A Hybrid Two-Level Support Vector Machine-Based Method for Churn Analysis

FERDI SARAC

Suleyman Demirel University, Isparta, Turkey

HUSEYIN SEKER

Birmingham City University, Birmingham, UK

MARCIN LISOWSKI

Bubo.AI Ltd., UK

ALAN TIMOTHY

Bubo.AI Ltd., UK

Customer churn is a central problem in almost every sector. Due to the diversity of the customers, products and services, and a massive amount of data being generated as a result of e-commerce tools and services, (big) data analytics and artificial intelligence-based methods have been developed and used for churn analysis in order to develop a strategy that is expected to understand the reasons behind the customer churn and subsequently to develop an effective and profitable customer retention programme. The analysis based on the data analytics and artificial intelligence methods focuses more on the profiling of customers, the classification of customer churn and identification of features that affect the churn. However, there doesn’t seem many studies that would be able to help understand how much a potential customer is likely to (or not likely to) pay for the products or services when churned or not, and to predict how much a particular customer or group of customers may have paid for the products or services. Therefore, in this study, a two-level churn analysis is proposed to (1) classify the customer churn or not, and (1) predict how much the customer has paid for the service. In order to achieve it, a machine learning method, namely support vector machine (SVM), was used for the classification part whereas a monthly service charge was predicted by using support vector regression (SVR) method. In order to select the most appropriate feature sub-set for both analyses, an unsupervised feature selection method, namely the multi-cluster feature selection method was utilized. The same feature selection method was used for both analyses for the sake of consistency to understand its performance over both analyses. The proposed hybrid approach was then applied in IBM’s Telcom data set with over 7000 customers in order to demonstrate the applicability and generalization ability of the proposed two-level approach. The SVM-based classification method has yielded AUC 85.6 and total classification accuracy of 81.5% being higher than those of a recent study where an aggressive set of the supervised classification methods was performed. The SVR-based prediction of the monthly charge has resulted in RMSE of 1.27, which is a reasonably acceptable outcome in the sector given the diversity of the ranges of charges as evidenced in its standard deviation. The approach presented in the study demonstrates that both the churn classification and charge prediction can be performed at the same time with a higher degree of accuracy. As the approach is open for further improvement, future analysis will be carried out to improve the accuracy for both analyses over other data sets to demonstrate its robustness and generalization ability.

CCS CONCEPTS • Computing Methodologies > Artificial Intelligence and Machine Learning

Additional Keywords and Phrases: Classification, Prediction, Feature Selection, Support Vector Machine, Support Vector Regression, Multi Clustering Feature Selection, Unsupervised Feature Selection, Churn Analysis, Customer Willingness to Pay, Customer Retention, IBM Telecom Data Set.

ACM Reference Format:

1 INTRODUCTION

Customer churn can be defined as the ratio of customers who did not continue using a company's service or buying its product, consequently stop paying for good or a certain period of time. The customer churn is one of the main factors that highly affect the company's operation, profitability and survivability. Therefore, the customer churn analysis is one of the major tasks in every sector as the cost of gaining new customers is generally higher than the cost of keeping existing customers [1, 2].

Due to the diversity of the customers, products and services, and a massive amount of data being generated as a result of e-commerce tools and services, (big) data analytics and artificial intelligence-based methods have been developed and used for churn analysis, outcome of which is used to help understand the reasons behind the customer churn and subsequently to develop an effective and profitable customer retention programme. In line with the technological advances, customer churn analysis has been enhanced to find more effective ways for customer retention by means of (big) data analytics, machine learning and artificial intelligence techniques [3, 4, 5, 6, 7].

The customer churn analysis based on the data analytics, machine learning and artificial intelligence methods focuses more on the profiling of customers, the classification of customer churn and identification of features that affect the churn [7]. However, there doesn't seem many studies that would be able to predict how much a particular or group of customers may have paid for the products or services, and how much a potential customer is likely to (or not likely to) pay for the products or services when churned or not.

There are generally two-stage method used for the customer churn analysis; (1) feature selection to identify which sub-set(s) of the features that characterise the customers would be better to find the churned or not, and (2) prediction of the churn YES/NO [3, 7]. In addition to this usual approach, this current study includes the prediction of the charge or fee that the customer has paid for.

Feature selection methods can be mainly divided into two different categories: unsupervised and supervised methods. In the supervised feature selection methods, sub-set of the features are selected based on their correlation with their desired outputs [8]. The unsupervised feature selection methods use only data inputs to identify discriminative attributes independent of their target values (desired outputs). It has been shown that the supervised feature selection methods generally pick up irrelevant and noisy features in addition to relevant variables as its aim is to increase its predictive accuracy regardless of the feature information retained in the set [9, 10]. In addition, it is based only on a specific predictive model but not independent of any such algorithmic architecture. Therefore, their results may not be applied to other predictive methods. In this study, an unsupervised feature selection method, namely Multi – Cluster Feature Selection (MCFS) [11] has been utilised for both classification and prediction tasks for the sake of consistency over these two approaches.

There have been several methods used for the churn classification. In order to generalise the methods for both classification and prediction tasks, a support vector-based model is generally utilised where support vector machine and support vector regression are two methods that have been used for the classification and prediction, respectively [12]. Unlike other studies that focus more on the churn classification, this study addresses to both the churn classification and the prediction of the product/service fee/charge. Therefore, an approach with a similar mathematical architecture, namely support vector-based method, has been utilised.

The rest of the paper is organised as follows. In Section 2, the methods and materials are explained including support vector machine and support vector regression methods, multi-cluster feature selection technique, IBM Telcom data set and techniques used for the statistical validation of all the methods studied. Section 3 presents the general characteristics of the data set and the results obtained using the support vector machine, support vector regression and multi-cluster feature selection methods where the discussions of the results are also given. Finally, Section 4 concludes the paper along with future works.

2 METHODS AND MATERIALS

This section includes the detail explanation of support vector machine, support vector regression, multi-cluster feature selection method, IBM Telcom data set and techniques used for the statistical validation of all the methods studied. General approach for the classification and prediction with all the features using SVM and SVR, respectively, is illustrated in Fig.1. The framework with the multi-cluster feature selection method is illustrated in Fig.2.

2.1 Support Vector Machine and Support Vector Regression

Support Vector Machine (SVM) can be considered as a statistical machine learning algorithm that has been used for both regression and classification purposes [13, 14]. The regression version of SVM is called Support Vector Regression (SVR). As the certain number of training instances is selected as support vectors to map the relationship between inputs and desired outputs, the weighted sum of these support vectors are then gained to construct the regression model. In this study, LIBSVM library [15] is used to implement both SVM and SVR.

2.2 Multi-Cluster Feature Selection Method

In this study, Multi-Cluster Feature Selection Method (MCFS) as an unsupervised feature selection method is used to select the most optimum sub-set of the features listed in Table 1 for the classification and prediction tasks. MCFS is one of the effective methods that exploit spectral regression with l1-norm regularization to identify relevant features [11]. As it is independent of target outputs, it can be used for both unsupervised and supervised feature selection purposes as well as both classification and prediction tasks.

2.3 IBM Telco Data Set

IBM Telco Data Set has 7032 samples (customers), each of which is represented with 21 features as detailed in Table 1. The features 19 and 21 are used as a desired target in the prediction (regression) and classification tasks, respectively. The “Churn” column includes the customers who left within the last month. Further information about the data set can be found in [16].

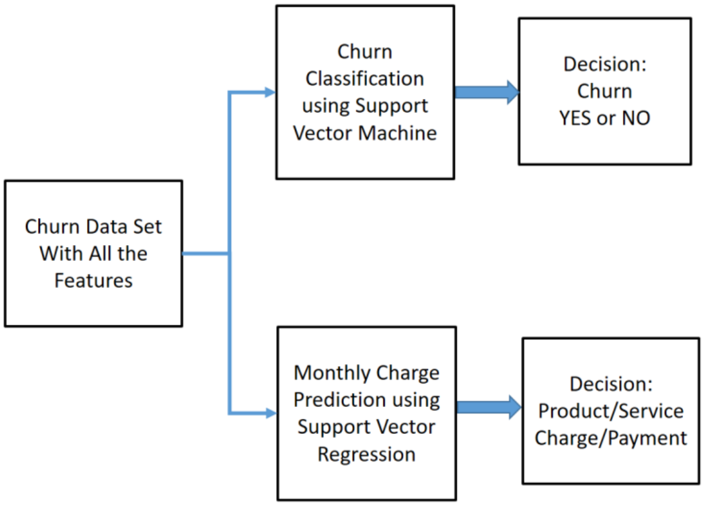
Churn Data Set With All the Features
Churn Classification using Support Vector Machine
Decision: Churn YES or NO
Monthly Charge Prediction using Support Vector Regression
Decision: Product/Service Charge/Payment

Fig.1. Churn Analysis with the Entire Future Set for the Classification of Churn (YES/NO) and the Prediction of Product/Service Charge/Payment.

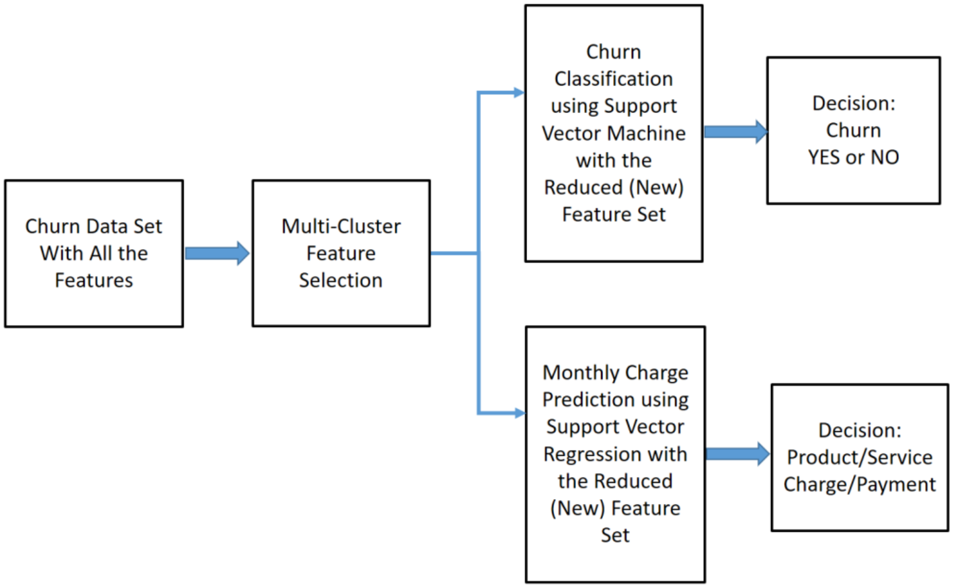
Churn Data Set With All the Features
Multi-Cluster Feature Selection
Churn Classification using Support Vector Machine with the Reduced (New) Feature Set
Decision: Churn YES or NO
Monthly Charge Prediction using Support Vector Regression with the Reduced (New) Feature Set
Decision: Product/Service Charge/Payment

Fig.2. Churn Analysis using MCFS-based Reduced (New) Feature Set for the Churn Classification (YES/NO) and the Prediction of Product/Service Charge/Payment.

2.4 Statistical Validation

For the classification task with Support Vector Machine, two metrics are used to assess the performance of the churn classifiers. They are Area Under Curve (AUC) and Total Accuracy (%). AUC can be interpreted as an aggregate measure of classifier performance over all possible classification thresholds. The total accuracy (%) can be formulated as

$$Total\ Accuracy\ (\%) = 100x \frac{Number\ of\ Correctly\ Classified\ Samples}{Number\ of\ All\ Samples\ in\ the\ Data\ Set}$$

For the prediction (regression) task with Support Vector Regression (SVR), Root Mean Squared Error (RMSE) is used to assess SVR's performance, namely the prediction of the monthly charge. It can be formulated as

$$RMSE = \sqrt{\frac{\sum_i^n (y_i - y_i')^2}{n}}$$

where *n* is the number of all the samples in a data set, y_i and y_i' are the actual (desired) value and predicted value, respectively.

In order to independently assess these two methods, the data set is divided into training and testing sets. The training set is used to train the models whereas the testing set is used to test the performance of the models. In order to perform this analysis, a 5-fold cross validation method is adapted for both the churn classification and monthly charge prediction tasks. This means that the data is randomly divided into 5 sub-sets with more or less equal number of samples, and while four of the sub-sets are used for training, the remaining sub-set is then used for testing the method. This process is repeated until all the sub-sets are tested independently without being used in the training process. The final result is the average over the testing results of these 5 sub-sets.

3 RESULTS AND DISCUSSIONS

The statistical analysis of the data set as presented in Tables 2 and 3 reveals the diversity of the customers and products/services. In particular, monthly and total charges represent a quite wide range of fees paid for as evidenced in their standard deviation values. The diversity seems to exist for both male and female customers. The result also indicates that it is a challenging prediction problem for a given customer. An interesting observation is that female customers seem to have paid for more than male customers although other features are more or less balanced for both genders.

The accuracy values for the classification task is presented in Table 4. The results shows that the proposed SVM-based method with 17 features selected by means of MCFS yields the highest predictive accuracy of 81.5% and AUC value of 85.6, which outperforms the results presented in [7].

The results for the prediction of the monthly charge are presented in Table 5. The SVR model with all the features yielded better RMSE than that of the SVR model with 10 features that were selected by means of MCFS.

4 CONCLUSIONS

In the paper, the churn analysis was performed for not only the classification of the churned customers but also the prediction of the monthly charge for the customers using SVM and SVR, respectively. The classification results were found to be higher than those presented in a recently published paper [7]. In addition to the increased classification accuracy, another unique aspect of this current study is the prediction of the monthly charge for the customers using SVR, which can potentially be used for several purposes including the profiling of the customers according to the charges for services/products and potential prediction of how much a customer is likely (or not likely) to pay for services/products.

Having demonstrated the robustness of both the classification and prediction tasks, future analysis will be carried out to develop machine learning-based methods that are more capable of improving the performance for both analyses to demonstrate its robustness and generalization ability using different feature selection and prediction methods as well as their applications over other data sets.

Table 1. The Features in IBM Telcom Data Set in the Churn Analysis [16]

# Features	Data Format	Description
1 CustomerID	String	Customer ID
2 Gender	(Female/Male)	Whether the customer is a male or a female
3 SeniorCitizen	(0/1)	Whether the customer is a senior citizen or not
4 Partner	(Yes/No)	Whether the customer has a partner or not
5 Dependents	(Yes/No)	Whether the customer has dependents or not
6 Tenure	Numerical	Number of months the customer has stayed with the company
7 PhoneService	(Yes/No)	Whether the customer subscribes a phone service or not
8 MultipleLines	(Yes/No /No phone service)	Whether the customer subscribes multiple lines or not
9 InternetService	(DSL/Fiber optic /No)	Customer's internet service provider
10 OnlineSecurity	(DSL/Fiber optic /No)	Whether the customer subscribes online security or not
11 OnlineBackup	(Yes/No /No internet service)	Whether the customer subscribes online backup or not
12 DeviceProtection	(Yes/No/No internet service)	Whether the customer subscribes device protection or not
13 TechSupport	(Yes/No/No internet service)	Whether the customer subscribes tech support or not
14 StreamingTV	(Yes/No/No internet service)	Whether the customer subscribes streaming TV or not
15 Streaming Movies	(Yes/No/No internet service)	Whether the customer subscribes streaming movies or not
16 Contract	(Month-to-Month /One Year /Two Year)	The contract term of the customer
17 PaperlessBilling	(Yes/No)	Whether the customer uses paperless billing or not
18 PaymentMethod	(Bank Transfer/Credit Card/Electronic)	The customer's payment method
19 Monthly Charges	numerical	The amount charged to the customer monthly
20 TotalCharge	numerical	The total amount charged to the customer
21 Churn	(Yes/No)	Whether the customer churned or not

Table 2. General Statistics of the IBM Telecom Data with respect to Churn YES/NO

	Number of Male Customers	Number of Female Customers	Average Monthly Charge (STDV)	Average Total Charge (STDV)	Number of Senior Citizen	Number of Multiple Line Users
Churn YES	930 (%49.8)	939 (%50.2)	74.44 (24.67)	1531.80 (1890.82)	476 (%25.5)	850 (%45.5)
Churn NO	2625 (%50.7)	2549 (%49.3)	61.27 (31.09)	2555.34 (2329.46)	666 (%12.9)	2121 (%41)

Table 3. Gender-specific Statistics of the IBM Telecom Data

	Churn	Number of Customers	Average Monthly Charge (STDV)	Average Total Charge (STDV)	Number of Senior Citizen	Number of Multiple Line Users
Male	YES	930	74.07 (25.18)	1440.98 (1777.93)	236 (%25.4)	434 (%46.7)
	NO	2625	60.88 (30.96)	2517.74 (2306.27)	338 (%12.9)	1051 (%40)
Female	YES	939	74.81 (24.16)	1623.50 (1995.12)	240 (%25.6)	416 (%44.3)
	NO	2549	61.67 (31.23)	2594.05 (2352.92)	328 (%12.9)	1070 (%41)

Table 4 – Performance of Churn YES/NO Classification

Method	Number of Features	AUC	Total Accuracy (%)
SVM with All the Features	All Features	79.2	80.1
SVM with MCFS	17	85.6	81.5
The highest result presented in [7]	-	84.5	80

Table 5 - Prediction of Monthly Charge

Method	Number of Features	RMSE
SVR with All the Features	All Features	1.27
SVR with MCFS	10	2.12

REFERENCES

[1] JAEHYUN AHN, JUNSIK HWANG, DOYOUNG KIM, HYUKGEUN CHOI, AND SHINJIN KANG, "A Survey on Churn Analysis in Various Business Domains", IEEE Access, December 2020, Vol.8, pp: 220816 – 220839.

[2] V. Umayaparvathi and K. Iyakutti, "A survey on customer churn prediction in telecom industry: Datasets, methods and metrics," International Research Journal of Engineering and Technology (IRJET), vol. 3, no. 04, 2016.

[3] A.Perisic and M. Pahor, "RFM-LIR feature framework for churn prediction in the mobile games market", *IEEE Transactions on Games*, March 2021, doi: 10.1109/TG.2021.3067114.

[4] P. Lalwani, M.K. Mishra, J.S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach", Computing, February 2021, doi: 10.1007/s00607-021-00908-y.

[5] M. Chandar, A. Laha, and P. Krishna, "Modeling churn behavior of bank customers using predictive data mining techniques," in National conference on soft computing techniques for engineering applications (SCT-2006), 2006, pp. 24–26.

[6] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data, vol. 6, no. 1, pp. 1–24, 2019.

[7] SHULI WU, WEI-CHUEN YAU, THIAN-SONG ONG, and SIEW-CHIN CHONG, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business", IEEE Access, April 2021, Vol.9, pp: 62118 – 62136.

[8] M. Dash and H. Liu, "Feature selection for classification," Intelligent Data Analysis, vol. 1, no. 3, pp. 131–156, 1997.

[9] Z. Zhao and H. Liu, "Semi-supervised feature selection via spectral analysis." in Proceedings of the 7th SIAM International Conference on Data Mining, 2007, pp. 641–646.

[10] Y. Liu and Y. Wang, "Unsupervised discriminative feature selection in a kernel space via l 2, 1-norm minimization," in Pattern Recognition (ICPR), 21st International Conference on, 2012, pp. 1205–1208.

[11] D.Cai, C. Zhang, and X. He, "Unsupervised feature selection for multi-cluster data," in Proceedings of the 16th ACM

SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010, pp. 333–342.

[12] Ferdi Sarac, Development of unsupervised feature selection methods for high dimensional biomedical data in regression domain, 2017, PhD Thesis, Northumbria University, UK.

[13] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. N. Vapnik, Support Vector Regression Machines, ser. Advances in Neural Information Processing Systems. MIT Press, vol.9, 1996.

[14] V. N. Vapnik, “An overview of statistical learning theory,” Neural Networks, IEEE Transactions on, vol. 10, no. 5, pp. 988–999, 1999.

[15] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no. 3, p. 27, 2011.

[16] IBM Telco Customer Churn Data Set, available at: https://www.kaggle.com/blastchar/telco-customer-churn.” [Last accessed 10 July 2021].

Authors' background

Your Name	Title*	Research Field	Personal website
Ferdi Sarac	Dr	Machine Learning	
Huseyin Seker	Dr	Data Science and Machine Learning	www.smartdatacrew.com
Marcin Lisowski	Mr		www.bubo.ai
Alan Timothy	Mr		www.bubo.ai

***This form helps us to understand your paper better, the form itself will not be published.**

***Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor**