



Enhanced predictive models for purchasing in the fashion field by using kernel machine regression equipped with ordinal logistic regression



Ali Fallah Tehrani*, Diane Ahrens

Technology Campus Grafenau, Hauptstrasse 3, 94481 Grafenau, Germany

ARTICLE INFO

Article history:

Received 1 December 2015

Received in revised form

19 May 2016

Accepted 22 May 2016

Keywords:

Ordinal logistic regression

Fashion products

Sales forecasting

Kernel machines

ABSTRACT

Identifying the products which are highly sold in the fashion apparel industry is one of the challenging tasks, which leads to reduce the write off and increases the revenue. In fact, beyond of sales forecasting in general a crucial question remains whether a product may sell well or not. Assuming three classes as substantial, middle and inconsiderable, the forecasting problem comes down to a classification problem, where the task is to predict the class of a product. In this research, we present a probabilistic approach to identify the class of fashion products in terms of sale. Thereafter, we combine kernel machines with a probabilistic approach to empower the performance of kernel machines and eventually to make use of it to predicting the number of sales. The proposed approach is more robust to outliers (in the case of highly sold products) and in addition uses prior knowledge, hence it serves more reliable results. In order to verify the proposed approach, we conducted several experiments on a real data extracted from an apparel retailer in Germany.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Recently applying data analytics techniques, e.g. machine learning methods on the use of business forecasting for tremendous and complex data have received considerable attention. Specifically for the goal sales forecasting these techniques attempt to recognize established patterns in the existing sales history. To extract sound patterns from data the forecasting models should take into account all deciding factors. Unlike expert knowledge based system, machine learning approaches provide more reliable and objective results, since they handle the data in an objective manner. Needless to say that the experts in apparel industry use some established rules,¹ which often may not consider any change over time, however from a customer point of view the so-called customer's interest is being changed notably over time. Therefore, there exists a danger of a wrong conclusion particularly due to the fact that the experts consider typically the same patterns. In contrast to this, however, machine learning techniques aim to bias observations in a reliable and convenient manner.

Typically in the fashion apparel industries ordering is

performed once in the each season; reordering usually is very costly and takes time, too. Seen from this view, an accurate ordering causes more benefits and also reduces write off and over-production. To this end, identifying products which may sell well can reduce drastically the over production. Naturally the whole procedure is affected mostly by highly sold products, which henceforth we refer them to *trendy products*. A trend is assigned to a fashion product when it exceeds a threshold in terms of sales. We discuss more in the details the way to categorize the products on the basis of number of sales. Since the trendy products effect on the majority of sales, detecting them may lead to enhancing the prediction accuracy.

More concretely, our objective is to improve the quality of forecast for an apparel retailer in Bavaria (Germany). Before the start of a new clothing season, the retailer sends catalogues to its customers from which it customers can order fashion products. Simultaneously or even earlier the retailer orders fashion products from other upstream retailers, most of which are located in Asia. Therefore, an accurate sales-ordering is desirable and leads to decrease in unnecessary expenditures. Note that, the most advertised fashion products are distinctive from year to year. Seen from this view, the estimator is required to recognize the proper patterns from the sales history. This research aims at enhancing the quality of forecast in two ways: firstly it deals with forecasting underlying probabilistic approaches to predict (classify) the class of new fashion products. Secondly on the basis of the classes

* Corresponding author.

E-mail addresses: ali.fallahtehrani@th-deg.de (A. Fallah Tehrani), diane.ahrens@th-deg.de (D. Ahrens).

¹ For instance, there have existed several established patterns for color combinations, which are not longer valid.

attempts to predict the number of sales for future. As is clear, the first challenge belongs to classification methodology, whereas the second challenge is counted as a regression problem. To this end, we propose a novel method for regression on the basis of kernel machine regression equipped with the ordinal class classification setting. Moreover, we show that by employing our proposed predictive model there is a chance to overcome the obstacles (outliers and noises), which indeed outperforms common regression methods. In the experimental part we compare our proposal versus ordinary linear regression, RBF-kernel and polynomial kernel to demonstrate the advantages and disadvantages.

The rest of the paper is organized as follows: after a short survey on existing methodologies in the next section, in Sections 3 and 4 the linear regression and the ordinal logistic regression are respectively presented. Section 5 is dedicated to the algorithm and our proposal. Finally the results are demonstrated in Section 6.

2. Related work

However, there is no comprehensive research available on fashion forecasting underlying machine learning techniques, there are some notable exceptions: Thomassey and Happiette (2007) employ neural network method, specifically Neural Clustering and Classification (NCC) to tackle the problem. Kit Wong and Guo (2010) apply a hybrid intelligent model underlying neural networks to make a prediction in terms of medium-term fashion. In this regard, Alon et al. (2001) serve a comparison between neural networks methods and conventional methods for the goal prediction retail sales. To analyze sales dynamically, Frank et al. (2003) deal with the forecasting in a different manner. In fact, they proposed online-learning, which basically supposes daily lag and weekly lag. Then they define on each lag auto correlations functions (ACF). Since feedbacks can be monitored dynamically, it is possible to improve the accuracy of prediction by neural network method. Apart from conventional neural network, Kuo et al. (2002) equip the common neural networks by fuzzy weights, called fuzzy neural networks. Beyond neural network methods, Thomassey et al. (2002a, 2002b) use classification methods to cope with mean-term forecasting. Happiette et al. (1996) apply partition technique to cluster similar trend and then on the basis of similar clusters identify similar characteristic and ultimately the products in each cluster have similar behavior in terms of sales. Sun et al. (2008) and Xia et al. (2012) employ extreme learning machine techniques for the goal sales forecasting in fashion retailing. Moreover, to make a prediction Bayesian techniques have been applied (Yelland and Dong, 2014). Apart from forecasting on demand, there are researches focused on trend forecasting for the case of color (Gu and Liu, 2010; Choi et al., 2014; Linton, 1994; Diane and Cassidy, 2009). In particular, Stansfield and Allan Whitfield (2005) use information in past on the use of prediction for future. Worth mentioning that in Liu et al. (2013) and Choi et al. (2014) a survey on the existing approaches w.r.t. the sales forecasting especially related to *artificial intelligence* are conveyed.

It would be worthwhile to mention that sales forecasting is mainly counted as a regression problem and mathematically is tractable by applying regression techniques. In this regard, Dai et al. (2015) and Lu (2014) apply state-of-the-art support vector machine regression (SRM) for the goal sales forecasting. Following the idea of support vector machine regression, Anandhi and Manicka Chezian (2013) use it for forecasting the demand of Pulpwood. Simutis et al. (2008) apply SRM to predicting the daily cash demand for automatic teller machine (ATM). Also for other purposes such as short-term wind energy forecasting (Kramer and Gieseke, 2011), SRM has been applied successfully. Roughly speaking, the support vector machine regression conducts a linear

regression by controlling the flexibility of learner underlying minimizing the risk. The common loss function used in this case is ϵ -insensitive loss function (Smola and Schölkopf, 2004). This setting takes advantage of quadratic programming. Beyond modeling dependencies linearly, decision tree regression is able to capture non-linear dependencies. In Atanackov and Boylan (2011) it has been proposed to use of decision tree regression for forecasting trended demand. In addition (Bala, 2010) deals with demand forecasting in retail sale. Kumar (2013) employs decision tree for wetter forecasting. Ulvila (1985) conveys a survey on decision trees underlying forecasting. Although, a well-known drawback of decision trees might be that a small change in input data can cause a very large change in output, which from a prediction point of view is not desirable. On contrary, kernel machine (ridge) regression is more robust to changes and delivers therefore more stable results in this regard. Specifically, non-linear kernels are able to model non-linear dependencies, which from an application point of view is much useful. In this regard, Hong (2013) applies kernel machine regression to forecast energy demand intelligently. Similarly kernel regression has been employed to tackle real-time building energy analysis (Brown et al., 2012), as well as it has been applied for estimating the demand in smart grids (Mirowski et al., 2014; Kramer et al., 2010). In Wolberg (2000) kernel regression has been used to model financial markets. Heerde et al. (2001) apply the same technique to estimate the deal effect. On contrary to kernel regression, from a statistical point of view a well-established model is (latent) Gaussian mixture regression (Falk et al., 2006), which presents the regression output for an instance $\mathbf{x}$ by a conditional expectation. In Tian et al. (2011) it has been used for human pose estimation. Kelly (2014) uses this technique to recognizing of ageing speakers. Hueber et al. (2015) use the same technique for training of an acoustic-articulatory model given a reference speaker to the voice of another speaker, in the case when amount of audio-only data is limited. In this paper, we propose a novel method which empowers kernel machine regression by underlying ordinal logistic regression.

The terminology prediction is a widely used term in machine learning, however, in this paper it refers to make a class prediction as well as the demand forecast of fashion products in the market.

3. Linear logistic regression (LLR)

Before introducing our proposal at this point we shall recall some preliminaries with respect to the logistic regression as well as the ordinal logistic regression.

In a simple linear regression given the data $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n \subset \mathbb{R}^{m+1}$, the goal is to find parameters $\mathbf{w}^* = (w_1^*, \dots, w_m^*) \in \mathbb{R}^m$, $\hat{w}^* \in \mathbb{R}$, such that,

$$(\mathbf{w}_1^*, \dots, \mathbf{w}_m^*, \hat{w}^*) \leftarrow \underset{\mathbf{w}_1, \dots, \mathbf{w}_m, \hat{w}}{\operatorname{argmin}} \sum_{i=1}^n \left\| \sum_{k=1}^m w_k x_i^k + \hat{w} - y_i \right\|^2.$$

Each x_i^k is called a *regressor* or *predictor* and y_i is called a *response*. Traditionally the linear regression is used for modeling the statistical expectation, namely the statistical mean can be expressed as follows:

$$\mu = w_1 x_1 + \dots + w_m x_m + \hat{w}.$$

The obvious advantage of this expression is to specify the effect of each regressor on the mean. But since the output of the linear regression ranges in the real numbers, imaging a probabilistic interpretation of the model is difficult, or even is impossible. The logistic regression, however, provides an embedding option to have a probabilistic interpretation, and even more, it provides a probabilistic tool for binary classification. Basically, for a *dichotomous* (binary response) case the logistic regression models the logarithm of odds ratio (the probability

Download English Version:

<https://daneshyari.com/en/article/7433778>

Download Persian Version:

<https://daneshyari.com/article/7433778>

[Daneshyari.com](https://daneshyari.com)