

RESEARCH

Open Access

Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques

Dennis Maaß, Marco Spruit* and Peter de Waal

* Correspondence: m.r.spruit@uu.nl
Department of Information and
Computing Sciences, Utrecht
University, Utrecht, The Netherlands

Abstract

Today's economy is characterized by increased competition, faster product development and increased product differentiation. As a consequence product lifecycles become shorter and demand patterns become more volatile which especially affects the retail industry. This new situation imposes stronger requirements on demand forecasting methods. Due to shorter product lifecycles historical sales information, which is the most important source of information used for demand forecasts, becomes available only for short periods in time or is even unavailable when new or modified products are introduced. Furthermore the general trend of individualization leads to higher product differentiation and specialization, which in itself leads to increased unpredictability and variance in demand. At the same time companies want to increase accuracy and reliability of demand forecasting systems in order to utilize the full demand potential and avoid oversupply. This new situation calls for forecasting methods that can handle large variance and complex relationships of demand factors.

This research investigates the potential of data mining techniques as well as alternative approaches to improve the short-term forecasting method for short-lifecycle products with high uncertainty in demand. We found that data mining techniques cannot unveil their full potential to improve short-term forecasting in this case due to the high demand uncertainty and the high variance of demand patterns. In fact we found that the higher the variance in demand patterns the less complex a demand forecasting method can be.

Forecasting can often be improved by data preparation. The right preparation method can unveil important information hidden in the available data and decrease the perceived variance and uncertainty. In this case data preparation did not lead to a decrease in the perceived uncertainty to such an extent that a complex forecasting method could be used. Rather than using a data mining approach we found that using an alternative combined forecasting approach, incorporating judgmental adjustments of statistical forecasts, led to significantly improved short-term forecasting accuracy. The findings are validated on real world data in an extensive case study at a large retail company in Western Europe.

Keywords: Demand forecasting; Sales forecasting; Consumer products; Fashion products; Short life-cycle products; Data mining; Predictive modeling; Big data; Sales forecast; Combined forecasting; Judgmental forecasting; Data preparation; Domain knowledge; Contextual knowledge; Demand uncertainty; Retail; Retail testing; Demand volatility; Impulsive buying

Background

Consumer products can be segmented into two different types of products regarding their demand patterns: basic or functional products and fashion or innovative products (Fisher & Rajaram 2000). Basic products have a long life-cycle and stable demand, which is easy to forecast with standard methods. Fashion products on the other hand have a short life-cycle and highly unpredictable demand. Due to their short life-cycles fashion products are often bought just once prior to a selling period (and not reordered after demand occurred which is usually the case for basic products) which makes them hard to forecast. Fashion products thus need different forecasting methods than basic products.

The problem of demand forecasting of fashion type products is described as being a problem of high uncertainty, high volatility and impulsive buying behavior (Christopher et al. 2004). Furthermore, Fisher & Rajaram (2000) describe it as a problem that is highly unpredictable. Several authors propose not to try to forecast demand for these products, but instead build an agile supply chain that can satisfy demand as soon as it occurs (e.g. Christopher et al. 2004). In practice this is very expensive solution and for our case even unfeasible due to the extremely short life-cycles.

Data mining and machine learning techniques have been shown to be more accurate than statistical models in real world cases when relationships become more complex and/or non-linear (Thomassey & Fiordaliso 2006). Classical models, like regression models, time series models or neural networks, are also generally inappropriate when short historic data is used that is disturbed by explanatory variables (Kuo & Xue 1999). Data mining techniques have already been successfully applied on demand forecasting problems (Fisher & Rajaram 2000; Thomassey & Fiordaliso 2006). In this paper we report on an analysis of demand forecasting improvements using data mining techniques and alternative forecasting methods in the context of a large retail company in Western Europe.

Problem description

The forecasting problem in this research is to predict the demand for each product in each outlet of the case company. The short-term demand forecast is used for distributing the products from the central warehouse to the outlets in the most profitable way, but not for determining the optimal buying quantity. In fact total product quantities are assumed to be fixed for this problem since products are only bought once in a single tranche prior to the selling period according to the outcome of a long-term forecasting process which is not discussed in this research.

The currently used forecasting method at the case company largely depends on retail testing. Retail tests are experiments in a small subset of the available stores, in which products are offered for sale under controlled conditions several weeks before the start of the main selling period. Additionally to demand also price elasticity is tested during the retail test. The measured price elasticity is then used in a dynamic pricing approach to maximize profits, given that total product quantities are fixed. The dynamic pricing approach optimizes the tradeoff between expected sales, already ordered quantity and change of expected sales through price alteration. For this purpose each product is presented at different prices to the customer. The allocation of price level to each product-outlet combination is done randomly but there are

always a fixed number of outlets having the same price for a given product. The random allocation scheme is used in order to minimize interaction effects between the different price levels of the products (high prices for a certain product could induce the customer to buy another cheaper product). The retail test is thus used to determine the sales potential and the price elasticity of each product. After the retail test the price for each product is set by a separate advisory board according to profit maximization goals (selling most of the bought quantity at the highest price possible within 4–6 weeks).

Literature review of existing forecasting methods and data mining techniques

Most of the standard forecasting methods for fashion type products are not able to deal with complex demand patterns or uncertainty. In the following we will present, next to data mining methods, those methods that have a potential to be useful for forecasting of fashion type products. Furthermore we will introduce data preparation methods which are especially important for this problem because they can transform the input data in such a way that uncertainty and volatility is reduced. This enables forecasting methods to deliver better results when they are applied on the transformed input data.

Data mining methods

Definition of data mining

Hand (1998) defines data mining as *“the process of secondary analysis of databases aimed at finding unsuspected relationships which are of interest or value to the database owner”*. He states that *“data mining [...] is entirely concerned with secondary data analysis”*, i.e. the analysis of data that was collected for other purposes but not the questions to be answered through the data mining process. This is opposed to primary data analysis where data is collected to test a certain hypothesis. According to Hand (1998) data mining is a new discipline that arose as a consequence of the progress in computer technology and electronic data acquisition, which lead to the creation of large databases in various fields. In this context data mining can be seen as a set of tools to unveil valuable information from these databases. With secondary data analysis there is the danger of sampling bias, which can lead to erroneous and inapplicable models (Pyle 1999).

Simoudis (1996) views data mining as *“the process of extracting valid, previously unknown, comprehensible, and actionable information from large databases and using it to make crucial business decisions”*.

A similar definition is given by Fayyad et al. (1996) although they use the term knowledge discovery in databases (KDD) instead of data mining. They use the term data mining only to denote the step of applying algorithms on data. Thus, their definition of knowledge discovery in databases is in fact also a definition of data mining: *“KDD is the process of using the database along with any required selection, preprocessing, sub sampling, and transformations of it; to apply data mining methods (algorithms) to enumerate patterns from it; and to evaluate the products of data mining to identify the subset of the enumerated patterns deemed ‘knowledge’”*.

Weiss & Indurkha (1998) state that data mining is *“the search for valuable information in large volumes of data”*. They also highlight that it is a cooperative effort of humans and

computers where humans describe the problem and set goals while computer sift through the data, looking for patterns that match with the given goals.

As can be seen from Table 1 definitions of data mining are very similar. One perceivable difference is that Hand (1998) sees relationships as the output of the data mining process instead of information or knowledge as the other authors. Although it appears to be different from the other definitions on the first view, both definitions can be seen as equal because information is created from the interpretation of the relationships between the variables (Pyle 1999). Overall we can say that there is no dispute or misconception about a definition of the term data mining

Despite the fact that data mining is seen as secondary data analysis (Hand 1998) the forecasting problem described in this case study is in fact (at least to large part) a primary data analysis since the case company actively conducts an experiment (the retail test) in order to determine the expected sales potential of their newly introduced products.

The application and success of the data mining (or knowledge discovery process) is largely dependent on data preparation techniques. As Weiss & Indurkha (1998) state: *“In many cases, there are transformations of the data that can have a surprisingly strong impact on results for prediction methods. In this sense, the composition of the features is a greater determining factor in the quality of results than the specific prediction methods used to produce those results.”* Thus, we cannot split the application of machine learning algorithms and the preceding data preparation tasks. Both processes are dependent on each other.

There are two main challenges one has to cope with during a data mining project: First, it is not known in the beginning of the data mining process what structure of the data and what kind of model will lead to the desired results. As Hand (1998) states: *“The essence of data mining is that one does not know precisely what sort of structure one is seeking”*. And second, the fact that many patterns that are found by mining algorithms will “simply be a product of random fluctuations, and will not represent any underlying structure” (Hand 1998).

Data mining process

Most authors describe the same general process of how to conduct a data mining task or project. It can be described by the steps of understanding the problem, finding and

Table 1 Definitions of data mining

Author	Type	Characteristics of type	Input	Characteristics of input	Output	Characteristics of output
Fayyad et al. (1996)	Process	Non-trivial, involves search or inference	Database	Larger data sets with rich data structures	Knowledge	Valid, novel, potentially useful, ultimately understandable
Hand (1998)	Process	Secondary data analysis	Database	Secondary data	Relationships	Unsuspected, of interest or value for database owner
Simoudis (1996)	Process	Extraction process	Database	Large scale	Information	Valid, previously unknown, comprehensible, actionable, useful for making crucial business decisions
Weiss & Indurkha (1998)	Search	Cooperative effort of humans and computers	Data	Large volume	Information	Valuable

analyzing data that can be used for problem solution, prepare the data for modeling, build models using machine learning algorithms, evaluate the quality of the models and finally use the models to solve the problem. Of course this is not a linear process, many steps have to be repeated and adapted when new insights were generated by another step. Exemplary for the general process we will present the CRISP-DM method (see Table 2) which was developed as a standard process model for data mining projects of all kinds across industries.

Each activity listed in Table 2 is further split into sub-activities which we will not present in detail here (for further information see www.crisp-dm.org).

Although the CRISP-DM method describes the general steps of a data mining project it does not describe what to do for specific problem types and how exactly it should be done. We will thus provide more details of the important steps of data mining in the following section. These steps are data preparation/data transformation, data reduction (called data selection in the CRISP-DM method) and modeling.

Data mining algorithms

For the discussed problem the specific characteristics of the data mining algorithm is not essential. The complexity of the concepts that can potentially be learned can be handled by almost all available algorithms. It is much more important to provide sufficiently prepared data in this case.

Table 2 Steps and activities of the crisp-dm method

Step	Activity
1. Business understanding	Determine business objectives
	Assess situation
	Determine data mining goals
	Produce project plan
2. Data understanding	Collect initial data
	Describe data
	Explore data
	Verify data quality
3. Data preparation	Select data
	Clean data
	Construct data
	Integrate data
	Format data
4. Modeling	Select modeling technique
	Generate test design
	Build model
	Assess model
5. Evaluation	Evaluate results
	Review process
	Determine next steps
6. Deployment	Plan deployment
	Plan monitoring and maintenance
	Produce final report
	Review project

Data preparation methods

Data transformation

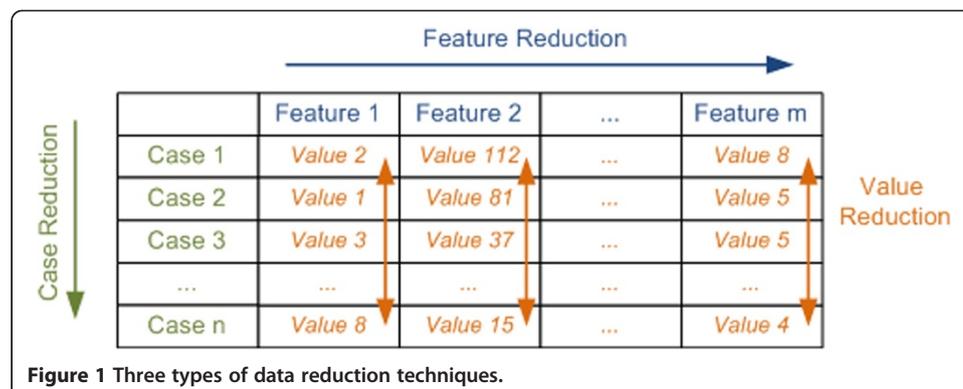
Many authors note the paramount importance of data preparation for the outcome of the whole data mining process (Pyle 1999; Weiss & Indurkha 1998; Witten & Frank 2005). The paramount importance of data preparation is due to the fact that prediction algorithms have no control over the quality of the features and must accept it as a source of error; “they are at the mercy of the original data descriptions that constrain the potential quality of solutions” (Weiss & Indurkha 1998). Pyle (1999) notes that data preparation cannot be done in an automatic way (for example with an automatic software tool). It involves human insight and domain knowledge to prepare the data in the right way. To goal of data preparation is to make the information which is enfolded in the relations between the variables of the training set “as accessible and available as possible to the modeling tool” (Pyle 1999).

Possible data preparation techniques are normalization, transformation of data into ratios or differences, data smoothing, feature enhancement, replacement of missing values with surrogates and transformation of time-series data. There are no rules that specify which techniques should be applied in a certain order given a specific problem type. The process to find the right techniques depends more on the insight and knowledge that is created during the process of data preparation and subsequent application of learning algorithms.

Data reduction

There are two good reasons for data reduction: First, although adding more variables to the data set potentially provides more information that can be exploited by a learning algorithm, it becomes, at the same time, more difficult for the algorithm to work through all the additional information (relationships between variables). That is because the number of possible combinations of relationships between variables increases exponentially, also referred to as the “combinatorial explosion” (Pyle 1999). Thus it is wise to reduce the number of variables as much as possible without losing valuable information. Second, reducing the number of variables and thus complexity can be very helpful to avoid overfitting of the learned solution to the training set.

There are three types of data reduction techniques: feature reduction, case reduction and value reduction (see Figure 1 for an overview). Feature reduction reduces the number of features (columns) in the data set through selection of the most relevant features



or combination of two or more features into a single feature. Case reduction reduces the number of cases in a data set (rows) which is usually achieved through specialized sampling methods or sampling strategies. Value reduction means reducing the number of different values a feature can take through grouping of values into a single category.

Possible feature reduction techniques are techniques such as principle components, heuristic feature selection with wrapper method and feature selection with decision trees. Examples for case reduction techniques are incremental samples, average samples, increasing the sampling period and strategic sampling of key events. For value reduction prominent techniques are rounding, using k-means clustering and discretization using entropy minimization.

Forecasting methods for demand with high uncertainty and high volatility

Not many forecasting methods can be applied in situations of high uncertainty and high volatility of demand. In the following we will thus give a short overview of methods that are applicable in this type of situation.

Judgmental adjustment of statistical forecasts

Sanders & Ritzman (2001) propose to integrate two types of forecasting methods to achieve higher accuracy: judgmental forecasts and statistical forecasts. They note that each method has strengths and weaknesses that can lead to better forecasts when they are combined. The advantage of judgmental forecasts is that they incorporate important domain knowledge into the forecasts. Domain knowledge in this context can be seen as knowledge about the problem domain that practitioners gain through experience in the job. According to Sanders & Ritzman (2001) “*domain knowledge enables the practitioner to evaluate the importance of specific contextual information*”. This type of knowledge can usually not be accessed by statistical methods but can be of high importance especially when environmental conditions are changing and when large uncertainty is present. The drawback of judgmental methods is their high potential for bias such as “optimism, wishful thinking, lack of consistency and political manipulation” (Sanders & Ritzman 2001). In contrast, statistical methods are relatively free from bias and can handle large amounts of data. However, they are just as good as the data they are provided with. Sanders & Ritzman (2001) propose the method “judgmental adjustment of statistical forecasts” to integrate judgmental with statistical methods. However, they also state that “judgmental adjustment is actually the least effective way to combine statistical and judgmental forecasts” because it can introduce bias. Instead an automated integration of both methods can provide a bias free combination of the methods. Sanders & Ritzman (2001) report that equally weighting of forecast leads to excellent results. However, in situations of very high uncertainty an overweighting of the judgmental method can lead to better results.

Transformation of time-series

Wedekind (1968) states that the type of time-series depends on the length of the time interval and that one type of time-series can be transformed into another type of time-series by changing the length of the considered time interval. We can thus transform a time-series that has trend and seasonal characteristics (time interval: month) into a time-series that has only trend characteristics by considering just intervals of annual length.

We can thus achieve a smoothing effect only by increasing the length of the time interval because we do not forecast the occurrence of a single event but of multiple events. The probability of the occurrence of a certain event is higher in a large time interval than in a small time interval. If we predict the average number of events our forecast then becomes more accurate (Nowack 2005).

Demand forecasting with data mining techniques

Thomassey & Fiordaliso (2006) propose a forecasting method for sales profiles (relative sales proportion of total sales over time) of new products based on clustering and decision trees. They cluster sales profiles of previously sold products and map new products to the sales profiles cluster via descriptor variables like price, start of selling period and life span. The mapping from descriptor variables to the sales profile cluster is learned using a decision tree. Although it is a useful approach, retail testing turns out to be much more precise than the proposed approach for the discussed problem.

Retail tests

Retail tests are “*experiments, called tests, in which products are offered for sale under carefully controlled conditions in a small number of stores*” (Fisher & Rajaram 2000). Such a test is used to test customer reaction to variables such as price, product placement or store design. If the test is used to predict season sales for a product it is called a depth test (Fisher & Rajaram 2000). In a depth test the test outlets are usually over-supplied in order to avoid stock-outs which usually distorts the forecast. The forecast is then used for the total season demand, which is ordered from a supplier before the start of the selling period.

Fisher & Rajaram (2000) report there exists no further academic or managerial literature describing how to design retail tests. In order to achieve optimal results with a retail test Fisher & Rajaram (2000) propose a clustering method to select test stores based on past sales performance. They found that clustering based on sales figures outperforms clustering on other store descriptor variables (average temperature, ethnicity, store type) significantly.

Fisher & Rajaram (2000) assume that customers differ in their preferences for products according to differing preferences for specific product attributes (e.g. color, style). Thus actual sales of a store can be thought of as a summary of product attribute preferences of the customers at that store. The clustering approach is thus based on percentage of total sales represented by each product attribute. Therefore stores are clustered according to their similarity in the percentage mix along the product attributes. Then one store from each cluster is selected as a test store to predict total season sales. The inference from the sales in the test stores to the population of all stores is done using a dynamic programming approach that determines the weights of a linear forecast formula such that the trade-off between extra costs of the test sale and benefits from increased accuracy is optimized.

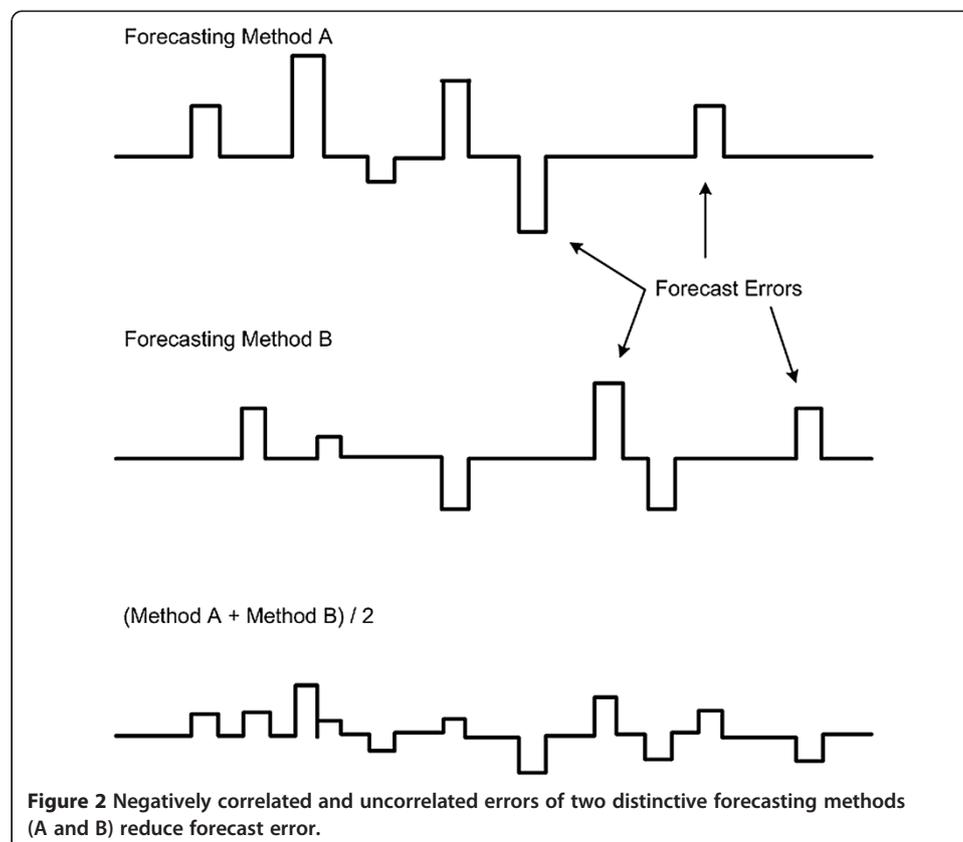
Combined forecasting

The idea of combined forecasting is to apply several different forecasting methods (or using several different data sources with the same forecasting method) on the same problem. Improvement in accuracy is achieved when the component forecasts contain

useful and independent information (Armstrong 2001). Especially when forecast errors are negatively correlated or uncorrelated the error might be canceled out or reduced and thus improve accuracy (see also Figure 2 for illustration).

The more distinct the methods or data sources used for the component forecasts are (the more they are independent from another) the higher is the expected improvement on forecasting accuracy compared to the best individual forecasts (Armstrong 2001).

It is a widely accepted and practiced method that very often leads to better results than a single forecasting method that is based on a single model (or data source) (Armstrong 2001). However, a prerequisite is that each component forecast is by itself a reasonably accurate forecast. Armstrong (2001) also states that combining forecasts can reduce errors caused by faulty assumptions, bias and mistakes in data. Combining judgmental and statistical methods often leads to better results. Armstrong (2001) quotes several studies that found that equal weighting of methods should be used unless precise information on forecasting accuracy of the single methods is available. Accuracy is also increased when additional methods are used for combined forecasting. Armstrong (2001) suggests using at least five different methods or data sources, provided this is comparatively inexpensive to achieve optimal results with combined forecasting. When more than five methods are combined accuracy is improved, but usually at a diminishing rate that becomes less and less notable. Armstrong (2001) states that combined forecasts are especially useful in situations of high uncertainty.



Methods

Data collection

The data used for our analysis originated from point of sale scanners at each outlet. The scanner data is loaded each night into a central data warehouse and archived for later analysis. Sales data is stored at the quantity per product per outlet per day granularity. For the purpose of this research we computed the cumulated sales sum until day 7 in order to reduce variance and uncertainty. We also limited the forecast horizon to the first seven days of the sales period in order to approximate a good measure for real demand. If we would extend the forecast horizon further the proportion of stock-outs would become too high and obscure real demand. During the first week stock-outs occur in fewer than 5% of the cases so we can assume that sales volumes for the first seven days are a sufficiently accurate approximation for real demand.

In a following step we cleaned the data for customer returns (negative sales numbers), oversized products that were delivered by an alternative logistic supplier (higher chance of stock-out than normal), products that were planned to be sold just in a subset of outlets and for products that were not tested in the retail test. The data set entails all remaining sales cases of the year 2009. For the development of forecasting models we limited the data set to weeks 14–51 because the case company used a different demand forecasting method and other replenishment cycles before week 14. We also excluded data from week 19 and 28 because here unsold products from earlier sales periods were sold without conducting another pilot sale beforehand. The remaining weeks were randomly split into two data sets. One was used for developing new forecasting methods and the other one was used for testing.

Currently used forecasting method

The currently used forecasting method at the case company (see Figure 3) is based on a calculation schema that consists of three components that are calculated separately. The first component is a measure for the overall sales potential of a product derived from the sales data of the retail test. It forecasts the total expected sales volume by

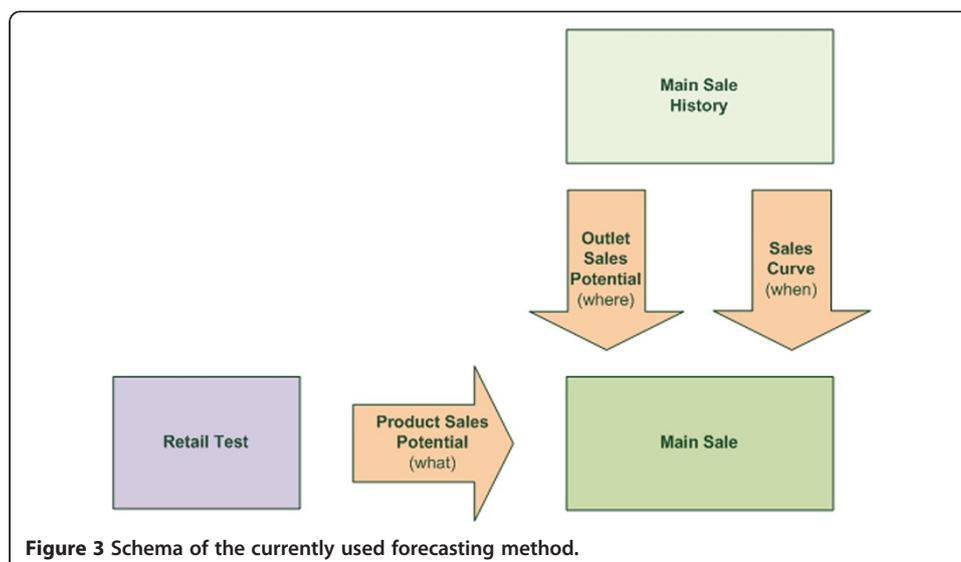


Figure 3 Schema of the currently used forecasting method.

extrapolating from the sample outlets to the whole population of outlets. The second component is a measure for the general (product independent) sales potential of each individual outlet which is derived from historical sales data. It determines how the forecasted total sales volume for a product is distributed among outlets. The third component is a measure for the sales curve over time which is calculated from historical sales data as the average sales curve for all outlets and all products using the sales data from several weeks. It determines how the forecasted total sales volume for a product in an outlet is distributed over time.

The measure for the overall sales potential of a product is influenced by experts that interpret the results of the retail test and adjust the product sales potential measure to special circumstances (like marketing campaigns for certain products or changed weather conditions). They also estimate price elasticity from the three different pricings of the retail test and adapt expected demand volumes to the sales price, which is set by a separate committee. In general the forecasting method makes strong use of aggregation in order to cope with high uncertainty and volatility in demand patterns. Sales are aggregated over all products regardless of product groups and common product features. It is also aggregated over time (average over several weeks) in order to reduce volatility.

A reduction of the aggregation level can lead to potentially more accurate forecasts since more complex forecasting methods (e.g. data mining techniques) can be applied. The question however is, if reducing the aggregation level is possible with the given level of volatility in the data. If volatility is too high the underlying effect which we want to measure is superimposed by noise and forecasting accuracy will decrease.

As it turns out reducing the aggregation level on the product dimension (calculating the sales potential for each product group separately instead of calculating the sales potential for all products combined) leads to a reduced forecasting accuracy in terms of increased misallocation with the current forecasting method (see Figure 4).

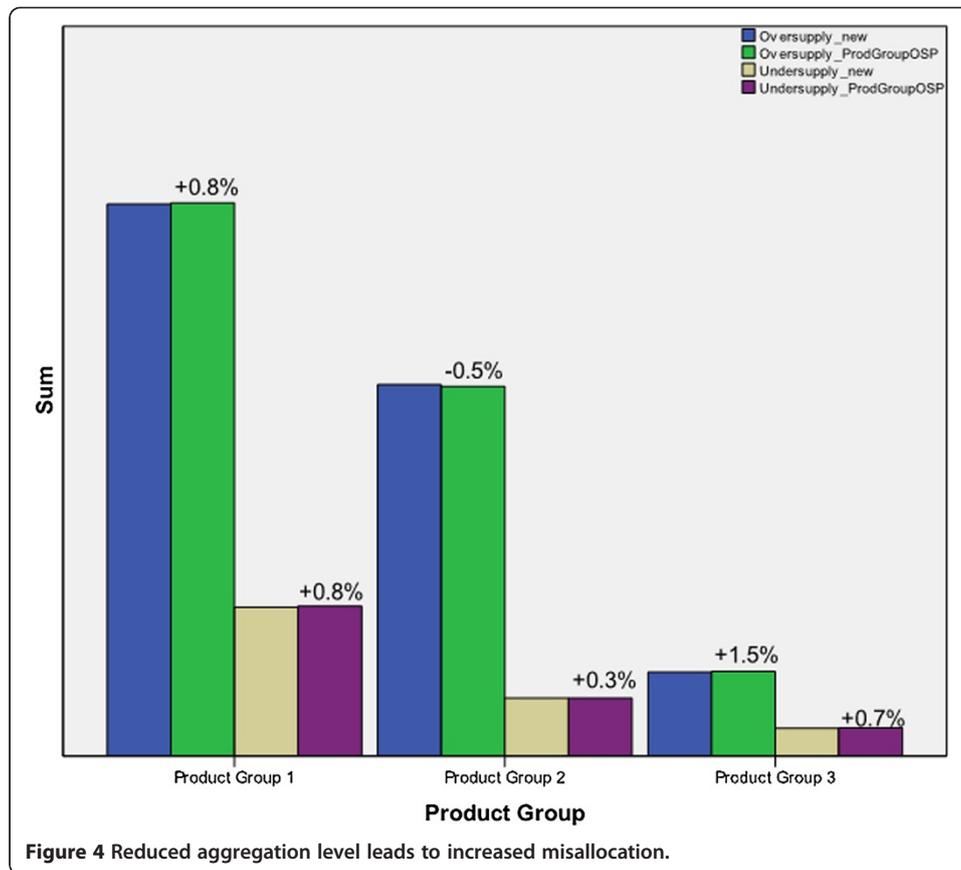
Reducing the aggregation level on the time dimension would reveal seasonal fluctuations in an outlet's sales proportion over the year but such an effect does not exist (at least no seasonality that is stronger than the general noise level) and would thus not lead to increased accuracy. The seasonal fluctuations of the total sales quantity is already captured in the sales forecast, since the retail test is conducted only several weeks before the selling period.

Why data mining techniques are not applicable in this case

This decrease in forecasting accuracy when the level of aggregation is reduced is the reason that data mining techniques are not applicable for the discussed problem. The advantage of data mining techniques is that its algorithms can capture more complex demand patterns compared to other forecasting methods. In this case however, more complex patterns can only be revealed when the level of aggregation is reduced. As this leads to lower forecasting accuracy (due to superimposition by noise) data mining techniques cannot unveil their potential to increase forecasting accuracy in this case.

Improved method

A possible way to reduce noise and uncertainty is to use multiple forecasting methods and combine their results. One promising approach is to combine judgmental forecasting and statistical forecasting as proposed by Sanders & Ritzman (2001). This approach



also satisfies the condition proposed by Armstrong (2001) that only the combination of distinct methods leads to improved results.

The forecasting method used at the case company can be seen as a method that strongly involves judgmental adjustment of statistical forecasts. The result of the retail test is always interpreted by experts and adjusted for special circumstances such as supply problems, weather conditions, competitor moves or special promotions. However, the process is strongly biased because there is a strong motivation to overestimate forecasts when the purchased quantity is larger than the expected sales volume. Furthermore the process itself, as well as the adjustment of the product sales potential to price changes, is unstructured which can lead to decreased accuracy as described by Sanders & Ritzman (2001).

We propose to increase forecasting accuracy by combining the current forecasting process at the case company with a purely procedural version (without involving human judgment) of the current forecasting method. This eliminates bias but does not take domain knowledge, contextual and environmental information into account.

Since the change in demand caused by an altered selling price is estimated by human judgment in the current forecasting process we further create a pricing function that estimates the pricing effect in a purely procedural manner. The product sales potential is then directly derived from the weighted sales figures of the retail sale without adjusting demand for the different (random) price settings in the test stores. Instead a linear price function is equally applied to all products. The price function determines the

increase or decrease in demand in dependence of the relative selling price change compared to the planning price which was decided on by the separate committee. The coefficients of the linear price functions (formula 1) were estimated by regression on the test data set such that the amount of misallocation in terms of oversupply and under-supply was not higher than with the original forecasting method.

$$\text{ChangeInDemand} = x_0 + x_1 \cdot \text{RelativePriceChange} \quad (1)$$

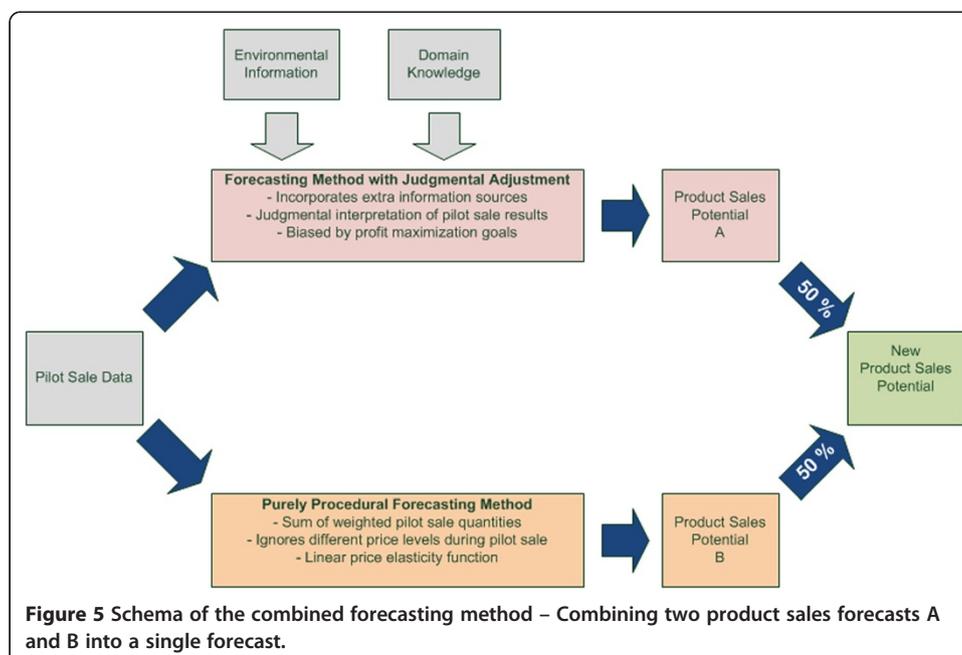
Two different price functions were estimated for each product this way: one price function for all cases in which the selling price was decreased compared to the planning price through the committee and one price function for all cases in which the selling price was increased compared to the planning price.

A schema of the combined forecasting method is shown in Figure 5. Both methods rely on the data of the retail test to estimate the product sales potential. But the retail test data is processed in two distinct ways. The judgmentally adjusted method uses extra information (domain knowledge, contextual and environmental information) but is biased. The purely procedural method is unbiased and uses a general linear price function. The results of each method are equally weighted with 50% as proposed by (Sanders & Ritzman 2001) and constitute the new product sales potential. A different weighting (75% judgmental, 25% mechanized) was also tried but lead to decreased forecasting accuracy. This finding is supported by Armstrong (2001) who states that the weighting of methods should only be different from an equal split if there is a plausible reason to do so.

Results

Evaluation method

In order to evaluate the used forecasting method and potential improvements a metric that measures the distance of the forecast to the real value (in this case real demand) is



defined. As stated above we can assume that realized sales quantities during the first seven days are sufficiently close to the real demand (sales that could have been realized with constant 0% stock-out rate). Thus the forecast error is the difference between the forecast sales quantity and the realized sales during the first week (see Figure 6).

In order to evaluate the quality of the forecasting method we will rely on the most practical and reasonable approach possible, that is to test forecasting methods “in situations that resemble the actual situation” (Armstrong 2001). In our case we measured the outcome of the forecasts in terms of oversupply and undersupply as it occurred in reality. We compared it with the amount of misallocation (in terms of oversupply and undersupply) that would have been generated if the company would have solely relied on the used forecasting method. In the comparisons we assume that each store only receives one delivery at the start of the selling period with the forecast quantity. In reality the case company is restocking the products several times a week in order to minimize the stock-out rate.

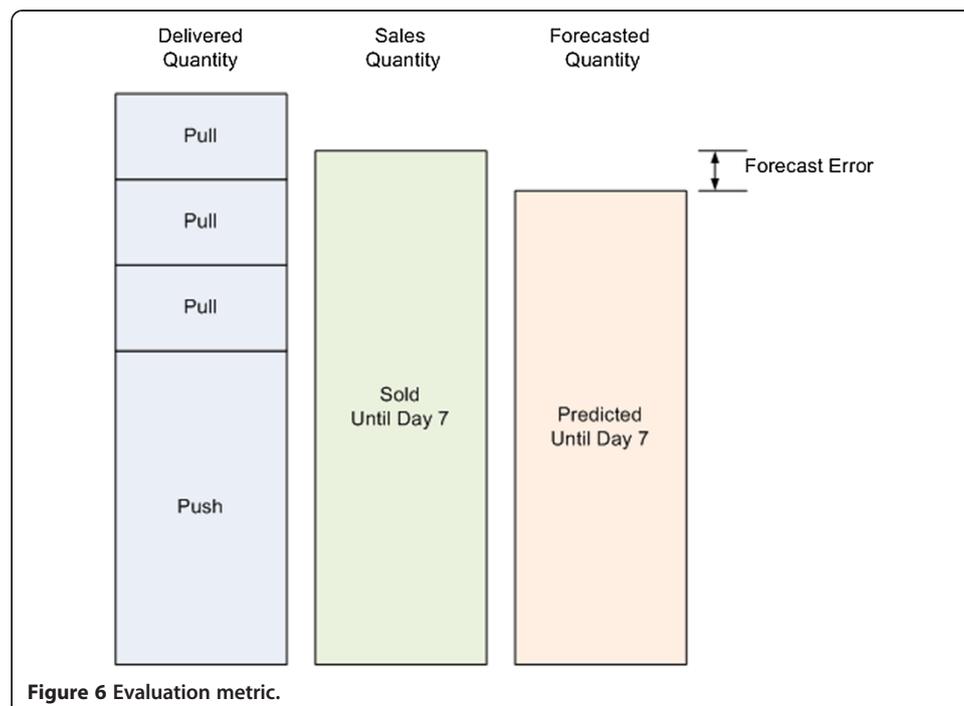
The evaluation is conducted on the test data set (randomly selected weeks 16,17,21,22,23,24,26,27,29,30,31,35, 38,41,43,44,45,47,51) while the forecasting method described in the previous section was developed on the training set to avoid overfitting.

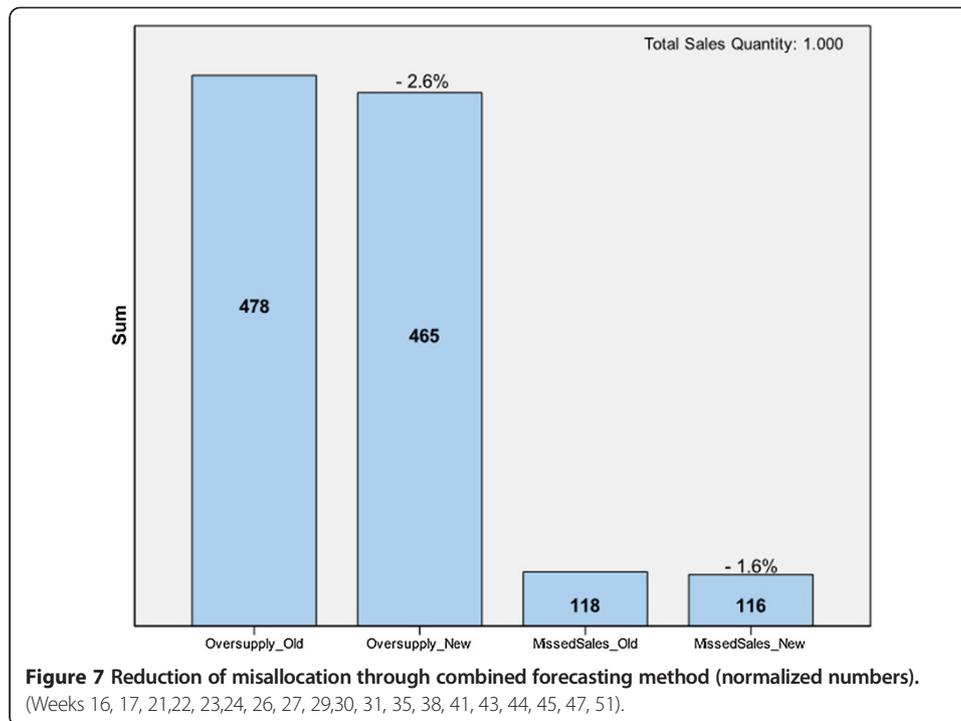
Results

Using the new combined forecasting method the amount of misallocation can be significantly reduced (as illustrated in Figure 7). Oversupply is reduced by 2.6% while undersupply is reduced by 1.6%. The reduction in misallocation has a reasonable cost saving impact through reduction of returning and restocking of unsold products.

Discussion and conclusion

For the problem type described in this research it is important to find ways to reduce noise in the data and to cope with volatility. We can derive three types of methods that





can be used to reduce noise and cope with volatility in the data: aggregation, using domain knowledge and combined forecasting. Aggregation can be applied over the three dimensions of the described problem: time, outlets and products. Aggregation is heavily used in the currently used forecasting method at the case company. Domain knowledge can be used in two ways: during model building and to adjust statistical forecasts. Using domain knowledge during model building means to use domain knowledge about the structure and causal relationships of the problem to prescribe the elementary building blocks of the model used for forecasting. There are in principle two ways to model the underlying concepts: first, to know the structure and interrelationships of the underlying concept through domain knowledge and theoretical knowledge and second, to leave the detection of underlying concepts of the problem to the learning algorithm in a data mining approach. The learning algorithm in turn can only detect those concepts that are not superimposed by noise. When the noise is large, fewer concepts can be detected by the learning algorithm. Thus, if concepts are known through domain knowledge they might be of more detail than any of the concepts a learning algorithm could possibly learn when noise level is large. Therefore the concepts known already should be implemented in the forecasting model manually. An example for such a concept known from domain knowledge is the concept of the price effect on demand. We know from other research that demand is almost always increased when the price is lowered. There are only very few special cases in which this relationship does not hold. With the domain knowledge about the products offered by the case company we can exclude these special cases and get to the conclusion that in the problem domain the demand is always increased or at least unchanged if the price is lowered and vice versa.

For the application of data mining algorithms it is essential that available domain knowledge is incorporated into data preparation. The domain knowledge about which

concepts might actually be there has to be transformed into an appropriate data preparation that makes the potential information accessible for the learning algorithm.

The third type of method is constituted by methods of combined forecasting. Combined forecasting means to apply several different forecasting methods on the same problem and use the average of the results as the forecast. Armstrong (2001) states that the results become usually better when the combined methods use distinct forecasting techniques or rely on distinct data sources.

One goal of this research was to examine if data mining techniques can be used to improve demand forecasting for products with high uncertainty and very short selling periods. We showed that in fact data mining algorithms can only be applied when noise and uncertainty in the data are comparatively low. Because the data at the case company comes with very high uncertainty and noise, aggregation has to be applied on the data to reduce the noise level so far that the data can be used for reliable forecasting. The problem here is that the extent of aggregation needs to be so high that the number of remaining relationships in the data is shrinking to a complexity level on which data mining algorithms need not be applied anymore. A single formula can be used to model the remaining relationships in the data. In order to apply data mining algorithms such that they can model more complex relationships the aggregation level has to be reduced to reveal additional relationships in the data. But we showed that a reduction of the aggregation level seems not possible because in this case noise is superimposing the information entailed in the data. Maybe a reduction of the aggregation level would be possible with another product group feature (such as style, novelty or usefulness), but it is questionable if such a feature can be found and it is also currently not captured in the data warehouse of the case company.

We showed in this research that combined forecasting is a useful approach to achieve better forecasting accuracy in situations of high uncertainty by developing an improved forecasting method that significantly increased forecasting accuracy. Next to combined forecasting judgmental adjustment of forecasts delivers a valuable source of information about the environment and the problem domain that is not entailed in the data. These findings encourage further research on how to integrate judgmental and contextual information with information from databases. Especially in the field of data mining there is almost no literature on a combined approach of data mining techniques with judgmental techniques which we believe will lead to much better results than relying on data mining techniques alone.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DM carried out the research and wrote the manuscript. MS supervised the research and reviewed the manuscript. PW co-supervised the research and gave recommendations for improvements. All authors read and approved the final manuscript.

Received: 2 September 2013 Accepted: 6 September 2013

Published: 19 February 2014

References

- Armstrong, JS. (2001). *Principles of forecasting: a handbook for researchers and practitioners*. New York, Boston, Dordrecht, London, Moscow: Kluwer Academic Publishers.
- Christopher, M, Lowson, R, & Peck, H. (2004). Creating agile supply chains in the fashion industry. *International Journal of Retail & Distribution Management*, 32(8), 367–376.

- Fayyad, U, Pietetsky-Shapiro, G, & Smyth, P. (1996). *Knowledge discovery and data mining: towards a unifying framework*. Menlo Park, CA: Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD-96). AAAI Press.
- Fisher, M, & Rajaram, K. (2000). Accurate retail testing of fashion merchandise: methodology and application. *Marketing Science*, 19(2), 266–278.
- Hand, DJ. (1998). Data mining: statistics and more? *The American Statistician*, 52(2), 112–118.
- Kuo & Xue. (1999). Fuzzy neural networks with application to sales forecasting. *Fuzzy Sets and Systems*, 108(2), 123–143.
- Nowack, A. (2005). Prognose bei unregelmäßigem bedarf. In P Mertens & S Rässler (Eds.), *Prognoserechnung* (pp. 61–72). Heidelberg: Physica-Verlag.
- Pyle, D. (1999). *Data preparation for data mining*. San Francisco, California: Morgan Kaufmann Publishers.
- Sanders, NR, & Ritzman, LP. (2001). Judgmental adjustment of statistical forecasts. In JS Armstrong (Ed.), *Principles of forecasting: a handbook for researchers and practitioners* (pp. 195–213). New York, Boston, Dordrecht, London, Moscow: Kluwer Academic Publishers.
- Simoudis, E. (1996). Reality check for data mining. *IEEE Expert Intelligent Systems and Their Application*, 11(5), 26–33.
- Thomassey, S, & Fiordaliso, A. (2006). A hybrid sales forecasting system based on clustering and decision trees. *Decision Support Systems*, 42, 408–421.
- Wedekind, H. (1968). Ein Vorhersagemodell für sporadische Nachfragemengen bei der Lagerhaltung. *Ablauf- und Planungsforschung*, 9, 1. et sqq.
- Weiss, AM, & Indurkha, N. (1998). *Predictive data mining – a practical guide*. San Francisco, California: Morgan Kaufmann Publishers.
- Witten, IH, & Frank, E. (2005). *Data mining – practical machine learning tools and techniques* (2nd ed.). San Francisco, California: Morgan Kaufmann Publishers.

doi:10.1186/2193-8636-1-4

Cite this article as: Maaß et al.: Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques. *Decision Analytics* 2014 1:4.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
