equivalence, proximity, similarity, indistinguishably or functionality (PEDRYCZ, 2005).

An FBeM-based model generated from data streams has a set of IF-THEN rules that is a granular representation of a system. The model is generated in a single pass of the data, i.e., storage of data samples is needless, which is an essential characteristic when dealing with data streams. As data streams may be infinite, scalability is a major issue (LEITE et al., 2012). Uncertainty handling is a distinctive feature of FBeM. In general, FBeM is able to process numerical, interval and fuzzy data simultaneously. Notwithstanding, the present study focus on numerical missing data only.

FBeM models do not need *a priori* information about the properties of the data to start learning and providing predictions. However, an initial dataset and expert information can be used to create an initial model if desired. Its rules and granules (local models) are generated as data samples become available. It means that rules and granules are dynamically created, updated, merged and deleted over time. For each granule in the data space, there exists a corresponding rule that governs or summarizes the general behavior of the elements that constitute the granule. The FBeM learning algorithm incrementally adjusts granules and other parameters associated to rules so that the fuzzy model captures recent occurrences without forgetting previous behaviors (LEITE et al., 2012).

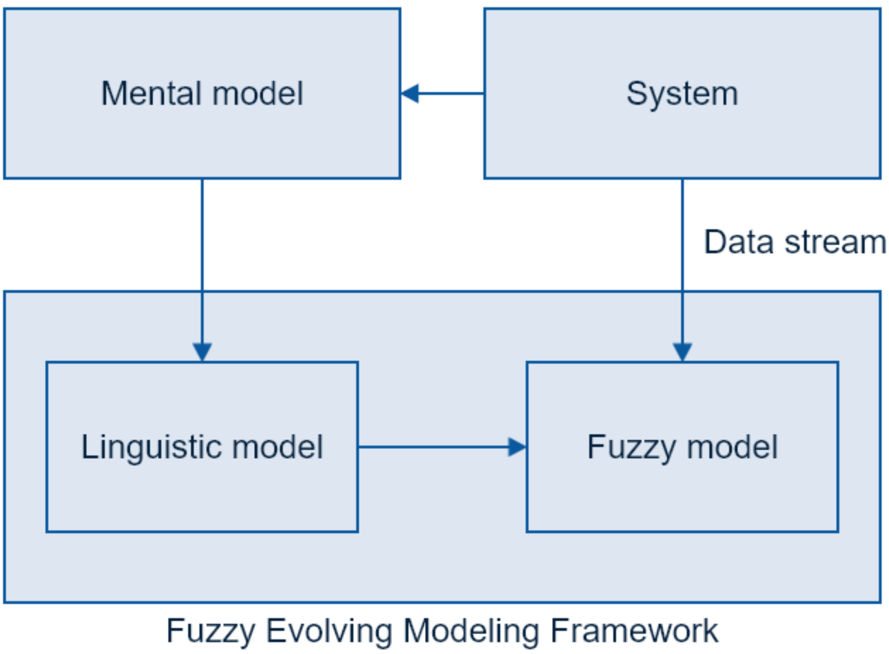
The antecedent part of FBeM rules are constituted by fuzzy hyperboxes (cylindrical extension of membership functions over the Cartesian product space), whereas the consequent part has functional and linguistic components. An example of an FBeM rule is

$$IF\ (x_1\ is\ A^i_j)\ AND\ ...\ (x_j\ is\ A^i_j)\ THEN\ \underbrace{(y\ is\ B^i)}_{Linguistic}\ AND\ \underbrace{(y\ =\ p^i(x_1,...,x_n))}_{Functional}$$

where x_j is an attribute of the current sample x; $A^i_j = (l^i_j, \lambda^i_j, \Lambda^i_j, L^i_j)$ is a trapezoidal membership function related to the i-th rule of a rule base and defined over the j-th universe of discourse; y is the output variable; $B^i = (u^i, \upsilon^i, \Upsilon^i, U^i)$ is a membership function in the output universe of discourse; and p^i is a linear of affine function. Notice that trapezoidal membership functions are cast by four parameters. The intermediate parameters form the core of the function; the boundary parameters form its support.

Figure 4.1 shows a general scheme on how an FBeM model can be generated. It can be observed that an FBeM model can be generated straightly from a data stream from a target system as well as from expert knowledge. Experts can provide linguistic rules about the system operation.

Figure 4.1 – FBeM general modeling scheme.
Source: the author, based on Leite et al. (2012)



4.2 evolving Fuzzy Granular Predictor (eFGP)

The proposal of this study is to develop an unsupervised missing data imputation method that can cope with data streams. This method was developed within the FBeM structure. The rules in the FBeM structure were adapted to handle missing data and impute a value, being able to handle missing data of MCAR and MAR types. Similarly to FBeM, eFGP is a supervised method as the output becomes available after the prediction is given. On the other hand, the imputation mechanism within eFGP is unsupervised.

The idea is to maintain additional consequent terms on the FBeM rules, which stand for first-order polynomials with one attribute less. For example, if a data sample has 9 attributes, then the affine functional consequent p^i has 10 coefficients. Additional affine functional consequents q^i_k, $k = 1, ..., 9$, with 9 coefficients each, excluded one attribute at a time, are also maintained.

Let $(x, y)^{[h]}$, $h = 1, ...,$ be the h-th observation of a data stream. The output $y^{[h]}$ is known given the input $x^{[h]}$ or will be known later (prediction problem). In this study, an attribute x_j of $\mathbf{x} = (x_1, ..., x_n)$ is a real value. The same holds for the output y. The pair (x, y) is a point in the product space $X \times Y$. Let γ^i, $i = 1, ..., c$, be the current set of granules built on the basis of (x, y). Granules γ^i are also defined in the product space $X \times Y$.

Rules R^i governing granules γ^i are given as

$$
\begin{aligned}
R^i\text{: IF } & (x_1 \text{ is } A^i_1) \text{ AND ... AND } (x_n \text{ is } A^i_n) \\
& \text{THEN } \underbrace{(y \text{ is } B^i)}_{\text{Linguistic}} \text{ AND } \underbrace{\bar{y} = p^i(x_1, ..., x_n)}_{\text{Functional}} \\
& \text{OR } \underbrace{\bar{y} = q^i_k(x_1, ..., x_{\theta-1}, x_{\theta+1}, ..., x_n),\ k = 1, ..., n}_{\text{Functional with reduced term}}
\end{aligned}
$$

where $A^i_j = (l^i_j, \lambda^i_j, \Lambda^i_j, L^i_j)\ \forall j$ and $B^i = (u^i, \upsilon^i, \Upsilon^i, U^i)$ are trapezoidal membership functions built from the data stream; x_θ is a missing value; and p^i and q^i_k are affine functions, where q^i_k has one term less than p^i. The set of rules R^i, $i = 1, ..., c$, is a fuzzy granular description of a system (LEITE et al., 2012) (YAGER, 2008). Initially, $c = 0$, i.e., no prior knowledge is assumed. A rule combines a linguistic and $n + 1$ functional consequents. Only one functional consequent, which gives a pointwise prediction of a time series, is taken into account for a given input x. The pointwise prediction is given by p^i, in case $x^{[h]}$ is complete, or by q^i_θ, in case x_θ is missing. The linguistic consequent offers prediction bounds and interpretability since the trapezoid B^i may come with a linguistic value.

Complete affine functions are given as

$$p^i(x) = \alpha_0^i + \sum_{j=1}^{n} \alpha_j^i x_j, \tag{4.1}$$

whereas the reduced-term functions are

$$q_k^i(x) = \beta_0^i + \sum_{j=1, j \neq \theta}^{n} \beta_j^i x_j, \tag{4.2}$$

$k = 1, ..., n$. Notice that as q_k^i has one term less than p^i, the coefficients α_j and β_j are uncorrelated. Coefficients α_j^i and β_j^i of the functions p^i and q_k^i $\forall k$ are adapted using the Recursive Least Squares algorithm as described in Åström & Wittenmark (2013) if R^i is the most active rule for a complete $x^{[h]}$; however, the coefficients β_k^i are adapted ignoring $x_k^{[h]}$ from $x^{[h]}$. For an incomplete $x^{[h]}$, i.e., if $x_j^{[h]}$ is missing, then only β_k^i are adapted.

As trapezoids A_j^i may overlap, overall pointwise prediction is found as the weighted mean value,

$$\bar{y} = \frac{\sum_{i=1}^{c} \Psi_{COM}^i \; p^i(x_1, ..., x_n)}{\sum_{i=1}^{c} \Psi_{COM}^i}, \tag{4.3}$$

for a complete x. Otherwise,

$$\bar{y} = \frac{\sum_{i=1}^{c} \Psi_{INC}^i \; q_\theta^i(x_1, ..., x_{\theta-1}, x_{\theta+1}, ..., x_n)}{\sum_{i=1}^{c} \Psi_{INC}^i} \tag{4.4}$$

is used if x_θ is missing. In the previous equations,

$$\Psi_{COM}^i \quad = \quad T(A_1^i(x_1), ..., A_n^i(x_n)) \tag{4.5}$$

and

$$\Psi^i_{INC} = T(A^i_1(x_1),...,A^i_{\theta-1}(x_{\theta-1}),A^i_{\theta+1}(x_{\theta+1}),...,A^i_n(x_n))$$

are the activation level of a rule R^i whether x is complete or not; T is any triangular norm, for example, the minimum.

When more than one attribute is missing, neither the complete functional consequent nor the reduced-term functional consequent can be used. Additional reduced-term functional consequents for the multiple missing values case are unfeasible, since the number of linear functions increases exponentially with the number of attributes.

Let $x = (x_1,\ ...,\ x_{\theta 1},\ ...,\ x_{\theta 2},\ ...,x_n)$ be an input vector with 2 missing values. Also, let γ^+ be the most active granule for x based on its similarity to those available attributes of x. Moreover,

R^+: IF (x_1 is A^+_1) AND...AND ($x_{\theta 1}$ is $A^+_{\theta 1}$) AND...AND ($x_{\theta 2}$ is $A^+_{\theta 2}$) AND...AND (x_n is A^i_n)
THEN (y is B^+) AND $\bar{y} = p^+(x_1,...,x_n)$
OR $\bar{y} = q^+_k(x_1,...,x_{\theta-1},x_{\theta+1},...,x_n),\ k = 1,...,n$

As R^+ is active for x, thus $x_{\theta 1}$ and $x_{\theta 2}$ are replaced by the midpoint (mp) of their respective membership functions $A^+_{\theta 1}$ and $A^+_{\theta 2}$. In other words, the fuzzy model inputs values for $x_{\theta 1}$ and $x_{\theta 2}$. The midpoint of a missing x_θ is defined as

$$mp(x_\theta) = \frac{(\lambda_\theta + \Lambda_\theta)}{2},$$

where λ_θ and Λ_θ assemble the core of the trapezoidal membership function in question. After the multiple imputation, the data sample is processed using full linear consequent functions.

Granular prediction is given by the convex hull of sets B^{i^*}, where i^* are indices of active granules. The convex hull of a set of trapezoids, say $B^1,...,B^c$, is given as

$$\begin{aligned} ch(B^1,...,B^c) = (&T(u^1,...,u^c),T(\upsilon^1,...,\upsilon^c),\\ &S(\Upsilon^1,...,\Upsilon^c),S(U^1,...,U^c)). \end{aligned} \tag{4.6}$$

where T and S may be any triangular norm and conorm. For instance the minimum and maximum provide the largest granular estimation while other norms yield narrower enclosures of p. Figure 4.2 shows the parameters of a trapezoidal membership function, while Fig. 4.3 explains graphically the convex hull operation between A_j^1 and A_j^2.

Figure 4.2 – Parameters of a trapezoidal membership function.
Source: The author, based on Leite et al. (2012).

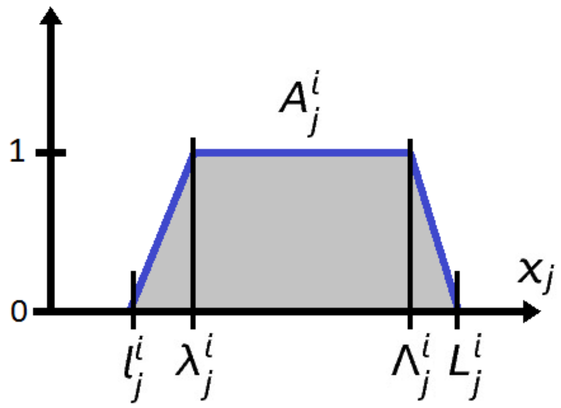
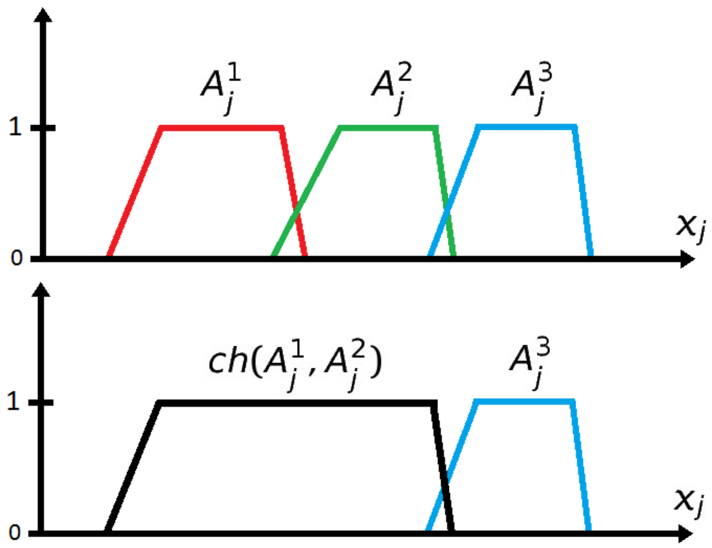


Figure 4.3 – Convex hull operation.
Source: The author, based on Leite et al. (2012).



The granular prediction given by $ch(.)$ provides a trapezoidal envelope around the pointwise prediction and may help decision making and improve model acceptability.

4.3 Incremental Learning

A new granule γ^{c+1} is created by adding a rule R^{c+1} to the collection of rules $R = \{R^1, ..., R^c\}$ of the eFGP model. A rule is created in case at least one input attribute, i.e. $x_j^{[h]}$ does not fit $A_j^i \forall i$ or $y^{[h]}$ does not fit $B^i \forall i$, where A^i and B^i are membership functions built as data are made available. T-norm aggregation suggests that $A_j^i \forall j$ and B^i should fit $(x,y)^{[h]}$ for the corresponding rule to be considered. On the contrary, a new rule is created to fit the new information. A new granule γ^{c+1} has trapezoidal memberships functions A_j^{c+1} and B^{c+1}, where $l_j^{c+1} = \lambda_j^{c+1} = \Lambda_j^{c+1} = L_j^{c+1} = x_j^{[h]}$ and $u^{c+1} = \upsilon^{c+1} = \Upsilon^{c+1} = U^{c+1} = y^{[h]}$. Initially, the coefficients of p^{c+1} are set as $\alpha_j^{c+1} = 0, j \neq 0$, and $\alpha_0^{c+1} = y^{[h]}$; the coefficients of q_k^{c+1} are set as $\beta_k^{c+1} = 0, k \neq 0$, and $\beta_0^{c+1} = y^{[h]}$.

Adaptation of a rule R^i happens by expanding or contracting the support and the core of antecedents A_j^i and consequent B^i to fit new data. Expansion must respect the parameters ρ and σ. Parameter ρ is the maximum width in the input space that trapezoidal membership functions are allowed to grow. Parameter σ plays the same role as ρ in the output space. Additionally, the coefficients of local complete and incomplete approximation functions p^i and q_k^i, respectively, are adapted.

Even when an instance (x,y) fits in a granule γ^i, a rule may be adapted. If an instance is close enough to the fuzzy hyperbox γ^i, the granule may expand to accommodate $(x,y)^{[h]}$. If the instance is inside the granule, parameters can be changed in the sense of contracting the granule.

When using trapezoidal membership functions, six situations may be happen according to the position of the instance. They are:

- If $x^{[h]} \in [mp(A_j^i) - \frac{\rho_j}{2}, l_j^i]$ then $l_j^i(new) = x^{[h]}$ (support expansion)
- If $x^{[h]} \in [l_j^i, \lambda_j^i]$ then $\lambda_j^i(new) = x^{[h]}$ (core expansion)
- If $x^{[h]} \in [\lambda_j^i, mp(A_j^i)]$ then $\lambda_j^i(new) = x^{[h]}$ (core contraction)
- If $x^{[h]} \in [mp(A_j^i), \Lambda_j^i]$ then $\Lambda_j^i(new) = x^{[h]}$ (core contraction)
- If $x^{[h]} \in [\Lambda_j^i, L_j^i]$ then $\Lambda_j^i(new) = x^{[h]}$ (core expansion)
- If $x^{[h]} \in [L_j^i, mp(A_j^i) + \frac{\rho_j}{2}]$ then $L_j^i(new) = x^{[h]}$ (support expansion)

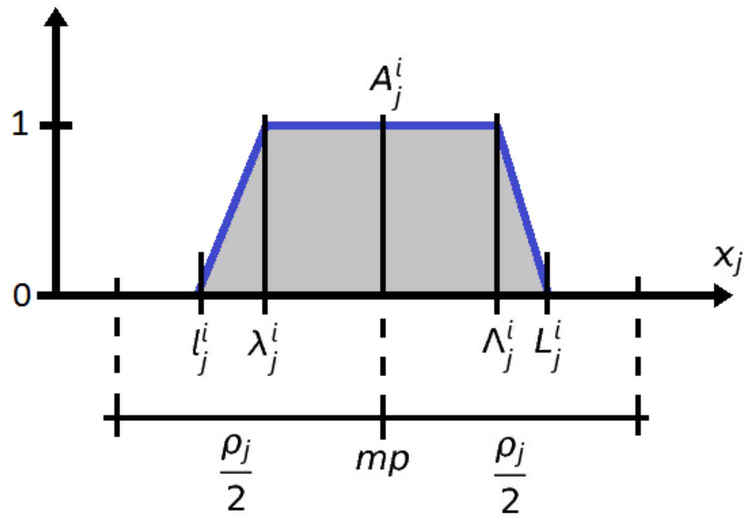
When operating on core parameters (λ_j^i and Λ_j^i), an adjustment of the midpoint of the granule is necessary. That is:

$$mp(A_j^i)(new) = \frac{\lambda_j^i + \Lambda_j^i}{2} \tag{4.7}$$

Adjusting the midpoint of a granule may adapt its support. Support contraction may happen in two situations, meaning that the membership function expanded excessively disrespecting ρ or σ. The situations are:

- If $mp(A_j^i)(new) - \frac{\rho_j}{2} > l_j^i$ then $l_j^i(new) = mp(A_j^i)(new) - \frac{\rho_j}{2}$
- If $mp(A_j^i)(new) + \frac{\rho_j}{2} < L_j^i$ then $L_j^i(new) = mp(A_j^i)(new) + \frac{\rho_j}{2}$

Figure 4.4 – Example of a trapezoidal membership function with associated parameters.
Source: The author, based on Leite et al. (2012)



In order to graphically understand the aforementioned situations, Fig. 4.4 is provided.

The adaptation of the fuzzy sets of rules consequents B^i uses data $y^{[h]}$. Polynomial coefficients α^i and β^i are updated using Recursive Least Squares algorithm (LEITE et al., 2012) taking advantage of the new instance that has activated granule γ^i and its corresponding rule.

eFGP uses a procedure to adjust parameters (maximum width, ρ and σ) of the granules over time, and it works as follows. Let δ be the number of existing rules after H steps of evolution. If the number of rules grows faster than a predefined threshold rate η, then ρ and σ are increased by a factor $\mu = (1 + (\delta - \eta)/H)$ during the next steps. On the other hand, if the number of rules grows at a rate smaller than η, then ρ and σ are decreased by μ. This procedure is useful when coping with data stream granularity and allows ρ and σ to "decide" values for themselves. The initial values for ρ and σ are defined over non-normalized data.

In some cases, fuzzy granules may overlap. Conflict resolution is needed to decide which granule is the most appropriate to fit a data sample. An approach for conflict resolution is selecting the granule with the highest specificity to be adapted, that is,

$$i* = argmax_i(sp(\gamma^{i*})) \tag{4.8}$$

Granule γ^{i*} provides a more precise description of the behavior of a process. The specificity of a fuzzy trapezoidal granule is obtained taking into account the specificity of the trapezoidal membership functions which forms the granules. The equation

$$sp(A_j^i) = 1 - \frac{1}{2}\left(\frac{(\Lambda_j^i - L_j^i) - (\lambda_j^i - l_j^i)}{T(L_j^i, ..., L_j^c) - T(l_j^i, ..., l_j^c)}\right), \tag{4.9}$$

measures the specificity of a trapezoidal membership function A_j^i. To calculate the specificity of fuzzy trapezoidal granule γ^i,

$$sp(\gamma^i) = T(sp(A_1^i)), ...sp(A_n^i), sp(B^i)) \tag{4.10}$$

must be used.

In order to reorganize the granules, there are some strategies for refinement of the granular mapping, that include combination of neighbor granules, covering gaps and deleting rules. This refinement happens after a certain amount of processing steps. It contributes to develop smoother approximands and keep model updated.

Merging neighbor granules, say γ^1 and γ^2, into a single granule formed by their convex hull, $\gamma^\Psi = ch(\gamma^1, \gamma^2)$, happens when they are placed close to each other so that γ^Ψ respects the limits given by ρ and σ. This strategy helps to reduce the number of rules and the number of overlapped granules covering similar information. Parameters of consequent functions of merged rules are taken from

$$\alpha_j^{\Psi} = \frac{\alpha_j^1 + \alpha_j^2}{2}, j = 0, ..., n, \tag{4.11}$$

and

$$\beta_j^{\Psi} = \frac{\beta_j^1 + \beta_j^2}{2}, j = 0, ..., \theta - 1, \theta + 1, ..., n. \tag{4.12}$$

Concept drift and shift may cause rules to become inactive. Rules are removed if they are not activated during h_r time steps. The purpose is to keep the rule base concise, with useful rules..

The learning procedure to evolve an eFGP model is described as follows:

Algorithm 1: evolving Fuzzy Granular Predictor (eFGP) Online Incremental Learning

1 **begin**
2 Define ρ, σ and h_r;
3 **do**
4 Read input $x^{[h]}, h = 1, ...$;
5 //Prediction
6 **if** $|\theta| = 0$ **then**
7 Give prediction $\bar{y}$ using complete functions p^i;
8 **else if** $|\theta| = 1$ **then**
9 Give prediction $\bar{y}$ using reduced-term functions q^i_θ;
10 **else if** $|\theta| > 1$ **then**
11 Multiple imputation;
12 Use complete functions p^i to give $\bar{y}$;
13 **end if**
14 //Model Adaptation ($y^{[h]}$ becomes available);
15 Create or adapt rules to fit $(x,y)^{[h]}$;
16 Delete sample $(x,y)^{[h]}$;
17 **if** $h = \alpha h_r$, $\alpha = 1, ...$ **then**
18 Delete inactive granules and rules;
19 Merge rules;
20 **end if**
21 **while** *1*;
22 **end**;

Notice that steps 4 and 16 indicate that samples are received and discarded one at a time, an essential feature of online data stream processing applications. Figure 5.1 shows a flowchart for a clear understanding.

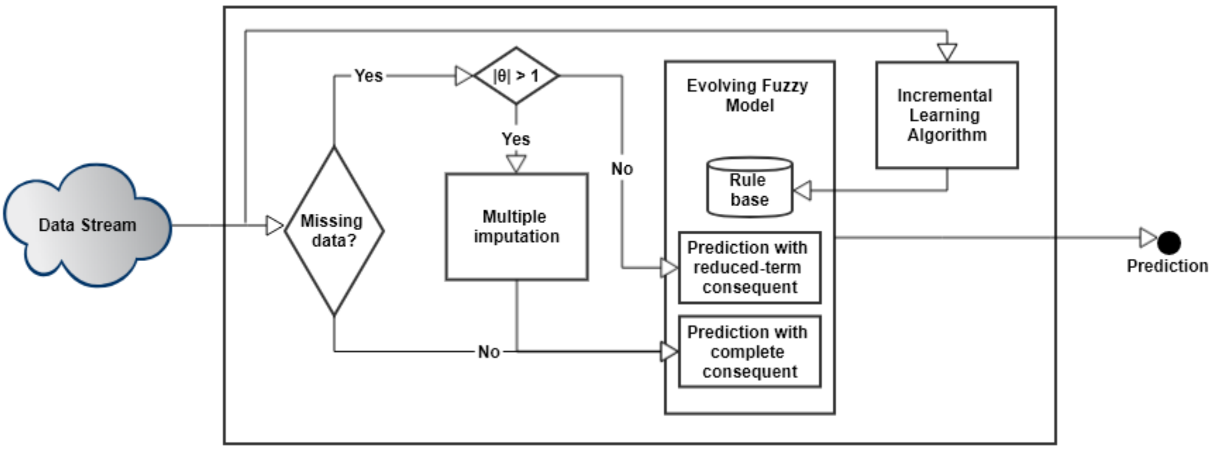
5 METHODOLOGY

Remarks about the online data stream environments and datasets are given in this chapter. A flowchart about the proposed fuzzy evolving framework able to handle missing data, and how datasets are adapted to emulate different types of missing data are also described. Statistics about the datasets, as well as other evolving methods to be used for comparisons, are outlined.

5.1 Considerations on the Environment

A general scheme of the online and the proposed evolving modeling framework is illustrated in Figure 5.1. The missing data imputation method receives the data and imputes missing values incrementally based on information from the current evolving fuzzy model. After processing a data sample through the fuzzy inference system and providing a prediction of the future value of a time series, the true value becomes available, which gives an idea about the system performance. Learning chooses new values for the parameters of the fuzzy model or adds, deletes or merges rules on the rule base.

Figure 5.1 – Evolving fuzzy modeling with missing data imputation mechanism
Source: the author



5.2 About the Datasets and Missing Data Generation

Benchmark datasets were chosen to evaluate the efficiency of the eFGP prediction model. The datasets contain no missing data in principle, which is convenient for the purpose of the experiments and comparative analyses. They are:

- Death Valley weather dataset[1]. Records of monthly mean temperature in degrees Celsius from 1901 to 2009 (1306 observations) is considered. A fixed time window of 12 months, with no exogenous inputs, is used for one-step prediction. Focus is on forecasting with missing values rather than on using information measures to select attributes;

- Bitcoin closing price dataset[2]. The lowest and highest prices reached during a day as well as bitcoin closing prices of previous days are used to predict the closing price of the next trading day. 4-attribute daily data samples over a 4-year period (starting on February 9th, 2014) are considered;

- Airfoil self-noise dataset[3]. The data are related to a series of aerodynamic and acoustic tests conducted in an anechoic wind tunnel at the National Aeronautics and Space Administration (NASA). This is a regression problem that contains 1503 samples, 5 attributes, namely, frequency (Hertz), angle of attack (degree), chord length (meter), free-stream velocity (meters per second), and suction-side displacement thickness (meter); and an output, the scaled sound pressure level (decibel). Samples were presented sequentially to a predicton model, as a data stream.

Prediction performance using MCAR and MAR types of missing values were assessed. For MCAR, the chance of occurrence of missing values is equal among the attributes. eFGP models were constructed and analyzed using datasets containing from 0% to 30% of missing figures – a range usually used in related studies (AYDILEK; ARSLAN, 2013)(AZIM; AGGARWAL, 2016)(SARAVANAN; SAILAKSHMI, 2015). In this study, the propensities for a value to be missing are: 1%, 5%, 10%, 20% and 30%.

For MAR data, a specific attribute is more inclined to receive an empty reading in relation to the other attributes. In this case, an attribute is taken at random to have its values more likely to be missing. eFGP prediction performance is evaluated for different percentages of chance that the single attribute and the rest of the attributes are missing. The MAR cases investigated are: (i) 5% – 1% (which means 5% of chance that the value of the chosen attribute is missing, and 1% of chance that each of the remaining values is missing); (ii) 10% – 1%; (iii) 10% – 5%; (iv) 20% – 5%; (v) 20% – 10%; (vi) 30% – 5%; and (vii) 30% – 10%.

[1] https://www.nps.gov/deva/planyourvisit/weather.htm
[2] https://bitcoincharts.com/charts/, accessed on 2018-Feb-7
[3] https://archive.ics.uci.edu/ml/datasets/airfoil+self-noise

The data were input one sample at a time, simulating an online environment where the input arrives continuously.

5.3 Evaluation and Comparison of Results

The results from the evolving Fuzzy Granular Predictor were compared with the results from the following intelligent methods under the same conditions:

- eTS according to Angelov & Filev (2004);
- xTS as in Angelov & Zhou (2006);
- eGNN as described in Leite, Costa & Gomide (2013).

Error metrics such as the root mean square error (*RMSE*) and the non-dimensional error index (*NDEI*) were chosen for performance comparison. The *RMSE* is given as

$$RMSE = \frac{1}{k_c} \sum_{k=1}^{k_c} \sqrt{(\bar{y}_{(k+1)} - y_{(k+1)})^2}, \tag{5.1}$$

where $\bar{y}$ is the predicted value and y is the expected value. The *NDEI* depends on the *RMSE* and standard deviation of the time series. It is given by

$$NDEI = \frac{RMSE}{std(y_{(k)\forall k})}, \tag{5.2}$$

NDEI is useful to compare the accuracy of predictors for data streams with different standard deviations.

6 RESULTS

The missing data imputation method and evolving modeling approach from data streams were tested on the problems described in Section 5.2. The prediction results were compared to those provided by well-known methods of the evolving intelligence literature, listed in Section 5.3. The results are organized in four Sections, showing prediction results with MCAR, predictions results with MAR, comparison with other approaches and statistical hypothesis testing.

6.1 eFGP Prediction Results for MCAR data

Average results for the Death Valley dataset considering different fractions of MCAR values and 10 runs of the eFGP learning algorithm for each case are shown in Table 6.1. The eFGP initial parameters are $\rho = \sigma = 0.3$, and $h_r = 150$. It is important to note that, as the percentage of missing data increases, the error indices increase monotonically. In a similar way, the number of rules in the model structure tends to increase with the number of missing values. Additional granules and fuzzy rules are needed to cover incomplete samples since correlation information is partially lost. As shown in Fig. 6.1, eFGP monthly temperature pointwise estimates for the roughest, 30%, MCAR scenario were able to track the peaks and valleys of the time series with a reasonable accuracy. Similarly, Fig. 6.2 shows part of the granular prediction provided by eFGP for a clearer view. Notice that the bounds of the granular prediction given by the support of active trapezoidal membership functions drawn in the output domain enclose the actual temperature values, as expected. If lower values for the granularity ρ and σ are chosen, then a narrower envelope can be achieved at the price of additional fuzzy rules. A tradeoff between model compactness, interpretability of rules, pointwise accuracy, and narrowness of the granular enclosure should be evaluated depending on the purpose of the model.

Table 6.1 – eFGP results for the Death Valley weather station assuming missing data of the MCAR type

MCAR	RMSE	NDE	Mean # of rules
0%	0.0579 +/- 0.0018	0.2239 +/- 0.0070	13.3 +/- 0.0
1%	0.0600 +/- 0.0014	0.2322 +/- 0.0055	13.9 +/- 0.5
5%	0.0636 +/- 0.0018	0.2461 +/- 0.0072	20.5 +/- 0.9
10%	0.0640 +/- 0.0020	0.2474 +/- 0.0080	22.9 +/- 0.8
15%	0.0703 +/- 0.0041	0.2719 +/- 0.0158	27.7 +/- 1.1
20%	0.0818 +/- 0.0059	0.3163 +/- 0.0223	28.8 +/- 1.1
30%	0.1142 +/- 0.0057	0.4180 +/- 0.0222	28.1 +/- 2.3

Figure 6.1 – Pointwise eFGP prediction for the mean monthly temperature of Death Valley considering 30% of MCAR values

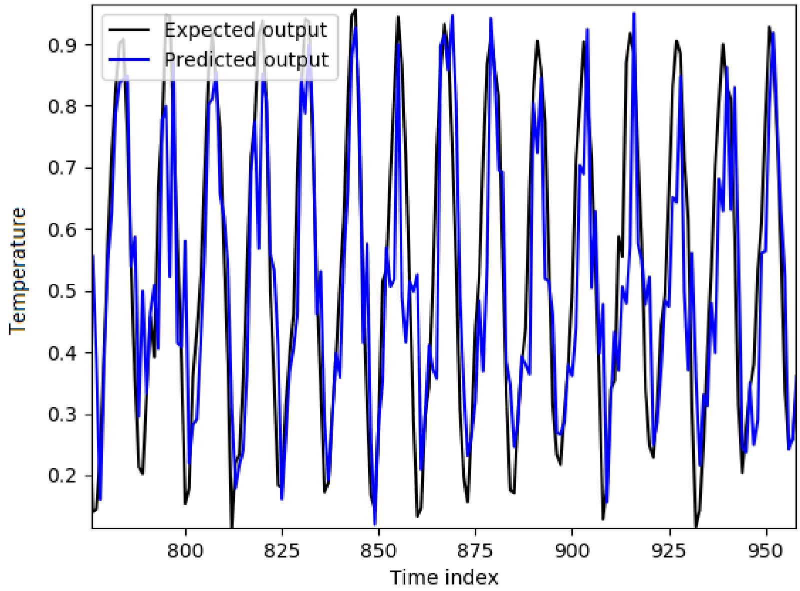
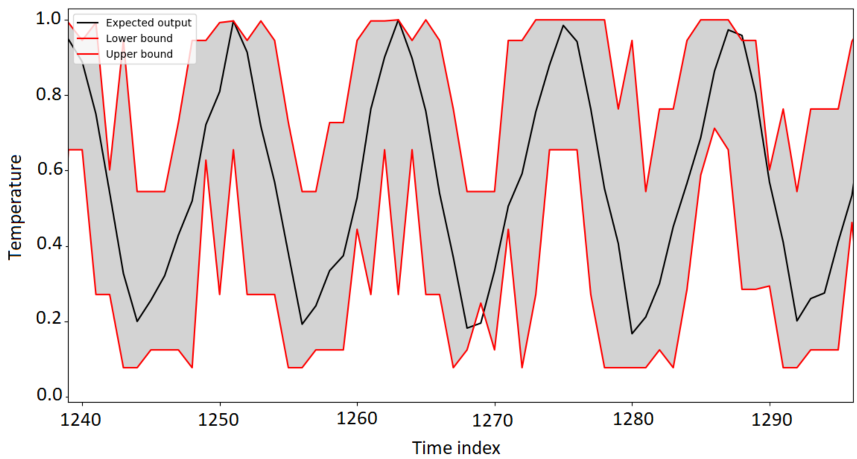


Figure 6.2 – Granular eFGP estimation for the Death Valley monthly mean temperature considering 30% of MCAR values



In eFGP, rules can be retrieved anytime. For example, at $h = 1295$, an example of active rule is:

R^i: IF (x_1 is $[0.54, 0.59, 0.70, 0.79]$) AND (x_2 is $[0.68, 0.68, 0.83, 0.87]$) AND
(x_3 is $[0.82, 0.83, 0.87, 0.95]$) AND (x_4 is $[0.78, 0.84, 0.89, 0.98]$) AND
(x_5 is $[0.73, 0.75, 0.80, 0.86]$) AND (x_6 is $[0.57, 0.64, 0.76, 0.77]$) AND
(x_7 is $[0.49, 0.55, 0.61, 0.61]$) AND (x_8 is $[0.19, 0.20, 0.34, 0.40]$) AND
(x_9 is $[0.15, 0.19, 0.25, 0.26]$) AND (x_{10} is $[0.16, 0.22, 0.32, 0.39]$) AND

$(x_{11}$ is $[0.27, 0.34, 0.43, 0.47])$ AND $(x_{12}$ is $[0.39, 0.39, 0.51, 0.61])$

THEN $(y$ is $[0.51, 0.56, 0.72, 0.74])$ AND

$\bar{y} = 0.14 + 0.17x_1 - 0.45x_2 + 0.64x_3 + 0.27x_4 - 0.28x_5 + 0.07x_6 - 0.28x_7 + 0.14x_8 - 0.09x_9 + 0.58x_{10} - 0.33x_{11} + 0.39x_{12}$ OR

$\bar{y} = 0.36 - 0.44x_2 + 0.55x_3 + 0.58x_4 - 0.43x_5 - 0.07x_6 - 0.24x_7 + 0.11x_8 + 0.19x_9 + 0.48x_{10} - 0.56x_{11} + 0.32x_{12}$ OR

$\bar{y} = 0.20 + 0.16x_1 + 0.45x_3 - 0.01x_4 - 0.55x_5 + 0.39x_6 - 0.18x_7 + 0.20x_8 - 0.11x_9 + 0.82x_{10} - 0.75x_{11} + 0.45x_{12}$ OR

$\bar{y} = 0.64 - 0.11x_1 + 0.35x_2 + 0.29x_4 - 0.32x_5 + 0.25x_6 - 0.31x_7 + 0.22x_8 + 0.40x_9 + 1.40x_{10} - 1.76x_{11} - 0.13x_{12}$ OR

$\bar{y} = 0.02 + 0.27x_1 - 0.32x_2 + 0.64x_3 - 0.27x_5 + 0.25x_6 - 0.28x_7 + 0.18x_8 - 0.28x_9 + 0.72x_{10} - 0.30x_{11} - 0.45x_{12}$ OR

$\bar{y} = 0.09 + 0.21x_1 - 0.53x_2 + 0.64x_3 + 0.26x_4 - 0.01x_6 - 0.39x_7 + 0.15x_8 - 0.14x_9 + 0.72x_{10} - 0.28x_{11} + 0.24x_{12}$ OR

$\bar{y} = 0.16 + 0.15x_1 - 0.51x_2 + 0.65x_3 + 0.35x_4 - 0.23x_5 - 0.30x_7 + 0.13x_8 - 0.05x_9 + 0.56x_{10} - 0.31x_{11} + 0.35x_{12}$ OR

$\bar{y} = 0.23 + 0.09x_1 - 0.21x_2 + 0.65x_3 + 0.24x_4 - 1.04x_5 + 0.30x_6 + 0.14x_8 - 0.03x_9 + 0.20x_{10} - 0.39x_{11} + 0.82x_{12}$ OR

$\bar{y} = 0.38 - 0.09x_1 - 1.06x_2 + 0.81x_3 + 1.13x_4 - 0.39x_5 - 0.55x_6 - 0.25x_7 + 0.35x_9 - 0.11x_{10} + 0.01x_{11} + 0.38x_{12}$ OR

$\bar{y} = 0.21 + 0.11x_1 - 0.45x_2 + 0.61x_3 + 0.37x_4 - 0.31x_5 + 0.01x_6 - 0.27x_7 + 0.13x_8 + 0.56x_{10} - 0.40x_{11} + 0.35x_{12}$ OR

$\bar{y} = 0.18 + 0.09x_1 - 0.68x_2 + 0.82x_3 + 0.55x_4 - 0.70x_5 - 0.05x_6 - 0.13x_7 + 0.07x_8 + 0.01x_9 + 0.01x_{11} + 0.70x_{12}$ OR

$\bar{y} = 0.01 + 0.25x_1 - 0.63x_2 + 0.78x_3 + 0.24x_4 - 0.22x_5 + 0.02x_6 - 0.29x_7 + 0.13x_8 - 0.23x_9 + 0.44x_{10} + 0.49x_{12}$ OR

$\bar{y} = 0.22 + 0.12x_1 - 0.51x_2 + 0.52x_3 + 0.41x_4 + 0.18x_5 - 0.13x_6 - 0.47x_7 + 0.14x_8 + 0.03x_9 + 0.91x_{10} - 0.55x_{11}$

If an ordered set of labels, such as cold, warm and hot, is given to the antecedent terms of each attribute, then the model comes with a level of interpretability as additional asset.

Results for the Bitcoin closing price with different fractions of MCAR data are show in Table 6.2. The best eFGP initial parameters in this problem were $\rho = \sigma = 0.15$, and $h_r = 50$. A behavior distinct from that observed in Table 6.1 for the error indices is noticed in Table 6.2. With the increase of the amount of missing data, the error indices do not exhibit a clear monotonic increasing trend. Nonetheless, the extreme values on Table 6.2, for 'no' and 30% of MCAR data, suggest that the estimation performance reduces slightly with the presence of missing completely at random data. The number of fuzzy rules tends to increase with more missing values.

Table 6.2 – eFGP results for the Bitcoin historical price assuming missing data of the MCAR type

MCAR	RMSE	NDE	Mean # of rules
0%	0.0219 +/- 0.0044	0.1410 +/- 0.0285	14.4 +/- 0.0
1%	0.0225 +/- 0.0033	0.1447 +/- 0.0212	14.3 +/- 0.8
5%	0.0220 +/- 0.0023	0.1411 +/- 0.0152	14.1 +/- 0.2
10%	0.0213 +/- 0.0034	0.1367 +/- 0.2184	6.0 +/- 0.3
15%	0.0231 +/- 0.0047	0.1486 +/- 0.0304	8.2 +/- 0.4
20%	0.0234 +/- 0.0041	0.1503 +/- 0.0266	18.9 +/- 1.9
30%	0.0252 +/- 0.0054	0.1619 +/- 0.0347	16.3 +/- 1.8

Figures 6.3 and 6.4 depict the numerical and granular eFGP estimates of the Bitcoin price for the hardest, 30%, MCAR case. An increasing growth rate can be perceived, especially in the last year (from sample 1000 to 1350). This distinguishes this time series from the previously analyzed weather series, which is predominantly seasonal in nature. Linear non-adaptive models would certainly have their performance reduced over time due to the concept drift. An accurate tracking of the original data can be observed in Fig. 6.3 by using the eFGP pointwise $\bar{y}$. Online learning and creation of local granules were key points to keep a reasonable prediction accuracy over time. Figure 6.4 elucidates the placement of output fuzzy granules relative to actual data. Bounds of granules (in red) refer to the support of active trapezoidal membership functions in the output space. If input data x are out of input granules' bounds according to ρ, then a new granule is created and evolved causing an apparent discontinuity of the granular prediction. Nonetheless, trapezia bounds give a range of uncertainty around the pointwise prediction that may help decision making regarding Bitcoin exchanges and purchases.

Figure 6.3 – Numerical daily prediction of the Bitcoin price considering eFGP modeling and 30% of MCAR data

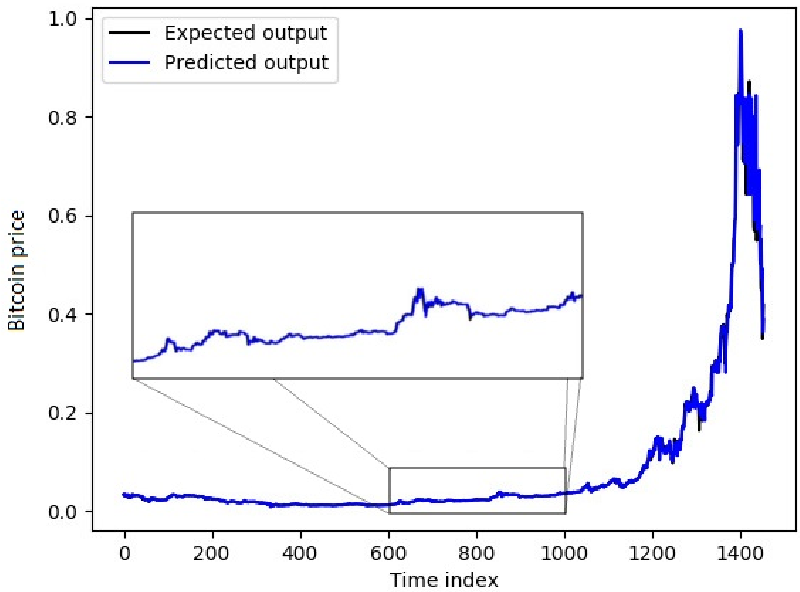
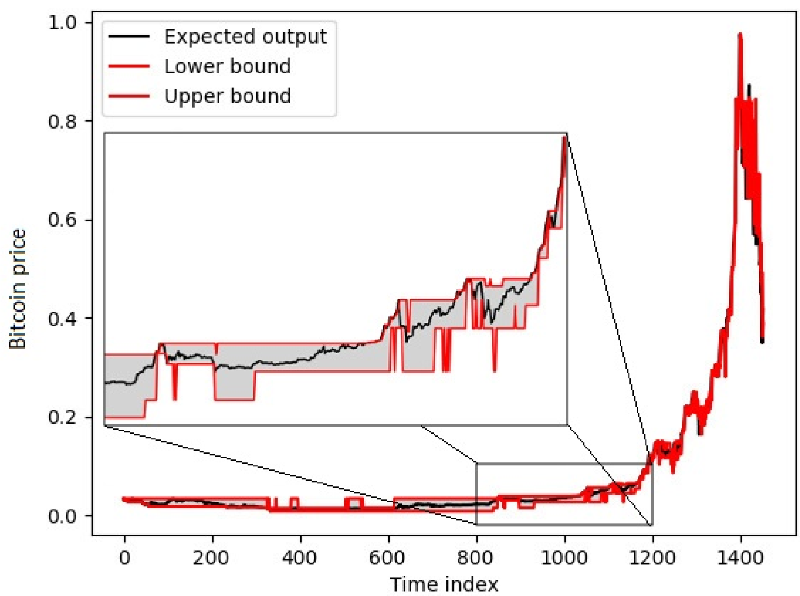


Figure 6.4 – Granular daily prediction of the Bitcoin price considering eFGP modeling and 30% of MCAR data



An active rule at $h = 1298$ is:

R^i: IF (x_1 is $[0.49, 0.49, 0.49, 0.53]$) AND (x_2 is $[0.54, 0.55, 0.56, 0.56]$) AND (x_3 is $[0.51, 0.52, 0.53, 0.53]$) AND (x_4 is $[0.50, 0.50, 0.50, 0.52]$)

THEN $(y \text{ is } [0.55, 0.56, 0.57, 0.57])$ AND

$\bar{y} = -0.16 - 0.81x_1 + 0.38x_2 + 1.1x_3 + 0.67x_4$ OR

$\bar{y} = -0.26 + 0.51x_2 + 0.88x_3 + 0.13x_4$ OR

$\bar{y} = -0.01 - 1.01x_1 + 1.34x_3 + 0.72x_4$ OR

$\bar{y} = -0.54 + 0.50x_1 + 1.39x_2 + 0.12x_3$ OR

$\bar{y} = -0.22 + 0.07x_1 + 0.44x_2 + 0.95x_3$

Results for the airfoil self-noise dataset considering MCAR data are summarized in Table 6.3. The eFGP parameters are $\rho = \sigma = 0.1$, and $h_r = 48$. Notice that, similar to the Death Valley problem, as the percentage of missing data increases, a monotonic increasing of the amount of fuzzy rules and error values are seen. The average RMSE and NDE indices are clearly higher in this problem not only because the standard deviation of the data is higher (as revealed by the larger NDE/RMSE value), but also due to both a lower average Pearson correlation between the five input attributes in question and the predicted sound pressure level; and a relatively faster and irregular (noisy) dynamical behavior. A stochastic component is intrinsic to the input data and this is reflected on the actual and predicted sound pressures as illustrated in Fig. 6.5 by the rapid amplitude variations. In spite of the noise, we notice that the eFGP approximation follows the trend of the data sequence and captures the change in the standard deviation over the iterations.

Table 6.3 – Summary of eFGP results for the airfoil self-noise dataset assuming MCAR data

MCAR	RMSE	NDE	Mean # of rules
0%	0.1114 +/- 0.0003	0.6059 +/- 0.0017	3.4 +/- 0.0
1%	0.1174 +/- 0.0085	0.6381 +/- 0.0465	5.5 +/- 0.4
5%	0.1251 +/- 0.0056	0.6801 +/- 0.0305	11.5 +/- 1.4
10%	0.1436 +/- 0.0046	0.7808 +/- 0.0249	12.2 +/- 0.9
15%	0.1476 +/- 0.0062	0.8022 +/- 0.0034	12.9 +/- 0.9
20%	0.1502 +/- 0.0073	0.8164 +/- 0.0397	20.4 +/- 1.6
30%	0.1674 +/- 0.0064	0.9102 +/- 0.0351	20.5 +/- 0.6

Figure 6.6 shows the evolution of the number of eFGP rules over time for a typical run of the learning algorithm. Initially, a number of rules are created to fit never-before-seen data. After about 50 iterations, the model structure became more stable and parameter adaptation prevailed over granule creation. The average number of rules was 15.7 on the simulation shown in Fig. 6.6. An interesting event to be observed in Figs. 6.5 and 6.6 happened at iteration 720.

A sudden reduction of the standard deviation of the actual sampling distribution required a new incremental growth of the fuzzy rule base. In other words, about 6 rules were added to the eFGP rule base for an appropriate handling of the concept shift.

Figure 6.5 – eFGP numerical approximation of the sound pressure level

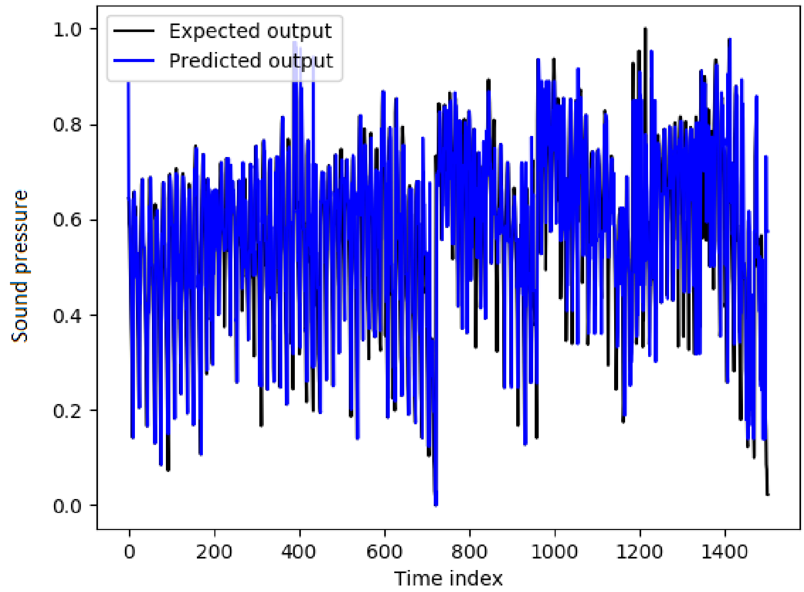
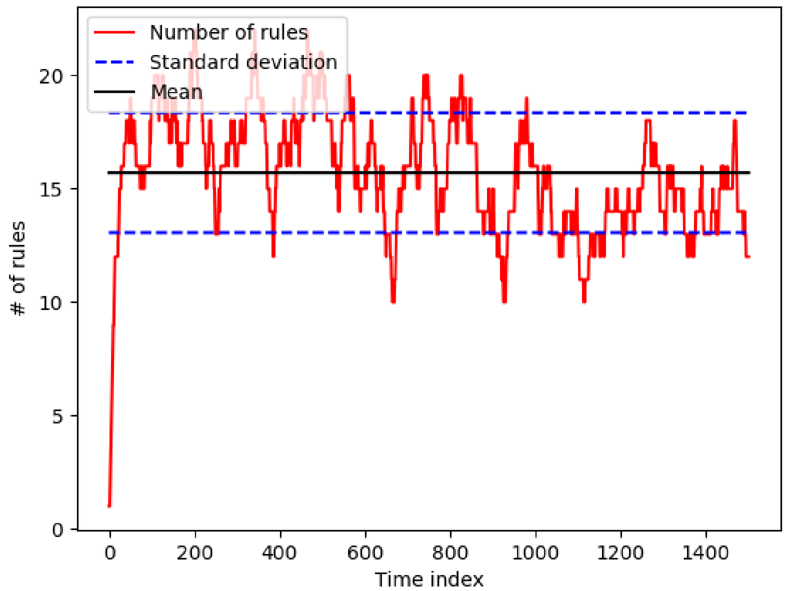


Figure 6.6 – Evolution of the fuzzy rule base



An example of rule at $h = 1502$ for the airfoil data is:

R^i: IF (x_1 is $[0.00, 0.02, 0.10, 0.15]$) AND (x_2 is $[0.15, 0.15, 0.15, 0.43]$) AND
(x_3 is $[0.27, 0.27, 0.27, 0.27]$) AND (x_4 is $[0.00, 0.19, 0.20, 0.20]$) AND
(x_5 is $[0.11, 0.13, 0.26, 0.50]$)

THEN (y is $[0.55, 0.76, 0.79, 0.97]$) AND

$$\bar{y} = 0.53 - 1.76x_1 - 0.49x_2 + 0.14x_3 + 0.75x_4 + 0.14x_5 \text{ OR}$$
$$\bar{y} = 0.24 - 0.31x_2 + 0.06x_3 + 0.75x_4 + 0.33x_5 \text{ OR}$$
$$\bar{y} = 0.14 - 1.24x_1 + 0.04x_3 + 0.73x_4 + 0.20x_5 \text{ OR}$$
$$\bar{y} = 0.57 - 1.76x_1 - 0.49x_2 + 0.75x_4 + 0.14x_5 \text{ OR}$$
$$\bar{y} = 0.96 - 1.76x_1 - 0.39x_2 + 0.26x_3 - 0.39x_5 \text{ OR}$$
$$\bar{y} = 0.69 - 2.07x_1 - 0.67x_2 + 0.19x_3 + 0.66x_4$$

Recall that the attributes are, respectively: frequency, angle of attack, chord length, free-stream velocity and suction side displacement thickness. Thus, the rule above can be read as: if frequency is ‘low’, angle of attack is ‘medium low’, chord length is ‘small’, free-stream velocity is ‘slow’ and suction side displacement thickness is ‘medium short’, then the sound pressure level is ‘medium high’.

6.2 eFGP Prediction Results for MAR data

Experiments with MAR data were conducted for the same datasets using the same initial eFGP parameters, as described in the previous section. Table 6.4 summarizes the results. In general, although a modest performance reduction is observed with the increase of the percentage of MAR values, the RMSE and NDE indices are essentially similar in all setups for a dataset. In addition, the size of the fuzzy rule base in most cases increases with the amount of MAR data. However, the increase of the fuzzy model structure is not in the same proportion as that observed for MCAR data. Therefore, MCAR data impose more challenges to incremental modeling than MAR data. In other words, if readings from a single attribute become partially available for some reason, the eFGP learning algorithm still benefits from the information of the other attributes to keep a reasonable prediction accuracy. Contrariwise, if the availability of readings of all attributes are limited, then the algorithm is required to use approximations from nearest granules to provide predictions. Nonetheless, the choice of the nearest granules becomes blurred when two or more values of a sample are simply unavailable.

Table 6.4 – eFGP results for the Death Valley weather station, Bitcoin price, and Airfoil self-noise considering MAR data

Death Valley monthly weather			
MAR	RMSE	NDE	Mean # of rules
5% – 1%	0.0604 +/- 0.0021	0.2335 +/- 0.0082	14.5 +/- 0.9
10% – 1%	0.0611 +/- 0.0010	0.2364 +/- 0.0041	14.4 +/- 0.4
10% – 5%	0.0637 +/- 0.0031	0.2464 +/- 0.0122	20.4 +/- 0.5
20% – 5%	0.0663 +/- 0.0029	0.2566 +/- 0.0112	19.7 +/- 1.6
20% – 10%	0.0651 +/- 0.0034	0.2516 +/- 0.0132	29.9 +/- 1.2
30% – 5%	0.0632 +/- 0.0015	0.2442 +/- 0.0059	20.7 +/- 0.8
30% – 10%	0.0678 +/- 0.0019	0.2622 +/- 0.0075	22.7 +/- 1.2
Bitcoin daily closing price			
MAR	RMSE	NDE	Mean # of rules
5% – 1%	0.0246 +/- 0.0037	0.1583 +/- 0.0241	8.8 +/- 0.1
10% – 1%	0.0229 +/- 0.0023	0.1474 +/- 0.0146	8.8 +/- 0.2
10% – 5%	0.0241 +/- 0.0054	0.1543 +/- 0.0347	8.1 +/- 0.1
20% – 5%	0.0237 +/- 0.0044	0.1523 +/- 0.0285	5.5 +/- 0.3
20% – 10%	0.0251 +/- 0.0039	0.1704 +/- 0.0343	5.8 +/- 0.5
30% – 5%	0.0214 +/- 0.0024	0.1374 +/- 0.0157	5.4 +/- 0.3
30% – 10%	0.0237 +/- 0.0035	0.1526 +/- 0.0284	10.9 +/- 0.4
Airfoil sound pressure			
MAR	RMSE	NDE	Mean # of rules
5% – 1%	0.1246 +/- 0.0127	0.6776 +/- 0.0694	4.1 +/- 0.1
10% – 1%	0.1200 +/- 0.0046	0.6523 +/- 0.0254	3.9 +/- 0.1
10% – 5%	0.1210 +/- 0.0060	0.0658 +/- 0.0329	6.8 +/- 0.6
20% – 5%	0.1242 +/- 0.0056	0.6752 +/- 0.0306	11.1 +/- 1.1
20% – 10%	0.1315 +/- 0.0040	0.7148 +/- 0.0219	12.4 +/- 1.0
30% – 5%	0.1205 +/- 0.0022	0.6552 +/- 0.0120	11.3 +/- 0.7
30% – 10%	0.1344 +/- 0.0032	0.7309 +/- 0.0175	11.6 +/- 0.8

The result for Death Valley shown in Table 6.4 portrays that the maximum difference in average RMSE is 0.0074 – a 10.91% increasing – considering the easiest 5%-1% and the roughest 30%-10% MAR cases. For the Bitcoin price and the Airfoil sound pressure, the same difference is of 14.74% and 10.71%, respectively. While 30% of missing values can be quite detrimental to other prediction models (as empirically shown in the next section), eFGP copes with MAR values in an evolving fashion, thus keeping reasonable RMSE and NDE rates. Notice also in Table 6.4 that the standard deviation of the error indices for the 10 runs of the algorithm for each MAR case is minimal. Therefore, from the empirical results it can be concluded that the eFGP learning approach provides prediction models that are robust to MAR values.

In summary, eFGP has shown to be able to handle nonstatinary data streams containing MCAR and MAR values at different rates. The behavior of the algorithm has been stable in different real missing data scenarios. MAR data are more easily dealt with by the algorithm compared to MCAR data. The latter requires a greater number of information granules and a larger expansion of the fuzzy rule base whereas parametric adaptation prevails in the former case.

6.2.1 Comparing Evolving Intelligent Models

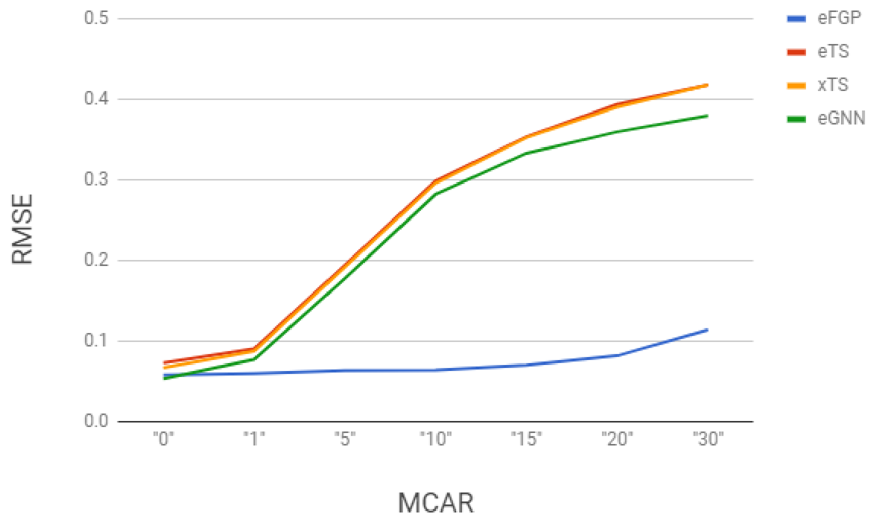
eFGP models are compared with evolving Granular Neural Networks (eGNN) (LEITE; COSTA; GOMIDE, 2013), evolving Takagi-Sugeno (eTS) (ANGELOV; FILEV, 2004) and extended Takagi-Sugeno (xTS) (ANGELOV; ZHOU, 2006) in this section. eGNN uses min-max aggregation neurons, $\rho^{[0]} = 0.2$, $h_r = 80$, $\zeta = 0.9$ (a constant useful for adapting connection weights), and $\eta = 0.5$ (a threshold for the number of neurons created in h_r time steps that is useful for adapting model granularity). Additionally, eTS uses $\Omega = 350$ and maximum radius length $r = 0.7$; and xTS employs $\Omega = 350$. As these methods are not equipped with mechanisms to impute missing data, incomplete samples are deleted and a zero-order hold approach is employed, i.e. the last prediction is replicated.

Figures 6.7 and 6.8 depict the *RMSE* indices of the predictors for the Death Valley monthly temperature data stream with different proportions of missing values. The eFGP's error index clearly increases less than that of the remaining methods due to the increasing of the percentage of MCAR values. eFGP takes full advantage of the information within incomplete samples instead of discarding them. The actual correlation matrix among attributes tends to become more distorted if entire samples are simply removed from the stream, which hampers more accurate predictions of the other predictors.

Figure 6.8 endorses the previous analysis and highlights that the percentage of missing values related to the attributes with less missing values is more influential to the predictions, specially for eTS, xTS and eGNN. The 20–10 and 30–10 MAR cases are the hardest scenarios for evolving methods with no procedures to deal with MAR data, whereas the eFGP approach demonstrates robustness to this issue. Table 6.5 gives a summary of the average results achieved for all methods and situations illustrated in Figs. 6.7 and 6.8.

For the Bitcoin dataset, eFGP models using $\rho = \sigma = 0.01$ and $h_r = 100$ were compared to (i) eGNN using min-max aggregation neurons, $\rho^{[0]} = 0.02$, $h_r = 1500$, $\zeta = 0.9$ and $\eta = 1.5$; (ii) eTS with $\Omega = 350$ and $r = 0.4$; and (iii) xTS using $\Omega = 100$. Figure 6.9 shows that eFPG is more robust to MCAR data across the range of analyzed values. Notice also that eTS

Figure 6.7 – Performance comparison on the prediction of the Death Valley monthly temperature with missing completely at random data (MCAR)



presented the lowest average RMSE for the complete Bitcoin dataset. However, removing a small percentage (between 1 and 5%) of random values from the data stream is enough for eFGP to overcome the eTS performance. The accuracies of xTS and eFGP are essentially similar up to the 15% MCAR case, when eFGP begins to prevail.

Figure 6.10 shows prediction performance for MAR data. A different and irregular evolution of the RMSE for all models is observed. Notably, comparing the extreme 5–1 and 30–10 cases, eFGP is the unique model to have its RMSE reduced, which argues in favor of its greater robustness. Although an irregularity on the prediction performance with the increase of missing data is not a usual pattern, the Bitcoin price have particularly witnessed a high daily volatility during the last 4 years. Moreover, the randomness of the choice of missing values contributes to the irregular trends shown in Fig. 6.10. The percentage of missing values related to the attributes with less missing values is more influential to the predictions.

Figure 6.8 – Performance comparison on the prediction of the Death Valley monthly temperature with missing at random data (MAR)

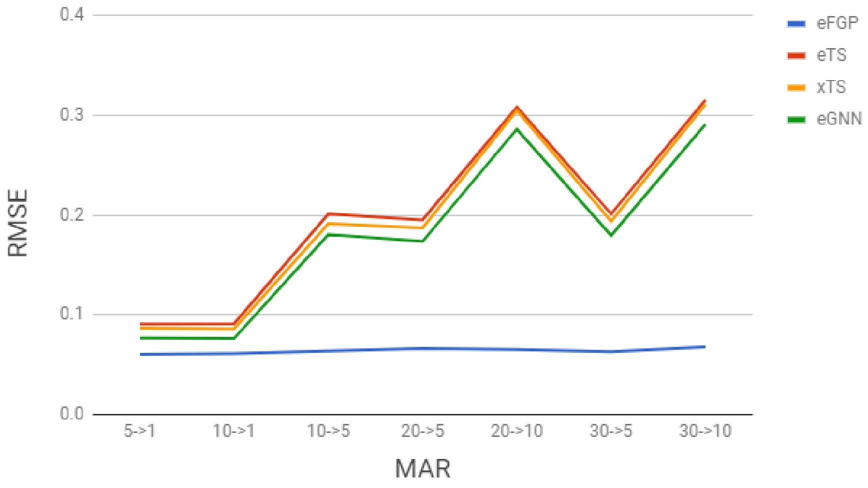


Figure 6.9 – Performance comparison on the prediction of the Bitcoin closing price with missing completely at random data (MCAR)

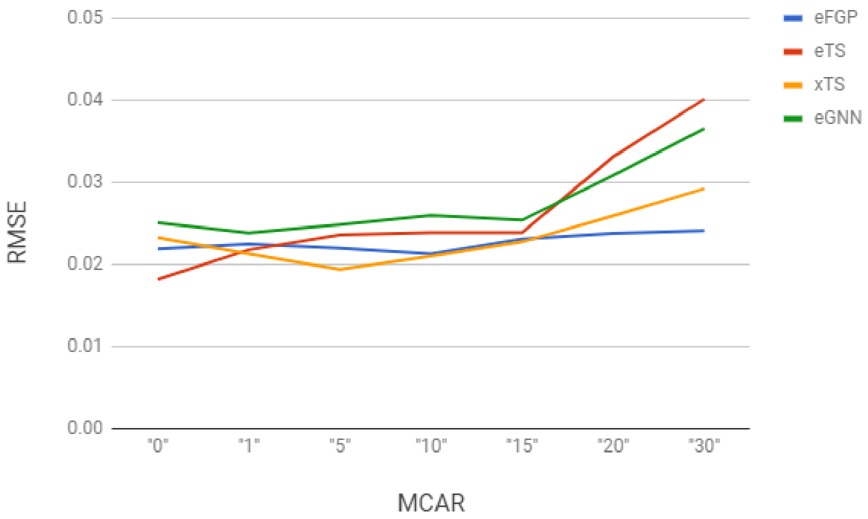


Table 6.5 – Summary results for Death Valley

Method	RMSE	NDE	# of Rules
Complete dataset			
eFGP	0.0579 +/- 0.0018	0.2239 +/- 0.0070	13.3 +/- 0.0
eGNN	**0.0532 +/- 0.0019**	**0.2055 +/- 0.0071**	14.0 +/- 0.8
eTS	0.0730 +/- 0.0000	0.2822 +/- 0.0000	11.0 +/- 0.0
xTS	0.0669 +/- 0.0000	0.2586 +/- 0.0000	**8.0 +/- 0.0**
Average results for MCAR scenarios			
Method	RMSE	NDE	# of Rules
eFGP	**0.1142 +/- 0.0057**	**0.4180 +/- 0.0222**	28.1 +/- 2.3
eGNN	0.3796 +/- 0.0237	1.4673 +/- 0.0917	8.9 +/- 0.7
eTS	0.4180 +/- 0.0095	1.6153 +/- 0.0368	6.4 +/- 0.9
xTS	0.4178 +/- 0.0432	1.6149 +/- 0.1670	**3.8 +/- 1.5**
Average results for MAR scenarios			
Method	RMSE	NDE	# of Rules
eFGP	**0.0678 +/- 0.0019**	**0.2622 +/- 0.0075**	22.7 +/- 1.2
eGNN	0.2906 +/- 0.0088	1.1233 +/- 0.0340	14.8 +/- 0.7
eTS	0.3150 +/- 0.0160	1.2174 +/- 0.0618	8.6 +/- 1.1
xTS	0.3106 +/- 0.0138	1.2004 +/- 0.0536	**5.4 +/- 1.1**

Figure 6.10 – Performance comparison on the prediction of the Bitcoin closing price with missing at random data (MAR)

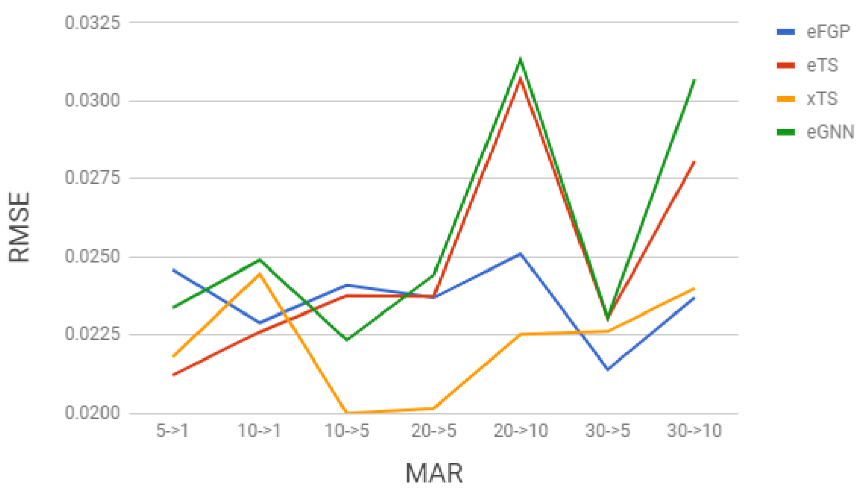


Table 6.6 summarizes the average performance and number of rules of the predictors. In this problem, eTS and eFGP are the most accurate predictors when the dataset is complete and incomplete, respectively. eGNN developed the most compact models.

In the Airfoil problem, eFGP was set with $\rho = \sigma = 0.6$ and $h_r = 48$; eGNN employs min-max aggregation neurons, $\rho^{[0]} = 0.2$, $h_r = 80$; $\zeta = 0.9$ and $\eta = 1.5$; eTS uses $\Omega = 100$ and $r = 0.9$; and xTS utilizes $\Omega = 100$. Figures 6.11 and 6.12 show the average performance of the

Table 6.6 – Summary results for the Bitcoin price

Method	RMSE	NDE	# of Rules
Complete dataset			
eFGP	0.0219 +/- 0.0044	0.1410 +/- 0.0285	14.4 +/- 0.0
eGNN	0.0251 +/- 0.0048	0.1731 +/- 0.0328	**8.0 +/- 0.0**
eTS	**0.0182 +/- 0.0000**	**0.1167 +/- 0.0000**	11.0 +/- 0.0
xTS	0.0233 +/- 0.0000	0.1497 +/- 0.0000	24.0 +/- 0.0
Average results for MCAR scenarios			
Method	RMSE	NDE	# of Rules
eFGP	**0.0252 +/- 0.0054**	**0.1619 +/- 0.0347**	16.3 +/- 1.8
eGNN	0.0366 +/- 0.0091	0.2346 +/- 0.0581	**5.5 +/- 0.7**
eTS	0.0401 +/- 0.0100	0.2574 +/- 0.0640	14.8 +/- 2.2
xTS	0.0292 +/- 0.0039	0.1877 +/- 0.0253	12.0 +/- 1.9
Average results for MAR scenarios			
Method	RMSE	NDE	# of Rules
eFGP	**0.0237 +/- 0.0035**	**0.1526 +/- 0.0284**	10.9 +/- 0.5
eGNN	0.0307 +/- 0.0088	0.1969 +/- 0.0566	**8.6 +/- 0.6**
eTS	0.0281 +/- 0.0056	0.1802 +/- 0.0361	16.4 +/- 2.0
xTS	0.0240 +/- 0.0039	0.1540 +/- 0.0252	18.4 +/- 1.1

models for MCAR and MAR cases. A monotonic trend is noticed for all models considering MCAR data and the fraction of MAR values associated to the attributes with less missing values. Clearly, eFGP is the most robust predictor in all scenarios.

Figure 6.11 – Performance comparison on the prediction of the Airfoil self-noise dataset with missing completely at random data (MCAR)

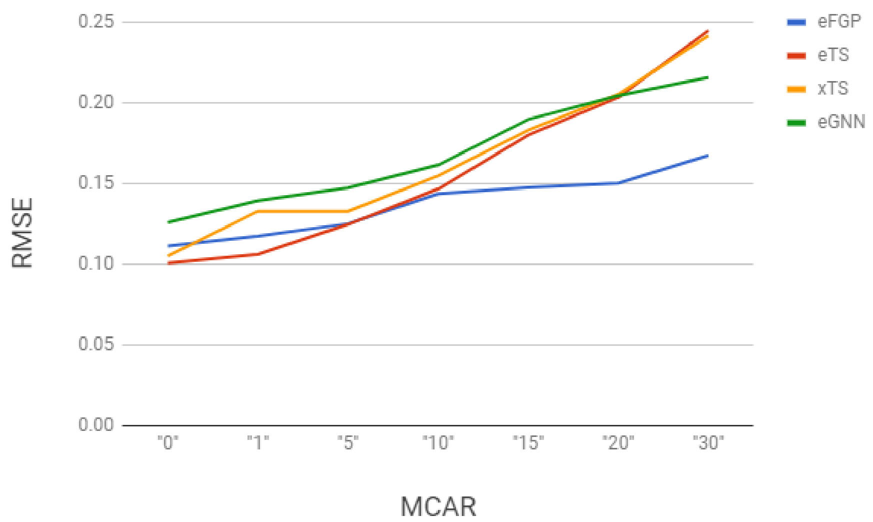


Figure 6.12 – Performance comparison on the prediction of the Airfoil self-noise dataset with missing at random data (MAR)

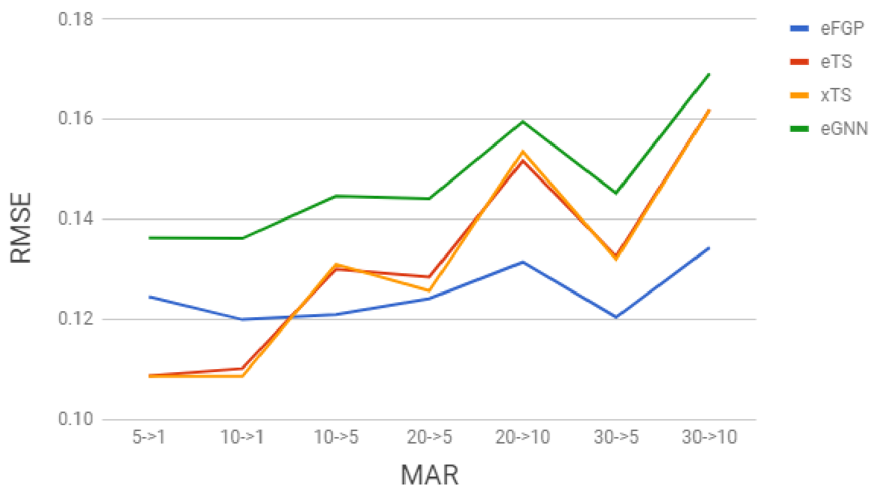


Table 6.7 compares eFGP, eGNN, eTS and xTS from the error indices and average number of rules during the models development. Even though eFGP uses only 3.4 fuzzy granules and rules in average against 9 and 12 clusters and rules of eTS and xTS for the complete Airfoil dataset, the RMSE indices of eTS (0.1006) and xTS (0.1050) are expressively better than that of eFGP (0.1114). Nonetheless, malfunction of sensors can quickly deteriorate their performance as shown in the table. eFGP overcame eGNN, eTS and xTS by 23.4%, 31.6% and 30.7% in MCAR scenarios and by 20.6%, 17.0% and 17.0% in MAR scenarios.

6.2.2 Statistical Hypothesis Testing

Balanced one-way ANalysis Of VAriance (ANOVA) (ANTONISAMY; PREMKUMAR; CHRISTOPHER, 2017) was used to compare the evolving predictors under the same roof, i.e., regardless of the application. ANOVA is able to determine whether or not there are statistically significant difference between the mean accuracy of more than 2 unrelated methods or models. The null hypothesis is that the mean accuracy of the methods is essentially the same. A cutoff value p less than 0.05 indicates that the accuracy of at least one of the methods is significantly different from the others. Considering the results of all experiments with complete datasets, a $p = 0.9982$ was obtained and, therefore, the null hypothesis holds true. In other words, any of the methods under comparison, eFGP, eGNN, eTS and xTS may provide the best estimates for a particular stream if no data is missing. eFGP is competitive with state-of-the-art evolving intelligent methods.

Table 6.7 – Summary results for the Airfoil sound pressure

Complete dataset			
Method	RMSE	NDE	# of Rules
eFGP	0.1114 +/- 0.0003	0.6059 +/- 0.0017	**3.4 +/- 0.0**
eGNN	0.1259 +/- 0.0003	0.6839 +/- 0.0019	14.9 +/- 0.0
eTS	**0.1006 +/- 0.0000**	**0.5467 +/- 0.0000**	9.0 +/- 0.0
xTS	0.1050 +/- 0.0000	0.5705 +/- 0.0000	12.0 +/- 0.0
Average results for MCAR scenarios			
Method	RMSE	NDE	# of Rules
eFGP	**0.1674 +/- 0.0064**	**0.9102 +/- 0.0351**	20.2 +/- 0.6
eGNN	0.2158 +/- 0.0058	1.1725 +/- 0.0316	45.5 +/- 1.4
eTS	0.2449 +/- 0.0183	1.3307 +/- 0.0994	12.0 +/- 5.3
xTS	0.2416 +/- 0.0199	1.3130 +/- 0.1079	**9.2 +/- 1.9**
Average results for MAR scenarios			
Method	RMSE	NDE	# of Rules
eFGP	**0.1344 +/- 0.0032**	**0.7309 +/- 0.0175**	11.6 +/- 0.8
eGNN	0.1693 +/- 0.0060	0.9197 +/- 0.0324	26.0 +/- 2.2
eTS	0.1620 +/- 0.0106	0.8804 +/- 0.0575	**9.4 +/- 0.6**
xTS	0.1620 +/- 0.0068	0.8801 +/- 0.0367	9.4 +/- 2.3

From the results of the MCAR and MAR experiments, the values $p = 0.0085$ and $p = 0.0148$, respectively, were gotten. The mean accuracy of the methods are not all the same, i.e., the null hypothesis is rejected.

The Tukey Honestly Significant Difference (HSD) test (ANTONISAMY; PREMKUMAR; CHRISTOPHER, 2017) was performed to compare pairs of methods. The Tukey test is optimal for balanced one-way ANOVA and for similar procedures with equal sample sizes. Table 6.8 shows the results of the Tukey HSD test for a 95% confidence interval (CI) for the true difference of the means.

A negative or positive difference of means denotes that the first or second method, respectively, outperformed the other. However, if the confidence interval, [LB, UB], in a row of Table 6.8 does contain 0, then the difference between the methods is not significant at the 0.05 level. In this case, for applications other than those evaluated, any of the methods can perform better and both should ideally be examined.

Notice from the MAR scenarios of Table 6.8 that eFGP models are statistically superior to the others. For MCAR scenarios, eFGP is statistically superior to eTS and xTS, and superior to eGNN. A minor positive gap on the upper bound of the confidence interval related to the MCAR eFGP-eGNN comparison prevents a statistical endorsement. The eGNN, eTS and xTS methods are similar between themselves in missing data scenarios.

Table 6.8 – Tukey Test Results

Tukey Test results with complete dataset				
Method 1	Method 2	CI LB	Diff of Means	CI UB
eFGP	eGNN	-0.3628	-0.0174	0.3279
eFGP	eTS	-0.3413	0.0041	0.3495
eFGP	xTS	-0.3474	-0.0020	0.3434
eGNN	eTS	-0.3238	0.0215	0.3669
eGNN	xTS	-0.3299	0.0154	0.3608
eTS	xTS	-0.3515	-0.0061	0.3393
Tukey Test results for MCAR scenarios				
Method 1	Method 2	CI LB	Diff of Means	CI UB
eFGP	eGNN	-0.5822	-0.2849	0.0123
eFGP	eTS	-0.6488	-0.3516	-0.0543
eFGP	xTS	-0.6318	-0.3346	-0.0373
eGNN	eTS	-0.3639	-0.0667	0.2306
eGNN	xTS	-0.3469	-0.0496	0.2476
eTS	xTS	-0.2802	0.0170	0.3143
Tukey Test results for MAR scenarios				
Method 1	Method 2	CI LB	Diff of Means	CI UB
eFGP	eGNN	-0.4462	-0.2265	-0.0068
eFGP	eTS	-0.4549	-0.2352	-0.0156
eFGP	xTS	-0.4463	-0.2266	-0.0069
eGNN	eTS	-0.2285	-0.0088	0.2109
eGNN	xTS	-0.2198	-0.0001	0.2196
eTS	xTS	-0.2110	0.0087	0.2284

LB: Lower Bound; UB: Upper Bound.

In addition to the preeminence of eFGP in terms of overall accuracy, the granular approximation of the time series and the linguistic description associated to its granular output are distinctive features to consider eFGP, especially if missing values may arise in a data stream setting.

7 CONCLUSION AND FURTHER WORK

In this study it was shed light onto the question of missing values in nonstationary data steams. An evolving granular fuzzy rule-based modeling method for function approximation and time series prediction in online settings where values may be missing at random and missing completely at random was described. eFGP models deal with single and multiple missing values using reduced-term consequent functions, partial similarity, and time-varying granules. Experimental results on actual weather, cryptocurrency and engineering applications considering from 1% to 30% of missing values have shown that the eFGP approach outperforms other evolving fuzzy and neuro-fuzzy modeling methods that resort to sample deletion and output replication. Moreover, these results are statistically significant on MAR and MCAR scenarios according to the ANOVA-Tukey test. A particular characteristic of eFGP models concerns the provision of a granular enclosure of the time series, which may assist decision making and may have a variety of interpretations in different areas.

Further study will discuss missing data imputation in semi-supervised multi-class classification of data streams. Semi-supervision presupposes part of the inputs instances are labeled and part unlabeled. In this situation online missing data imputation becomes even more challenging.

The development of fuzzy granules with different geometries and incremental adaptation of the parameters of aggregation operators will be discussed considering streams of nonstationary data subject to missing values.

Developing and adapting online active learning approaches, where the system decides whether or not it should use the input sample to evolve, will also be studied. The key challenge is how to determine if inputs with missing values are important for the model in an online recursive fashion, with all the constraints that online environments impose.

8 PUBLICATIONS

During the course of this research, some publications were produced based on or somehow related to the content of this manuscript. They are listed below, split taking into account the degree of relation with this study.

8.1 Direct contributions

- **Garcia, C.**; Leite, D.; Škrjanc, I. Incremental Missing Data Imputation for Evolving Fuzzy Granular Prediction, 12p., 2018 (To be submitted to Expert Systems With Applications);

- **Garcia, C.**; Leite, D.; Esmin, A. Evolvable Fuzzy Systems from Data Streams with Missing Values: With Application to Cryptocurrency Prediction, 2p., 2018 (To be submitted to the IET Electronics Letters).

8.2 Strongly related contributions

- Soares, E.; **Garcia, C.**; Pouças, R.; Camargo, H.; Leite, D. Evolving Fuzzy and Cloud Unsupervised Classifiers to Online Spam Data Stream, 7p., 2018. (Submitted to IEEE Latin America);

- Leite, D.; **Garcia, C.**; Pouças, R.; Ferreira, S. Classificação Online De Spam via Modelos Inteligentes Evolutivos Baseados em Nuvens e Granulos de Dados, 8p., 2018 (Submitted to the XXII Brazilian Conference on Automation).

8.3 Weakly related contributions

- **Garcia, C. M.**; Catalano, M. D.; Soares, E. A.; Barbosa, B. H. G. Teaching-Learning-Based Optimization no Treinamento de Redes Neurais Artificiais para Problemas de Classificação. XIII Brazilian Conference on Computational Intelligence, 2017, Niterói / RJ;

- **Garcia, C. M.**, Abreu, A. M.; Abilio, R.; Leite, D. Inteligência Computacional aplicada à Evasão Escolar na Educação Superior, 8p., 2018 (To be submitted to Revista Brasileira de Computação Aplicada).

REFERENCES

AL-KHATEEB, T. et al. Recurring and novel class detection using class-based ensemble for evolving data stream. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 28, n. 10, p. 2752–2764, 2016.

AMIRI, M.; JENSEN, R. Missing data imputation using fuzzy-rough methods. **Neurocomputing**, Elsevier, v. 205, p. 152–164, 2016.

ANDONOVSKI, G. et al. Evolving model identification for process monitoring and prediction of non-linear systems. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 68, p. 214–221, 2018.

ANGELOV, P. **Autonomous Learning Systems: from Data Streams to Knowledge in Real-time**. [S.l.]: John Wiley & Sons, 2012.

ANGELOV, P. Anomaly detection based on Eccentricity analysis. In: IEEE. **Evolving and Autonomous Learning Systems (EALS), 2014 IEEE Symposium on**. Orlando, FL, USA, 2014. p. 1–8.

ANGELOV, P.; GU, X.; KANGIN, D. Empirical Data Analytics. **International Journal of Intelligent Systems**, Wiley Online Library, v. 32, n. 12, p. 1261–1284, 2017.

ANGELOV, P.; LUGHOFER, E.; ZHOU, X. Evolving fuzzy classifiers using different model architectures. **Fuzzy Sets and Systems**, Elsevier, v. 159, n. 23, p. 3160–3182, 2008.

ANGELOV, P. et al. On-line evolution of Takagi-Sugeno fuzzy models. **IFAC Proceedings Volumes**, Elsevier, v. 37, n. 16, p. 67–72, 2004.

ANGELOV, P.; ZHOU, X. Evolving fuzzy systems from data streams in real-time. In: IEEE. **Evolving Fuzzy Systems, The 2006 International Symposium on**. Ambleside, UK, 2006. p. 29–35.

ANGELOV, P. P.; FILEV, D. P. An approach to online identification of Takagi-Sugeno fuzzy models. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, IEEE, v. 34, n. 1, p. 484–498, 2004.

ANGELOV, P. P.; GU, X.; PRÍNCIPE, J. Autonomous learning multi-model systems from data streams. **IEEE Transactions on Fuzzy Systems**, IEEE, 2017.

ANTONISAMY, B.; PREMKUMAR, P. S.; CHRISTOPHER, S. **Principles and Practice of Biostatistics**. [S.l.]: Elsevier Health Sciences, 2017.

ANTSAKLIS, P. J.; PASSINO, K. M. **Introduction to intelligent control systems with high degrees of autonomy**. [S.l.]: Kluwer Academic Publishers, 1993.

ÅSTRÖM, K. J.; WITTENMARK, B. **Adaptive control**. [S.l.]: Courier Corporation, 2013.

AYDILEK, I. B.; ARSLAN, A. A novel hybrid approach to estimating missing values in databases using k-nearest neighbors and neural networks. **International Journal of Innovative Computing, Information and Control**, v. 7, n. 8, p. 4705–4717, 2012.

AYDILEK, I. B.; ARSLAN, A. A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm. **Information Sciences**, Elsevier, v. 233, p. 25–35, 2013.

AZIM, S.; AGGARWAL, S. Using fuzzy c means and multi layer perceptron for data imputation: Simple v/s complex dataset. In: IEEE. **Recent Advances in Information Technology (RAIT), 2016 3rd International Conference on**. Dhanbad, Jharkhand, India, 2016. p. 197–202.

BATISTA, G. E.; MONARD, M. C. An analysis of four missing data treatment methods for supervised learning. **Applied Artificial Intelligence**, Taylor & Francis, v. 17, n. 5-6, p. 519–533, 2003.

BENGIO, Y.; GINGRAS, F. Recurrent neural networks for missing or asynchronous data. In: **Advances in Neural Information Processing Systems**. Denver, Colorado, USA: [s.n.], 1996. p. 395–401.

BERINGER, J.; HÜLLERMEIER, E. Efficient instance-based learning on data streams. **Intelligent Data Analysis**, IOS Press, v. 11, n. 6, p. 627–650, 2007.

BEZDEK, J. C.; EHRLICH, R.; FULL, W. FCM: The fuzzy C-means clustering algorithm. **Computers & Geosciences**, Elsevier, v. 10, n. 2-3, p. 191–203, 1984.

BOUCHACHIA, A.; GABRYS, B.; SAHEL, Z. Overview of some incremental learning algorithms. In: IEEE. **Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International**. [S.l.], 2007. p. 1–6.

BOYLES, S. Comparison of Interpolation Methods for Missing Traffic Volume Data. In: **Transportation Research Board 90th Annual Meeting**. [S.l.: s.n.], 2011.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to time series and forecasting**. [S.l.]: Springer, 2016.

CASTILLO, P. Rey-del; CARDEÑOSA, J. Fuzzy min–max neural networks for categorical data: application to missing data imputation. **Neural Computing and Applications**, Springer, v. 21, n. 6, p. 1349–1362, 2012.

DITZLER, G.; POLIKAR, R. Incremental learning of concept drift from streaming imbalanced data. **ieee transactions on knowledge and data engineering**, IEEE, v. 25, n. 10, p. 2283–2301, 2013.

DOMINGOS, P. M.; HULTEN, G. Catching up with the Data: Research Issues in Mining Data Streams. **Data Mining and Knowledge Discovery**, 2001.

DOVŽAN, D.; LOGAR, V.; ŠKRJANC, I. Implementation of an evolving fuzzy model (efumo) in a monitoring system for a waste-water treatment process. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 23, n. 5, p. 1761–1776, 2015.

DOVŽAN, D.; ŠKRJANC, I. Predictive functional control based on an adaptive fuzzy model of a hybrid semi-batch reactor. **Control Engineering Practice**, Elsevier, v. 18, n. 8, p. 979–989, 2010.

FARHANGFAR, A.; KURGAN, L. A.; PEDRYCZ, W. A novel framework for imputation of missing values in databases. **IEEE Trans Syst Man Cybern Part A**, IEEE, v. 37, n. 5, p. 692–709, 2007.

GAMA, J. **Knowledge discovery from data streams**. [S.l.]: CRC Press, 2010.

GAMA, J. et al. A survey on concept drift adaptation. **ACM Computing Surveys (CSUR)**, ACM, v. 46, n. 4, p. 44, 2014.

GARCÍA-LAENCINA, P. J. et al. Multi-task neural networks for dealing with missing inputs. In: SPRINGER. **International Work-Conference on the Interplay Between Natural and Artificial Computation**. La Manga del Mar Menor, Spain, 2007. p. 282–291.

GRZYMALA-BUSSE, J. W. et al. Handling missing attribute values in preterm birth data sets. In: SPRINGER. **International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing**. Regina, SK, Canada, 2005. p. 342–351.

GUNN, I. A.; ARNAIZ-GONZÁLEZ, Á.; KUNCHEVA, L. I. A taxonomic look at instance-based stream classifiers. **Neurocomputing**, Elsevier, v. 286, p. 167–178, 2018.

JANG, J.-S. R.; SUN, C.-T.; MIZUTANI, E. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence. **IEEE Transactions on Automatic Control**, IEEE, v. 42, n. 10, p. 1482–1484, 1997.

JENSEN, R.; CORNELIS, C. Fuzzy-rough nearest neighbour classification. **Transactions on Rough Sets XIII**, Springer, p. 56–72, 2011.

KASABOV, N. K. **Evolving Connectionist Systems: the Knowledge Engineering Approach**. [S.l.]: Springer Science & Business Media, 2007.

KASABOV, N. K.; SONG, Q. DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 10, n. 2, p. 144–154, 2002.

KUPPUSAMY, V.; PARAMASIVAM, I. Grey fuzzy neural network-based hybrid model for missing data imputation in mixed database. **International Journal of Intelligent Engineering and Systems**, v. 10, n. 3, p. 146–155, 2017.

LEITE, D. **Evolving Granular Systems**. Tese (Doutorado) — School of Electrical and Computer Engineering, University of Campinas, 2012.

LEITE, D. et al. Evolving fuzzy granular modeling from nonstationary fuzzy data streams. **Evolving Systems**, Springer, v. 3, n. 2, p. 65–79, 2012.

LEITE, D.; COSTA, P.; GOMIDE, F. Evolving granular neural network for semi-supervised data stream classification. In: IEEE. **Neural Networks (IJCNN), The 2010 International Joint Conference on**. Barcelona, Spain, 2010. p. 1–8.

LEITE, D.; COSTA, P.; GOMIDE, F. Evolving granular neural network for fuzzy time series forecasting. In: IEEE. **Neural Networks (IJCNN), The 2012 International Joint Conference on**. [S.l.], 2012. p. 1–8.

LEITE, D.; COSTA, P.; GOMIDE, F. Interval approach for evolving granular system modeling. In: **Learning in Non-stationary Environments**. [S.l.]: Springer, 2012. p. 271–300.

LEITE, D.; COSTA, P.; GOMIDE, F. Evolving granular neural networks from fuzzy data streams. **Neural Networks**, Elsevier, Brisbane, Australia, v. 38, p. 1–16, 2013.

LEITE, D. et al. Evolving granular fuzzy model-based control of nonlinear dynamic systems. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 23, n. 4, p. 923–938, 2015.

LI, D.; GU, H.; ZHANG, L. A hybrid genetic algorithm–fuzzy c-means approach for incomplete data clustering based on nearest-neighbor intervals. **Soft Computing**, Springer, v. 17, n. 10, p. 1787–1796, 2013.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical Analysis with Missing Data**. New York, NY, USA: John Wiley & Sons, Inc., 1986. ISBN 0-471-80254-9.

LOPES, P. de A.; CAMARGO, H. de A. Fuzzstream: Fuzzy data stream clustering based on the online-offline framework. In: IEEE. **Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on**. Naples, Italy, 2017. p. 1–6.

LUENGO, J.; GARCÍA, S.; HERRERA, F. On the choice of the best imputation methods for missing values considering three groups of classification methods. **Knowledge and Information Systems**, Springer, v. 32, n. 1, p. 77–108, 2012.

LUENGO, J.; SÁEZ, J. A.; HERRERA, F. Missing data imputation for fuzzy rule-based classification systems. **Soft computing**, Springer, v. 16, n. 5, p. 863–881, 2012.

LUGHOFER, E. D. FLEXFIS: A robust incremental learning approach for evolving Takagi–Sugeno fuzzy models. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 16, n. 6, p. 1393–1410, 2008.

LV, Y. et al. Traffic flow prediction with big data: a deep learning approach. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 16, n. 2, p. 865–873, 2015.

MACIEL, L.; BALLINI, R.; GOMIDE, F. Evolving fuzzy modelling for yield curve forecasting. **International Journal of Economics and Business Research**, Inderscience Publishers (IEL), v. 15, n. 3, p. 290–311, 2018.

MORSHEDIZADEH, M. et al. Application of imputation techniques and adaptive neuro-fuzzy inference system to predict wind turbine power production. **Energy**, Elsevier, v. 138, p. 394–404, 2017.

MOSHTAGHI, M. et al. Evolving fuzzy rules for anomaly detection in data streams. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 23, n. 3, p. 688–700, 2015.

NANNI, L.; LUMINI, A.; BRAHNAM, S. A classifier ensemble approach for the missing feature problem. **Artificial Intelligence in Medicine**, Elsevier, v. 55, n. 1, p. 37–50, 2012.

NELWAMONDO, F. V.; GOLDING, D.; MARWALA, T. A dynamic programming approach to missing data estimation using neural networks. **Information Sciences**, Elsevier, v. 237, p. 49–58, 2013.

NUOVO, A. G. D. Missing data analysis with fuzzy c-means: A study of its application in a psychological scenario. **Expert Systems with Applications**, Elsevier, v. 38, n. 6, p. 6793–6797, 2011.

PEDRYCZ, W. **Knowledge-based Clustering: from Data to Information Granules**. [S.l.]: John Wiley & Sons, 2005.

PEDRYCZ, W. Evolvable fuzzy systems: some insights and challenges. **Evolving Systems**, Springer, v. 1, n. 2, p. 73–82, 2010.

PIGOTT, T. D. A review of methods for missing data. **Educational Research and Evaluation**, Taylor & Francis, v. 7, n. 4, p. 353–383, 2001.

PRATAMA, M. et al. Panfis: A novel incremental learning machine. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, v. 25, n. 1, p. 55–68, 2014.

RAZAVI-FAR, R.; SAIF, M. Imputation of missing data using fuzzy neighborhood density-based clustering. In: IEEE. **Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on**. Vancouver, BC, Canada, 2016. p. 1834–1841.

RUBIO, J. de J. USNFIS: Uniform Stable Neuro Fuzzy Inference System. **Neurocomputing**, Elsevier, v. 262, p. 57–66, 2017.

SAHA, S. et al. An improved fuzzy based missing value estimation in dna microarray validated by gene ranking. **Advances in Fuzzy Systems**, Hindawi Publishing Corp., v. 2016, p. 1, 2016.

SAMAT, N. A.; SALLEH, M. N. M. A Study of Data Imputation Using Fuzzy C-Means with Particle Swarm Optimization. In: SPRINGER. **International Conference on Soft Computing and Data Mining**. Bandung, Indonesia, 2016. p. 91–100.

SARAVANAN, P.; SAILAKSHMI, P. Missing value imputation using fuzzy possibilistic c means optimized with support vector regression and genetic algorithm. **Journal of Theoretical & Applied Information Technology**, v. 72, n. 1, 2015.

SILVA, J. A. et al. Data stream clustering: A survey. **ACM Computing Surveys (CSUR)**, ACM, v. 46, n. 1, p. 13, 2013.

SOARES, E. et al. Ensemble of evolving data clouds and fuzzy models for weather time series prediction. **Applied Soft Computing**, Elsevier, 2017.

SUBBIAN, K.; AGGARWAL, C. C. Mining social streams: Models and applications. **Bulletin of the IEEE Computer Society Technical Committee on Data Engineering**, 2017.

TAN, W. et al. Social-network-sourced big data analytics. **IEEE Internet Computing**, IEEE, v. 17, n. 5, p. 62–69, 2013.

TANG, J. et al. A hybrid approach to integrate fuzzy c-means based imputation method with genetic algorithm for missing traffic volume data estimation. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 51, p. 29–40, 2015.

TROYANSKAYA, O. et al. Missing value estimation methods for dna microarrays. **Bioinformatics**, Oxford University Press, v. 17, n. 6, p. 520–525, 2001.

TSAI, M.-H.; AGGARWAL, C. C.; HUANG, T. S. Towards classification of social streams. In: SIAM. **Proceedings of the 2015 SIAM International Conference on Data Mining**. Vancouver, BC, Canada, 2015. p. 649–657.

WALKER, S. J. **Big Data: A Revolution that will transform how we live, work, and think**. [S.l.]: Taylor & Francis, 2014.

WU, S.; ER, M. J.; GAO, Y. A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks. **IEEE Transactions on Fuzzy Systems**, IEEE, v. 9, n. 4, p. 578–594, 2001.

YAGER, R. R. Intelligent social network analysis using granular computing. **International Journal of Intelligent Systems**, Wiley Online Library, v. 23, n. 11, p. 1197–1219, 2008.

ZADEH, L. A. Fuzzy sets. **Information and Control**, Elsevier, v. 8, n. 3, p. 338–353, 1965.

ZHANG, L.; BING, Z.; ZHANG, L. A hybrid clustering algorithm based on missing attribute interval estimation for incomplete data. **Pattern Analysis and Applications**, Springer, v. 18, n. 2, p. 377–384, 2015.

ZHANG, L. et al. A global clustering approach using hybrid optimization for incomplete data based on interval reconstruction of missing value. **International Journal of Intelligent Systems**, Wiley Online Library, v. 31, n. 4, p. 297–313, 2016.

ZHU, H. et al. Case-deletion measures for models with incomplete data. **Biometrika**, Oxford University Press, v. 88, n. 3, p. 727–737, 2001.

APPENDIX A – Terms and Definitions

- **Adaptive**: in the context of intelligent systems, *adaptive* refers to the ability of a model to have its parameters updated over time.

- **Data stream**: general (potentially endless and unbounded) flow of information, characterized by nonlinearity, nonstationarity and heterogeneity, subject to changes of several types (BERINGER; HÜLLERMEIER, 2007)(LEITE, 2012).

- **Discrete time series**: set of observations x_t in which each observation is recorded at a particular time t. In discrete times series, observations are read at fixed time intervals (BROCKWELL; DAVIS, 2016).

- **Dynamic system**: system generally described by differential or difference equations, which highlights not only the steady state but also the transient response of variables.

- **Evolving**: in the intelligent systems literature, *evolving* is the ability of a model to have its parameters and structure (rule base or number of neurons) updated over time.

- **Nonstationary**: adjective given to something changing. In case of online environments, it means that the statistical distribution that governs the data changes over time.

- **Time series vs. Data stream**: time series convey static objects that can be analyzed in an offline fashion in general. A data stream requires model adaptation/evolution of its structure in online mode (LEITE, 2012).