Application of Data Mining in Term Deposit Marketing

Q.R. Zhuang, Y.W. Yao, and O. Liu

Abstract—**Term deposits are facing challenges from both economic pressure and marketing competition. There are a number of valuable studies concerning bank and deposit marketing. These studies mentioned the significance of customers and customer segmentation in bank and deposit marketing. However, problems like obsolescence of data, inadequate maps, lack of data and specific methods encounter in practical application of deposit market segmentation. This research adopts data mining techniques through SPSS Modeler to predict customers' term deposit subscription behaviors and understand customers' features to improve the effectiveness and accuracy of bank marketing.**

Index Terms—**data mining, customer segmentation, term deposits, SPSS Modeler**

I. Introduction

BANKING industry is an important sector of social economy. Bank sectors provide various products and services for clients. Deposits constitute one of the most traditional and fundamental operations of banks and meanwhile, deposits are a primary source of bank financing [1]. There are many types of deposit accounts and some major types, including checking accounts, savings accounts, term deposit accounts and money market deposit accounts [2]. This study will especially focus on term deposit accounts, because term deposit accounts provide bank sectors with the most stable sources of credit and profit. However, the global financial crisis in 2008 raised people's distrusts on banks and the suspiciousness result in deposits shrank [3]. In addition, due to the rapid development of capital market, the emergence of a large amount of financial intermediation and financial instruments provides more investment channels and opportunities for residents. Both economic pressure and marketing competition drive bank sectors to improve the effectiveness of marketing campaigns.

There are two typical marketing campaigns for companies to promote services and/or products, including mass campaigns and direct marketing [4]. Mass campaigns aim at general indiscriminate public and direct marketing campaigns are implemented with the target of a specific group. According to a study implemented by [5], positive responses to mass campaigns are less than 1%; conversely, direct marketing campaigns are more effective. As a result, this research will mainly concentrate on direct marketing campaigns of term deposit accounts.

Nevertheless, direct marketing might cause negative attitudes toward banks due to the intrusion of privacy. Therefore, pinpointing the target customer groups is the most important marketing strategy when adopting direct marketing campaigns. This study will predict customers' term deposit subscription behaviors and understand customers' features to improve the effectiveness and accuracy of bank marketing.

II. Literature Review

There are a number of valuable studies concerning bank and deposit marketing. Different recommendations are put forward from different marketing aspects based on qualitative methods or quantitative analysis. Data mining techniques have been widely applied in bank marketing as well. Wu came up the idea that the association rules can be applied in cross-selling of bank products and customer risk control [6]. However, many studies just compare the performance of different classification algorithms on predicting the success rate of bank marketing campaigns. For example, Moro, Cortez and Laureano used the rminer Package and R Tool to test three classification models (Decision Trees, Naïve Bayes and Support Vector Machines) and compare their performance through Receiver Operating Characteristic curve (ROC) and Lift curve analysis [3]. Similarly, Moro, Cortez and Rita tested four data mining models, including logistic regression, decision trees (DT), neural network (NN) and support vector machine [7]. After evaluating area of the receiver operating characteristic curve (AUC) and area of the LIFT cumulative curve (ALIFT), neural network presented the best performance. Nachev combined cross-validation and multiple runs to partition the data set into train and test sets [8]. He also explored the impact of performance caused by different neural network designs.

All researches above focus on predicting customers' behaviors resulting from bank marketing. In order to avoid marketing campaigns being annoying rather than attractive, the right promotional messages should be delivered to right customer groups. As early as 1974, Robert put forward the idea of the use of census data in bank marketing [9]. He mentioned that census data can be applied into location analysis and marketing segmentation. Wang, Song and Fang

Manuscript received January 1, 2018; revised February 1, 2018.
Q.R. Zhuang is with the International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Suzhou, China. (e-mail: Qiuran.Zhuang12@student.xjtlu.edu.cn).
Y.W. Yao is with Department of Accountancy, Hang Seng Management College, Hong Kong. (e-mail: yiweiyao@hsmc.edu.hk).
O. Liu is with the International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Suzhou, China. (corresponding author, e-mail: Owen.Liu@xjtlu.edu.cn).

ISBN: 978-988-14048-8-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

mentioned that the banking industry lacks scientific marketing management and banks generally adopt some traditional marketing methods, including relationship marketing (use employees' personal relationship to find deposit clients), self-interest marketing (obtain deposits by satisfying clients' individual interests, such as gifts), passive marketing (attract customers to increase deposit by offering warm and thoughtful counter service) and simple service marketing (attract deposits by meeting the low-level requirements of customers, such as providing door-to-door services) [10]. They came up with the idea that carrying out market segmentation of deposit marketing and selecting the marketing target is the scientific way of marketing management. However, problems like obsolescence of data, inadequate maps, lack of data and specific methods encounter in practical application of deposit market segmentation.

This study will adopt data mining techniques to predict customers' term deposit subscription behaviors and understand customers' features to improve the effectiveness and accuracy of bank marketing. In order to achieve this objective, the following questions will be addressed.

I. How to predict whether a bank client will subscribe to a term deposit or not?

II. Which determinants would indicate a client is ready to subscribe to a term deposit through direct marketing?

III. How to segment term deposit market?

IV. Are there any common features of clients who have subscribed to a term deposit?

III. Methodology

In this research, classification models and clustering models will be built through SPSS Modeler. A number of machine learning algorithms and modeling techniques are included in IBM SPSS Modeler for different types of problems solving.

Classification algorithms are used to establish predictive model by learning and discovering the relationship between a set of feature variables and a target variable. Two phases are typically contained in classification algorithms [11]. In the first phase, models are constructed from the training instance. In the second phase, unlabelled testing instances can be predicted and assigned through the model established in the training phase. Several indicators are typically used to evaluate the performance of a binary classifier. For example, accuracy is used to describe outcomes that are predicted correctly. Moreover, AUC is the area under the ROC (Receiver Operating Characteristic) curve, which is a probability [12]. Furthermore, Gini coefficient is related to AUC that Gini=2*AUC-1. A Gini coefficient above 60% corresponds to a good classification model.

Clustering algorithms are applied to customer segmentation. Instances can be divided into natural groups through clustering techniques, which is an unsupervised learning scheme [13]. Instances with strong resemblance will be in the same cluster. There are different types of clustering algorithms, including portioning approaches, hierarchical methods, density-based methods, grid-based methods, model-based methods, etc. The quality of clustering algorithms can be evaluated by the average silhouette coefficient of all instances [14]. A higher silhouette coefficient indicates that the instances are better matched to its own clusters.

IV. Data Understanding

A secondary dataset related to direct marketing campaigns on term deposit accounts of a Portuguese banking institution is obtained from the Internet [15]. The dataset contains 41188 observations and 21 variables. The detailed attribute information is shown in the table below.

TABLE I
ATTRIBUTE INFORMATION

Name	Data type	Description
	Bank Client Data	
age	numeric	age
job	categorical	type of job
marital	categorical	marital status
education	categorical	education background
default	categorical	has credit in default?
housing	categorical	has housing loan?
loan	categorical	has personal loan?
	Contact/ Campaign Data	
contact	categorical	contact communication type
month	categorical	last contact month of year
day_of_week	categorical	last contact day of the week
duration	numeric	last contact duration, in seconds
campaign	numeric	number of contacts performed during this campaign and for this client
pdays	numeric	number of days that passed by after the client was last contacted from a previous campaign
previous	numeric	number of contacts performed before this campaign and for this client
poutcome	categorical	outcome of the previous marketing campaign
	Social and Economic Context Attributes	
emp.var.rate	numeric	employment variation rate - quarterly indicator
cons.price.idx	numeric	consumer price index - monthly indicator
cons.conf.idx	numeric	consumer confidence index - monthly indicator
euribor3m	numeric	euribor 3 month rate - daily indicator
nr.employed	numeric	number of employees - quarterly indicator
	Output Variable	
y	binary	has the client subscribed a term deposit?

V. Modeling

Classification and clustering models are established on the processed data.

A. Classification

Classification algorithms are used to establish a predictive model of whether a client will subscribe to a term deposit or not. Auto Classifier node of SPSS Modeler enables to automatically create and compare multiple different classification models. As a result, C5.0 model shows the best performance with the highest accuracy.

ISBN: 978-988-14048-8-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Therefore, a boosted C5.0 model is built to further improve the performance of the C5.0 model. Figure 1 presents that the boosted C5.0 improve the accuracy of the model to 97%. In addition, both AUC and Gini coefficient indicate that the boosted C5.0 classifier generate a more accurate classification results and a better predictive model.

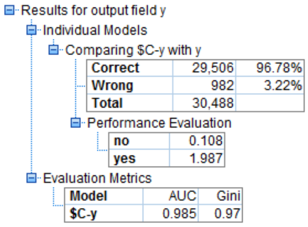

Fig. 1. Boosted C5.0 Model

In this boosted C5.0 classification predictive model, contact duration ('duration') is the most important predictor. Combining the rules generated by the boosted C5.0, one of the rules predicts that if the contact duration is no more than 77 seconds, the clients will not subscribe to a term deposit.

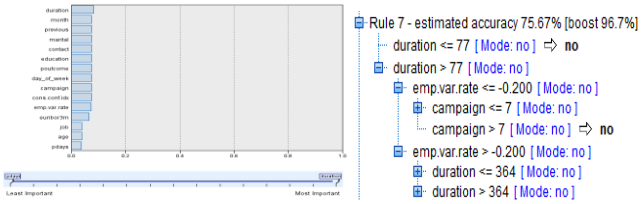

Fig. 2. Significant Predictor

B. *Clustering*

Clustering algorithms are applied to segment clients who have subscribed to a term deposit. Therefore, the dataset is filtered by the condition that y= yes, which includes 3859 instances. In order to discover and understand customers' behaviors and characteristics, social and economic context attributes and output variable will not be used to generate clusters. Because according to the classification results, economic context attributes have less impact on clients' deposit subscription behavior.

Firstly, Auto Cluster node is attached to compare different clustering models. Two clusters are automatically created by TwoStep mode with the highest silhouette coefficient (0.389). Significant differences between the two clusters are displayed in Table Ⅱ.

TABLE Ⅱ
DISTINCT ATTRIBUTES OF TWO CLUSTERS

Attribute	New Customer (Cluster 1)	Regular Customer (Cluster 2)
pdays	999 (not previously contacted)	281.96
poutcome	nonexistent	success
previous	0.05	1.58
duration	long(598.74)	short(370.56)
campaign	more(2.16)	less(1.72)
contact	telephone	cellular

Clients in the first cluster were not previously contacted and there was basically no marketing campaign offered for clients in this cluster. Clients in the second cluster were in the opposite situation. Therefore, the TwoStep model segments clients into two clusters: new customers and regular customers. The duration of contact with new customers is much longer than regular customers. Meanwhile, more number of contacts are performed for new customers during the marketing campaign with the average of 2.16 times. Furthermore, usual communication type for new customers is telephone and commonly used communication type for regular customers is cellular.

However, it seems that there is no significant difference between the two clusters when comparing the bank client attributes. As a result, clustering techniques are re-applied in bank client data to only focus on the customers' characteristics. Five clusters are generated by K-means clustering algorithms with the highest silhouette of 0.37. Comparing the five clusters, differences between each cluster can be summarized as the distinct attributes shown in this table.

TABLE Ⅲ
DISTINCT ATTRIBUTES OF FIVE CLUSTERS

Clusters	Distinct Attributes
Cluster 1	No housing loan; Married or divorced; Age: 43
Cluster 2	4-year basic education background; Retired, housemaid; Divorced or married; Age: 66
Cluster 3	Single; No housing loan; Student; Age:30
Cluster 4	Single or divorced; Housing loan; administration; Age: 33
Cluster 5	Married; Housing loan; Entrepreneur, management; Illiterate; Age: 40

In the first cluster, people with stable lives and income have no pressure of housing loan. They may want to subscribe to a term deposit to advance prepare for their future retirement or for their children. In the second cluster, people are mainly retired or housemaid. Their jobs, ages and education background may lead to their habit of adopting risk-averse investment and preference of saving in banks. In the third cluster, typical groups are students who are about to step into the society and they may start to subscribe to a deposit to prepare for their future life. In the fourth cluster, major groups are young people who have jobs and housing loan as well. The pressure of lives may drive them to choose a more secure way to invest, which is saving in banks. In the fifth cluster, people are generally not well educated and they become entrepreneurs or managers at their middle age. People in this cluster may have experienced challenges and understand lives are not that easy. They may like subscribing to a term deposit to cherish their gains.

VI. CONCLUSION

To conclude, term deposits of bank sectors are facing the challenges from both economic pressure and marketing competition. This study adopts data mining techniques to predict customers' term deposit subscription behaviors and understand customers' features to improve the effectiveness and accuracy of bank marketing.

The results generated by the application of classification algorithms and clustering algorithms have practical meaning for the objectives of this research. Some feasible suggestions are put forward as followings. Firstly, marketing staffs should be patient when implementing direct marketing, especially telemarketing for new customers. The contact duration has significant impact on the success rate of telemarketing. Secondly, the number of contacts performed during the marketing campaign should be controlled. It is better to control the number of contacts less than 3 times; otherwise, too frequent contacts may cause aversion. Thirdly, it is better to call customers' telephone numbers (such as office number) rather than their cellular to try to avoid the feeling of intrusion of privacy, especially telemarketing for new customers. Fourthly, bank sectors can launch targeted marketing campaigns to attract specific customers in accordance with the results of clustering, such as children's growth deposit scheme, retirement term deposit scheme, student term deposit scheme, housing loan deposit scheme, etc.

REFERENCES

[1] Khir, K., Gupta, L., & Shanmugam, B. (2008) 'Islamic banking: A practical perspective'.

[2] Islam, M.A. and Ghosh, P. (2014) 'A comparative analysis of deposit products in banking industry: an opportunity for eastern bank Ltd.', Journal of Investment and Management, 3(1), January, pp.7-20.

[3] Moro, S., Cortez, P. & Laureano, R. (2013) A data mining approach for bank telemarketing using the rminer package and r tool [Online]. Available from: https://www.researchgate.net/publication/256464440_A_data_mining_approach_for_bank_telemarketing_using_the_rminer_package_and_r_tool?enrichId=rgreq-ef9c19b19ab77f6e64e62c02ff6bdc5c-XXX&enrichSource=Y292ZXJQYWdlOzI1NjQ2NDQ0MDtBUzoxMTkyMzAyMzgzMDIyMTFAMTQwNTQzODExMjgxNQ%3D%3D&el=1_x_2&_esc=publicationCoverPdf (Accessed: 4 September 2017).

[4] Ling, X. and Li, C. (1998) 'Data Mining for Direct Marketing: Problems and Solutions'. Proceedings of the 4th KDD conference, AAAI Press, pp.73–79.

[5] Ou, C., Liu, C., Huang, J. & Zhong, N. (2003) 'On Data Mining for Direct Marketing'. Proceedings of the 9th RSFDGrC conference, 2639, pp.491–498.

[6] Wu, Q.H. (2008) 'Some Issues with Applying Association Rules in Commercial Bank', Journal of System Simulation, 20 (8), April, pp.2206-2209.

[7] Moro, S., Cortez, P. & Rita, P. (2014) A Data-Driven Approach to Predict the Success of Bank Telemarketing [Online]. Available from: https://pdfs.semanticscholar.org/4a27/709545cfa225d8983fb4df8061fb205b9116.pdf (Accessed: 14 September 2017).

[8] Nachev, A. (2015) Application of data mining techniques for direct marketing [Online]. Available from: http://www.foibg.com/ibs_isc/ibs-30/ibs-30-p09.pdf (Accessed: 14 September 2017).

[9] Predue, R.T. (1974) 'SOME TYPICAL USES OF CENSUS DATA IN BANK MARKETING RESEARCH', Review of Public Data Use, 2(2), pp.31-36.

[10] Wang, B.Z., Song, J.L., & Fang, C. (2002) 'Opinions on Deposit Marketing of Commercial Banks', Financial Theory and Practice, 2002 (9), August, pp.32-33.

[11] Aggarwal, C.C. (2015) 'Data Classification Algorithms and Applications', *CRC Press*, EBSCOhost [Online]. Available from: http://10.7.1.204:81/read.php?resid=99673992 (Accessed: 25 November 2017).

[12] Fawcett, T. (2006) 'An introduction to ROC analysis', Pattern recognition letters, 27(8), pp. 861-874.

[13] Witten, I., Frank, E. & Hall, M.A. (2011) Data Mining – Pratical Machine Learning Tools and Techniques. Burlington: Elsevier.

[14] Chen, X. and Li, Z. (2013) 'Effectiveness Analysis of The Application of Clustering in Student Grouping', *International Conference on Education Technology and Information System*, Atlantis Press, pp.988-991.

[15] UCI Machine Learning Repository (2014). Bank Marketing Data Set [Online]. Available from: http://archive.ics.uci.edu/ml/datasets/Bank+Marketing (Accessed: 4 September 2017).

ISBN: 978-988-14048-8-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)