MECAI 2021
I Workshop de Matemática, Estatística e Computação Aplicadas à Indústria

Bayesian Networks applied to sales forecasting in a large Brazilian fast-food chain

Robson Fernandes[1], Alneu de Andrade Lopes[2]
ICMC, University of Sao Paulo, Sao Carlos, Brazil

1 Abstract

The fast-food segment has become a very competitive market with large companies. Artificial intelligence methods can offer numerous benefits for this market, such as allowing the development of computational models for decision-making. This work aims to develop probabilistic models based on Bayesian Networks to make sales predictions and analyze the causality between variables that influence the sales process of certain groups of products in this segment. We consider the adoption of the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology in the stages of data science, as well as the application of structure learning algorithms in Bayesian Networks from data, and parameter learning. Finally, we performed cross-validation and evaluation based on the Mean Absolute Percentage Error (MAPE) metric. The techniques proposed achieved promising results in terms of forecasting and causality analysis.

Key words: Bayesian Networks, Structure Learning, Sales Forecasting, Causal Analysis.

2 Introduction

The fast-food system is generally associated with large food chains, although it is also a type of consumption found in some more elaborate restaurants. The most promising cafeteria chains took over the world from the 1970s onwards and became one of the symbols of capitalism during the Cold War. As examples of greater prominence, we can mention McDonald's, the largest of all fast-food chains, followed by Burger King, KFC, Subway, and Pizza Hut.

Techniques related to Artificial Intelligence can offer numerous benefits in this niche, which usually respond quickly to market reactions, given the degree of competition. We highlight techniques related to sales forecasting and product causality are among the most common to assist in decision making. Thus, we believe that this research favors a causality analysis in products sold by

[1]robson.fernandes@usp.br
[2]alneu@icmc.usp.br

large fast-food chains, as well as assisting in sales projection, considering the use of probabilistic models through Bayesian networks.

This research work was based on studies on Bayesian networks developed by [3], describing them as a graphic structure that allows us to represent and reason about an uncertain domain, as well as research by [1], in which they approach that Bayesian Networks are a knowledge representation model that works with uncertain and incomplete knowledge through the Bayesian probability theory, published by the mathematician Thomas Bayes in 1763.

3 Objective and method

The general objective of this research was to develop probabilistic models, with a focus on Bayesian Networks, to infer the sales of a fast-food franchise chain and analyze the causal relationships between the variables that influence marketing certain product groups. Among the specific objectives, we seek to generate Bayesian Networks from structure learning algorithms based on punctuation and restrictions, and subsequently evaluate them for causal representation and prediction.

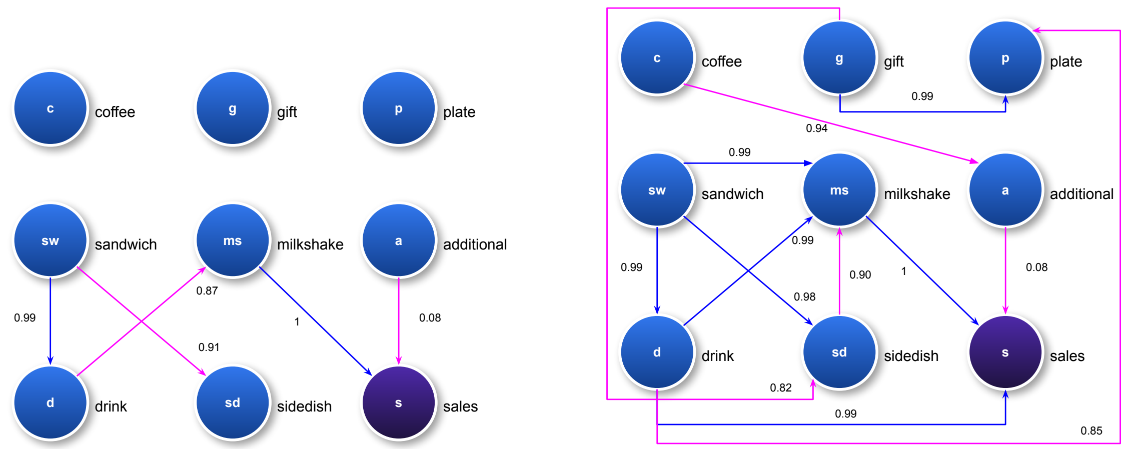
The methodology of this research was based on the CRISP-DM (Cross-industry standard process for data mining) which provides a structured approach to data mining processes, is widely used due to its powerful practicality, flexibility and usefulness when using the analysis to solve complex business problems [5]. We consider a data set acquired from a Brazilian *fast-food* chain that has about 1100 associated stores, of these, stores belonging to the state of Sao Paulo were used, in addition, variables from sales groups were evaluated in the period from 2010 to 2017. In the development of the prediction and causality models, Bayesian Network techniques were adopted with different approaches to structure learning, being: Bayesian Networks with structure learning based on restrictions, through the *Grow Shrink* (GS) algorithm [4], and learning structure based on punctuation, using the *Hill-Climbing* (HC) algorithm [4]. They were evaluated using the metric *Mean Absolute Percentage Error* (MAPE) and cross-validation, thus, we divided the total data set into two mutually exclusive subsets, one for training (parameter estimation) and another for testing (validation), considering the period from 2010 to 2016 for training, and, one year for testing, using the sample from the year 2017.

4 Results and Discussion

In Figure 1a, we have the graphical representation of the learning of the structure in Bayesian networks based on a constraint method, using the *Grow Shrink* (GS) algorithm [4]. The edges in magenta colors indicate the relations with less force between the vertices, while the edges in blue colors indicate the greatest strength. We observe that the variable *Sandwich* directly influences the variables *Drink, SideDish* and the variable *Drink* in turn, influences the variable *Milkshake*, which are related to the composition of a combo of products, and consecutively these variables influence the variable *Sales*.

In Figure 1b, we have the graphical representation of the learning of the structure in Bayesian networks based on a scoring method, using the *Hill Climbing* (HC) algorithm [4]. We observe that the variable *Sandwich* directly influences the variables *Milkshake* and *Drink*, which are generally related to the composition of a product combo, and consecutively these variables influence the variable *Sales*.

For both models, the edges have weights ranging from 0 to 1, and indicate the strength of the edge, with the edges in blue colors, indicating forces equal to or greater than 0.95, and the edges in shades of magenta, indicate forces less than 0.95.



(a) Grow Shrink Structure Learning

(b) Hill Climbing Structure Learning

Figure 1: Model - Bayesian Network. Source: Elaborated by the author.

The Bayesian Network model with scoring-based structure learning, using the *Hill-Climbing* (HC) algorithm, obtained the best results between the difference between the actual and predicted values, and its *Mean Absolute Percentage Error* (MAPE), resulted in 2.40%, while the network generated by the *Grow Shrink* (GS) algorithm obtained a MAPE of 5.04%. We used parameter adjustment using the *Maximum-Likelihood Estimation* (MLE) method [2], and later we transcribed the equations obtained described below. In the Equation (1), we have the set of mathematical models that describe the sales forecast, by adjusting the Bayesian Network with a structure-based learning *Hill-Climbing* (HC).

$$f_{\textbf{drink}}(sandwich) = -0.681 + (0.996 * ln(sandwich)) + 0.061 \tag{1a}$$

$$\begin{aligned} f_{\textbf{sidedish}}(sandwich, gift) = &-2.202 + (0.996 * sandwich) \\ &+ (0.091 * ln(gift)) + 0.084 \end{aligned} \tag{1b}$$

$$\begin{aligned} f_{\textbf{milkshake}}(sandwich, drink, sidedish) = &\ 0.913 + (-0.209 * ln(sandwich)) \\ &+ (1.005 * drink) + (0.248 * sidedish) + 0.147 \end{aligned} \tag{1c}$$

$$f_{\textbf{sales}}(milkshake, drink, additional) = e^{1.583+(0.455*milkshake)+(0.475*drink)+(0.071*ln(additional))+0.023} \tag{1d}$$

In Figure 2, the graph shows the real values compared to values expected during the 2017 period that were used for testing. The values of the variables a *priori Sandwich, Gift* and *Additional* were collected from the test set to estimate the values of the variables *Drink, Sidedish, Milkshake* and consecutively the variable *Sales*.

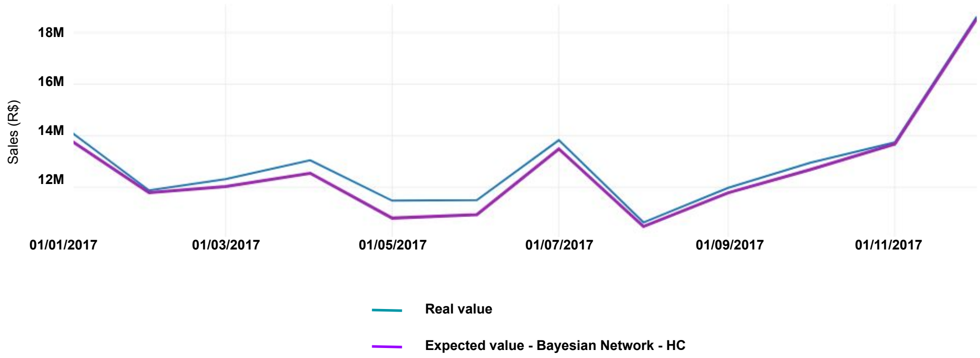


Figure 2: Sales Forecasting - Bayesian Network - *Hill-Climbing* (HC)

5 Conclusions

In this research, Bayesian Network applications applied to forecasting and causality in product sales were comparatively analyzed. It was observed that the *Hill-Climbing* (HC) structure learning algorithm found the best relationships, as well as resulting in the lowest MAPE, being 2.40%. The application of these techniques proved to be effective in terms of causality analysis, allowing managers of fast-food chains to better understand the products that influence the sales process, as well as to make sales projections.

References

[1] ISBA - International Society for Bayesian Analysis, `http://www.bayesian.org, 2019.`

[2] J. Jae. *Tutorial on maximum likelihood estimation.* Journal of mathematical Psychology 47.1 (2003): 90-100.

[3] K. B. Korb and A. E. Nicholson. *Bayesian Artificial Intelligence*. CRC press, 2010. DOI: 10.1201/b10391.

[4] D. Margaritis. *Learning Bayesian network model structure from data*. Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science, 2003.

[5] R. Wirth and J. Hipp. *CRISP-DM: Towards a standard process model for data mining.* Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. Vol. 1. London, UK: Springer-Verlag, 2000.