Forecasting Store Foot Traffic using Facial Recognition, Time Series and Support Vector Machines

Paulo Cortez[1], Luís Miguel Matos[1], Pedro José Pereira[1], Nuno Santos[2], and Duarte Duque[3]

[1] ALGORITMI Centre, Department of Information Systems, University of Minho, 4804-533 Guimarães, Portugal
`pcortez@dsi.uminho.pt`, WWW home page: `http://www3.dsi.uminho.pt/pcortez`
[2] EXVA Technologies, Rua do Comércio, 6, 4710-820 Braga, Portugal
`nuno@exva.pt`
[3] ALGORITMI Centre, DIGARC - Polytechnic Institute of Cavado and Ave, 4750-810 Barcelos, Portugal
`dduque@ipca.pt`

Abstract. In this paper, we explore data collected in a pilot project that used a digital camera and facial recognition to detect foot traffic to a sports store. Using a time series approach, we model daily incoming store traffic under three classes (all faces, female, male) and compare six forecasting approaches, including Holt-Winters (HW), a Support Vector Machine (SVM) and a HW-SVM hybrid that includes other data features (e.g., weather conditions). Several experiments were held, under a robust rolling windows scheme that considers up to one week ahead predictions and two metrics (predictive error and estimated store benefit). Overall, competitive results were achieved by the SVM (all faces), HW (female) and HW-SVM (male) methods, which can potentially lead to valuable gains (e.g., enhancing store marketing or human resource management).

Keywords: data mining, facial recognition, time series forecasting, support vector machine

1 Introduction

In this age of big data, the passive collection of human activities (without their explicit intervention), is becoming commonplace. In effect, there is a large number of Information and Communications Technologies (ICT) that facilitate such passive collection, including Radio-frequency identification (RFID) technology, digital cameras and other sensors. These technologies open room for what is known as modern retail management [15]. By data mining [19] the behavior of people, it is possible to detect interesting patterns and build data-driven models that allow the prediction of human activities. This data-driven knowledge can be used to enhance the management of shopping centers or individual stores.

This work focuses on a particular modern retail application, where a digital camera is used to detect foot traffic to a store and then time series and data mining methods are used to build store traffic forecasting models. Such models can have a potential benefit in diverse store management areas, including security, marketing campaigns and promotions to attack customers, store window and layout design, management of stocks and human resources.

There are several works that adopted digital cameras for tracking and predicting human trajectories. For instance, an intelligent system was proposed in [7] and that is able to detect in advance if a person is going to enter a region of interest (restricted area) by analyzing human movements in a stationary video camera that captures a wide outdoor scene. This system was devised for surveillance but it could also be used to track a shopping center or store incoming traffic by defining the physical entrance as the region of interest. A similar approach was adopted in [11], where digital cameras were combined with state-of-the-art semantic scene methods in order to forecasting of plausible human destination paths in outside car parks near buildings. More recently, [1] predicted crowd mobility in a transport hub that included a shop. The prediction was based on social affinity maps created from a large number of monitoring cameras placed through the hub main corridors.

Time series forecasting predicts the behavior of a phenomenon based on the past patterns of the same event and has been applied in distinct domains, such as grocery store sales prediction [12] and Internet traffic [5]. In the 60s and 70s, several statistical methods were proposed for time series forecasting, including the popular Holt-Winters (HW) or AutoRegressive Integrated Moving Average (ARIMA) methodology [13]. More recently, Soft Computing methods, such as Support Vector Machines (SVMs) have also been proposed for time series prediction, often achieving competitive results when compared with statistical methods [16].

Recently, in 2013, we have executed a pilot project during a eight month period. In such pilot, a digital camera was set to capture the frontal view of the entrance of a Portuguese sports store. The camera was linked with a human facial recognition system that allowed the automatic detection of human faces. Each time a face is detected, the system also estimates its gender (female or male) and generates an event. In this paper, we explore these event data, in terms of daily counts. Thus, rather than modeling human trajectories, we predict store foot traffic in terms of three time series (all faces, female and male). Six forecasting methods are compared: a simple baseline, HW, ARIMA, a recent SVM approach [16] and two newly proposed hybrids HW-SVM and SVM-SVM that combine time series data with other features (e.g., weather and special events). The comparison assumes a robust and realistic rolling window evaluation scheme that considers from 1 to $h = 7$ daily ahead predictions and two metrics (predictive error and estimated store benefit).

The paper is organized as follows. Section 2 presents the collected data, forecasting methods and evaluation procedure. Next, Section 3 describes the ex-

periments conducted and analyzes the obtained results. Finally, Section 4 draws the main conclusions and also mentions future work directions.

2 Materials and Methods

2.1 Collected Data

The data was collected from a pilot project conducted in a sports store, placed inside a shopping center in a Portuguese city. From April to December of 2013 (total of eight months), a digital camera captured the frontal view of the store unique entrance. Due to privacy issues, the camera did not record the video data. Yet, the video was fed to a facial recognition system, partially based in the technology proposed in [7] and that is property of the company EXVA Technologies (`http://www.exva.pt/?lang=en`). This system uses the Viola-Jones framework [18] to detect faces and a machine learning classifier to distinguish gender. The system has a high accuracy in facial and gender recognition and is is capable of detecting multiple faces within a single video frame. However, due to commercial issues, the full recognition system details cannot be disclosed. Since the goal of this paper is to measure the predictive accuracy of store foot traffic time series models, and not to validate the computer vision system, we assume that the collected facial event data is correct.

Each time a face is detected, the recognition system stores a face event that includes: a gender type {female, male, unknown}; and uptime, the total time (in seconds) that the face was detected. The unknown class is assigned when gender is more difficult to detect. This situation occurs very rarely. In effect, the percentages of collected gender types are: 46% for female, 49% for male and 5% for unknown. In terms of data preprocessing, we discarded 235 events related with uptime values higher than 300 seconds and that might be related with anomalies (e.g., face in a marketing ad board). The result was a dataset with 888,376 event records. Then, we built the daily time series by aggregated the female, male and all faces events (including unknowns). After this aggregation, we detected a total of 16 days (12 consecutive days in September) with missing data and that were related with pilot system failures. The missing data only occurred during the training data period and were replaced by the previously known values for the same day of the week (e.g., a missing Monday value is replaced by the value from the previous Monday). The final three time series (all faces, female and male) include a total of 257 daily entries. For demonstration purposes, the all faces series is plotted in the right of Figure 1.

To test the value of other data features, we also collected six additional daily variables. These include exterior weather conditions (maximum wind speed, average temperature, humidity and rain) in the city (from `https://www.wunderground.com/`). Also, it included special daily events (freeday – if weekend or holiday; special – if there is a major sports or entertainment event in the city), retrieved using Internet queries.
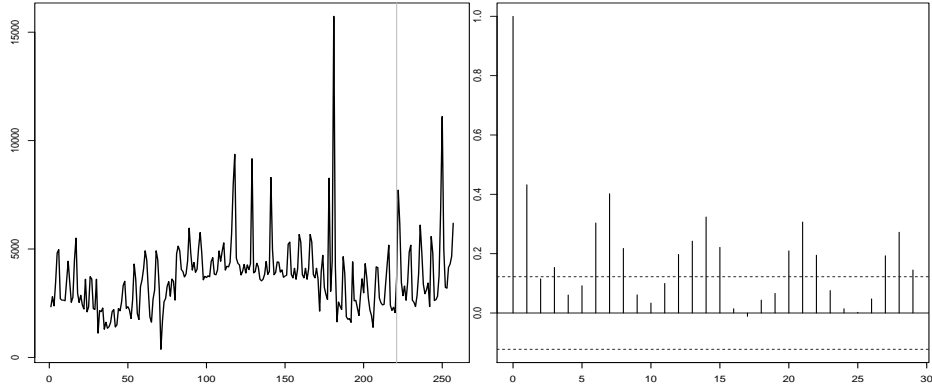
Fig. 1. Daily all faces time series (right, x-axis denotes the time period, y-axis the number of daily faces, vertical gray line splits the training and test data) and its autocorrelation values (left, x-axis denotes the time lags and y-axis the autocorrelations)

2.2 Forecasting Methods

A time series includes time ordered observations $(y_1, y_2, \ldots, y_t)$, where t is the time period. A time series model predicts a value for current time t based on past observations: $\hat{y}_t = f(y_{t-k_I}, ..., y_{t-k_1})$, where f is the forecasting function and the k_i values denotes the sliding window with I time lags. The k_i values can include any elements from the set $\{1, ..., t-1\}$. The horizon h is defined by the ahead time in which a prediction is executed. Multi-step ahead forecasts can be built by iteratively using 1-ahead predictions as inputs [5].

For example, let us consider the series $5_1, 10_2, 14_3, 17_4$ (y_t values). If the $\{1, 3\}$ window is adopted, then the prediction for time $t = 4$ is built using $\hat{y}_4 = f(5_1, 14_3)$. Multi-step ahead forecasts can be built by iteratively using 1-ahead predictions as inputs. If $h = 2$ and $t = 4$, then $\hat{y}_5 = f(10_2, \hat{y}_4)$.

In this work, we compare six time series approaches. Daily series often present a weekly seasonal period ($K = 7$), which is the case of the analyzed series. For example, the left of Figure 1 shows higher autocorrelation values for the multiples of 7, which confirms a weekly cycle for the series. Thus, the baseline method is set to predict the future values (from $h = 1$ up to h=7) using the known values from the previous week. It is assumed that this is the easiest method that could be used by the store manager. The HW is from the family of exponential smoothing methods and it is very popular for predicting trended and seasonal time series. We adopt the HW additive model with $K = 7$. For ARIMA, we adopt its multiplicative seasonal version, also known as SARIMA ($K = 7$).

Regarding the SVM model, we use the popular regression variant based on the Gaussian kernel and ϵ-insensitive loss function. The performance of SVM is affected by the choice of sliding window time lags (k_i, the model inputs) and the model hyperparameters (λ – the kernel parameter; C – a trade-off between fitting the errors and the flatness of the mapping; and ϵ – the width of the insen-

sitive tube). In particular, we adopt a fast two stage input and hyperparameter selection methodology, similar to what was proposed in [3, 16] and that has outperformed multilayer perceptrons in the forecasting of seasonal series [3]. Before fitting any SVM model, the data are first standardized to a zero mean and one standard deviation [9]. The first stage (input selection) is applied only once, using the first training dataset (of size W) with an internal ordered holdout split (with 70% for fitting the model and 30% for validation). Then, a fast backward selection feature selection procedure is used to discard irrelevant time lags. The feature selection starts with all k_i time lags ($i \in \{1, ..., I\}$) and computes its associated validation error. At this stage, the SVM hyperparameters are set to their default values ($\gamma = 2^{-4}, C = 1, \epsilon = 0.1$). Then, the sensitivity analysis proposed in [4] (e.g., DSA method, AAD measure of importance, 7 sensitivity levels) is used to measure input relevance and discard the least relevant time lag. This sensitivity analysis is capable of measuring the true global importance of an input, even when it includes interactions effects with other inputs. At the end of the first stage, the set of time lags with the best validation error is fixed and used during the next stage. The second stage (hyperparameter selection) is executed in each iteration of the rolling windows evaluation scheme. Using the training data of each iteration, a similar ordered holdout split is used to execute a grid search of the best SVM hyperparameters (λ from 2^{-8} to 2^0, C from 2^{-1} to 2^6, ϵ from 2^{-8} to 2^{-1}), under a total of 13 searches as provided by a uniform design strategy [2].

Two additional hybrid models are proposed, with the goal of combining time lagged variables with other types of features (e.g., daily weather conditions). Since HW provided better predictions (when compared with ARIMA, see Section 3) we adopted it for the first hybrid, while the second hybrid is based on two SVM models. The HW-SVM model uses first HW to get the $\hat{y}_t$ predictions for the training data. The $\hat{y}_t$ input variable is merged with the six non lagged features (wind speed, temperature, humidity, rain, freeday, special) in order to create a dataset. Then, a SVM is trained (using the same hyperparameter selection procedure) to predict y_t based on the seven input features. When performing up to h ahead predictions, the HW-SVM model assumes the up to h ahead HW forecasts and previous knowledge for the other features (e.g., weather forecasts, special event schedules). The SVM-SVM hybrid uses the same input selection procedure of SVM to get the best set of time lags (k_i). Then, these time lags are merged with the six non lagged features and the combined dataset is used to train another SVM (set with the same hyperparameter selection). When executing predictions for test data, iterative 1-ahead predictions are used as inputs for the lagged inputs, as in SVM.

2.3   Evaluation

To measure the quality of the predictions, we applied a fixed-size rolling windows scheme [17] that allows a realistic training and testing of a large number of forecasting models. The rolling window assumes a training set of size W, from which from $h = 1$ to $h = H$ ahead predictions are executed at time $t - 1$ (first

iteration, $i = 1$). Then training window is slided by discarding its oldest element and adding the value of t in order to retrain the model and predict from $h = 1$ to $h = H$ ahead predictions at time t (second iteration, $i = 2$), and so on. Thus, for a series of length L the rolling windows will have $U = L - (W + H - 1)$ model updates/iterations. For each forecasting method, we execute the rolling windows and store the $\hat{y}_{i,h}$ forecasts, where i denotes the rolling iteration.

To measure the predictive error, we selected the Normalized Mean Absolute Error (NMAE) metric [8]:

$$
\begin{aligned}
MAE_h &= \frac{1}{U} \sum_{i=1}^{U} |y_{T+i+h-1} - \hat{y}_{i,h}| \\
NMAE_h &= \frac{MAE_h}{y_{\text{High}} - y_{\text{Low}}}
\end{aligned}
\tag{1}
$$

where T is the last known time period for $i = 1$, y_{High} and y_{Low} are the highest and lowest target values in the test set (of size U). The NMAE metric has the advantage of easy interpretation, since it expresses the error as a percentage of the full target scale. Also, it is a scale independent metric, which is a relevant issue since the all faces, female and male series have distinct scale ranges. Statistical significance of the forecasting differences will be evaluated by the Diebold-Mariano (DM) test [6], under pairwise comparison between the best method and the baseline for each h value. When comparing distinct forecasting methods, we use the Average $NMAE_h$ (ANMAE) values (h from 1 to H).

Also, we propose a new metric, the Estimated Store Benefit (ESB). The metric assumes that the store manager executes at time $t - 1$ from $h = 1$ to $h = H$ ahead store traffic predictions, allowing her/him to set a plan with a better management of the store in the next seasonal cycle (with H days). For each individual daily absolute error, there is an average cost of c EUR, related with a missed opportunity (e.g., better marketing or human resource management). For a particular forecasting method M and plan i (related with iteration i of the rolling window), we define:

$$
\begin{aligned}
LOSS_i(M) &= c \times \sum_{h=1}^{h=H} |y_{T+i+h-1} - round(\hat{y}_{i,h})| \\
ESB_i(M) &= LOSS_i(\text{baseline}) - LOSS_i(M)
\end{aligned}
\tag{2}
$$

where $round$ is the rounding function, since we assume that face estimates are treated as integer numbers by the store manager (e.g., if $y_t - \hat{y}_t < 1$ then the loss is 0). Similarly to the NMAE metric, when comparing several time series methods, we use the Average $ESB_i(M)$ (AESB) values (i from 1 to U).

3   Results

All experiments were conducted using the R tool [14]. The statistical time series methods were executed using the **forecast** package [10], while the SVM based methods were implemented using the **rminer** package [2].

The rolling windows was set with $W = 220$ and the maximum number of ahead forecasts was set to $H = 7$ (one week). For each time series method, this allows the execution of $U = 31$ model trainings and predictions. For HW

and ARIMA, the forecasting model is rebuilt using the W past observations. Regarding the ARIMA, we adopted the `auto.arima` function of the **forecast** package, which executes the automatic SARIMA model identification for each i-th iteration.

For the SVM time lagged models, we set the maximum number of time lags to $I = 28$ (four times the seasonal period). As explained in Section 2.2, the time lag model selection is only executed for the first rolling window iteration ($i = 1$), in order to reduce the computational effort. Once the best set of time lags is found, the hyperparameter selection is executed for each i-th iteration (similarly to the statistical methods). Under this setup, the SVM computation effort is relatively fast and affordable by current computers. For instance, the execution time for the all faces time series and full rolling window procedure was just around 16 seconds in a 2.5 GHz Intel Core i7 machine.

The obtained prediction errors are shown in Table 1, in terms of the ANMAE values. All methods present predictive errors lower than 13%, which confirms that the three time series are predictable. In particular, the best forecasting methods obtain ANMAE values lower than 10%: SVM for all faces (8.32% error), HW for female (9.32%) and HW-SVM for male (9.56%). The predictive performance for these best models is detailed in Table 2, in terms of the NMAE values for distinct ahead predictions (h values). In this table, the selected time lags (k_i values) are shown for the SVM based methods. In both cases (all faces or male), feature selection procedure has kept time lags related with the seasonal period or its multiples (e.g., $k_i = 7$, $k_i = 14$). Table 2 also shows the p-values ($<5\%$ or $<10\%$) when comparing the best method with the baseline using the DM test. For several horizon values (e.g., $h=4$, $h=5$), the differences are significant.

Table 1. Comparison of the forecasting errors ($ANMAE$ values, in %; best values in **bold**)

Series	Baseline	HW	ARIMA	SVM	HW-SVM	SVM-SVM
all faces	11.74	9.22	11.42	**8.32**	10.59	9.81
female	11.37	**9.32**	11.25	9.36	10.86	11.53
male	11.25	9.69	11.44	11.15	**9.56**	12.82

When estimating the impact of using these forecast models to improve store management, we opt two reasonable (and maybe conservative) scenarios, where the loss of one wrong visitor prediction is $c=0.50$ and $c=1.00$ EUR. Such impact is presented in Table 3, in terms of the AESB values. As expected, the best forecasting methods from Table 1 provide the best gains. When compared with the baseline method, the average weekly gain is: all faces (SVM), 1050 ($c=0.50$) or 2099 ($c=1.00$) EUR; female (HW), 311 ($c=0.50$) or 622 ($c=1.00$) EUR; and male (HW-SVM), 261 ($c=0.50$) or 522 ($c=1.00$) EUR.

Table 2. Forecasting errors for the best methods and distinct horizons ($NMAE$ values, in %; * – p-value < 5%, ° – p-value < 10%)

Series	Method	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$
all faces	SVM ($k_i \in\{1,4,6,7,9,14,17,18\}$)	9.31°	8.85*	7.62*	7.50*	7.92*	8.54°	8.50°
female	HW	9.71	10.32	9.29	8.80°	8.93*	8.88*	9.33*
male	HW-SVM ($k_i \in\{1,14,16\}$)	11.13	10.91	9.61	8.46°	8.06°	9.18	9.58

Table 3. Comparison of the estimated store benefit gain ($AESB$ values, in EUR; best values in **bold**)

Series	c	Baseline	HW	ARIMA	SVM	HW-SVM	SVM-SVM
all faces	0.50	0	774	97	**1050**	353	593
all faces	1.00	0	1547	193	**2099**	707	1186
female	0.50	0	**311**	17	305	78	-25
female	1.00	0	**622**	34	609	155	-50
male	0.50	0	241	-28	15	**261**	-241
male	1.00	0	482	-56	31	**522**	-483

For demonstration purposes, Figure 2 shows the last rolling window iteration weekly ahead forecasts (i=31, h from 1 to 7) for the all faces (left) and male (right) series. In both cases, the plots show an improved performance of the best forecasting methods when compared with the baseline.
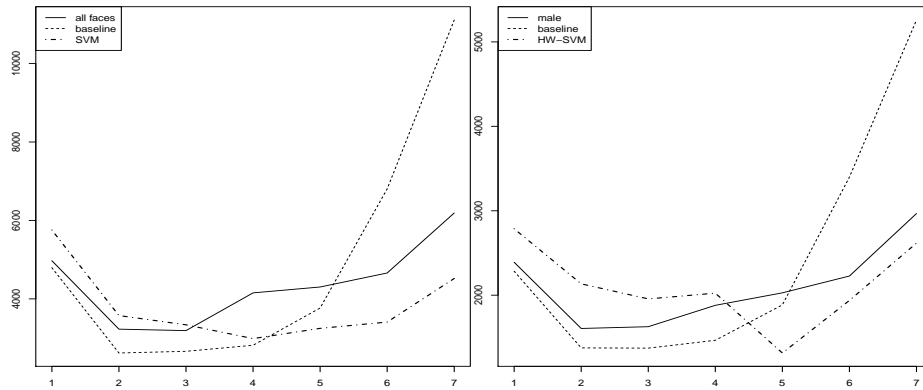

Fig. 2. Last rolling window weekly forecasts (i=31) for all faces (right) and male (left) series (x-axis denotes the horizon, y-axis the number of daily faces)

4 Conclusions

In this paper, we explore six time series methods to predict foot traffic to a sports store, placed inside a shopping center of a Portuguese city. The data was collected from a pilot project, where a digital camera captured the store entrance during an eight month period, from April to December of 2013. The video data was fed into a facial recognition system that had a high accuracy when detecting faces and their genders, generating events. We preprocessed these events in order to get the daily incoming store traffic counts of all faces, female and male.

The obtained time series were modeled using: a simple baseline (use of last week values); statistical methods, namely Holt-Winters (HW) and AutoRegressive Integrated Moving Average (ARIMA); a Support Vector Machine (SVM) model; and two hybrid methods (HW-SVM and SVM-SVM), that combined time lagged variables with other data features (e.g., weather conditions, major sport events). The forecasting models were compared using a robust and realistic rolling windows evaluation that assumes up to $h = 7$ ahead predictions and two metrics: predictive error and estimated store benefit.

The best results were obtained by SVM for the all faces series (8.32% error), HW for the female data (9.32% error) and HW-SVM for male foot traffic (9.56% error). We believe that these models can support a better store management (e.g., in terms of marketing or human resources). In effect, we have estimated a valuable impact with weekly gains of 1050 (c=0.50) or 2099 (c=1.00) EUR (SVM), 311 or 622 EUR (HW) and 261 or 552 EUR (HW-SVM).

This paper is the first forecasting study that used the store pilot data. While competitive and valuable results were achieved, the forecasting models were built offline, after the store pilot execution. As such, in the future, we intend to apply the proposed forecasting models such that they could be embedded into a friendly decision support system for real usage in a store environment. This would allow us to get a valuable feedback from store managers. Also, we plan to adapt the forecasting methods to other time scales (e.g., hourly). Finally, further analysis will be devoted to measure the impact of the non lagged features (e.g., sport matches), in order to understand why HW-SVM gets better results for male store visits.

Acknowledgments

References

1. A. Alahi, V. Ramanathan, and L. Fei-Fei. Socially-aware large-scale crowd forecasting. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2211–2218. IEEE, 2014.

2. P. Cortez. Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool. In P. Perner, editor, *Advances in Data Mining – Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining*, pages 572–583, Berlin, Germany, July 2010. LNAI 6171, Springer.

3. P. Cortez. Sensitivity Analysis for Time Lag Selection to Forecast Seasonal Time Series using Neural Networks and Support Vector Machines. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2010)*, pages 3694–3701, Barcelona, Spain, July 2010. IEEE.

4. P. Cortez and M. J. Embrechts. Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, 225:1–17, Mar. 2013.

5. P. Cortez, M. Rio, M. Rocha, and P. Sousa. Internet Traffic Forecasting using Neural Networks. In *Proceedings of the 2006 International Joint Conference on Neural Networks (IJCNN 2006)*, pages 4942–4949, Vancouver, Canada, July 2006. IEEE.

6. F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 2012.

7. D. Duque, H. Santos, and P. Cortez. Prediction of abnormal behaviors for intelligent video surveillance systems. In *CIDM*, pages 362–367. IEEE, 2007.

8. K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151, 2001.

9. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, NY, USA, 2001.

10. R. Hyndman and Y. Khandakar. Automatic time series forecasting: the forecast package for r 7, 2008. *URL http://www. jstatsoft. org/v27/i03*, 2007.

11. K. Kitani, B. Ziebart, J. Bagnell, and M. Hebert. Activity forecasting. *Computer Vision–ECCV 2012*, pages 201–214, 2012.

12. S. Ma, R. Fildes, and T. Huang. Demand forecasting with high dimensional data: the case of sku retail sales forecasting with intra-and inter-category promotional information. *European Journal of Operational Research*, 249(1):245–257, 2016.

13. S. Makridakis, S. Weelwright, and R. Hyndman. *Forecasting: Methods and Applications*. John Wiley & Sons, New York, USA, third edition, 1998.

14. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

15. Z. Segetlija et al. New approaches to the modern retail management. *Interdisciplinary Management Research*, 5:177–184, 2009.

16. M. Stepnicka, P. Cortez, J. P. Donate, and L. Stepnicková. Forecasting seasonal time series with computational intelligence: On recent methods and the potential of their combinations. *Expert Systems with Applications*, 40(6):1981–1992, 2013.

17. L. Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Forecasting Journal*, 16(4):437–450, 2000.

18. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.

19. I. Witten, E. Frank, and M. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Franscico, USA, San Francisco, CA, 3rd edition, 2011.