Paper

Incorporating Customer Preference Information into the Forecasting of Service Sales

Piotr Rzepakowski

Abstract— Customers change their preferences while getting more familiar with services or being motivated to change their buying habits. Different sources of motivation induce customers to change their behavior: an advertisement, a leader in a reference group, satisfaction from services usage and other experiences, but usually those reasons are unknown. Nevertheless, people vary in susceptibility to suggestions and innovations, and also in preference structure change dynamics. Historical information about the preference structure gives additional information about uncertainty in forecasting activity. In this work the conjoint analysis method was used to find customer preference structure and to improve a prediction accuracy of telecommunication services usage. The results have shown that prediction accuracy increases about by one percent point, what results in a 20 percent increase after using proposed algorithm modification.

Keywords— conjoint analysis, consumer behavior, decision analysis, forecasting, marketing tools, multiple criteria analysis, preference measurement.

1. Introduction

The goal is to forecast services usage without complete knowledge and deep understanding of the domain, including lack of knowledge about predictor variables and intervention effects. Some of intervention effects [1] like customer relationship management activities are usually known but that information can be difficult to obtain. On the other hand, other factors, such as: an advertisement, an influence of a leader in a reference group, satisfaction from services usage, and other experiences can change a customer behavior, but usually this information is unavailable or the influence is unidentified. Taking into account this lack of konowledge, we make an assumption that an analyst has only substantial knowledge about business relationships and constraints which affect the customer activity. His knowledge must be good enough to identify which attributes describing users behavior differentiate them.

Usually, forecasting of time series, when only historical time series are known, are solved by univariate time series models which describe the behavior of a variable in terms of its own past values. Mostly, the exponential smoothing models (ESM) with or without seasonal effects are used [2]. In this work we consider user preference information to improve the exponential smoothing forecasting algorithm. Moreover, we make an assumption that data which were used to create time series are those which can be used for forecasting and for forecasting improvement.

In the summary of the progress made over the past quarter century with respect to methods reducing a forecast error [3] we can find seven well-established approaches which had been shown to improve prediction accuracy. The four of them: combing forecasts, Delphi, causal models, and trend-damping help with time series data. Additionally, other methods such as: segmentation, rule-based forecasting, damped seasonality, decomposing by causal forces and a damped trend with analogous data, were mentioned to be promising for those data. The author indicates also relatively untested methods: prediction markets, a conjoint analysis, diffusion models, and game theory. One of the conclusions from the summary is that, in general, the methods that have ignored theory, prior evidence, and domain knowledge have had a poor record in forecasting. That is why the general structure of the data should be analyzed.

Let us consider two promising methods: segmentation and decomposition by causal forces. The segmentation method is presented as an advantageous one because forecasting errors in different segments may offset one another. The author stresses also problems that can occur, if segments are based on small samples and noisy data, segment forecasts might contain very large errors. However, three reported comparative studies on segmentation that had been conducted since 1975 brought good results. The causal forces method also seems to be worth considering in the analysis of complex series. Complex series are defined as those in which causal forces derive series in opposite directions. If components of a complex series can be forecast more accurately than global series, it helps to decompose the problem by causal forces.

We have combined those two methods with the conjoint analysis to improve ESM models. It is known that forecasting in subgroups shouldn't bring worse prediction accuracy as long as values come from a stationary stochastic process [2]. Furthermore, if time series is known to follow a univariate autoregressive integrated moving-average (ARIMA) model, a forecast made using disaggregated data is, in terms of a mean square error (MSE), at least as good as using aggregated data. However, analyzed stochastic processes are not stationary and the disaggregation can deteriorate accuracy. On the other hand, a good subgroup selection can also improve forecasting exactness [4].

As a consequence, the main idea is to perform forecasting in subgroups defined dynamically by the customers' preference information gained from the conjoint analysis.

The proposed method has been verified on artificialy generated telecommunication services usage data. The best conjoint analysis model was chosen from models definied to identify telecommunication customers' preferences, and run on behavioral data [5]. The following values are forecasted: the number and duration of voice calls, the number of short message service (SMS) usages, the number of multimedia messaging service (MMS) usages, and the number and amount of general packet radio service (GPRS) usages. All the above mentioned values must be predicted within dimensions defined further in table in Subsection 4.2. The 18 months' history of the original telecommunication behavioral data – call data records (CDR) – are aggregated monthly by attributes defined in Table 1.

Table 1
Attributes of call data

Attribute	Levels
Service	Voice
	SMS
	MMS
	GPRS
Location	Home
	Roaming
Net	To on-net
	To off-net (mobile operators)
	To fixed operators
	To international operators
Tariff	Tariff [1–120]
Day type	Working days
	Weekend or holiday
Duration class	0 seconds
	15 seconds
	60 seconds
	240 seconds
Volume	Real values
Count	Integer values

In Section 2 the exponential smoothing models are introduced. Next, in Section 3 the preference identification method is described. In Section 4, a forecasting improvement is proposed. Results are presented in Section 5 and in Section 6 conclusions are drawn, and a plan for future work is proposed.

2. Exponential Smoothing Models

An exponential smoothing is a pure time series technique. This means that the technique is suitable when data have only been collected for series that are going to be forecasted. The exponential smoothing can therefore be applied when there are not enough variables measured to achieve good causal time series models, or when the quality of data is such that causal time series models give poor forecasts. In comparison, more general multivariate ARIMA models allow to predict values of a dependent time series with a linear combination of its own past values, past errors (also called shocks or innovations), and current and past values of other time series. Exponential smoothing takes the approach that recent observations should have relatively more weight in forecasting than distance observations. "Smoothing" implies predicting an observation by a weighted combination of previous values and "exponential" smoothing means that weights decrease exponentially as observations get older. In exponential smoothing only the slowly changing level is being modeled, nevertheless, it can be extended to different combinations of trend and seasonality:

- simple,
- double (Brown),
- linear (Holt) trend,
- damped-trend linear,
- no seasonality,
- additive seasonality,
- multiplicative seasonality.

Additionally, transformed versions of these models can be defined:

- logarithmic,
- square root,
- logistic,
- Box-Cox.

Given a time series $Y_t : 1 \leq t \leq n$, the underlying model assumed by the smoothing models has the following (additive seasonal) form:

$$Y_t = \mu_t + \beta_t t + s_p(t) + \varepsilon_t, \quad (1)$$

where:
μ_t – represents the time-varying mean term,
β_t – represents the time-varying slope,
$s_p(t)$ – represents the time-varying seasonal contribution for one of the p seasons,
ε_t – are disturbances.
Different smoothing models are presented in Table 2.

Table 2
Exponential smoothing models

Smoothing model	Equation
Simple	$Y_t = \mu_t + \varepsilon_t$
Double (Brown)	$Y_t = \mu_t + \beta_t t + \varepsilon_t$
Linear (Holt)	$Y_t = \mu_t + \beta_t t + \varepsilon_t$
Damped-trend linear	$Y_t = \mu_t + \beta_t t + \varepsilon_t$
Seasonal	$Y_t = \mu_t + s_p(t) + \varepsilon_t$
Winters – additive	$Y_t = \mu_t + \beta_t t + s_p(t) + \varepsilon_t$
Winters – multiplicative	$Y_t = (\mu_t + \beta_t t) s_p(t) + \varepsilon_t$

2.1. Smoothing State and Smoothing Equations

The smoothing process starts with an initial estimate of the smoothing state, which is subsequently updated for each observation using the smoothing equations. Depending on the smoothing model, the smoothing state at time t will consist of the following:
L_t – smoothed level that estimates μ_t,
T_t – smoothed trend that estimates β_t,
$S_{t-j}, j = 0, ..., p-1$, are seasonal factors that estimate $s_p(t)$.
The smoothing equations determine how the smoothing state changes as time progresses. Knowledge of the smoothing state at time $t-1$ and that of the time series value at time t uniquely determine the smoothing state at time t. The smoothing weights determine the contribution of the previous smoothing state to the current smoothing state. The smoothing equations for each smoothing model are listed in Table 3.

Table 3
Equations for the smoothing models

Smoothing model	The error-correction form, The k-step prediction equation
Simple	$L_t = L_{t-1} + \alpha\varepsilon_t$
	$\hat{Y}_t(k) = L_t$
Double (Brown)	$L_t = L_{t-1} + T_{t-1} + \alpha\varepsilon_t$
	$T_t = T_{t-1} + \alpha^2\varepsilon_t$
	$\hat{Y}_t(k) = L_t + ((k-1) + 1/\alpha)T_t$
Linear (Holt)	$L_t = L_{t-1} + T_{t-1} + \alpha\varepsilon_t$
	$T_t = T_{t-1} + \alpha\gamma\varepsilon_t$
	$\hat{Y}_t(k) = L_t + kT_t$
Damped-trend linear	$L_t = L_{t-1} + \phi T_{t-1} + \alpha\varepsilon_t$
	$T_t = \phi T_{t-1} + \alpha\gamma\varepsilon_t$
	$\hat{Y}_t(k) = L_t + \sum_{i=1}^{k} \phi^i T_t$
Seasonal	$L_t = L_{t-1} + \alpha\varepsilon_t$
	$S_t = S_{t-p} + \delta(1-\alpha)\varepsilon_t$
	$\hat{Y}_t(k) = L_t + S_{t-p+k}$
Winters – additive	$L_t = L_{t-1} + T_{t-1} + \alpha\varepsilon_t$
	$T_t = T_{t-1} + \alpha\gamma\varepsilon_t$
	$S_t = S_{t-p} + \delta(1-\alpha)\varepsilon_t$
	$\hat{Y}_t(k) = L_t + kT_t + S_{t-p+k}$
Winters – multiplicative	$L_t = L_{t-1} + T_{t-1} + \alpha\varepsilon_t/S_{t-p}$
	$T_t = T_{t-1} + \alpha\gamma\varepsilon_t/S_{t-p}$
	$S_t = S_{t-p} + \delta(1-\alpha)\varepsilon_t/L_t$
	$\hat{Y}_t(k) = (L_t + kT_t)S_{t-p+k}$

In order to use the multiplicative version of Winters method, the time series and all predictions must be strictly positive. Additionally, coefficient α, δ, γ must fulfill stability conditions [6].
Almost all exponential smoothing models have ARIMA equivalents presented in Table 4. ARIMA is more general than ESM and allows to predict values of a dependent time series with a linear combination of its own past values,

Table 4
ARIMA equivalent models

Smoothing model	ARIMA equivalent
Simple	ARIMA $(0,1,1)$
Double (Brown)	ARIMA $(0,2,2)$
Linear (Holt)	ARIMA $(0,2,2)$
Damped-trend linear	ARIMA $(1,1,2)$
Seasonal	ARIMA $(0,1,p+1)(0,1,0)_p$
Winters – additive	ARIMA $(0,1,p+1)(0,1,0)_p$
Winters – multiplicative	No equivalent

past errors (also called shocks or innovations), and current and past values of other time series (predictor time series).

2.2. Prediction Errors

Predictions are made based on the last known smoothing state. Predictions made at time t for k steps ahead are denoted $\hat{Y}_t(k)$ and the associated prediction errors are denoted $\varepsilon(k) = Y_{t+k} - \hat{Y}_t(k)$.
The one-step-ahead predictions refer to predictions made at time $t-1$ for one time unit into the future, that is $\hat{Y}_{t-1}(1)$, and the one-step-ahead prediction errors are more simply denoted $\varepsilon_t = \varepsilon_{t-1}(1) = Y_t - \hat{Y}_{t-1}(1)$. The one-step-ahead prediction errors are also the model residuals, and the statistic related to the one-step-ahead prediction errors is the objective function used in smoothing weight optimization.

Table 5
The variance of the prediction errors

Smoothing model	$\varepsilon_t(k)-$ variance
Simple	$\mathrm{var}(\varepsilon_t)[1+\sum_{j=1}^{k-1} \alpha^2]$
Double (Brown)	$\mathrm{var}(\varepsilon_t)[1+\sum_{j=1}^{k-1} (2\alpha+(j-1)\alpha^2)^2]$
Linear (Holt)	$\mathrm{var}(\varepsilon_t)[1+\sum_{j=1}^{k-1} (\alpha + j\alpha\gamma)^2]$
Damped-trend linear	$\mathrm{var}(\varepsilon_t)\left[1+\sum_{j=1}^{k-1} (\alpha+\frac{\alpha\gamma\phi(\phi^j-1)}{(\phi-1)})^2\right]$
Seasonal	$\mathrm{var}(\varepsilon_t)\left[1+\sum_{j=1}^{k-1} \psi_j{}^2\right]$
Winters – additive	$\mathrm{var}(\varepsilon_t)\left[1+\sum_{j=1}^{k-1} \psi_j{}^2\right]$
Winters – multiplicative	$\mathrm{var}(\varepsilon_t)\left[1+\sum_{i=0}^{\infty} \sum_{j=1}^{p-1} (\frac{\psi_{j+ip} S_{t+k}}{S_{t+k-j}})^2\right]$

The variance of the prediction errors counted as presented in Table 5 is used to calculate the confidence limits.

3. Conjoint Analysis for Preference Identification

For preference identification, wchich are going to be used for spliting customers into homogenous segments, we used the conjoint analysis method running on behavioral data [5].

The conjoint analysis process consists of:

- selection of utility factors,
- conjoint measure definition,
- conjoint model definition,
- questionnaire preparation,
- questionnaire data acquisition,
- statistical analysis,
- data interpretation.

For utility factors we get some attributes from the behavioral data. The questionnaire preparation step is not required because the historical data are analyzed. Hence, the questionnaire data acquisition step changes to the behavioral data preparation one.

3.1. Selection of Utility Factors

Attributes differentiating the cost of services mostly were chosen to be utility factors. Among them there are: service, location, network, day types, and duration class attributes with categories presented in Table 6. Original CDR were transformed to determine chosen attributes. Next, the data were aggregated and statistics of call frequencies for each aggregation were calculated.

Table 6
Utility factors

Attribute	Levels
Service	Voice
	SMS
	MMS
	GPRS
Location	Home
	Roaming
Net	To on-net
	To off-net (mobile operators)
	To fixed operators
	To international operators
Day type	Working days
	Weekend or holiday
Duration class	0 seconds
	15 seconds
	60 seconds
	240 seconds

3.2. The Conjoint Measure Definition

The dependency between utility factors is defined by the conjoint measure. It consists of intercept coefficient μ and part-worth utilities associated with attributes. If some attributes are correlated then the interaction between those attributes are added to the conjoint measure. Interactions between pairs usually suffice but sometimes interactions of higher orders, for example, between three variables are used. For presented telecommunication task, the conjoint measure is defined by Eq. (2). In that example part worth utilities are presented by α vectors of utilities for attribute values, β vectors of utilities for all combinations of values associated with two attributes and γ vector of utilities for a combination of values taken from service, net, and day type attributes. For the presented telecommunication task, we used a measure consisting of linear terms and correlation between all pairs of attributes extended by interactions between three attributes. Finally, the conjoint measure consists of factors presented in Table 7 and is defined as follows:

$$
\begin{aligned}
y = \mu & \\
& + \alpha_{service} + \alpha_{location} + \alpha_{net} + \alpha_{day\ type} + \alpha_{duration\ class} \\
& + \beta_{service*location} + \beta_{service*net} + \beta_{service*duration\ class} \\
& + \beta_{service*day\ type} + \beta_{location*net} + \beta_{location*duration\ class} \\
& + \beta_{location*day\ type} + \beta_{net*day\ type} + \beta_{net*duration\ class} \\
& + \gamma_{service*net*day\ type} \\
& + \varepsilon. \qquad (2)
\end{aligned}
$$

Table 7
Conjoint measure factors

Attribute	Levels
Service	4
Location	2
Net	4
Day type	2
Volume	4
Service*location	8
Service*net	16
Service*day type	8
Service*duration class	7
Location*net	8
Location*day type	4
Location*duration class	8
Net*day type	8
Net*duration class	16
Service*net*day type	32
Total	131

3.3. Conjoint Model Definition

The conjoint model is a statistical model which represents dependencies between utility of a profile and its attributes and is defined by Eq. (3). Now α coefficient represent utilities associated with all conjoint factors α, β and γ defined earlier. Because all of attributes of conjoint measure are categorical, dummy variables x created to represent no met-

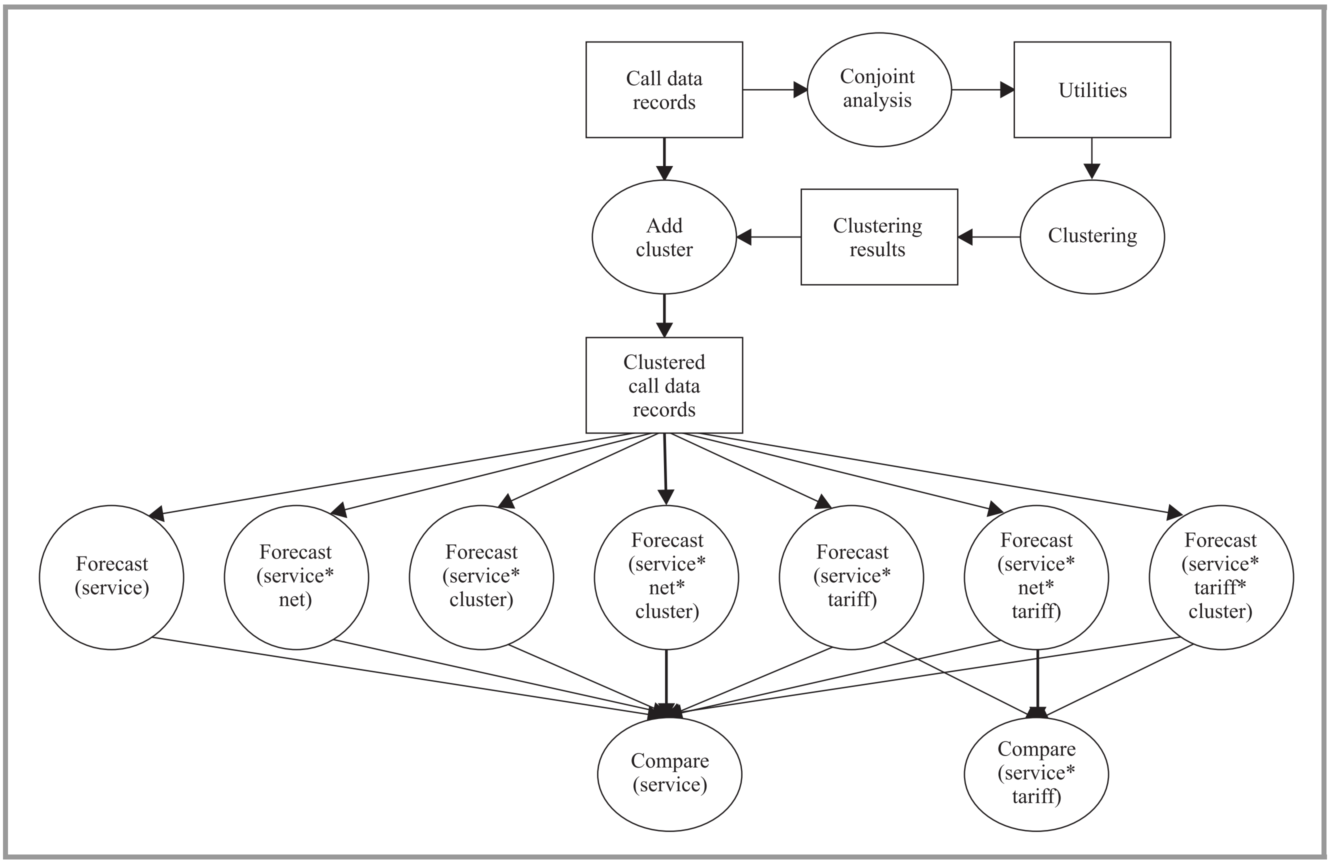

Fig. 1. Forecasting procedure.

ric information. One attribute with k levels was replaced by $k-1$ binary attributes

$$y = \alpha^T x + \varepsilon. \tag{3}$$

After adding dummy variables, regression techniques can be used for part worth utilities identification. Dependent variable y in the regression model represents the utility of a profile. In the analyzed problem it was calculated as the number of events, which means that it has binomial distribution. That problem cannot be solved simply by linear regression as regression techniques require normal distribution of dependent variable. However, binomial distribution can be simply transformed to the normal one by logarithmic function. In consequence, general linear model (GLM) was defined as

$$\ln(y) = \alpha^T x + \varepsilon. \tag{4}$$

4. Forecasting Improvement

Forecasting improvement is done by data disaggregation and the criteria of splitting the data are the main point of this improvement. In fact, the data are split using information about customer preferences. This proposition is supported by hypothesis which states, that customers who have similar preferences behave similarly and the variance of a service usage in a group is lower than in the whole population. In the presented method, preferences come from a behavioral data and can be treated as aggregated representation of the way in which customers use services. As a consequence of this idea, analyses are carried out as follows:

– at first, customer segmentation is done on preferences to a service usage;

– next, forecasts are made in segments;

– finally, a forecast in the whole population is calculated as a sum of forecasts in subgroups.

In conduct analysis, forecasts using different disaggregation methods are compared on two levels: on the service aggregation level and the combinations of service and tariff aggregations. The process of forecasting using various disaggregation methods is presented in Fig. 1. At first, the CDR data are used to find coustomers part-worth utilities. After that, customers are clustered into homogenous groups using calculated utilities. Information about a customer group is added to each record in the CDR. Then a customer segment identifier can be used in data aggregation to make forecasting in subgroups.

4.1. Customer Segmentation on Preferences to Service Usage

Consumer preferences were determined by running a conjoint analysis procedure on behavioral data as it has been shown in Section 3. Those preferences were computed on 12 months' data. Next, clustering was done to split consumers into homogenous groups.

There are two types of clustering: partition clustering and hierarchical clustering. Partition clustering attempts to directly decompose data set into a set of disjoint clusters. Hierarchical clustering, on the other hand, proceeds successively by either merging smaller clusters into larger ones, or by splitting larger clusters. For a huge amount of data hierarchical clustering is not practically applicable, thus we used partition clustering implemented in statistic analytical software (SAS) as a FASTCLUS procedure. In the used partition clustering, the number of clusters has to be given as an input to the procedure. The procedure was run many times to make: 200, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000 groups clustering.

4.2. Forecasting

To check how preference clustering influences the forecasting accuracy, we made comparisons of forecasting made in aggregations presented in Table 8.

Table 8
Dimension intersections

Intersection	Number of forecasts
Service	4
Service*net	16
Service*tariff	800
Service*net*tariff	3200
Service*cluster	4*clusters
Service*net*cluster	16*clusters
Service*tariff*cluster	800*clusters
Service*net*tariff*cluster	3200*clusters

The high-performance forecasting (HPF) procedure from SAS was used for forecasting. This procedure provides an automatic way to generate forecasts for each time series.

The best model is automatically choosen from the exponential smoothing models presented in Section 2. And the mean absolute percent error (MAPE) good-of-fit statistic is used to measure how models fit data:

$$MAPE = \frac{100}{\mathrm{n}} \sum_{t=1}^{n} \left| \frac{\mathrm{y_t} - \mathrm{\hat{y}_t}}{\mathrm{y_t}} \right|. \tag{5}$$

The summation ignores observations, where $y_t = 0$.

5. Analytical Results

Analitycal results are summarized in two subsections. The first one concerns the conjoint analysis and the second one the forecasting process.

5.1. The Conjoint Analysis

The conjoint analysis was performed on 12 months' data. Statistics R^2 presented in Table 9 show that the model is well fitted to the data. The average value of R^2 is 95% and the standard deviation is very low.

Table 9
Analysis of variance for the coujoint model

Statistic	Avg	Std
R^2	0.95	0.11
$adj-R^2$	0.82	0.29
p-value	0.05	0.15

Table 10
Relative importance statistics in population [%]

Attribute/statistic	Avg	Std
Service	12.0	10.2
Location	2.2	5.8
Net	10.9	9.2
Day type	7.2	9.8
Duration class	13.3	17.6
Service*location	2.2	5.7
Service*net	10.1	9.3
Service*day type	5.2	5.4
Service*duration class	3.9	6.2
Location*net	1.6	4.5
Location*day type	0.9	2.9
Location*duration class	1.3	4.0
Net*day type	5.3	5.1
Net*duration class	13.3	8.8
Service*net*day type	5.8	7.4

Comparing standard deviations to average values of importances illustrated in Table 10, we find that customers have different menners and different features of services are important for them. These statistics show, that spliting customers into more homogenous grups is worth considering, what is also shown in Subsection 5.2.

5.2. Forecasting Comparison

Prediction accuracy in clusters has been compared to forecasting made in data disaggregated by attributes available a priori (Figs. 2–12): the tariff plan and the net including intersections defined in Table 8. An optimal number of clusters were found from figures presenting prediction accuracy of statistics at the total level drawn in different number of clusters. In three out of five time series, clustering brougth good results and only prediction of GPRS

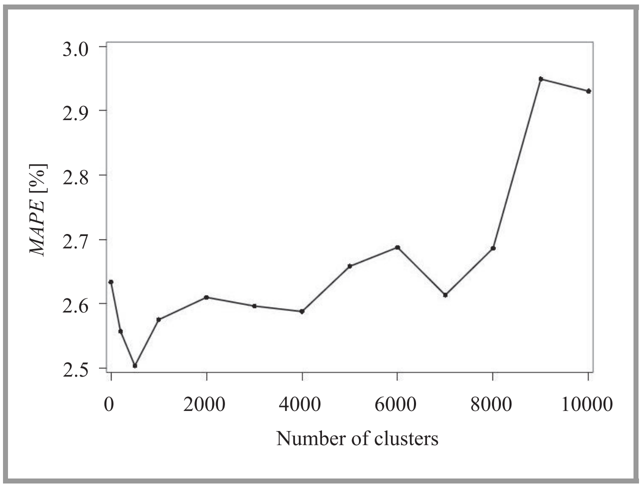

Fig. 2. Prediction accuracy of the total duration of voice events for different number of clusters verified on 15 month data.

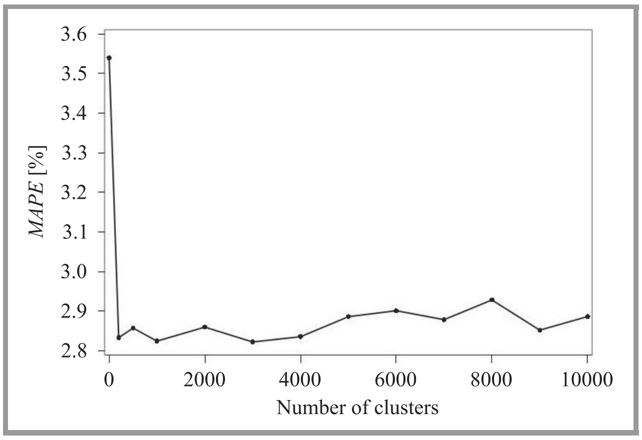

Fig. 3. Prediction accuracy of the total number of voice events for different number of clusters verified on 15 month data.

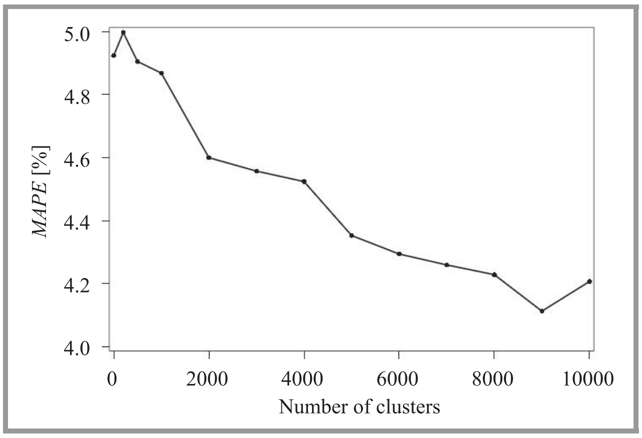

Fig. 4. Prediction accuracy of the total number of SMS events for different number of clusters verified on 15 month data.

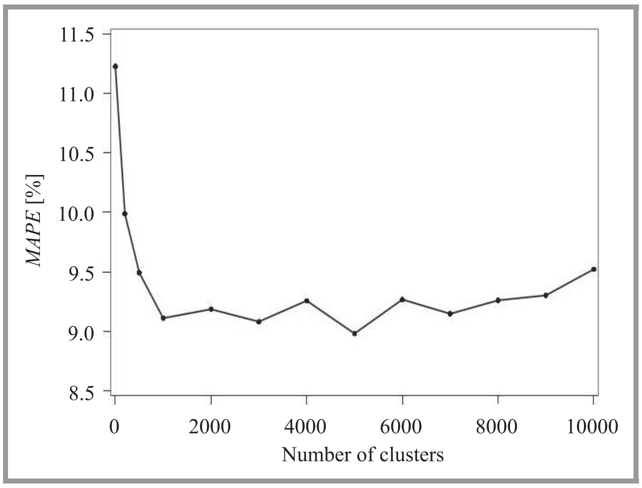

Fig. 5. Prediction accuracy of the total number of MMS events for different number of clusters verified on 15 month data.

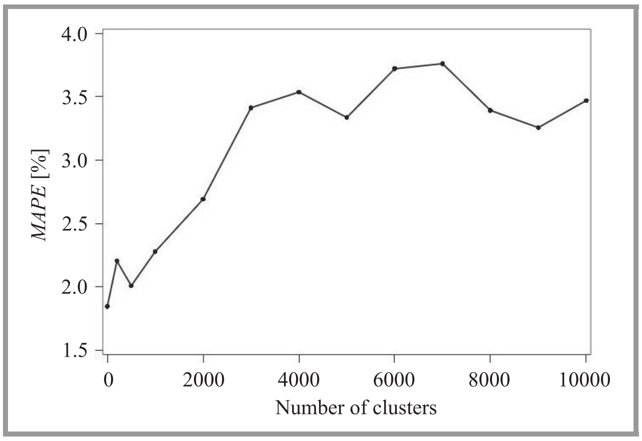

Fig. 6. Prediction accuracy of the total number of GPRS events for different number of clusters verified on 15 month data.

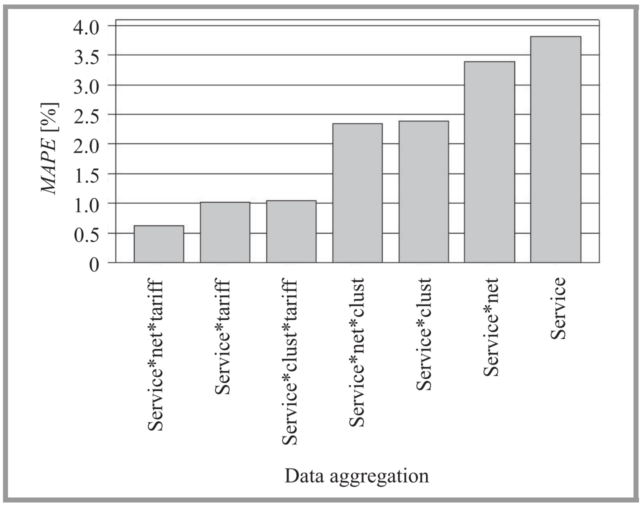

Fig. 7. Prediction accuracy of the total number of voice events (500 clusters).

usages (Fig. 6) and duration of voice calls (Fig. 2) in clusters brought worse results. Probably this is caused by the conjoint analysis model not properly suited to GPRS data. Prediction of the total number of voice calls is better without clustering (Figs. 7–10), however, when predictions of the same value are compared at tariff aggregations, results

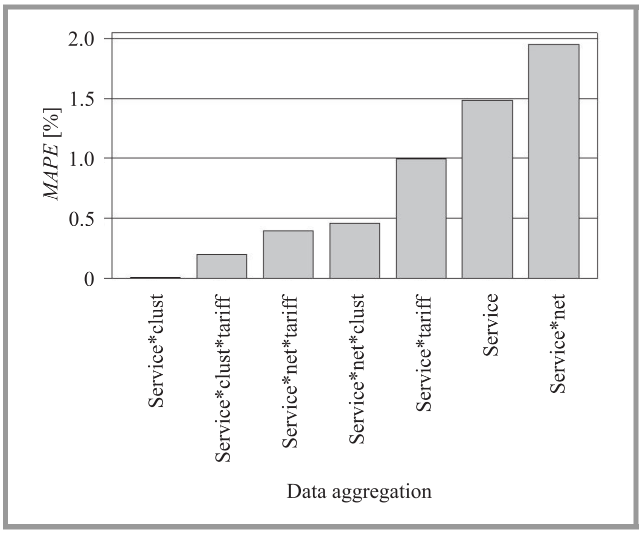

Fig. 8. Prediction accuracy of the total duration of voice events (500 clusters).

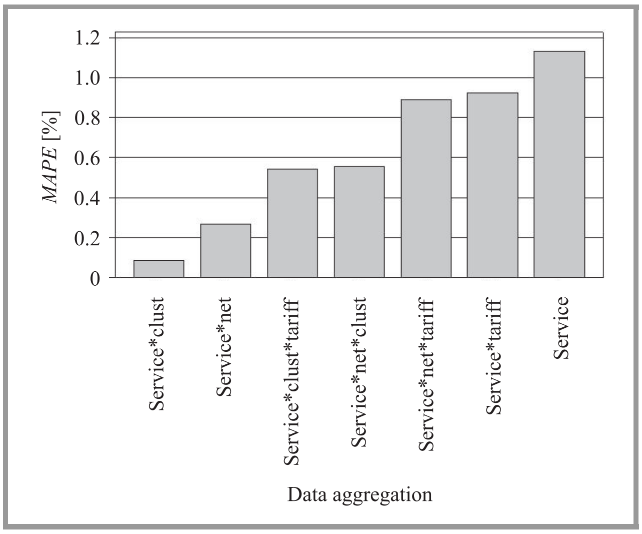

Fig. 9. Prediction accuracy of the total number of SMS events (10000 clusters).

are much better what is shown in Fig. 12 and accordingly in Fig. 11 for voice duration. An optimal number of clusters for prediction statistics at the total level as well as a prediction accuracy increase are summarized in Tables 11 and 12.

Table 11
Optimal number of clusters

Service	Optimal number of clusters	Accuracy increase [p.p.]
Voice – duration	500	1.5
Voice – count	500	0.7
SMS	10000	0.8
MMS	1000	2.0
GPRS	0	0

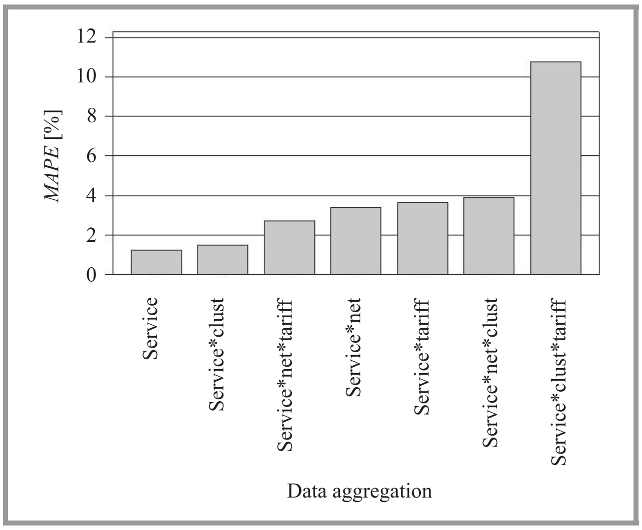

Fig. 10. Prediction accuracy of the total number of the MMS events (1000 clusters).

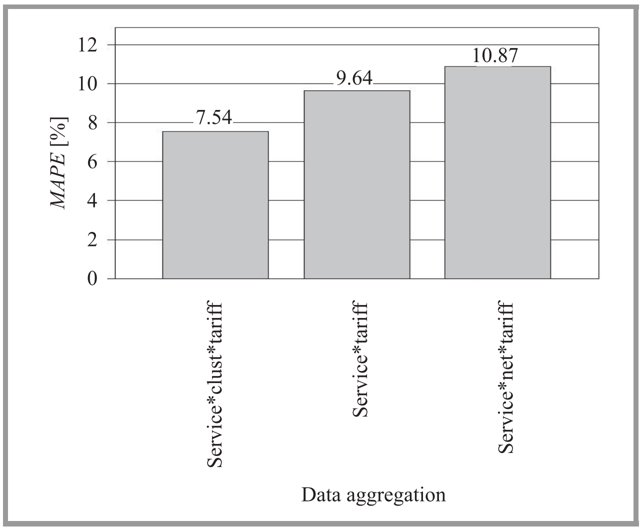

Fig. 11. Prediction accuracy of the total duration of voice events calculated at the tariff level (500 clusters).

From Fig. 13, which presents what kind of forecasting models were used, we can find that only logarithmic transformation is applied to telecommunication data and the number of transformations decreases while the number of clusters

Table 12
Total service usage prediction

Service	Best data aggregation	Accuracy change after clustering [p.p.]
Voice – duration	Service*cluster	0.4
Voice – count	Service*net*tariff	–0.4
SMS	Service*cluster	0.2
MMS	Service*cluster	1.2
GPRS	Service*net	–1.6

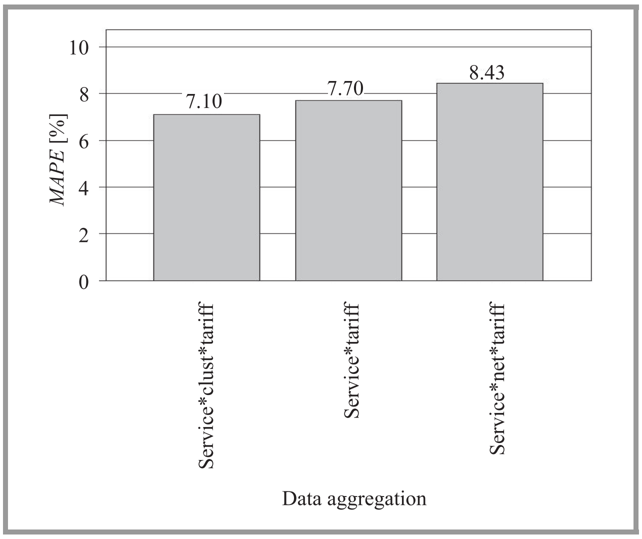

Fig. 12. Prediction accuracy of the total number of voice events calculated at the tariff level (500 clusters).

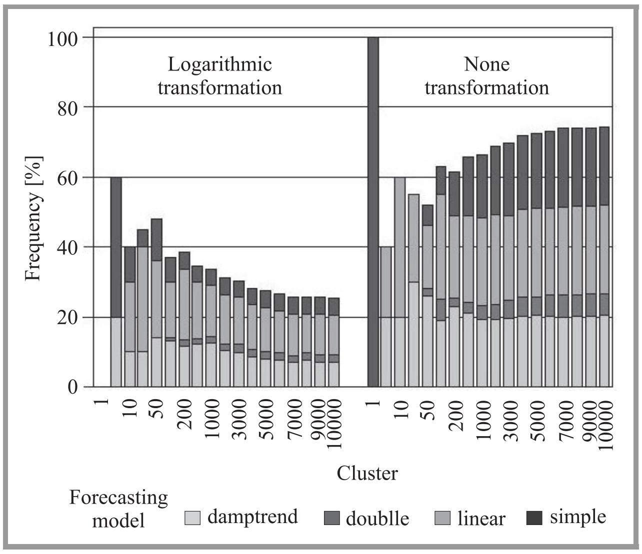

Fig. 13. Statistics of forecasting models applied to the number of voice events prediction.

increases. We can also observe that the number of models called simple, increases as the number of clusters is going up. It shows that forecasting in disaggregated data results is simple and usually more accurate models.

6. Conclusions and Future Research

Analytical results have shown that clustering with the optimal number of clusters, can increase model prediction accuracy. However, good results can be achieved only when the preference model used to identify customers' preferences describes dependencies in data appropriately. The used conjoint measure is not sufficiently suited to data and does not describe GPRS usage properly. The week conjoint measure causes a lack of the prediction accuracy increase in the GPRS time series. On the other hand, poor prediction made after clustering at the top level does not have to cause poor prediction at the lower level. This feature was shown on the number of voice call prediction example, where predictions at the tariff level were much better then at the service level.

In future work more sophisticated forecasting model should be considered. It would be worth to knowing if the multivariate ARIMA models get better results then ESM with proposed improvement. Probably customer preferences could be incorporated into ARIMA models as intervention effects and would also give positive results, what is going to be verified in future work.

References

[1] A. Vag, "Simulating changing consumer preferences: a dynamic conjoint model", *J. Busin. Res.*, vol. 60, no. 8, pp. 904–911, 2007.

[2] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting", *Int. J. Forecast.*, vol. 22, no. 3, pp. 443–473, 2006.

[3] J. S. Armstrong, "Findings from evidence-based forecasting: methods for reducing forecast error", *Int. J. Forecast.*, vol. 22, pp. 583–598, 2006.

[4] M. J. del Moral and M. J. Valderrama, "A princilpal component approach to dynamic regression models", *Int. J. Forecast.*, vol. 13, no. 2, pp. 237–244, 1997.

[5] P. Rzepakowski, "Supporting telecommunication product sales by conjoint analysis", *J. Telecommun. Inform. Technol.*, no. 3, pp. 28–34, 2008.

[6] SAS Institute Inc., Cary, NC: SAS Institute Inc. SAS/ETS® 9.2: User's Guide, 2008.

Piotr Rzepakowski received the M.Sc. degree in computer science from the Warsaw University of Technology, Poland, in 2003. Currently he is a Ph.D. candidate in computer science at the Warsaw University of Technology (Institute of Control and Computation Engineering). He is employed by National Institute of Telecommunications in Warsaw. Has taken part in projects related to data warehousing and analysis for a telecommunication operator. His research focuses on modeling, decision support, data mining, and customer preferences identication issues.

e-mail: Piotr.Rzepakowski@elka.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

e-mail: P.Rzepakowski@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland