



UNIVERSITÀ DI TRENTO

Department of Information Engineering and Computer Science

Bachelor's Degree in Informatica

FINAL DISSERTATION

TITLE
Subtitle (optional)

Supervisor

.....

Student

.....

Academic year 2019/2020

Contents

Summary	2
1 State of art	2
1.1 Time series forecasting	2
1.1.1 Exponential smoothing	3
1.1.2 Artificial Neural Networks	3
1.1.3 ARIMA models	3
1.1.4 TBATS models	3
1.2 Queueing theory	4
1.3 Literature review	4
2 Data analysis	5
2.1 Datasets description	5
2.2 Inflow rate, dwell time and service time distributions	5
3 Methodology	5
3.1 Model overview	5
3.2 Inflow rate forecast	5
3.2.1 Persistence model	5
3.2.2 Naïve model	5
3.2.3 Artificial Neural Network model	6
3.2.4 ARIMA model	6
3.3 Arrival rate forecast	6
3.4 Queue length forecast	6
3.4.1 Service rate approximation	6
3.4.2 Outflow rate approximation	6
3.5 Checkouts optimization	6
3.5.1 Opening/closing fluctuations control	6
4 Results	6
5 Conclusions	6
References	6

Summary

Sommario è un breve riassunto del lavoro svolto dove si descrive l'obiettivo, l'oggetto della tesi, le metodologie e le tecniche usate, i dati elaborati e la spiegazione delle conclusioni alle quali siete arrivati.

Il sommario dell'elaborato consiste al massimo di 3 pagine e deve contenere le seguenti informazioni:

- contesto e motivazioni
- breve riassunto del problema affrontato
- tecniche utilizzate e/o sviluppate
- risultati raggiunti, sottolineando il contributo personale del laureando/a

1 State of art

This chapter gives an overview of the available methods of time series forecasting and queue analysis. First, the available methods for time series forecasting are presented, then an overview of queueing theory techniques is given and finally different approaches from similar researches are described.

Terminology:

- *Inflow/outflow rate*: number of customers that enter/exit the supermarket in a given interval.
- *Arrival rate*: number of customers that arrive at the checkouts in a given interval.
- *Service rate*: maximum number of customers that can be served by each checkout in a given interval.
- *Dwell time*: time spent by a customer in the supermarket.

1.1 Time series forecasting

A *time series* is a collection of observations made sequentially through time. *Time series forecasting* is the use of a model to predict future values based on previous observations.

A time series can be decomposed into four components:

- *Trend*: Linear/nonlinear increasing or decreasing behavior of the series over time.
- *Seasonality*: Repeating patterns or cycles behavior over time.
- *Noise*: Variability of the observations not explainable by the model.

The following sections describes different techniques that can be used to obtain a forecast of the future values.

1.1.1 Exponential smoothing

In the *Simple exponential smoothing* technique, the forecasted values are based on a weighted average of the previous values, where the most recent observations are given more importance using larger weights [6].

A more complex implementation is the *Holt-Winters exponential smoothing* approach, which is able to decompose the time series into level, trend and seasonal component, giving a more precise forecast. However, this methods are unable to model more complex series with multiple seasonality patterns, as in a supermarket inflow rate.

Taylor's *double seasonal exponential smoothing* method [8] was developed to forecast time series with two seasonal cycles: a short one that repeats itself many times within a longer one.

1.1.2 Artificial Neural Networks

Artificial Neural Networks (ANN) allows complex nonlinear relationships between the target variable and its predictors. With time series data, lagged values of the time series can be used as inputs. With seasonal data, it is also useful to also add the last observed values from the same season as inputs [6]. In general, the input can be expressed as:

$$(y_{t-1}, y_{t-2}, \dots, y_{t-p}, y_{t-w}, y_{t-2w}, \dots, y_{t-Pw})$$

where p is the number of previous lagged values and P is the number of lagged values from previous seasonal cycles to consider. w is the size in terms of number of intervals of the seasonality. For forecasting two steps ahead, the result of one-step forecast can be used in the input.

RNN c
layer

1.1.3 ARIMA models

ARIMA (AutoRegressive Integrated Moving Average) models provide another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting, and provide complementary approaches to the problem. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.[6]

AutoRegressive means the model use the relationship between an observations and some number of lagged previous observations to generate a linear regression model. *Integrated* refers to the use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary. A stationary time series has mean, variance and autocorrelation constant over time. *Moving Average* means that the model uses the dependency between an observation and the residual error from a moving average model applied to lagged observations.

The parameters of the ARIMA model are defined as follows:

- p : The number of lagged observations included in the model, also called the lag order.
- d : The number of times that the raw observations are differenced, also called the degree of differencing.
- q : The size of the moving average window, also called the order of moving average.

A simple version of the ARIMA model can be used to forecast only non-seasonal time series, but it can be extended to model seasonal patterns (SARIMA, Seasonal ARIMA). However, only a single seasonal effect can be modelled with SARIMA. A solution is given by the SARIMAX model, that supports exogenous variables. The seasonal effects can be captured by Fourier terms and used as exogenous variables. This gives a better approximation and allows other exogenous variables to be considered (e.g. weather) to further improve the accuracy.

1.1.4 TBATS models

An alternative approach proposed by De Livera et al. [2] uses a combination of Fourier terms with an exponential smoothing model and a Box-Cox transformation, in a completely automated manner. The main advantage of a TBATS model over a SARIMAX model with Fourier terms is that the seasonality

is allowed to change slowly over time, while harmonic regression terms force the seasonal patterns to repeat periodically without changing. However, one drawback of TBATS models, is that they are slow to estimate, since they will consider various alternatives and fit different models.

1.2 Queueing theory

Queueing theory is the mathematical study of waiting lines, or queues. A queueing model is constructed so that queue lengths and waiting time can be predicted.

There are various types of queueing models, but the best representation of a supermarket queue is probably given by the $M/M/c$ model, that represent a system with an infinite queue capacity, where the inflow rate follow a Poisson distribution and the service times of the c servers follow an exponential distribution. The customers are served in FCFS order (First Come First Served).

The *utilization factor* $\rho = \frac{\lambda}{c\mu}$ describes the proportion of total service capacity being used in the system. If $\rho \geq 1$, i.e. the arrival rate exceeds the service capacity, then the queue will grow indefinitely, but if $\rho < 1$ the system has a stationary distribution and the average queue length can be calculated.

The main issue with this stationary approach in the context of supermarkets is that the arrival rate and service rate cannot be expressed with probability distributions with constant mean, since they are strongly time-dependent. Moreover, it could be possible that the utilization factor exceeds 1 for a short period of time, in which the queue length increases, and after that the system returns slowly to its equilibrium, when $\rho < 1$. Green et al. (2001) [4] shown that the commonly used SIPP approach (Stationary Independent Period by Period), where an $M/M/s$ model is created for each interval and solved independently, is inaccurate for parameter values corresponding to many real situations.

Rider [7] proposed an approximation of the average queue length for a $M/M/1$ model with time-varying parameters. He found that the average queue length $Q(t)$ satisfies the equation:

$$Q(t) = \lambda(t) - \mu(t)[1 - P_0(t)]$$

where $P_0(t)$ is the probability of the server being in idle. This equation can be interpreted to mean that the queue will increase with the arrival rate and decrease with the service rate multiplied by an "efficiency factor" $1 - P_0(t)$, i.e. the probability that the server is busy at time t . $P_0(t)$ is not a priori known, thus an approximation must be used. .

Newell [3] proposed...

1.3 Literature review

There are various studies that try to analyze and predict the queue length in different settings.

Berman et al. (2004) [1] developed a system that manages the switching of workers between the "front room", where the checkouts are located, and the "back room", that is the shop. This system process real-time data about the count of customers either in front and back rooms, with the goal of minimizing the customer's waiting time.

Aksu H. (2018) [5] main goal was to optimize the idle time of the staff operating at the checkout while maintaining the waiting queue length under a predetermined threshold. He proposed a model based on the inflow of customers at the entrance, the number of customers entering the checkout area, the current waiting queue length and number of available checkouts. He divided the supermarket area into a shop and a checkout area and used a video-based system to count the customers moving in each area. He then developed a prediction model to calculate a realistic forecast of customer dwell times. The inward flow into the checkout area was predicted using the inward flow into the shop, since the inward flow into the shop appear with a delay at the checkout area. This delay is the dwell time. To consider non-standard events, the model was adjusted with real-time measured data, to increase the prediction accuracy. Queueing theory is then used to predict the average waiting time at the checkouts. Checkouts can be opened or closed based on a minimum and maximum acceptable queue length values. To avoid closing and opening too frequently, the system is able to decide in how many cases the defined waiting queue length can be exceeded, for example when it is exceeded for a short

time.

2 Data analysis

This chapter describes the analysis conducted.

2.1 Datasets description

2.2 Inflow rate, dwell time and service time distributions

3 Methodology

This chapter describes the methods used to build the prediction model.

3.1 Model overview

The final model is composed by different parts that cooperates to obtain the final result.

First, the measured inflow rate for the recent past is combined with a forecast of the inflow rate for the immediate future. This values, combined with a dwell time prediction, are used to obtain a prediction of the arrival rate at the checkouts. The arrival rate is then used with queueing theory methods to calculate the expected queue length for a given interval.

3.2 Inflow rate forecast

The inflow rate can be expressed as a time series, hence time series forecasting methods were used. As described in Chapter 1.1, different forecasting methods were tested and the best was selected based on the performance obtained.

3.2.1 Persistence model

The first model tested was used to define a baseline in performance. This gives an idea of how well the other models could perform on the time series, and if a model's performance are worse than this baseline it should no be considered. The baseline was obtained by a *persistence model*, were the last measured value y_{t-1} is used as forecast:

$$\hat{y}_t = y_{t-1}$$

3.2.2 Naïve model

As seen in Chapter 2, the arrival rate time series presents a strong seasonal component, so the forecast should be based on this repeating pattern. The forecast value was calculated as the average of the values in the previous weeks for the same day and time:

$$\hat{y}_t = \hat{f}(t) = \frac{1}{N} \sum_{i=1}^N y_{t-iW}$$

where W is the number of intervals in a week and N is the number of previous week considered.

The model performance were worse than the baseline. Looking at the results, it is clear that this model is not able to consider the variations in the inflow rate that can be caused by external factors, e.g. holidays. To take in account this variations, a local drift was calculated using the forecasted

errors value of the immediate past:

$$\hat{y}_t = \hat{f}(t) + \frac{1}{M} \sum_{i=1}^M \hat{y}_{t-i} - \hat{f}(t-i)$$

where M is the number of previous step considered.

3.2.3 Artificial Neural Network model

3.2.4 ARIMA model

3.3 Arrival rate forecast

3.4 Queue length forecast

3.4.1 Service rate approximation

3.4.2 Outflow rate approximation

3.5 Checkouts optimization

3.5.1 Opening/closing fluctuations control

4 Results

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam

5 Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

Bibliography

- [1] Larson R. Berman O. A queueing control model for retail services having back room operations and cross-trained workers. *Computers and Operations Research*, 31:201–222, 02 2004.
- [2] Snyder R. D. De Livera A. M., Hyndman R. J. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527, 2011.
- [3] Newell G. F. Queues with time-dependent arrival rates I, the transition through saturation. *Journal of Applied Probability*, 5(2):436–451, 1968.
- [4] Soares J. Green L. V., Kolesar P. J. Improving the SIPP approach for staffing service systems that have cyclic demands. *Operations Research*, 49:549–564, 08 2001.
- [5] Aksu H. Dwell time forecast and checkout optimisation in supermarkets. -, 2018.
- [6] Athanasopoulos G. Hyndman R. J. *Forecasting: Principles and Practice*. OTexts, 2nd edition, 2018.
- [7] Rider K. L. A simple approximation to the average queue size in the time-dependent M/M/1 queue. *J. ACM*, 23(2):361–367, 04 1976.
- [8] James W. Taylor. Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19(4):715–725, 2003.