Feature selection in bipolar disorder episode classification using cost-constrained methods

Olga Kaminska[1] [0000−0003−3940−9471], Tomasz Klonecki[2] [0000−0002−0216−3685], and Katarzyna Kaczmarek-Majer[1] [0000−0003−0422−9366]

[1] Systems Research Institute Polish Academy of Sciences, Warsaw, Poland
[2] Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland
o.kaminska@ibspan.waw.pl
tomasz.klonecki@ipipan.waw.pl
k.kaczmarek@ibspan.waw.pl

Abstract. An important step in the classification process of bipolar disorder episodes is feature selection process indicating the most relevant factors in patients' behavior. The features in this task are associated with costs. Besides basic (low-cost) information about patients' phone calls and text messages, we are studying the impact of acoustic features (high-cost) on classifying patients' states. Unlike in previous papers, now we take the costs into account and thus we apply cost-constrained methods. The purpose of this paper is to examine whether the cost-constrained feature selection procedure is capable of improving the performance of the classification model while reducing the cost of making predictions. Moreover, we are trying to determine whether the reduced number of expensive features maintains a relatively high performance. We use a filter feature selection method that applies information theory. In the cost-constrained modification, we add a cost factor parameter that controls the trade-off between feature importance and its cost. The experiments were performed on a large medical database collected from patients with bipolar disorder during their daily mobile calls. The results indicate that the cost-constrained method allows to achieve better results than traditional feature selection when the budget is limited.

Keywords: bipolar disorder episode classification · mobile calls acoustic features · feature selection · information theory · costly features

1 Introduction

In Bipolar Disorder (BD) there are plenty of factors that have a significant impact on current patients' mood. The habit of using mobile phones is one of them. Both the behavioral features related to the frequency of using a mobile phone and the acoustic parameters from the patient's recordings seem to be reasonable markers that can help determine the current state of the patient. The selection of the most relevant features is a fundamental stage of data modeling and should be given a lot of attention from the very beginning. In medical problems, acquiring important features is associated with a financial cost. Current research aims to analyze whether the acoustic features (with the highest acquisition cost) are all important and demanding to maintain the high performance of classifying the BD state. Moreover, we want to indicate the most significant features that are capable of distinguishing the phase of BD patients. This research is an extension of our latest paper presented during the AIME'19 workshops [8], where the process of feature selection was conducted using manual analysis. It is an important step in the classification process of BD episodes, therefore we apply a cost-constrained automatic feature selection (FS) method, a model-independent filter algorithm. Its objective is to select a subset of features that are the most relevant and at the same time, their summarized cost does not exceed the user-defined budget. Each feature is related to a specific cost that is comparable to the financial cost of acquiring its value for a single observation. In this type of algorithm, it is critical to find the trade-off between the relevancy of the feature subset and its cost.

Some related research indicates that behavioral data collected by mobile phones [4] as well as acoustic features [9] based on mobile recordings are essential factors in the classification of BD episodes. Both authors indicate important features for particular BD states, regardless of the cost of acquiring them. The field of cost-constrained feature selection has primarily focused on the adaptation of traditional algorithms to be cost-constrained. There are modifications to the various filter methods [1,7,11]. In this work, we focus on the filter methods based on the information theory, essentially due to their model-independent nature and ability to detect non-linear interactions between features. For a general review of FS methods, we refer to [3] and in a particular case of a medical project that collects a large amount of data, we refer to [10].

2 Methodology

2.1 Data preprocessing

The used BDMON [8] dataset received with OpenSmile [5] library describes patients diagnosed with BD based on their daily mobile phone calls. The final dataset contains 86 data streams with the main acoustic characteristics of the voice, such as loudness, voice energy or pitch and 11 behavioral features such as the number of incoming/outgoing calls per day or the average length of characters in daily text messages. That set of features was next smoothed in the time domain by applying an average of the day's measurements. Labels that were collected during patients' visits represent 4 bipolar phases: Depression, Mania, Euthymia and Mixed. They were extended for 7 days before the visit and 2 days after the visit [6].

We assume that two types of features: behavioral and acoustic have two different costs assigned. The standard application library[3] was able to collect basic phone call statistics at the time the research started (2017) therefore, we assume that the cost of this feature type is equal to 1. For commercial use,

[3] Android Developer: https://developer.android.com/

the OpenSmile library is an expensive tool. We initially set the cost of each acoustic feature to 5 concerning the time that experienced programmers would need to extract them from mobile recordings. The assigned cost values could be controversial, thus we decided to compare two variants of them (5 and 10). Then compare the results based on the AUC score using Wilcoxon signed rank test, if there is a relevant difference between contentious costs values. We received p value > 0.05, so it indicates that there is no significant difference between the two distributions of firstly selected features therefore, we decided to use the cost equal 5 as a strategy in this research.

2.2 Algorithm

In this paper, we focus on the greedy forward selection algorithm based on information theory [11]. Let $X_1, \ldots, X_p$ be the features, $F = \{1, \ldots, p\}$ be the set of features indexes, Y be a multiclass variable that we consider as a target and $c(k)$ be the cost of the k -th feature. In each step of the algorithm, we add the index of the candidate feature X_k to the set of features already selected in the previous steps $S \leftarrow S \cup \{k^*\}$ by maximizing the following equation:

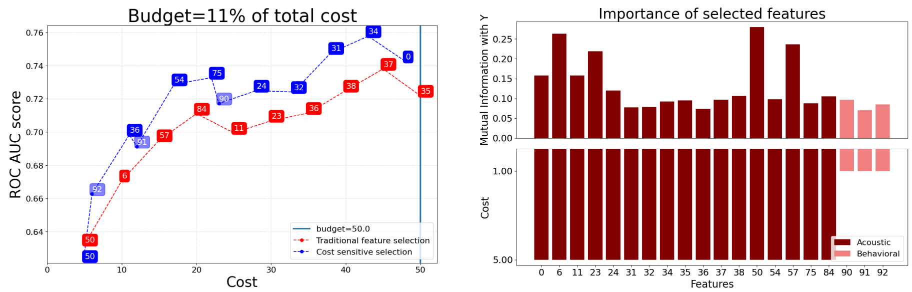
$$k^* = \arg\max_{k \in F \setminus S} [J(X_k, Y|X_S) - \lambda c(k)], \tag{1}$$

where $J(X_k, Y|X_S) = I(Y, X_k) + \sum_{j \in S}[I(X_k, Y|X_j) - I(X_k, Y)]$ measures the informativeness of the added feature X_k in the context of already selected features and $\lambda c(k)$ is the penalty for the cost of the added feature. The $I(Y, X_k)$ is the mutual information metric and it measures the dependence between Y and X_k, $I(X_k, Y|X_j)$ is the conditional mutual information, whereas $I(X_k, Y|X_j) - I(X_k, Y)$ is the interaction information that measures the interaction between X_k and X_j in predicting Y. The general idea behind Equation (1) is that it allows us to find the trade-off between the relevance of the feature and its costs and λ controls the balance. The optimal parameter λ value can be calculated by minimizing the loss function with the cross-validation on the training set.

3 Preliminary results

The main goal of the experiments was to compare the cost-constrained algorithm for feature selection and to compare it with its traditional counterpart. By a traditional algorithm, we mean the method recalled in Equation 1 with $\lambda = 0$. First, the parameter λ was optimized separately for each budget. Then we run a traditional and cost-constrained feature selection algorithm for $\lambda = 0$ and $\lambda = \lambda_{opt}$ respectively. Finally, we scored the Random Forest [2] model trained on the selected features with the AUC score. The optimization process, feature selection and model training were launched on the training data (50%) and the AUC metric is calculated on the test dataset (50%).

Figure 1a presents the AUC of a model trained on features selected and in Figure 1b we can see the importance of all selected features measured as mutual



(a) FS results for budget = 50. (b) Selected features importance

Fig. 1: Feature selection results.

information with the target variable. Variables in the acoustic group result in greater mutual information with the target class, but at the same time, they are very expensive. On the other hand, we have behavioral features, which are 5 times cheaper, but their mutual information with the target class is lower. In the first step of both algorithms, the feature number 50 (magnitude of fast Fourier transform coefficients in band 250-650Hz) from the acoustic group is selected. That is obvious because it is the most informative feature. In the second step cost-constrained method chooses feature number 92 (ratio of outgoing calls to all mobile calls per day) from the behavioral group and the traditional method feature number 6 (another magnitude of fast Fourier transform coefficients) from the acoustic group. The cost-constrained method in the third step selects feature number 36 (energy in the specific band), which combined with two previous features results in 0.70 AUC score, at the same time the traditional method results in 0.67 for almost the same budget. As the budget increases, both methods select more features, but the cost-constrained method tends to result in a higher AUC score. The cost-constrained method continues to be slightly better for classifying BD episodes than the traditional method.

4 Conclusions and future plans

The preliminary results conducted in this paper have shown that using the cost-constrained method let us select features that yield a more accurate classification model when restrictions on the budget are imposed in the BD episode classification problem. The difference in predictive power can be seen especially when the budget is low, for higher budgets both traditional and cost-constrained methods tend to equalize.

The current study points out that the implementation of only a small part of acoustic features (which are time/cost related for researchers) could be sufficient to receive first-worth results. The next steps for that study will be related to an extended version of the cost-constrained feature selection method and a study on datasets and models available for patients with mental disease.

Acknowledgment

Olga Kamińska and Katarzyna Kaczmarek-Majer are supported by the Small Grants Scheme (NOR/SGS/BIPOLAR/0239/2020-00) within the research project: "Bipolar disorder prediction with sensor-based semi-supervised Learning (BIPOLAR)". BDMON data were collected in the CHAD project − entitled "Smartphone-based diagnostics of phase changes in the course of bipolar disorder" (RPMA.01.02.00-14-5706/16-00) financed from EU funds in 2017-2018.

References

1. Bolón-Canedo, V., Porto-Díaz, I., Sánchez-Maroño, N., Alonso-Betanzos, A.: A framework for cost-based feature selection. Pattern Recognition **47**(7), 2481–2489" (2014)
2. Breiman, L.: Random forests. In: Machine Learning. pp. 5–32. 45 (2001)
3. Brown, G., Pocock, A., Zhao, M.J., Luján, M.: Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. Journal of Machine Learning Research **13**(2), 27–66 (2012), http://jmlr.org/papers/v13/brown12a.html
4. Dominiak, M., K.Kaczmarek-Majer, Antosik-Wojcinska, A.Z., Opara, K.R., Wojnar, M., Olwert, A., Radziszewska, W., Hryniewicz, O., Swiecicki, L., Mierzejewski, P.: Behavioural data collected from smartphones in the assessment of depressive and manic symptoms for bipolar disorder patients: Prospective observational study. Journal of Medical Internet Research (2021)
5. Eyben, F., Weninger, F., Gross, F., Schuller, B.: Recent developments in opensmile, the munich open-source multimedia feature extractor. In: Proc. of the 21st ACM Int. Conf. on Multimedia. pp. 835–838 (2013)
6. Grünerbl, A., Muaremi, A., Osmani, V.: Smartphone-based recognition of states and state changes in bipolar disorder patients. IEEE Journal of Biomedical and Health Informatics **19(1)** (2015)
7. Jagdhuber, R., Lang, M., Stenzl, A., Neuhaus, J., Rahnenführer, J.: Cost-constrained feature selection in binary classification: adaptations for greedy forward selection and genetic algorithms. BMC bioinformatics **21**(1), 1–21 (2020)
8. Kamińska, O., Kaczmarek-Majer, K., Opara, K., Jakuczun, W., Dominiak, M., Antosik-Wójcińska, A., Święcicki, Ł., Hryniewicz, O.: Self-organizing maps using acoustic features for prediction of state change in bipolar disorder. Artificial Intelligence in Medicine: Knowledge Representation and Transparent and Explainable Systems (2020)
9. Kamińska, O., Kaczmarek-Majer, K., Hryniewicz, O.: Acoustic feature selection with fuzzy clustering, self organizing maps and psychiatric assessments. Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2020, Lisbon (2020)
10. Remeseiro, B., Bolon-Canedo, V.: A review of feature selection methods in medical applications. Computers in Biology and Medicine **112**, 103375 (2019)
11. Teisseyre, P., Klonecki, T.: Controlling costs in feature selection: Information theoretic approach. In: Proceedings of the International Conference on Computational Science. pp. 483–496. ICCS'21 (2021)