

Customer Profile Classification Using Transactional Data

Edward T. Apeh^{*†}, Bogdan Gabrys^{*‡}, Amanda Schierz^{*§}

^{*}Smart Technology Research Centre, School of Design, Engineering and Computing

Bournemouth University

Dorset BH12 5BB

Email: [†]eapeh@bournemouth.ac.uk,

[‡]bgabrys@bournemouth.ac.uk, [§]aschierz@bournemouth.ac.uk

Abstract—Customer profiles are by definition made up of factual and transactional data. It is often the case that due to reasons such as high cost of data acquisition and/or protection, only the transactional data are available for data mining operations. Transactional data, however, tend to be highly sparse and skewed due to a large proportion of customers engaging in very few transactions. This can result in a bias in the prediction accuracy of classifiers built using them towards the larger proportion of customers with fewer transactions. This paper investigates an approach for accurately and confidently grouping and classifying customers in bins on the basis of the number of their transactions. The experiments we conducted on a highly sparse and skewed real-world transactional data show that our proposed approach can be used to identify a critical point at which customer profiles can be more confidently distinguished.

Index Terms—Data mining, data preprocessing, decision support systems, classification algorithms, industry applications.

I. INTRODUCTION

The problem of identifying different types of customers in order to adequately meet their product/service needs is of great importance to businesses who want to remain competitive in the current economic climate.

Algorithms from data mining have been used extensively for explorative and predictive purposes to model business problems. For instance, clustering techniques are often the first techniques used in market segmentation [1, 2] while classification techniques such as K -NN have been used in customer profile personalization in on-line retail systems [3, 4].

Using a data mining algorithm to discover the best model for a business problem involves processing historical data with the goal of identifying the relevant independent variables which will best minimize the error for predicting unseen future instances.

Classification techniques, which are concerned with arranging the data into predefined groups are often

used to perform the predictive data mining task. The classification process essentially is characterized by well-defined classes, and a training set consisting of pre-classified examples. The task is to build a model that can be applied to confidently and accurately classify future data instances.

Pre-processing techniques are used in almost all data mining applications to efficiently perform the data mining task. The decision to apply a pre-processing technique may be driven by the need to generate a model from a dataset that is too large to process in full (data reduction) [5], handle missing values/inconsistent data (data cleansing) [6], combine data from multiple sources into a coherent store (data integration) [7], and normalize data so that it can be more efficiently processed (data transformation) [8].

This paper presents an investigation of a data mining approach that combines the unsupervised data binning pre-processing technique with classification to identify different types of customer profiles using their transactions.

Customers with sparse transactions, which tend to make up the bulk of transactional data, are difficult to distinguish and accurately classify. This problem is even more pronounced when the sparse transactions are mixed with dense transactions, as the classifier performance tends to be biased towards the larger number of customers with sparse transactions.

Our proposed approach groups customer profiles into bins on the basis of the number of items transacted so as to more accurately and confidently classify a customer given their transactions.

The rest of this paper is organized as follows: Section II describes the problem of constructing customer profiles from transactional data, and provides background knowledge of our work. Section III gives a description of our proposed approach while Section IV

gives a description of the data used for the proof-of-concept experiments as well as the experiment evaluation measure used. Section V presents an analysis of the results obtained from the experiments. The paper concludes in Section VI with directions for future work.

II. PROBLEM DESCRIPTION AND BACKGROUND KNOWLEDGE

A customer profile is defined as a collection of data describing an individual user or a group of users. There are three main types of customer profiles; factual, rule-based and collaborated/verified [9, 10].

The construction of profiles for individual customers is a major concern for businesses who desire to build and effectively manage the relationship they have with their valued customers. Well constructed customer profiles provide businesses with vital information such as [9]:

- 1) who their valuable customers are, and
- 2) how they behave.

The information used for building a customer profile can be sourced electronically, in which case the customer profile is said to be based on the customer’s on-line behaviour and transactional histories, which can be captured using registration forms, log files, cookies, and collaborative software. Non-electronic/off-line information for customer profile construction can be sourced from marketing events, and in-store transactional history which can be captured at point of sale [11].

This paper presents an approach for analysing transactions of collaborated/verified customer profiles for the purpose of classification. For our experiments, we use verified customer profiles for **Electricians** and **Plumbers/Heat Fitters** (-referred to henceforth as **PlumbHeaters**) sourced from Screwfix - a large UK retailer. The customer profiles contain the trade-types (the only demographic attribute) and transactional data consisting of the number of items (categorised into Topics) transacted over a period of 30 months. Details of Screwfix’s customer data and the transactional data used for our experiment can be found in Section IV-A. The rest of this section defines the transactional data problem and discusses the nature of transactional data together with the challenges encountered when analysing such data for the purpose of building customer profile models. An overview of data binning techniques is also presented.

A. Problem Statement

Formally, given a set of transactions \mathbf{T} categorized into d product item topics:

$$\mathbf{T} = \begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,d} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,d} \\ \vdots & \vdots & \dots & \vdots \\ t_{N,1} & t_{N,2} & \cdots & t_{N,d} \end{bmatrix}$$

we define a set of n customer profiles, $\mathbf{P} = \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_n\}$ where $\mathbf{P}_j = (t_1, t_2, \dots, t_M)^T$ and $t_i \in \mathbf{T}_j$ is an aggregated transaction from the set \mathbf{T}_j of all the transactions of the j -th customer.

Goal: We seek to identify distinctive groups, defined by the number of items purchased, for which the predictive error of classifying unseen customer profiles is minimal.

B. The Issue with Transactional Data

Transactional data apart from being traditionally large, is also inherently sparse, due mainly to the underlying process from which they are generated. For example, in retail transactional data, where it is usual for customers to purchase only a very small fraction of products, the average size of a basket (i.e., the collection of items that a customer purchases in a typical transaction) might contain just 3-4 products out of 1,000s of products in the retailer’s catalogue/inventory. Such a transaction when represented in an attribute-vector representation will have an average of 3-4 out of 1000s of product attributes that are not null. This implies that the fraction of non-zero attributes on the table (i.e. the sparsity factor) will be 3/1000 - 4/1000, or 0.3 - 0.2%. The sparsity factor of the transactional data used in our work here can be seen from the relationship between the number of transactions and the number of items transacted in Table I.

Pre-processing techniques such as sampling [12], clustering [13], data binning [5, 14], etc. are often used to prepare the dataset used in building and maintaining data mining based business models.

For sparse data, conventional sampling may not work well, because most of the samples are zeros [15]. Likewise, sampling fixed dataset columns from the dataset, as is done in some cases [16], is also inflexible because different rows may have very different sparsity factors leading to each sampled data instance conveying little or no information for accurate inference.

In order to address the issue of sparsity present at an individual transaction level in our work here, the transactions over the 30 months period for a given customer were aggregated.

C. Overview of Data Binning

Data binning is an unsupervised discretization method in which the data is grouped into either *Equal Interval*

Width or Equal Frequency Intervals.

The equal-width data binning algorithm work by determining the minimum and maximum values of the attribute of interest and then divides the range into a user-defined number of equal width bin intervals. This approach to data binning is however vulnerable to outliers that may drastically skew the range [17].

The equal-frequency data binning algorithm, on the other hand, determines the minimum and maximum values of the attribute of interest, sorts all values in ascending order, and divides the range into a user-defined number of intervals so that every interval contains the same number of sorted values.

Kerber [18] asserts that since binning, like many unsupervised methods, do not utilize instance labels in setting partition boundaries, it is likely that classification information will be lost by binning as a result of combining values that are strongly associated with different classes into the same bin. This can result in effective classification to be much more difficult to perform in some cases.

Chiu et al. [19], Chmielewski and Grzymala-Busse [20] use a variation of equal frequency intervals - maximal marginal entropy - to adjust the boundaries so as to minimise entropy at each interval. Holte [21] presented an example of a simple supervised data binning approach, in which his *Information Retrieval* (IR) algorithm divides the domain of every continuous variable into pure bins, each containing a strong majority of one particular class with the constraint that each bin must include at least some pre-specified number of instances. This approach appears to work reasonably well when used with the IR induction algorithm.

The binning process presented in this paper ranks and sequentially assigns the customer profiles, \mathbf{P} , in ascending order of number of items transacted, into bins based on an arbitrarily chosen maximum bin size q . The proposed method is described in Section III, while Section IV describes the set up of the experiment used for evaluating the proposed method. The results are discussed in Section V.

III. METHODOLOGY

A typical data binning process broadly consists of four steps [20]:

- 1) sorting the continuous values of the feature to be binned,
- 2) evaluating a cut-point for splitting or adjacent intervals for merging,
- 3) according to some criterion, splitting or merging intervals of continuous value, and
- 4) finally stopping at some point.

One key parameter of concern in the data binning process is determining the best “cut-point” to split a range of continuous values or the best pair of adjacent intervals to merge. Entropy based-and/or-statistical based evaluation function have been used to determine an appropriate “cut-point” with varying results [14, 18, 22, 23, 24].

As discussed in Section II-B, transactional data tends to be skewed towards the large number of customers who make fewer purchases. This makes distinguishing them for classification purposes difficult.

To improve the classification performance, we regroup the customer transactions into bins defined by the number of items per transactions and choose a “cut-point” that is a large fraction of the total number of customer profile transactions. This heuristic ensures that the bins with the fewer items per transaction have a large enough representation of examples to make up for the sparseness of the transactions, $\mathbf{t}_i \in \mathbf{T}_j$.

The proposed approach as shown in Algorithm 1, takes in as input a labelled transactional data set and a maximum bin size threshold value; and returns the bins and their ROC values.

IV. EXPERIMENTAL EVALUATION

A. Data Description

To evaluate the performance of the proposed approach on real-world transactional data, a series of experiments were performed using transactional data provided by Screwfix of **Electricians** and **PlumbHeaters** covering a period of 30 months.

The provided **Electricians** and **PlumbHeaters** trade-types used for the experiments in this paper are two of the trade-types identified and recorded by Screwfix. They were verified with third-party trade bodies as many of the recorded customers’ trade types do not reflect the recorded label (trade-type) of the customers mainly due to changes in the transaction behaviour of customers overtime. Table I shows the total number of the verified **Electricians** and **PlumbHeaters** trade-types used for the evaluation experiment together with the number of transactions and items transacted by the aforementioned trade-types over the period under consideration.

B. Bin Evaluation using the AUC score

As can be seen in Table I the **Electricians** and **PlumbHeaters** are imbalanced by a ratio of approximately 1:3. The performance of data mining algorithms is typically evaluated using predictive accuracy. However, this is not appropriate when the data is imbalanced and/or the costs of different errors vary markedly [25, 26]. This is mainly because the large difference in representation between the classes can lead to a bias in which even a simple

TABLE I
CUSTOMER PROFILE DATA

Profile Name	No. of Customers	No. of Transactions	No. of Items Transacted
Electricians	1537	32063	111730
PlumbHeaters	4135	68715	230542

Input: Customer Profiles \mathbf{P} , bin_size threshold n

Output: Sets of bins each assigned a ROC model accuracy estimate

```

sumT[i] ← Row-wise sum of each transactions  $t_i$ 
in data set  $P$ ;
sortT[j] ← sort( $i$ );
uniqueSum[u] ← Unique( $j$ );
initialize bins [ ];
initialize bin_ModelROC_Values [ ];
//Group  $\mathbf{P}$  into bins defined by the
number of items bought with each
bin having a maximum size =  $q$ ;
foreach unique summed transaction value  $u$  do
    while bin_size <  $q$  do
        bin[k] ← Find transactions in  $T$  with sums
        equal to those in segment  $u$ ;
    end
     $k \leftarrow k + 1$ 
end
//Compute ROC performance values
for each bin;
foreach bin do
    bin_ModelROC_Values[k] ← Compute the
    ROC performance value using 10-fold cross
    validation;
     $k \leftarrow k + 1$ 
end
Return bins and bin_ModelROC_Values as
output

```

Algorithm 1: Transactional Data Binning and ROC Computation Algorithm

default strategy of guessing would give a high predictive accuracy to the majority class [27].

The Receiver Operating Characteristic (ROC) curve is a standard technique for summarizing classifier performance over a range of trade-offs between true positive and false positive error measures. It is not influenced by decision biases and prior probabilities, and it places the performance of diverse systems on a common, easily interpreted scale [28].

The Area Under the Curve (AUC) is an accepted traditional performance metric for a ROC curve [29, 30, 31].

The ROC convex hull can also be used as a robust method of identifying potentially optimal classifiers [32].

V. DISCUSSION OF EXPERIMENT RESULTS

The goal of the experiment was to identify a range of required items per transaction to more accurately classify unseen customers to a customer profile. The transactional data binning part of Algorithm 1 was implemented using Matlab while an evaluation of the ROC performance on the identified bins was performed using Weka's implementation of C4.5 [33](implemented as J48 in Weka), linear discriminant classification [34](implemented as Classification via Regression in Weka), Naive Bayes [35] and SVM [36, 37](implemented as SMO in Weka).

Table II shows the classification performance on the identified bins while Table III shows the classification performance on the whole un-binned data of customers. Figure 1 shows the comparative ROC performance of the 4 classifiers on both the identified bins and the whole un-binned dataset.

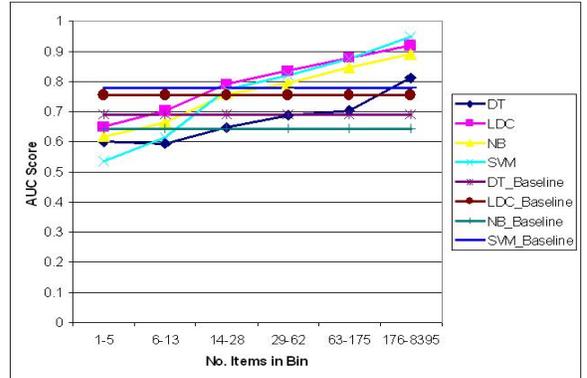


Fig. 1. Plot showing the performance of 4 classifiers on both binned and un-binned Customer Profiles

The difficulty in classifying customers with few transactions can be observed from Table II and Figure 1 in which the ROC classification performance values increase as the number of items transacted increase. The effect of the sparsity and skewness of the transactional data on classification performance can also be seen from Tables II and III where the classification performance is drawn more to the majority PlumbHeaters class as reflected by the true positive rate measures. Thus, a

TABLE II
CLASSIFICATION PERFORMANCE ON IDENTIFIED BINS

Bins	No. of Electricians	No. of PlumbHeaters	Bin Size	TP (E)	FP (E)	TP (P)	FP (P)	Accuracy	ROC
1-5	303	775	1078	0.221	0.132	0.868	0.779	0.686	0.599
6-13	302	805	1107	0.325	0.199	0.801	0.675	0.671	0.594
14-28	249	784	1033	0.442	0.18	0.82	0.558	0.729	0.648
29-62	256	761	1017	0.516	0.184	0.816	0.484	0.74	0.687
63-175	288	713	1001	0.59	0.147	0.853	0.41	0.777	0.702
176-8395	139	297	436	0.748	0.101	0.899	0.252	0.851	0.813

(a) C4.5 Decision Tree

Bins	No. of Electricians	No. of PlumbHeaters	Bin Size	TP (E)	FP (E)	TP (P)	FP (P)	Accuracy	ROC
1-5	303	775	1078	0.096	0.041	0.959	0.904	0.716	0.649
6-13	302	805	1107	0.129	0.037	0.963	0.871	0.735	0.704
14-28	249	784	1033	0.289	0.07	0.93	0.711	0.775	0.791
29-62	256	761	1017	0.402	0.075	0.925	0.598	0.794	0.836
63-175	288	713	1001	0.573	0.069	0.931	0.427	0.828	0.879
176-8395	139	297	436	0.633	0.047	0.953	0.367	0.851	0.92

(b) Linear Discriminant Classification

Bins	No. of Electricians	No. of PlumbHeaters	Bin Size	TP (E)	FP (E)	TP (P)	FP (P)	Accuracy	ROC
1-5	303	775	1078	0.294	0.183	0.817	0.706	0.67	0.617
6-13	302	805	1107	0.351	0.217	0.783	0.649	0.665	0.666
14-28	249	784	1033	0.522	0.162	0.838	0.478	0.762	0.76
29-62	256	761	1017	0.59	0.171	0.829	0.41	0.769	0.795
63-175	288	713	1001	0.625	0.143	0.857	0.375	0.79	0.846
176-8395	139	297	436	0.41	0.047	0.953	0.59	0.78	0.892

(c) Naive Bayes

Bins	No. of Electricians	No. of PlumbHeaters	Bin Size	TP (E)	FP (E)	TP (P)	FP (P)	Accuracy	ROC
1-5	303	775	1078	0.04	0.017	0.983	0.96	0.718	0.535
6-13	302	805	1107	0.066	0.024	0.976	0.934	0.728	0.616
14-28	249	784	1033	0.309	0.071	0.929	0.691	0.779	0.773
29-62	256	761	1017	0.445	0.072	0.928	0.555	0.806	0.821
63-175	288	713	1001	0.625	0.077	0.923	0.375	0.837	0.877
176-8395	139	297	436	0.784	0.061	0.939	0.216	0.89	0.95

(d) SVM

TABLE III
BASELINE CLASSIFICATION PERFORMANCE ON UN-BINNED CUSTOMER PROFILES

Classifiers	No. of Electricians	No. of PlumbHeaters	Total No. of Instances	TP (E)	FP (E)	TP (P)	FP (P)	Accuracy	ROC
C4.5	1537	4135	5672	0.472	0.163	0.837	0.528	0.738	0.69
LDC	1537	4135	5672	0.114	0.007	0.993	0.886	0.755	0.756
Naive Bayes	1537	4135	5672	0.181	0.049	0.951	0.819	0.742	0.644
SVM	1537	4135	5672	0.24	0.033	0.967	0.76	0.77	0.779

classifier built using the entire transactional data will be less accurate and less confident in its classification of the customers with larger number of items per transaction as can be seen in Figure 1. The bins for which the ROC classification performance becomes better than that obtained from the baseline, (bin 63-175 for C4.5, bin 14-28 for LDC, bin 6-13 for Naive Bayes and bin 29-62 for SVM), can be interpreted as the “critical point” identifying the minimum number of purchased items required for the classifiers to more confidently classify

a customer profile than if the whole data was used. This minimum number of items to be bought does not only have implications on the accuracy with which different customers can be classified, but also on the ability to track potential changes overtime which is of great interest for businesses.

VI. SUMMARY, CONCLUSION AND FUTURE RESEARCH WORK

This paper presented an approach for identifying and accurately classifying groups of customers on the basis of the number of items transacted. It commenced in Section II by describing the problem of identifying and classifying customer profiles where only their highly sparse and skewed transactional data is available. The issues inherent in transactional data which make using them for classification purposes challenging were further discussed in Section II-B, while an overview of the data aggregation process of data binning was presented along with their drawbacks in Section II-C. The proposed transactional data binning algorithm was then presented in Section III as an effective method used in the process of mining transactional data for the purpose of identifying and accurately classifying customers. The experimental results obtained from 4 classifiers, as discussed in Section V, showed that the proposed approach can be used to identify a critical point at which customer profiles can be more confidently distinguished by a classifier compared to that obtained when the whole data is used for building a single classifier.

The transactional data for the customer profile used for the experiments in this paper were accumulated over a period of 30 months. We are currently investigating methods for extending the proposed approach described in Section III to track customer buying behaviour overtime with the goal of quickly identifying when customers change their buying behaviour and move from one profile to another.

Also, it can be noticed from the AUC performance scores in Figure 1 that the bins containing customers with fewer items per transaction are not only difficult to distinguish, but the accuracy performance of all 4 classifiers are distorted in favour of the PlumbHeaters majority class as reflected in the true positives and false positive measures in Table II.

This distortion in the predictive accuracy of a classifier in favour of the majority class is often referred to as *the class imbalance problem* [25, 27], and typically occurs when, in a classification problem, there are many more instances of some classes than others. In such cases, standard classifiers tend to be overwhelmed by the large classes and ignore the small ones. A number of solutions to the class imbalance problem have been proposed in the literature at both the data and algorithmic levels [25, 27, 38]. Future work will investigate the application of the aforementioned class imbalance solutions in boosting the overall classification performance of the

bins with the fewer items per transaction.

*REFERENCES

- [1] M. J. A. Berry and G. S. Linoff, *Mastering Data Mining: The Art and Science of Customer Relationship Management*. John Wiley & Sons, January 2000.
- [2] —, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons, April 2004.
- [3] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, “Improving the effectiveness of collaborative filtering on anonymous web usage data,” in *Proceedings of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization (ITWP01)*, Seattle, August 2001.
- [4] A. S. Das, M. Datar, A. Garg, and S. Rajaram, “Google news personalization: scalable online collaborative filtering,” in *Proceedings of the 16th international conference on World Wide Web*, ser. WWW '07. New York, NY, USA: ACM, 2007, pp. 271–280.
- [5] J. Han and M. Kamber, *Data Mining*, 2nd ed. Morgan Kaufmann, June 2006.
- [6] T. C. Redman, *Data quality: management and technology*. New York, NY, USA: Bantam Books, Inc., 1992.
- [7] J. Han, “Olap mining: An integration of olap with data mining,” in *In Proceedings of the 7th IFIP 2.6 Working Conference on Database Semantics (DS-7)*, 1997, pp. 1–9.
- [8] D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*. The MIT Press, 2001.
- [9] M. Bazik, “Defining your customer profile - an essential tool,” *Journal of Extension (JOE)*, vol. 37, no. 6, December 1999.
- [10] P. S. Yu, “Data mining and personalization technologies,” in *Proceedings of the Sixth International Conference on Database Systems for Advanced Applications*, ser. DASFAA '99. Washington, DC, USA: IEEE Computer Society, 1999, pp. 6–13.
- [11] B. Mobasher, “Data mining for web personalization,” *The Adaptive Web*, vol. 4321, pp. 90–135, 2007.
- [12] J. Kivinen and H. Mannila, “The power of sampling in knowledge discovery,” in *Proceedings of the thirteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, ser. PODS '94. New York, NY, USA: ACM, 1994, pp. 77–85.
- [13] A. Berson, S. Smith, and K. Thearling, *Building Data Mining Applications for CRM*, 1st ed. McGraw-Hill Professional, 1999.

- [14] I. H. Witten and E. Frank, *Practical Machine Learning Tools and Techniques with Java Implementations*, ser. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science & Technology, October 1999, vol. 1.
- [15] K. W. Church, P. Li, and T. J. Hastie, "Conditional random sampling: A sketch-based sampling technique for sparse data," in *In NIPS*, 2006, pp. 873–880.
- [16] R. Gemulla, "Sampling algorithms for evolving datasets," Ph.D. dissertation, Technische Universitt Dresden, 2008.
- [17] J. Catlett, "On changing continuous attributes into ordered discrete attributes," in *Proceedings of the European working session on learning on Machine learning*. New York, NY, USA: Springer-Verlag New York, Inc., 1991, pp. 164–178.
- [18] R. Kerber, "Chimerge: Discretization of numeric attributes," in *Proc. Ninth Int'l Conf. Artificial Intelligence*, 1992, pp. 123–128.
- [19] D. K. Y. Chiu, A. K. C. Wong, and K. C. C. Chan, "Synthesis of statistical knowledge from time-dependent data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 265–271, March 1991.
- [20] M. R. Chmielewski and J. W. Grzymala-Busse, "Global discretization of continuous attributes as preprocessing for machine learning," in *Proceedings of the Third International Workshop on Rough Sets and Soft Computing*, San Jose, CA, November 10-12 1995, pp. 474–480.
- [21] R. C. Holte, "Very simple classification rules perform well on most commonly used datasets," *Machine Learning*, vol. 11, pp. 63–90, 1993.
- [22] R. Kohavi and M. Sahami, "Error-based and entropy-based discretization of continuous features," in *Proc. 13th Int'l Joint Conf. Artificial Intelligence*, 1996, pp. 1022–1027.
- [23] R.-P. Li and Z.-O. Wang, "An entropy-based discretization method for classification rules with inconsistency checking," in *Proceedings of the International Conference on Machine Learning and Cybernetics, 2002.*, vol. 1, 2002, pp. 243 – 246.
- [24] X. Liu and H. Wang, "A discretization algorithm based on a heterogeneity criterion," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 9, pp. 1166–1173, September 2005.
- [25] F. Provost, "Machine learning from imbalanced data sets 101 (extended abstract)," Technical Report WS-00-05, The AAAI Press, Menlo Park, California, Tech. Rep., 2000.
- [26] N. Japkowicz, "Learning from imbalanced data sets: A comparison of various strategies," AAAI Technical Report WS-00-05., Tech. Rep., 2000.
- [27] N. Chawla, *Data Mining for Imbalanced Datasets: An Overview*. Springer US, 2005, pp. 853–867.
- [28] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, no. 4857, pp. 1285–1293, June 1988.
- [29] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.
- [30] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern Recogn.*, vol. 30, pp. 1145–1159, July 1997.
- [31] S. S. Lee, "Noisy replication in skewed binary classification," *Comput. Stat. Data Anal.*, vol. 34, pp. 165–191, August 2000.
- [32] F. Provost, D. Jensen, and T. Oates, "Efficient progressive sampling," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '99. New York, NY, USA: ACM, 1999, pp. 23–32.
- [33] J. R. Quinlan, *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- [34] E. Frank, Y. Wang, S. Inglis, G. Holmes, and I. H. Witten, "Using model trees for classification," *Mach. Learn.*, vol. 32, pp. 63–76, July 1998.
- [35] F. Pernkopf, "Discriminative learning of bayesian network classifiers," in *Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications*. Anaheim, CA, USA: ACTA Press, 2007, pp. 422–427.
- [36] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, ser. COLT '92. New York, NY, USA: ACM, 1992, pp. 144–152.
- [37] N. Cristianini and J. Shawe-Taylor, *An introduction to support Vector Machines: and other kernel-based learning methods*. New York, NY, USA: Cambridge University Press, 2000.
- [38] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: special issue on learning from imbalanced data sets," *SIGKDD Explor. Newsl.*, vol. 6, pp. 1–6, June 2004.