

An Intelligent Model for Internet Advertising Selection Based on User-Profile

Siamak Mirzaei, Ashkan Farrokh Hayati

College of Science and Engineering, Flinders University, Adelaide, Australia
Advanced Computing Research Center, Unisa University, Adelaide, Australia
e-mails: siamak.mirzaei@flinders.edu.au
ashkan.hayati@unisa.edu.au

Abstract - With the growth of using the internet advertising, display error rate has been subsequently increased. As an instance of display error rate, it can be referred to advertisement inappropriate to user demand of modifying wrong advertising display. The most important problem related to marketing and advertising is to absolutely consider advertising true or false. To cope with such a problem, personalized advertising is made with respect to users' profile and behavior in order that accurate internet advertising is selected, and each user receives her/his favorite internet advertising. In this study, we presented a new profile with the internet advertising in an online bookstore to students and gathered their responses. Then, we used decision tree in data mining applications and modeled two separated datasets in two states of with a profile and without a profile. The results obtained for both datasets revealed that users profile can highly influence proper classification of the internet advertising.

Keywords: advertising, the internet advertising, data mining, decision tree, personalization

1. Introduction

Electronic commerce, particularly marketing and internet advertising has a relatively low cost [1, 2]. With the growth of the internet advertising usage, display error rate has been subsequently increased. One of the instances of display error is to blindly display unwanted internet advertising, namely false advertising. In various definitions, the common point of false display is their unwanted property. According to the agreed-upon definition, satisfaction or dissatisfaction is judged but not content [4].

False display is followed by various problems. Some of them directly cause economic losses such as traffic, bandwidth loss and some others cause to high waste of time to separate superfluous advertising by users. Additionally, some of false displays cause mental annoyance, uncertainty and insecurity and finally, legal problems such as economic fraud and pyramid advertising [2], [6].

Typically, machine learning and statistical methods are employed to select good-accuracy display. However, there is no unique method and each of existing methods has many defects and errors such as FN (false negative) and FP (false positive). Here, we consider false display class as negative class and valid internet advertising as positive class [4], [7, 8]. The mentioned methods have some defects such as great data, data dependency assumption and high impressionability of various environment and data [7, 8].

Another problem in selecting advertising display regarding marketing and advertising through is an intangible advertisement. For example, in the internet advertising, purchasing automobile may be considered false display for someone who tends to purchase automobile while it may be considered false display for someone who does not tend to purchase. Such the internet advertising is also called gray advertising [9, 10]. To cope with this problem, a personal profile is built with respect to users' behavior [8], [10]. Due to the importance of this issue, we need to have a correct view regarding e-commerce to design advertising selection. Lack of such a view in design prevents the coordination between advertising selection and marketing and internet advertising service providers. Therefore, a high percentage of the internet advertising published by these service providers are selected false, imposing a high cost on the service providers [1,2], [11]. In this study, we investigated the effect of profile and behavior of users on displays personalization for better classification and refinement in the internet advertising through tree decision. In the following, we present the related studies, profile design, data generation and its statistical population, the proposed method, and the obtained results and conclusion.

2. Related Literature

Marketers carefully have investigated virtual communities, trying to determine behavioral process of customers in purchasing a certain product. Due to the access to user information in social network portals, it is possible to target marketing messages even through common methods such as web banners (banner advertising). Such a form of online advertising leads to institutionalizing advertising in a web page and this advertising has been built of an image. When visitors click on the banner, they are directed into website advertized in the banner. Banner-based advertising activities in social network portals can be observed in real time and they may

be targeted based on the interests of visitors. It is possible since users of virtual community are identified with their login and present personal information (age, gender, education, etc.) as well as behavioral information (sending or receiving invitation, comments, and times of use). Access to behavioral information has a certain competitive advantage for online social networks compared to other web portals. In the present research, we firstly investigate the potential superiority of behavioral data mining for web banners-based marketing campaign management. Then, we select the most appropriate data mining techniques for this particular issue.

The main problem is the optimization of banner advertising campaigns in marketing through targeting a proper user and the maximization of response analysis through the number of clicks. The issue of response analysis rate and marketing campaign optimization has been widely explained in data mining course book [12, 13] and recently in online social network content [14].

2.1 Class Imbalance Problem

Class imbalance problem refers to a situation in which the number of objects of a class (a class of dependent variables) is obviously less than the number of other class objects. This problem is highly important particularly in response analysis in which customer's reaction (in this case, a click on banner) is significantly less than the number of messages (displays). Regarding marketing, churn models refer to gaining customers; while in other fields, they refer to fraud detection, medical diagnosis and so forth. Coping with this problem, there are two main approaches [15]: learning sample's (sampling techniques) structure change-based and cost-sensitive algorithms. Researchers propose one class learning in case of strong class imbalance problem [16]. This problem is due to the fact that gathering information about other class is sometimes difficult and domain nature automatically suffers from imbalance. Sometimes, creating classifiers using the items belonging to a class is successful sometimes. Some writers [17] distinguish cost-sensitive learning and ensemble classifiers, i.e. bootstrap procedure (bagging and random forests). Although this approach can include cost-sensitive learning algorithms, they are based on CART algorithm [18] (classification and regression trees) and employ misclassification costs and probably, CART.

2.2 Sampling Techniques of the Imbalanced dataset

Up-sampling (or over-sampling) is to reiterate items which belong to the minority class. This fact can occur randomly, directly or through synthetic cases, e.g. SMOTE algorithm [19]. Downsampling (under-sampling or down-sizing) is to decrease the number of cases which belong to the majority class. Sometimes, over-represented cases related to redundant samples [20] are omitted based on Tomek's link [21].

2.3 Cost-sensitive Learning

Cost-sensitive learning is another approach which can contribute to overcoming class imbalance problem. The purpose of building such classifications is to increase the accuracy of predicting cases which belong to the given class. Researchers should allocate various costs to objects misclassification. [22] have detected two classes of cost-sensitive learning. One of them is a set of direct algorithms such as cost-sensitive decision tree and the other is cost-sensitive meta learning methods including CSC (cost-sensitive classification), ET (empirical threshold) or cost-sensitive naive Bayes. The two classes are different in facing bias data when they define misclassification costs.

For example, TN stands for true negative; that is, an object which belongs to negative class has been classified as negative. Since TN and TP refer to correct classification, costs are allocated to FN and FP. Creating classifiers for a dichotomous dependent variable often offer researchers to focus on positive class; therefore, the cost for FN should be higher than FP.

In other words, it is very important to decrease the error of positive class misclassification. If a higher cost is allocated to FN, the individual considers refusing to classify a positive object as a negative object. [23] emphasizes that costs cannot be merely monetarily considered.

2.4 Classification Methods

Data mining models such as single classification tree (CART algorithm), RF (random forest) and gradient tree boosting are widely used in marketing to evaluate selection. All these methods can employ a cost misclassification and detect prior probabilities.

[24] proposed CAERT which is a recursive partitioning algorithm. This algorithm is used to build a classification tree, in case of the presence of nominal dependent variable, and a regression tree, in case of a continuous dependent variable. The purpose of the test is to predict customers' responses, which means to develop a classification model. To sum up, a graphic model of a tree can be presented as a set of if-then rules.

Visualizing a model is a very important advantage of this analysis approach in marketing. Prediction is an important task for marketing managers, but knowledge is vital in the considered area. Although CART has been proposed about 30 years ago, important features such as prior probabilities and misclassification costs cause to be useful in cost-sensitive learning.

For prediction, [24] also proposed RF which is a set of regression or classification trees. RF includes a number of classifiers which have been built by a different set of independent variables. In each stage of tree construction process, a set of explanatory variables are randomly selected for each tree (in each node). The number of the selected variables is determined by m but the total number of variables is determined by M . The best split of a node is based on these m predictors ($m < M$). Each tree is constructed without pruning to the maximum extent possible. Finally, the tree vote for the class of an object. RFs are constructed from learning sample using the bootstrap sample. Accordingly, classic algorithms such as CART or C4.5 are better performed.

Gradient tree boosting is based on the incremental concept proposed by [26, 27]. Briefly, a decision tree attempts to allocate an object to the given class. After the first effort to predict, the cases which belong to a class with weak classification (usually, minority class) are given more weight. In the next step, a classifier uses weighted learning sample and once again, give more weight to those cases which have been misclassified. During the recursive process, many trees are constructed and sample voting procedure is applied through expanding model-based testing. In other words, the predictions of one decision-making tree are combined to achieve the best output. Each classifier derives from its own booster which has been randomly obtained from the entire the sample learning:

3. The Proposed Model

Investigating the introduced methods reveals that there are four main stages to design and generate advertising selector (Figure 1).

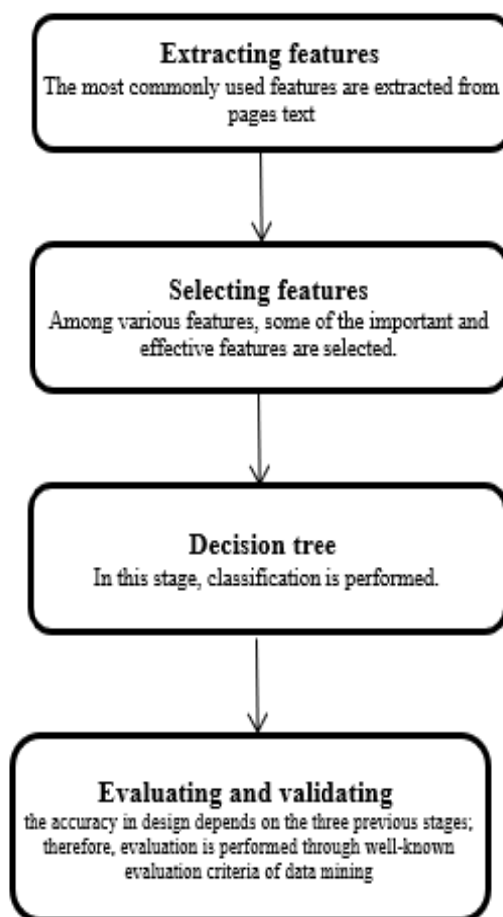


Figure 1. The steps of designing selector

3.1 Extracting features

For the purpose of the study, the internet advertising with users profile and label are firstly gathered. Since such information is often inaccessible, it should be simulated just like related articles. Firstly, investigating the related studies, we find out which features can be extracted from the internet pages. Three types of basic features are as following:

- Page title
- Page content
- Page structure

3.2 Advertising design

We select some of the online bookstore advertising. Using the experts' opinions, we determine marketing and advertising principles based on the references [1] and [10] to provide the content for each of the mentioned features. Table 1 shows the extracted features of advertising with content selection.

Table 1. The selected features based on area

Total feature	Available alternatives
Page title	1. Experienced user 2. New user
Advertising subject	1. Free books 2. The newest book 3. The bestselling book
Page content	1. Engineering and basic sciences 2. Medicine 3. Humanities 4. Art 5. Other
Page structure	1. Animated advertising 2. Textual advertising

3.3 Profile design

The next and the most important stage of extracting features is users' profile. Most of the studies have used standard profiles existing in sites such as job, gender, age, education, field of study, and so forth [7]. Since personalizing is performed based on users' profile, it is necessary to consider other alternatives to increase accuracy. To this end, we gather two alternatives from various articles and put into our profile [4], [7]. The first alternative is the number of times a person announces it false after receiving advertising. This alternative is different for different individuals such that a person may announce an internet advertisement false in the first stage; however, another person may announce the same internet advertising false in high frequencies. The main cause of placing such an alternative in users' profile is gray internet advertising. The second alternative which is placed and questioned in users' profile is the ratio of person's tolerable errors in selecting which can be acceptable.

In fact, different persons have different behavioral features. Some of them state that none of their valid internet advertising should be falsely selected and in return, they accept receiving some daily false displays. On the contrary, some individuals are not willing to receive any false display although some of their valid internet advertising is falsely selected. In fact, different individuals can be behaviorally detected using these two alternatives. This part is in accordance with the first step, i.e. extracting features from the internet advertising.

Table 2 shows the output of predicting accuracy coefficient of the model based on the predictor features in target variable row and evaluative adaptive matrix.

Table 2. The features of user profile

Property	Significance coefficient to predict target variable changes
User type (student and free)	0.37
Purchasing history (purchased, not purchased)	0.25
Using place (academic, non-academic)	0.11
Educational degree (BA, MA or PhD)	0.01

3.4 Statistical Population

Since our purpose was to refine and classify the internet advertising, it is necessary that the content of advertising is in the same regard. To carefully conduct the study, we first consider our statistical population and then, provide the content of the internet advertising. The statistical population includes academic community and students. Therefore, it is necessary to select the area in which the population has information and willingness. Accordingly, we select online bookstore advertising as the statistical sample. An internet advertising sample is obtained from the Cartesian product of the values presented in Table 1 such that $3 \times 5 \times 5 = 150$ of the internet advertising frame is obtained. The number of profile items is 10 and each item can have a different value. Each item of simulated advertising include users' profile and 150 designed internet advertising and response label. After designing through the web, the items of simulated advertising are answered by 150 students. After cleaning, 98 people were used (60 females and 38 males). In the following, we randomly split the gathered data and to evaluate the framework, we use the two following datasets:

- The first type dataset: 2500 internet advertising which includes 1800 false displays.
- The second type dataset: 8000 internet advertising which includes 1700 false displays.

3.5 User Behavior Classification

The final objective of the present research is to map the customers into two groups of avoidant (a user who leaves the page) or non-avoidant (a user who stay in the page). In the study, "object", refers to the customer and "class", refers to persistent planar or customer avoidance.

3.6 Constructing Decision Tree

Considering the fact that the patterns extracted from decision tree model are as sequences of if-then rules provides more efficient context to formulate marketing strategies for each of customers' class according to their demographic and behavioral features. Therefore, "decision tree" has been selected as the optimal alternative for the purpose of the study.

A decision tree can be constructed using various algorithms. We have used CHAID algorithm to construct the decision tree model. This algorithm organizes internal nodes of tree based on the correlation rate of each feature with target variable. To create leaf nodes, considering discrete, qualitative and divalent target variable of the research model, we have used independent test between the target variable and each of observation features to attribute each of the model observations to one of the two classes of the target variable (avoidant and persistent) based on observation features. To perform this test, we have formed an agreed-upon table for each feature. In the tables, the number of lines is correspondent to the rows of the feature and its two columns are correspondent to the rows of the target variable. We have computed the test statistic using the following formula:

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Where O_{ij} indicates the expected frequency for the cell located in row i and column j ; e_{ij} indicates the observed frequency of the cell located in row i and column j , and R indicates the number of table rows and C refers to the number of rows.

4. Results

After implementing the constructed model, it is the turn of the fourth step. In this step, we compare and evaluate the obtained results using the data obtained from the implementation. Figures 2 and 3 show the results obtained from the model for both sets.

In this table, we use common criteria of data mining to evaluate and compare. To prevent limiting comparison and evaluation merely to accuracy and involve two other types of error in the comparison, we employ some criteria such as FP Rate, Spam Recall and Spam Precision.

- a: a false display which has been predicted as false display
- d: a valid advertising which has been predicted as valid advertising
- b: a false display which has been predicted as valid advertising (FN)
- c: a valid advertising which has been predicted as false display (FP)

$$\text{Accuracy} = (a+d)/(a+d+b+c) \tag{1}$$

$$\text{Accuracy} = 1 - \text{Error Rate} \tag{2}$$

$$\text{FP Rate} = c / (d+c) \tag{3}$$

$$\text{Spam Recall} = a/(b+a) \tag{4}$$

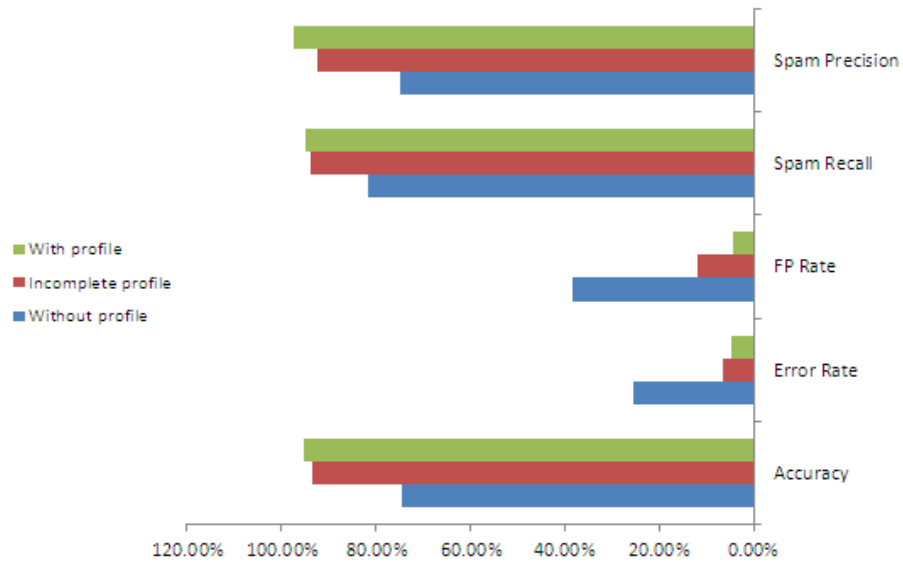


Figure 2. The results for the first group

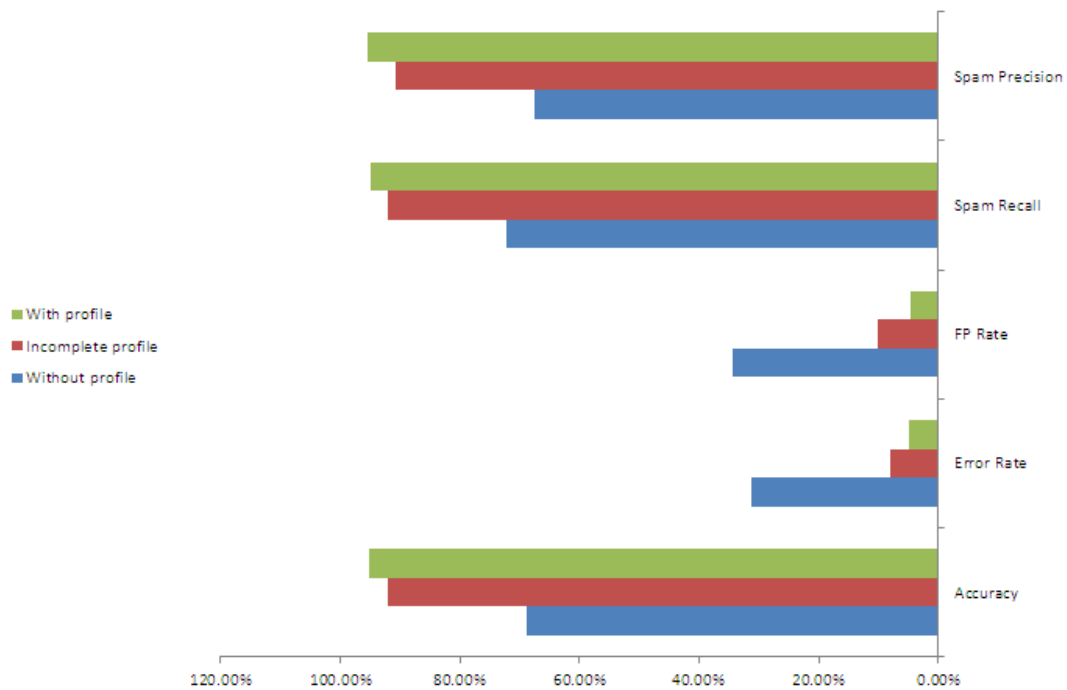


Figure 3. The results for the first group

Table 3 to Table 11 measure the model prediction accuracy for each of the two states of the target variable in the adaptive matrix and totally evaluate the model predictions.

Table 3. Zero group

Classes	Relative frequency (%)
Avoidant	93.59
Persistent	06.41
Total	100

Zero group indicates a 60% exit rate of customers from the study.

Table 4. Free users

Classes	Relative frequency (%)
Avoidant	62.89
Persistent	37.11
Total	100

The group one includes free users what 90% of them leave the page after seeing advertising.

Table 5. Student users

Classes	Relative frequency (%)
Avoidant	96.40
Persistent	03.60
Total	100

The group two includes free students whose exit rate, compared to the previous group, is reduced by half.

Table 6. Users without purchasing history

Classes	Absolute frequency (%)
Avoidant	65.24
Persistent	34.92
Total	100

Table 7. Users with purchasing history

Classes	Relative frequency (%)
Avoidant	25.38
Persistent	74.62
Total	100

Comparing Tables 6 and 7, we observe that users with purchasing history have very less avoidant.

Table 8. Non-academic using place

Classes	Relative frequency (%)
Avoidant	23.36
Persistent	76.64
Total	100

Table 9. Academic using place

Classes	Relative frequency (%)
Avoidant	58.52
Persistent	41.48
Total	100

Comparing Tables 8 and 9, we observe that academic users have very less avoidant.

Table 10. Bachelor degree

Classes	Relative frequency (%)
Avoidant	66.63
Persistent	33.37
Total	100

Table 11. MA and PhD degree

Classes	Relative frequency (%)
Avoidant	59.90
Persistent	40.1
Total	100

Comparing Tables 10 and 11, we observe that users with high educational degrees have very less avoidance.

5. Conclusion

Generally, there are few studies conducted on refining and classifying the internet advertising regarding marketing and advertising. Therefore, the presented research attempted to create a personalized advertising selector to estimate the importance the internet advertising and classifying users with respect to their behavior and profile. Classifying and refining through users' profile not only increases accuracy, but also decrease FP and FN errors. To implement the model, we used two separate datasets. In the selection step, features were determined such that our two proposed alternatives have the second and fourth selection. To implement each dataset, the internet advertising with a profile, we employed incomplete profile and without a profile. Then, comparing the determined evaluation criteria and the results obtained from implementing two separate datasets, we revealed that classifying the internet advertising with a profile has the highest accuracy. Other criteria mentioned in this comparison revealed that the increase of accuracy leads to the decrease of FP and FN errors. Decreasing these errors does not incur extra costs for advertising companies and users also receive their favorite internet advertising. In other words, some sort of compatibility is created among advertising selector, advertising companies and users' interest.

References

- [1] Blanzieri, E., Bryl, A., "A survey of learning-based techniques of email spam filtering", *Artif Intell Rev*, vol.29,pp.63–92,2008.
- [2] Cukier, W. L., Cody, Susan., Nesselroth, E. J., "Genres of Spam: Expectations and Deceptions", *Proceedings of the 39th Hawaii International Conference on System Sciences*, 2008
- [3] Sousa, p., et al. "A Collaborative Approach for Spam Detection", *Second International Conference on Evolving Internet*, IEEE, 2010.
- [4] Ying, K.C., et al., "An ensemble approach applied to classify spam e-mails", *Expert Systems with Applications*, vol. 37, pp 2197–2201, 2010.
- [5] Kim, J., Dou, D., Liu, H., Kwak, D., "Constructing a User Preference Ontology for Anti-spam Mail Systems", *Canadian AI 2007, LNAI 4509*, pp. 272–283, 2007.
- [6] Youn, S., McLeod, D., "Spam Decisions on Gray E-mail using Personalized Ontologies", *Proceedings of the 2009 ACM Symposium on Applied Computing (SAC)*, Honolulu, Hawaii, USA, pp. 1262-1266, 2009.
- [7] Rossiter J. R., Bellman S., "Marketing Communications". Prentice Hall, 2005.
- [8] Raad, M., et al., "Impact of spam advertisement through e-mail: A study to assess the influence of the anti-spam on the e-mail marketing", *African Journal of Business Management*, Vol. 4(11), pp. 2362- 2367, 2010.
- [9] Guzella, T.S., Caminhas, W.M., "A review of machine learning approaches to Spam filtering" , *Expert Systems with Applications*, vol. 36 ,pp.10206–10222, 2009.
- [10] Renuka, D.K., Hamsapriya, T., Chakkaravarthi, M.R. , Surya, P. L., "Spam Classification based on Supervised Learning using Machine Learning Techniques", *Process Automation, Control and Computing (PACC)*, 2011 International Conference on, Coimbatore, 2011.
- [11] Silva, R. M., Almeida, T. A., Yamakami, A., "Artificial Neural Networks for Content-based Web Spam Detection", *The 2012 International Conference on Artificial Intelligence (ICAI'12)*
- [12] R. Nisbet, J. Elder, and G. Miner, "Handbook of Statistical Analysis and Data Mining Applications," Elsevier, Amsterdam, 2009.
- [13] S. Chiu and D. Tavella, "Data mining and market intelligence for optimal marketing returns," Elsevier, Amsterdam, 2008.
- [14] J. Surma and A. Furmanek, "Improving marketing response by data mining in social network," *The 2nd International Workshop on Mining Social Networks for Decision Support* Odense, 2010.
- [15] C. Chen, A. Liaw, and L. Breiman, "Using random forest to learn unbalanced data," *Technical Report 666*, Statistics Department, University of California at Berkeley, 2004. *IMMM 2011 : The First International Conference on Advances in Information Mining and Management* Copyright (c) IARIA, 2011. ISBN: 978-1-61208-162-5
- [16] B. Raskutti and A. Kowalczyk, "Extreme rebalancing for SVMs: a case study," *SIGKDD Explorations*, 2004.
- [17] S. Hido and H. Kashima, "Roughly balanced bagging for imbalanced data," *In SDM 2008*, SIAM, 2008, pp. 143-152.
- [18] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees," Belmont, CA: Wadsworth International Group, 1984.
- [19] D. A. Cieslak and N. V. Chawla, "Learning decision trees for unbalanced data," *ECML/PKDD*, 2008.
- [20] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: one-sided selection," *Proc. 14th Intl. Conf. on Machine Learning*, 1997, pp. 179–186.
- [21] I. Tomek, "Two modifications of CNN. *IEEE Trans. On Systems, Man and Cybernetics* 6, 1976, pp. 769–772.
- [22] C. X. Ling and V. S. Sheng, "Cost-Sensitive Learning and the Class Imbalance Problem", *Encyclopedia of Machine Learning*. C. Sammut (Ed.). Springer Verlag, Berlin, 2008.
- [23] C. Elkan, "The Foundations of Cost-Sensitive Learning," *In Proc. of the Seventeenth International Joint Conference of Artificial Intelligence*, Seattle, Washington, Morgan Kaufmann, 2001, pp. 973-978.
- [24] L. Breiman, "Random Forests," *Machine Learning*, 45, 5–32, Kluwer Academic Publishers, 2001, pp. 5-32.
- [25] Y. Freund and R. Shapire, "Experiments with a new boosting algorithm," *Machine Learning, Proc. of the Thirteenth International Conference 1996*, pp. 148-156.
- [26] J. H. Friedman, "Greedy Function Approximation: a Gradient Boosting Machine," *Technical Report*, Department of Statistics, Stanford University, 1999.
- [27] J. H. Friedman, "Stochastic Gradient Boosting," *Technical Report*, Department of Statistics, Stanford University, 1999.