A Two-step Bayes Method for Spatial Drift in Consumer Classification

Liancai Hao
School of Management, Harbin Institute of Technology, P. R. China, 150001
Email: haolc@hit.edu.cn

Peng Zou
School of Management, Harbin Institute of Technology, P. R. China, 150001
Email: zoupeng@hit.edu.cn

Yijun Li
School of Management, Harbin Institute of Technology, P. R. China, 150001
Email: liyj@hit.edu.cn

***Abstract*—In the field of data mining, spatial drift refers to the data used to develop the model consists of only one part of the population, and the differences among the samples or between sample and population are unknown. This paper proposes a two-step Bayes method to improve adaptability for different region samples, which also maintains high model accuracy. The new method first groups region based on similarity, second, sets a model structure without parameters for populations or large samples with good data quality, and then trains parameters using samples in same region group. This method builds a estimation model, proving the method by showing how it can to some extent solve the uncertainty of consumer classification.**

***Index Terms*—spatial drift, consumer classification, bayes network**

I Introduction

Consumer classification refers to that firms classify consumers by their attributes, behaviors, preferences, values, etc. al [1]. However, the original rules can not adapt to encompass new samples, a phenomenon referred to concept drift. This issue has been studied under a variety of names [2]. Concept drift is the difference of distribution of population sample [3]. In on-line learning, concept drift describes the problem of changes occurring during the learning process, while statistics uses concept drift to refer to "population drift" [4].

Concept drift impacts data mining in two ways, one is temporal drift due to population are not static, specifically, consumer behavior changes over time. The other one is spatial drift. This drift exists because the data for training model consists of only partial population, and the differences among the samples are unknown [5–7]. In statistics, the minimum estimation variance (MEV) measures the differences between samples [8]. For example, we can have customer data from Shenzhen (a city in southern China) as one sample, while the customer data of Wuhan (a city in central China) is another sample. "Shenzhen" and "Wuhan" are the labels of the two samples.

Spatial drift limits the extent use of patterns or rules determined through data mining. In this instance, a model derived from one consumer sample can not be used to estimate other consumer sample [1] because the differences in consumer characteristics in different regions lead to differences in classification. These unstable results prevent firms from developing reliable marketing strategies. A significant number of studies have analyzed this problem [9]. Michael found that, when building a model estimating regional consumer responses to marketing promotion, significant errors occurred if a model on some regional consumers was applied to estimate the of consumers' responses in other regions [10]. How to solve the difference uncertainty between the two samples and predict consumer behavior in the various regions still remains.

Our basic algorithm process is from [11]. The improvement work is how to choose samples for estimating model structure and parameters. In literature [11], samples for estimating model structure and samples for estimating model parameters are chosen randomly as long as they have same feature vectors, but in practice, there are still so many sample sets (i. g. Beijing, Shanghai and other cities) that have same feature vectors and no criteria to pair samples for estimating model structure and samples for estimating model parameters. Inspirited by literature [12] and [13], we implement a more accurate method to pair samples for estimating model structure and samples for estimating model parameters, which first groups regions based on similarity in the composition of segmentation, and then we choose samples for estimating model structure and samples for estimating model parameters in each city group. The new method can not only reduce the searching cost of pair of samples for estimating model structure and samples for estimating model parameters, but also enhance the accuracy of segment estimation. The validation test shows that our

doi:10.4304/jcp.8.11.2966-2971

improving way gets better result.

II LITERATURE REVIEW

Recent works have focused on spatial drift. Schlimmer and Granger induce the Stagger algorithm to identify drift [8]. Kevin B. Pratt designed a system to effectively recognize the drift in data streaming [14], while Gerhard proposed the influential FLORA2 algorithm to study drift [15]. Ralf Klinkenberg studied drift in information filtering systems [16–19]. Klinkenberg suggests that the information filtering system must adapt to these changes to find the real interests of those customers. In addition, drift phenomenon has been studied for fields such as stocks, weather prediction, changing consumer interests, and illness diagnosis [20].

Literatures propose alternative ideaes to deal with spatial drift. One is that Modeling for each sample. Li Jian-ping presented a credit scoring method via a principal component analysis linear-weighted comprehensive assessment. The merit of this method is that it easily adapts to the different economic & cultural environments in different regions, and to drift in the same region, causing disparity in credit scoring [21]. However, this way is not feasible. The cost of modeling process is too prohibitive, even for models with more granular datasets. If the analysis dimensions are low granularity—such as, at the city level—the modeling cost becomes much higher.

Li, Xu, and Shi (2004) proposed market segmentation individually for each city by dissimilarity of customer segmentation in various cities when solving multi-city problems [22]. However, it is time consuming and ineffective. it is hard to generalize the results for application to other cities; thus the marketing strategies devised for one city cannot be directly applied to another city. Consequently, the method has limited applicability.

The other one is modeling the population. Arbia proposed a method to determine sub-optimal, model-based area sampling strategies that sequentially selects sampling locations, with each selection minimizing the estimation error with respect to already selected sites. This means that, given a super population model and an optimal unbiased linear homogeneous estimator model of the population mean, the MEV algorithm returns a purposive design, i.e. a sequential procedure that develops through various steps [8]. The critical element of this algorithm is to determine a proper sample size.

The shortcoming of the second strategy is that the model on the population. But the model presents the distribution of whole population, and sometime differs from particular sample. This difference can arise because a sample may be just a small part of the whole data, and the population model can hardly describe a pattern for just part of the data. Therefore, the difference between the population model and the pattern of the part makes the estimation for a specific sample fail.

Previous research assessing spatial drift in market segmentation proposed a two stage framework [23][24]. Both studies first grouped regions according to demographics and economic figures, then found consumer segmentations in each country group. The methods assumed that demographics and economic figures are important and valid factors for grouping countries. Unfortunately, no study presents clear connection between the countries. Moreover, some of the figures may be costly or hard to acquire. Bijmolt et al. (2004) proposed a multilevel latent class analysis for attaining simultaneous country and consumer segments for financial product ownership. Domain-specific behavioral variables such as product usage and ownership were then used to form both country and consumer segments (i.e. the grouping of countries) was based on the composition of consumer segments. In other words, two countries were clustered into the same segment if they had similar within-country structures of consumer segmentation [9].

Enlightened by these works in two stage modeling, this paper improves the algorithm to measure the difference among samples. The tactic used in this work for accounting for spatial drift is to set a model on part of the data at the first step and then modify the model by a new sample to get model estimating new sample.

Consumer classification generally labels classes according to their characters. Customer credit assessment is typical classification used for customer credit records, so Bayes [25] and nearest-neighbor algorithm [26][27] used for credit assessment can supply some reference.

Bayes method is competitive. One reason is that it can solve for uncertain classifications, based on the likely characteristics of individuals from different classes. Note, however, that the classes cannot exclude each other absolutely. The Bayes method can describe the uncertainty of classification through probability, and choose the appropriate class based on the greatest posterior probability. Another is that the Bayes model can effectively use prior knowledge and sample data information. The Bayes method invokes two important concepts: 1) prior probability is probability from history, which does not need to be verified; and 2) posterior probability is the probability modified by Bayes learning and verified by data [28].

A Bayes network is a graphical pattern that represents the dependent probability of variables, which can represent consequence and find potential relationships in data. In a Bayes network, nodes are variables and directional lines represent dependence. If network G〈S，P〉is composed of topological structure S and the set of probability distribution P, then S is a set of connections of nodes and P is a set of parameters measuring the net. Training a Bayes network consists of two factors. The training structure consists of searching a network structure to represent the dependent relationship among variables that most exactly matches a data set, while the training parameter consists of computing the dependent probability of the variables [29].

A data sample can relate to other samples through a combination of prior knowledge and sample information. For example, though the estimation matching the sample of a certain region is not suitable for other regions, the

attribute relationships of the different samples have some similarity because of similar economic, cultural, and etc. al. The main difference is in the intensity of the dependent relationships between some attributes.

The nearest-neighbor algorithm is a non-parameter method, which can give individuals who need assessment the same class label as individuals whose attributes are most similar to the individual. By adding new individuals and deleting individuals out of date, the algorithm can update the assessment system to overcome population temporal drift. Our hypothesis is that by using new data to modify the model developed for existing data, we can use this same algorithm to compensate for spatial drift. However, the algorithm does suffer from excessive computing requirements and inaccuracy.

III Spatial Drift-Oriented Bayes Algorithm

A. Idea Description

The problem is as follows：

(1) Every sample set is labeled by an n-dimensional vector, X=(x1, x2, … xn), depicting n attributes, i.e., A1, A2, … An.

(2) Given a set V1 and m classes c1, c2 … cm, build classifier G<S, P1> presenting the relation between X and class on V1, of which the output is the network structure S with nodes A1, A2,… Ak（k<n).

(3) Given another set V2 and m classes c1, c2 … cm (having the same feature vector as V1), the task is to estimat which class x (in V2 and having no class label) should belong to.

Given a Bayes network structure S fits V1, it is prior knowledge of the classifier for V2. Then V2 can train the parameters of S to output G'(S, P2), of which P2 is a parameter set of G' learning from V2. At this point G' will assign x to the class ci having the highest posterior probability.

Our basic algorithm process is from [11]. The improvement work is how to choose V1 and V2. In [11], V1 and V2 are chosen randomly as long as they have same feature vectors, but in practice, there are still so many sample sets that have same feature vectors and no criteria to pair V1 and V2. Inspirited by [12], we implement a more accurate method to pair V1 and V2, which first groups regions based on similarity in the composition of segmentation, then we choose V1 and V2 in each city group. The new method can not only reduce cost of searching the pair of V1and V2, but also enhance the accuracy of segment estimation. The validation test shows that our improving way gets better result. This algorithm process is described in Fig. 1.

$$C_{i\,MAP} = \arg\max P\left(X_{c_i}\right)$$

B. Learning Bayes Network Structure

The first step in the proposed method is building a classifier on the sample. The algorithm to learn the structure of the Bayesian network of V1, is as follows:

Procedure
begin
Input: $V_1=\{X_1; \ldots X_n; C_m\}$
//available variables and class label//
Output: *G* //Bayes network classifier//
Procedure Learning Bayes network structure
begin
Initialize network to simple Bayes
Evaluate the current network
while classification accuracy improves
Consider adding every legal arc to the current structure of the classifier and evaluate the classifier
if there is an arc addition which improves the accuracy
then
Add the arc giving the largest classification accuracy improvement among all the possible arcs to the current network
else
Return current network
end
end
end

Each possible arc from Xi to Xj (Xi≠ Xj, Xj is a node without another attribute as parent) is evaluated within the while-loop. If the arc enhances the classification accuracy estimate with respect to the current network, then the algorithm updates the current network with that arc giving the largest improvement. Otherwise, if no arc results in an enhancement of the classification performance, the algorithm returns the current classifier. This algorithm concludes the search if there is no arc resulting in an enhancement of the classification accuracy estimate, or if no attribute without another attribute as parent is available.

C. Learning Bayes Network Parameters

The algorithm can process parameter learning, i.e. computing the class conditional probabilities table (CPT) for each node variable in the network, after learning the Bayes network structure. These probabilities can be estimated from the sample V_2. The CPT for variable Z specifies the conditional distribution $P\left(Z\middle|Parents(Z)\right)$, where *Parents(Z)* are the parents of *Z*. the joint probability of any tuple $(z_1, \ldots z_n)$ corresponding to the variables or attributes $Z_1, \ldots Z_n$ is computed by

$$P\left(z_1,\ldots z_n\right)=\prod P\left(z_i\middle|Parents\left(Z_i\right)\right)$$

Thus, we get the classifier $G'(S, P_2)$, which can estimate the class that *x* should belong to.

IV The Example

The goal in this example is classifying customers using an existing data set of customer profitability, and then predicting the customer's unknown profitability according to their features. The data consists of customer records from the data warehouse of a Chinese commercial bank. We first group cities as Mo etc. al (2010) method, we find Shenzhen and Wuhan are in one group, and then randomly selected 47,000 customers from Shenzhen City and 18,000 customers from Wuhan

city, all from the year 2004. **Table 1** provides an example of the original data, while **Table 2** shows pre-processed data (including discrete, generalized, and deleted values). The sample from Shenzhen is divided into three parts: 16,000 records for training set, 15,000 records for validations set, and 16,000 records for test set.

We set up a Bayes classifier structure on the Shenzhen sample, and computed the parameters of the CPT required modifying the classifier on the Wuhan sample. Fig. 2 is the resulting Bayes classifier structure, which shows that the main relevant attributes are: A) education level, B) income, D) age, C) gender, and E) occupation. The learning parameters on the Wuhan customer sample are 600 training set records, 600 validation set records and 500 test set records, **Table 3** is the CPT for this dataset, and the modified classifier structure is provided in **Fig. 3**.

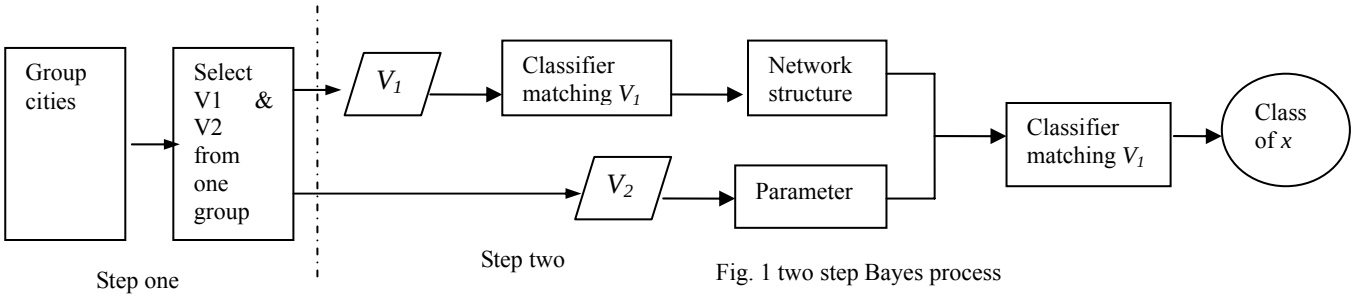

Fig. 1 two step Bayes process

TABLE1
SAMPLE OF THE ORIGINAL DATA

ID	Gender	Age	occupation	Income	Education level	…
13733624	male	34	manager	230000	M.A	
11905976	male	28	teacher	150000	B.A	…
11689270	female	36	government staff	80000	High middle school	…
…	…	…	…	…	…	…

TABLE2
SAMPLE OF THE PRE-PROCESSED DATA

ID	Gender	Age	occupation	Income	Education level	…
13733624	1	2	1	1	1	…
11905976	1	3	3	2	2	…
11689270	2	2	2	3	4	…
…	…	…	…	…	…	…

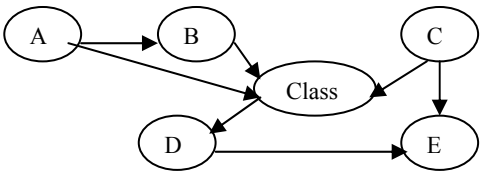

Fig. 2 Bayes classifier structure for Shenzhen sample

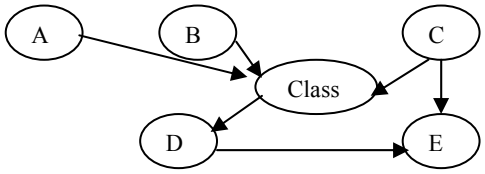

Fig. 3 Bayes classifier structure for the Wuhan sample

TABLE 3
CONDITIONAL PROBABILITIES TABLE

Class	P=(A=1 | B)	P=(A=2 | B)	P=(A=3 | B,)
1	0.1349	0.1157	0.7493
1	0.0445	0.09993	0.8562
1	0.1274	0.2274	0.6452
2	0.0476	0.9048	0.0476
2	0.0476	0.0476	0.9048
2	0.0049	0.0049	0.9902

We then used the modified classifier to estimate customer profitability of customers in the Wuhan sample, i.e. $x=(B=1,C=1,D=2,E=1)$, which is a data sample of V_2. The result is $P(Class=1)=0.85$, $P(Class=2)=0.15$, so Wuhan customers who are male, 30-40 years old, married, and have 20,000 yuan annual income represent a higher profit for the bank (Class=1) than other profit levels (Class=2).

We used a 5-fold cross validation to compare the

accuracy of our proposed Bayes classifier with that of other methods. We build classifier using the methods of [11][12][30], and contrasted the results with the results from this work. The results presented in **Table 4** shows that our new method is better that others.

Accuracy = （TP+ TN）/（TP+ TN+ FP+ FN）

TABLE 4
RESULTS OF CLASSIFICATION METHODS

method	TP	TN	FP	FN	accuracy
decision-tree	62	74	10	14	85%
Bayes	66	77	6	11	89.4%
two-stage clustering	64	75	8	13	86.9%
Two-step Bayes	69	78	3	10	91.9%

(TP: the number of estimated Class=1 customers who are Class=1; TN: the number of estimated Class=2 customers who are Class=2; FP: the number of estimated Class=1 customers who are Class=2; FN: the number of estimated Class=2 customers who are Class=1)

V CONCLUSIONS

The rapid growth of economy in China represents both an opportunity and a threat to banks in the country. How to correctly identify and target potential customers in such a large country with a diverse consumer base residing across different regions of the nation is a challeng. Previous research in marketing shows that using classifying rules derived from one area to another area creates unpredictable rule performance. However, it is too time consuming and costly to perform individual market segmentation for each city without the possible similarities in underlying consumer classification among different cities, especially for a country like China.

The method proposed in this paper improved model adaptability for different samples while maintaining a high accuracy. This new method also solves the problem of differences among populations and samples caused by spatial drift. Nowadays, the population features from various regions are quite diverse in China, but this proposed method is suitable for confronting this issue. However, more work is needed to improve the method's estimating ability in situations where there are more than two segments.

ACKNOWLEDGMENT

This research was partially funded by a research grant from the National Science Foundation of China under project No. 70802019, and through the Fundamental Research Funds for Central Universities (Grant No. HIT.NSRIF.2010079).

REFERENCE

[1] Bernard Liautaud, Mark Hammond. "*Business Intelligence*". Publishing House of Electronics Industry, 2002: pp186-205.

[2] G. Widmer. "Learning in dynamically changing domains: Recent contributions of machine learning" Perner Ed. *Proceedings of the MLNet Workshop On Learning In Dynamically Changing Domains: Theory Revision And Context Dependence Issues*, Prague, Czech Republic, 1997: pp 286-292.

[3] M.G. Kelly, D. J. Hand, N. M. Adams. "The impact of changing populations on classifier performance" Volkmar ed. Conference on Knowledge Discovery in Data, Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining., San Diego, California, 1999: pp 332-339.

[4] Liang Shi-dong, Fang Zhao-ben. "Consumer credit assessment analysis review". *System Engineering*, 2001, 11: pp 9-15.

[5] Arbia G., Lafratta G. "Anisotropic spatial sampling designs for urban pollution". J. Roy. *Statist. Soc. C Ser*. 2002, 51 (2), pp 223–234

[6] Ferguson C.C. "Techniques for Handling Uncertainty and Variability in Risk Assessment Models". Umweltbundesamt. Berlin 1998.

[7] Juang, K.W., Chen, Y.S., Lee, D.Y., "Using sequential indicatorsimulation to assess the uncertainty of delineating heavy-metal contaminated soils". *Environ. Pollut*. 2004, 127: pp 229–238

[8] J.C. Schlimmer, R.H.Granger. "Incremental learning from noisy data". *Machine Learning*, 1986, 13(3): pp297-309.

[9] Bijmolt, H. T., Paas, L. J., & Vermunt, J. K. "Country and consumer segmentation: Multi-level latent class analysis of financial product ownership". *International Journal of Research in Marketing*, 2004, 21(4), pp 323–340.

[10] J. Michael, G.S. Linoff. "*Data Mining Techniques: for MarketingSales, and Customer Relationship Management*". New York: John Wiley & Sons, Inc, 2004, 120-150.

[11] Li, Yi-jun, Peng Zou and Qiang Ye. "Customer Sample Difference-oriented Bayes Segmentation Algorithm" *Proceeding of Management Science and Engineering*, Harbin, China 2006

[12] Jiahui Mo, Melody Y. Kiang, Peng Zou, Yijun. "A two-stage clustering approach for multi-region segmentation". *Expert Systems with Applications* 37 2010, pp 7120–7131

[13] Ping Ling, Xiangsheng Rong, Xiangyang You, Ming Xu. "Novel Three-Phase Clustering based on Support Vector Technique" *Journal of Software*, 2013, 8 (4), 955-962.

[14] K. B. Pratt, G. Tschapek. "Visualizing concept drift Petra ed. Conference on Knowledge Discovery in Data", *Proceedings of the Ninth ACM SIGKDD* International Conference on Knowledge Discovery and Data Mining, 2003: pp412-419.

[15] G. Widmer, M. Kubat. "Learning flexible concepts from streams of examples: FLORA2" Fink ed. *Proceedings of the European Conference on Artificial Intelligence*.1992: pp189-195.

[16] R. Klinkenberg, S. Ruping. "Concept drift and the importance of examples" Wysotzki F. ed. *Text Mining — Theoretical Aspects and Applications*, PhysicaVerlag, Heidelberg, Germany, 2003:78-84.

[17] R. Klinkenberg. "*Learning drifting concepts with partial user feedback*" Beiträge zum Treffen der ed. GIFachgruppe Maschinelles Lernen, 1999:31-36.

[18] R. Klinkenberg, I. Renz. "*Adaptive information filtering: learning drifting concepts*" Geibel P. ed Germany, Fachbereich Informatik, TU Berlin, 1998.

[19] R. Klinkenberg. "Meta-Learning, model selection, and example selection in machine learning domains with concept drift". Schadler K. ed. Annual workshop of the special interest group on machine learning, knowledge

discovery, and data mining of the German Computer Science Society (GI), Saarbrucken, Germany, 2005, pp79-84.
[20] H.Wang, W. Fan, P.S.Yu, J. Han. "Mining concept-drifting data streams using ensemble classifiers". Johannes and Grieser ed. Proc. 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, ACM Press, 2003:pp1002-1007
[21] LI Jian-ping, XU Wei-xuan, SHI Yong. "Credit Scoring & Principal Component Analysis Linear-weighted Comprehensive Assessment and Application". *System Engineering*, 2004, 8: pp64-68.
[22] Li, J. P., Xu, W. X., & Shi, Y. "Credit scoring and principal component analysis linear-weighted comprehensive assessment and application". *System Engineering*, 2004, pp64–68.
[23] Kotabe, M., & Helsen, K. *Global marketing management* (2nd ed.). New York: Wiley, 2001.
[24] Hofstede, T. F., Steenkamp, J. B., & Wedel, M. Identifying spatial segments in international markets. Marketing Science, 2002, 117–160.
[25] Zhang Zhong-xiu. "*Personal credit*". Beijing: Publishing House of Chinese Union of Industry and Commerce, 2002 (in Chinese)
[26] Henley W. E, Hand D. J. "A k-nearest-neighbor classifier for assessing consumer credit risk. Statistician," 1965, 45: 77- 95.
[27] Ping Ling, Xiangsheng Rong, Xiangyang You, Ming Xu. "Vector-Distance and Neighborhood Development for High Dimensional Data". *Journal of Software*, 2012, 7 (12), 2832-2839
[28] Shi Zhong-zhi "*Knowledge discovery*". Beijing: Pubishing House of tsinghua, 2002.
[29] Tom M. Mitchell: "*Machine Learning*". McGraw-Hill Companies, Inc, 1997
[30] Xinmeng Zhang, Shengyi Jiang. "A Splitting Criteria Based on Similarity in Decision Tree Learning" *Journal of Software*, 2012, 7,(8), 1775-1782

Liancai Hao: Doctoral student in the school of management, Harbin Institute of Technology.

Peng Zou is Associate professor in the school of management, Harbin Institute of Technology.

Yijun Li is professor in Harbin Institute of Technology.