Detect and Correct Abnormal Values in Uncertain Environment: Application to Demand Forecast

Éric Villeneuve, Cédrick Béler, Laurent Geneste

Detect and correct abnormal values in uncertain environment : application to demand forecast

Éric Villeneuve, Cédrick Béler, Laurent Geneste

Laboratoire Génie de Production (LGP), INPT-ENIT, Université de Toulouse, Tarbes, FRANCE

Abstract This article presents the first results of a study which deals with the detection and the correction of abnormal values in data series intended to forecast demand. This work fits in the broader context of performance management for proximity retailers. Indeed, when this kind of point of sales (POS) is studied, sales volumes are often too small to be effectively exploited by statistical processing methods. It is therefore useful to consolidate the information with expertise and additional knowledge resulting from similar POS. It is also relevant to take into account the inherent uncertainty of such information. The proposal of this paper is a methodological contribution which uses consolidated knowledge to detect and correct abnormal values and to improve the quality of data used to implement forecast methods.

Keyword: Possibility theory, Combination rules, Similarity measures, Forecast

1 Context of the study

The sector of proximity retail of so-called "high tech" products is currently undergoing a major transformation. Indeed, the increasing competition with online sales and supermarkets offers a significant challenge for the retailers. French telecommunications vendors are a good example of these changes. The introduction on the French market of a new stakeholder with a very aggressive marketing policy based mainly on online sales pushed long-established operators to revise their offers. These operators have responded by shifting their activity to online sale and therefore have increased the pressure on retailers which were previously their privileged partners. The main consequence of this market changes for retailers is a significant decrease in their incomes. The need for better demand forecasting is essential to ensure the survival of their points of sales (POS).

1.1 Demand Forecast

Conventional forecast methods [1,2] produce suitable results but improving the forecast accuracy is a difficult challenge. To achieve this goal, it is necessary to take account of many parameters, such as competition, commercial and hazard management or POS typology. In addition, for retailers who want control more precisely their activity (sale forecast of a particular product by a particular seller or in a particular POS), the amount of available data is often insufficient to obtain reliable statistics. This work is a follow-up of previous studies that have begun to offer a forecast methodology adapted to these activities when the

amount of data is insufficient [3]. The proposed methodology is based on similarity measures and human expertise to consolidate data and compensate them for the lack of statistical data. The goal is to use data from other POS to inject knowledge into the model. This knowledge must be corrected to take account of the differences between POS. The correction can be done by considering similarity measures used to compare POS contexts according to criteria defined by experts. Subsequently, this consolidated knowledge allow to build a complete solution for decision support system designed to control POS commercial activity.

1.2 Abnormal Values

One of the challenges of this work lies in the selection and the formatting of the data used to produce a forecast. The main difficulty to format the data concerns the correction of abnormal values that affect the quality of the forecast. These abnormal values are the result of exceptional events (whether they have positive or negative outcomes) that disrupt the traditional sale process of a POS. For example, the release of a product which was long-awaited by the public will cause a temporary increase of sales for a given period that will not be repeated on a regular basis. This increase can disrupt all the future forecasts. It is therefore necessary to identify and correct these abnormal values to obtain better forecasts. The method studied here aims to automate the detection and the correction of abnormal values. To achieve this goal, several time series are available. They represent the sales of different POS of a retailer. Each time series is characterized by its context which is used to compare the studied series to the results of other POS using a similarity measure. Then, each new value of the series is tested to identify abnormal value. Finally, abnormal values are corrected using similar time series adapted to the context of the studied series. To avoid expert overload, the expert is only engaged in the description of the context of each point of sale and in the choice of indicators used to forecast demand. This article focuses only on the mechanisms related to formalize knowledge from multiple POS and to merge this knowledge to detect and correct abnormal values.

1.3 Issues Addressed

The study of mechanisms to detect and correct abnormal values in a time series generates several issues:

(a) How to formalize the time series to deal with the lack of data?
(b) How to take into account contextual differences between data series, i.e. the similarity level between the studied series and other series in the model?
(c) How to merge information of different POS time series?
(d) How to identify abnormal values according to resulting merged information?
(e) How to correct abnormal values by taking into account the information from different POS and similarity measures between contexts?

This article will therefore seek to answer these issues. The next part will detail the proposal. Then, the methodology will be illustrated on a case study and the results will be discussed. Finally, a conclusion with prospects will be presented.

2 Proposal

The methodology developed to address the issues mentioned above, consists in a five steps process described in figure 1.

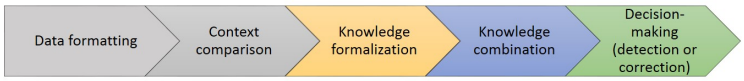


Figure 1: The proposed methodology

The first two steps of this process are not detailed in this article where the choice was made to focuses on the last three steps. The first step aims at formatting the data to be able to compare information sources. The second step enables to compare the contexts of each source to obtain similarity measures [4] to be used as confidence indicators during the combination step.

2.1 Knowledge formalization

The main problem identified for knowledge formalization lies in taking into account the uncertainty related to the lack of data. The choice was made to use possibility theory to address this issue. This section introduces the formalism and how it was used to meet identified requirements.

Possibility theory: The problem of incompleteness of information (lack of data) is widely discussed in the literature related to the more general notion of imperfect information. According to [5], incompleteness is an aspect of knowledge uncertainty. Uncertainty is defined as the fact that the information source is unable to distinguish the veracity of an information. It therefore measures a degree of conformity of information to reality. It is possible to distinguish two kinds of uncertainty. The random uncertainty is induced by the variability of an entity in a population and is the result of random experiments. The epistemic uncertainty is due to lack of knowledge and therefore is related to the notion of incompleteness. Taking into account incompleteness requires the use of a representation formalism taking into account the epistemic uncertainty. Probability theory is the most widespread representation formalism of uncertainty but it does not allow to unambiguously represent the epistemic nature of the information [6]. Therefore, possibility theory was chosen. By extending the fuzzy set theory [7], Zadeh [8] and Dubois and Prade [9] introduced the possibility theory to represent imprecise but also uncertain knowledge. In this theory, from a possibility distribution, $\pi(A)$, it is possible to construct the possibility(Π) and the necessity (N) measures thanks to the following relations:

$$N(A) = 1 - \max_{x \notin A} \pi(x) \text{ and } \Pi(A) = \max_{x \in A} \pi(x) \quad (1)$$

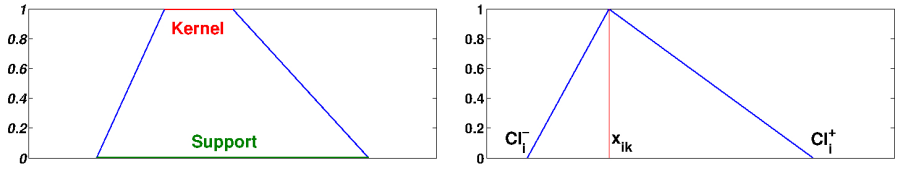
$$\max \pi(x) = 1 \quad (2)$$

It is therefore possible to characterize the uncertainty of an event, not with a value as in the context of probability theory, but with two values representing the possibility and necessity of the event. Figure 2a illustrates the principle of this theory to represent the information "*I am sure that the parameter is in [1, 5] (Support), but the values of [2, 3] (Kernel) seem the most likely*".

Building of possibility distributions: The objective is to associate possibility distribution for each value representing the studied period. For example, to detect or correct an abnormal value related to the first quarter of the current year, a possibility distribution is created for each value representing the first quarter of the past years stored in the data history of the studied source but also for all the other information sources.
To build a possibility distribution, the confidence interval of the data set which contains the values corresponding to the studied period in previous years for each information source is calculated. The confidence interval, CI_i, for each source, i, of the data series which contains the values of n last years, $X_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$, with a confidence level, $1-\alpha$, can be determined by the following formula [10]:

$$CI_i = \left[\overline{x_i} - t_{1-\frac{\alpha}{2}}(n-1) \times \frac{\sigma(X_i)}{\sqrt{n}},\ \overline{x_i} + t_{1-\frac{\alpha}{2}}(n-1) \times \frac{\sigma(X_i)}{\sqrt{n}}\right] \quad (3)$$

with $\overline{x_i}$, the mean of the values of X_i, $t_{1-\frac{\alpha}{2}}(n-1)$, the fractile of the Student law at the level $1-\frac{\alpha}{2}$ with $n-1$ freedom degrees, and $\sigma(X_i)$, the standard deviation of the sample.
Then a triangular possibility distribution is associated for each data series value by selecting as support, the confidence interval of the source, and as kernel, the concerned value (Figure 2b). The result is composed of n distributions by source.



(a) Classical possibility distribution (b) Possibility distribution built with confidence interval

Figure 2: Possibility distributions

2.2 Knowledge combination

After the knowledge formalization step, it is necessary to combine different sources to obtain a global information gathering all the available knowledge. Two levels of fusion should be made. The first level concerns the fusion of all information about one source taking into account the context evolution of this source throughout years (using similarity measures) resulting in a combined possibility distribution for each source. The second level relates to the fusion of these combined possibility distributions by taking into account the contextual differences between sources (using similarity measures). The result of this two successive fusion is a global possibility distribution which synthesize the entire knowledge about the problem.
The literature identifies many combination rules applicable to the possibility theory. The majority of these merging operators are based on t-norms and t-conorms, generalizing respectively the intersection and union in the context of fuzzy set theory [11]. The t-norm combination rules are difficult to use because they produce results difficult to interpret due to a very high sensitivity to changes

in initial possibility distributions [12]. The t-conorm fusion rules are also difficult to use, particularly for decision-making problems because of the too uncertain results they provide [12]. Therefore, adaptive rules of combination have been developed to obtain results more adapted to the reality of the studied problems. Among the existing adaptive rules, the Dubois and Prade rules [13,14] or the Deveughele rule [15] are quite common but the Delmotte rules [16,12] were chosen because they allow to explicitly take into account the confidence in information sources. This notion of confidence incorporates the similarity between information sources during the combination. The more the context of a source is similar to the context studied, the more the source is considered reliable (with a high confidence).
In this article, only the first rule of Delmotte [16] is presented. The use of the rule developed in [12] is a prospect of this work. The first rule of Delmotte [16] requires the use of a reliability $t_i \in [0,1]$ associated with each source i. The combination between sources is done with the following formula:

$$\pi_0^*(x)=\left(1-\prod_{i=1}^{n}(1-t_i)\right)\times\left(1-\prod_{i=1}^{n}t_i\right)\times\max_{i=1}^{n}(t_i\times\pi_i(x))+\prod_{i=1}^{n}t_i\times\min_{i=1}^{n}(t_i\times\pi_i(x)) \quad (4)$$

This combination behaves as a conjunction when all sources are reliable and as a disjunction of the most reliable sources when no source is completely reliable. The resulting distributions have to be normalize to respect the constraint (2).

2.3 Decision-making

The studied problem generates two kind of decision. The first decision is about determining whether a value is abnormal or not by using the possibility distribution which represents the fusion of all opinions from the different sources. It is enough to determine the possibility level of the tested value and, if its level is below a user threshold, the value is therefore considered abnormal.
The second decision is related to the correction of an abnormal value and is a classic problem of decision-making under uncertainty. Indeed, decision-making in the context of possibility theory usually consists in removing the uncertainty of a possibility distribution to end up with a unique and precise value for the studied variable. This is exactly the purpose of the correction value which aims to obtain a precise value, by using a possibility distribution representing the combined opinion of all information sources. This corrected value will be used to make the forecast. There are several "defuzzification" methods for such purpose [17]. One of the most common, Mean Of Maximum (MOM) method, has been used in this study and consists in choosing the mean value of the set of maximum possibility level values.

3 Application to a case study

To illustrate the proposal, a realistic case study has been built. The studied retailer has four POS and has a history of quarterly sales for each POS during the last three years. The objective is to consolidate the data of the POS n°1 using data from other POS to check if the value of the year 4 first quarter is abnormal and, if necessary, correct it.

Data formatting: To facilitate the comparison between POS, the expert chose to work with seasonal coefficients (SC) (ratio of sale volume in the period divided by the mean sale volume for the year). This choice allows to easily compare POS with similar seasonal variations. To test the first quarter SC of the year 4, all the first quarter SC in database are selected (Table 1).

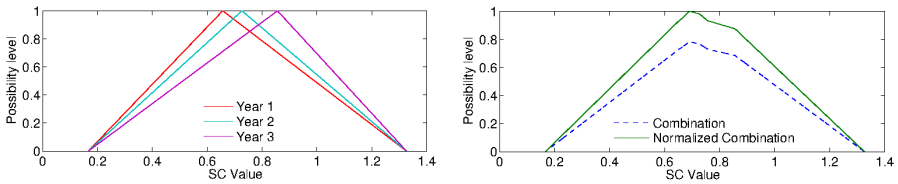
Context comparison: The Contextual Similarity measures (CS) between the studied POS and the other POS are made by using different attributes of their contexts. For example, differences in POS locations (downtown, commercial area, ...) influence the evolution of sale volume and therefore must be taken into account in the similarity measures between POS.
Also using an Annual Similarity (AS) allows to reduce the influence of older data coming from previous years to deal with the fact that the older the data are, the less relevant they are. Table 1 summarizes the information on the similarity measures (elicited by experts).

Table 1: Initial seasonal coefficients and similarity measures

	POS n°1 $CS_1 = 1$		**POS n°2** $CS_2 = 0.95$		**POS n°3** $CS_3 = 0.6$		**POS n°4** $CS_4 = 0.7$	
	AS	**SC**	**AS**	**SC**	**AS**	**SC**	**AS**	**SC**
Year 1	0.8	0.5116	0.8	0.4812	0.8	0.6565	0.8	0.3526
Year 2	0.9	0.5227	0.9	0.4907	0.9	0.7268	0.9	0.3636
Year 3	1	0.5322	1	0.4963	1	0.8560	1	0.4000

Knowledge formalization: Formalization starts with the calculation of the confidence interval for each POS using the equation 3. In a conservative approach, the confidence levels of the interval can be set to 99% ($\alpha = 1\%$). For example, the confidence interval of POS n°3 is [0.1666, 1.3263]. Then the triangular possibility distributions are built using these confidence intervals. Figure 3 shows the three possibility distributions (corresponding to three years of data set) for the POS n°3.

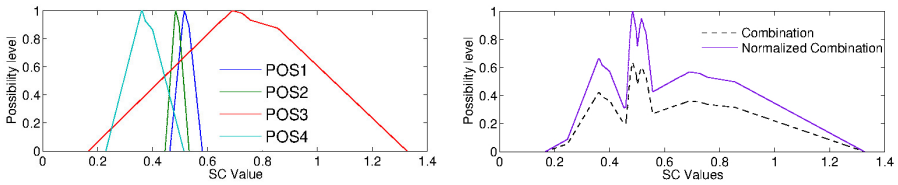
Knowledge combination: Possibility distributions must be combined to obtain a global information. The first combination is done at the POS level. The first Delmotte rule of combination (equation 4) uses AS (considered as confidence levels) to achieve combination. The result is then normalized. The figure 3 shows the result of this combination for the POS n°3.



(a) Initial possibility distributions (b) Combinations

Figure 3: Possibility distributions of POS n°3

Once each POS has a combined possibility distribution, it is necessary to combine opinions of different POS by taking now into account CS. These are, in the same manner as AS, considered as reliabilities in Delmotte rule. The result of this combination is also normalized. The figure 4 shows the result of the global combination for the case study.



(a) POS combined possibility distributions (b) Global combination

Figure 4: Global combination for the case study

Decision-making: As described above, there are two levels of decision-making. The first is to determine whether a value is abnormal or not based on a threshold set by the user. In the above example, if the threshold is set at 0.7, the value must be in the interval [0.4707, 05407] to be validated. The second level is to determine the most appropriate value to correct an abnormal value. Given that there is only one maximum, the MOM method allows to chose this value for the correction. The value in this example will be corrected to 0.4840.

Discussion: These results raise several issues, including the choice of the combination rule. Tests were conducted with other combination rules but none takes naturally into account the source confidence. In addition, the adaptive behavior of the Delmotte rule is an advantage for the treatment of complex and varied problems as studied here. The other issue is related to the defuzzification method. The choice of the MOM method is questionable and requires further work. Ideally, the result should not be subjected to defuzzification and uncertainty should be fully propagated to the final result (i.e. the forecast result). This choice would retain all the information and improve the quality of the results provided to the decision maker, not with a more accurate result but with a realistic result that will allow to make a decision by better taking into account the uncertainty. However, the uncertainty propagation requires important computing resources and therefore is difficult to implement. This is why it should be relevant to work on appropriate methods to make a partial defuzzification of the result without introducing too much bias, especially considering side effects.

4 Conclusion and prospects

This article describes a method to detect abnormal values in data used in demand forecast. This method also allows to correct these abnormal values by consolidating the initial data. This consolidation is done by using data from similar data sources together with human expertise (for the contextual analysis of the sources and the model parameterization). The implementation of a case study have illustrated the feasibility of this method. A prospect of this work is the replacement of the first version of the Delmotte combination rule by its

improved version [12] to enhance the fusion accuracy. However, this rule requires the setting of two additional parameters to adjust the combination sensitivity. A test campaign must be conducted to optimize the parameter adjustments. The next step is to apply the improved method to real data from business partners.

References

1. Bourbonnais, R., Usunier, J.C.: Prévision des ventes. Economica, Paris (2007)
2. Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: Time Series Analysis: Forecasting and Control. Wiley (2008)
3. Malo, A., Villeneuve, E., Geneste, L., Martinez, O.: Consolidation des données statistiques par expertise et similarité pour la prévision des ventes. In: 10ème Congrès International Pluridisciplinaire en Qualité et Sûreté de Fonctionnement : QUALITA 2013, Compiègne, France. (2013)
4. Bisson, G.: Why and how to define a similarity measure for object based representation systems. In: 2nd International Conference on Building and Sharing Very Large-Scale Knowledge Bases. (1995)
5. Dubois, D., Prade, H.: Formal representation of uncertainty. In: Decision-Making Process, UK and Wiley (2009)
6. Denoeux, T.: Introduction to belief functions. In: First BFTA Spring School on Belief Functions, Autrans, France. (2011)
7. Zadeh, L.A.: Fuzzy sets. Information and Control (1965)
8. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems (1978)
9. Dubois, D., Prade, H.: Possibility theory. In: Wiley Encyclopedia of Electrical and Electronics Engineering, John Wiley and Sons, Inc. (2001)
10. Kendall, M.G., Stuart, A.: The Advanced Theory of Statistics, Vol.1 : Distribution Theory. Griffin and Co (1943)
11. Weber, S.: A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms. Fuzzy Sets and Systems (1983)
12. Delmotte, F.: Detection of defective sources in the setting of possibility theory. Fuzzy Sets and Systems (2007)
13. Dubois, D., Prade, H.: Adaptive combination rules for possibility distributions. In: 2nd European Congress on Intelligent Technics and Soft Computing. (1994)
14. Dubois, D., Prade, H.: La fusion d'informations imprécises. Traitement du Signal (1994)
15. Deveughele, S., Debuisson, B.: The influence of a conflict index in the frame of the adaptive combination. In: CESA'96 IMACS Multiconference: computational engineering in systems applications. (1996)
16. Delmotte, F., Borne, P.: Modeling of reliability with possibility theory. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans (1998)
17. Runkler, T.: Selection of appropriate defuzzification methods using application specific properties. Fuzzy Systems, IEEE Transactions on (1997)