A Feature Selection Method based on Feature Correlation Networks

Miloš Savić[1], Vladimir Kurbalija[1], Mirjana Ivanović[1], and Zoran Bosnić[2]

[1] University of Novi Sad, Faculty of Sciences, Department of Mathematics and Informatics, Serbia
{svc,kurba,mira}@dmi.uns.ac.rs
[2] Univeristy of Ljubljana, Faculty of Computer and Information Science, Slovenia
zoran.bosnic@fri.uni-lj.si

Abstract. Feature selection is an important data preprocessing step in data mining and machine learning tasks, especially in the case of high dimensional data. In this paper we present a novel feature selection method based on complex weighted networks describing the strongest correlations among features. The method relies on community detection techniques to identify cohesive groups of features. A subset of features exhibiting a strong association with the class feature is selected from each identified community of features taking into account the size of and connections within the community. The proposed method is evaluated on a high dimensional dataset containing signaling protein features related to the diagnosis of Alzheimer's disease. We compared the performance of seven widely used classifiers that were trained without feature selection, with correlation-based feature selection by a state-of-the-art method provided by the WEKA tool, and with feature selection by four variants of our method determined by four different community detection techniques. The results of the evaluation indicate that our method improves the classification accuracy of several classification models while drastically reducing the dimensionality of the dataset. Additionally, one variant of our method outperforms the correlation-based feature selection method implemented in WEKA.

Keywords: feature selection, feature correlation networks, community detection, Alzheimer's disease.

1 Introduction

The feature selection problem has been studied by the data mining and machine learning researchers for many years. The main aim of feature selection is to reduce the dimensionality of data such that the most significant aspects of the data are represented by selected features. Consequently, feature selection has become an important data preprocessing step in data mining and machine learning tasks due to the rise of high dimensional data in many application domains. Feature selection usually leads to better machine learning models in terms of prediction accuracy, lower training time and model comprehensibility [29]. The two most

dominant types of feature selection approaches are filter and wrapper methods [9, 15]. Wrapper methods rely on performance of some prespecified classifier to evaluate the quality of selected features. In contrast to wrapper methods, filter methods are independent of learning algorithms. Those methods usually rely on some efficiently computable measure for scoring features considering their redundancy, dependency and discriminative power.

In this paper we present a novel graph-based approach to feature selection. Our feature selection approach belongs to the class of filter-based methods. The main idea of the proposed approach is to select relevant features considering community structure of *feature correlation networks*. A feature correlation network is a weighted graph where nodes correspond to features and links represent their strongest correlations. Feature correlation networks used in our feature selection method are conceptually similar to weighted correlation networks used in the analysis of genomic datasets [11] with one important difference: a class variable (a special feature denoting example classes) is not included as a node in the feature correlation network, but to each node in the feature correlation network is associated a number which specifies the strength of association between the corresponding feature and the class variable.

A *community* (cluster, module or cohesive group) within a weighted network is a subset of nodes such that the weight of links among them is significantly higher than with the rest of the network [17]. We say that a network has a community structure if the set of nodes can be partitioned into communities. The existence of communities is a typical feature of complex networks in various domains [2, 16]. Various community detection techniques enable automatic identification of communities in complex networks [7]. Uncovering communities helps to understand the structure of complex networks on a higher level of abstraction by constructing and analyzing their coarse-grained descriptions (networks of communities). Our approach to feature selection relies on community detection techniques to identify communities of features such that correlations within a community are stronger than correlations between features belonging to different communities. Then, one or more features strongly associated to the class variable is selected to represent each of identified communities taking into account the number of nodes and connections within communities.

The paper is structured as follows. Related work is presented in Section 2. The proposed method for feature selection is described in Section 3. The evaluation of the method is given in Section 4. The last section concludes the paper and gives directions for possible future work.

2 Related Work

Feature selection is a common data mining preprocessing step, which aims at reducing the dimensionality of the original dataset. Adequate selection of features has numerous advantages [24] like: simplification of learning models, improving the performance of algorithms, data reduction (avoidance of curse of dimensionality), improved generalization by reducing overfitting etc.

Wrapper-based feature selection methods estimate usefulness of features using the selected learning algorithm. These methods usually give better results than filter methods since they are adapting their result to a chosen learning algorithm. However, since a learning algorithm is employed to evaluate each subset of features, wrapper methods are very time consuming and almost unusable for high dimensional data. Furthermore, since the feature selection process is tightly interconnected with a learning algorithm, wrappers are less general than filters and have the increased risk of overfitting. On the other hand, filter methods are independent of learning algorithm. They are based only on general features like the correlation with the variable to predict. These methods are generally many times faster than wrappers and robust to overfitting [10]. Recently, some embedded methods are introduced [14] which try combine the positive characteristics of both previous methods.

Relying on the characteristics of data, filter models evaluate features without utilizing classification algorithms. Usually, a filter algorithm has two steps: it ranks features based on certain criteria and it selects the features with highest rankings [6]. Considering the first step, a number of performance criteria have been proposed for filter-based feature selection. Correlation based Feature Selection (CFS) is a simple filter algorithm that ranks features according to a feature-class correlation [10]. The fast correlated-based filter (FCBF) method [29] is based on symmetrical uncertainty, which is defined as the ratio between the information gain and the entropy of two features. The INTERACT algorithm [31] uses the same goodness measure as FCBF filter, but it also includes the consistency contribution as an indicator about how significantly the elimination of particular feature will affect accuracy. The original RELIEF [12] and extended ReliefF [22] algorithms estimate the quality of attributes according to how well their values distinguish between instances that are near to each other but belonging to different classes.

Recently, several approaches proposed feature clustering in order to avoid selection of redundant features [3, 13, 26]. The authors in [25] proposed Fast clustering-bAsed feature Selection algoriThm (FAST). Here, the features are divided into clusters by using graph-theoretic clustering methods and the final subset of features is selected by choosing the most representative feature that is strongly related to target classes from each cluster. Similarly, the approach in [30] proposed hyper-graph clustering to extract maximally coherent feature groups from a set of objects. Furthermore, this approach neglects the assumption that the optimal feature subset is formed by features that only exhibit pairwise interactions. Instead of that, they are using multidimensional interaction information which includes third or higher order dependencies feature combinations in final selection.

Compared to existing graph-based and clustering-based feature selection methods, our approach leans on community detection techniques to cluster graphs that describe the strongest correlations among features. Additionally, the approach takes into account the size of identified communities. In contrast to traditional graph partitioning and data clustering techniques, a majority of

community detection techniques are not computationally demanding and they do not require to specify the number of clusters in advance [7].

3 FSFCN: Feature Selection based on Feature Correlation Networks

The method for feature selection proposed in this paper, denoted by FSFCN, is based on the notion of feature correlation networks. A feature correlation network describes correlations between features in a dataset that are equal or higher than a specified threshold. To formally define feature correlation networks, we will assume that a dataset is composed of data instances having numeric features and a categorical class variable. The below stated definition of feature correlation networks can be adapted in a straightforward manner for other types of datasets (categorical features, a mix of categorical and numeric features, continuous target variable) by taking appropriate correlation measures.

Definition 1 (Feature Correlation Network). Let D be a dataset composed of data instances described by k real-valued features $f_1, f_2, \ldots, f_k \in \mathbb{R}$ and a categorical class variable c. Let $C_f : \mathbb{R} \times \mathbb{R} \rightarrow [-1, 1]$ denote a correlation measure applicable to features (e.g the Pearson or Spearman correlation coefficient) and let C_c be a correlation measure applicable to a feature and the class variable (e.g. the mutual information, the Goodman-Kruskal index, etc.). The feature correlation network corresponding to D is an undirected, weighted, attributed graph $G = (V, E)$ with the following properties:

- The set of **nodes** V corresponds to the set of features ($f_i \in V$ for each i in $[1 \,..\, k]$).
- Two features f_i and f_j, $i \neq j$, are connected by an **edge** $e_{i,j}$ in G, $e_{i,j} \in E$, if $|C_f(f_i, f_j)| \geq T$, where T is previously given threshold indicating a significant correlation between features. The weight of $e_{i,j}$ is equal to $|C_f(f_i, f_j)|$.
- Each node in the network has a real-valued **attribute** reflecting its association with the class variable which is measured by C_c.

The features in D can be ranked according to the C_c measure and highly ranked features can be considered as the most relevant for training a classifier.

Definition 2 (Subset of Relevant Features). A subset F_r of the set of features F is called *relevant* if $(\forall f \in F_r)\, C_c(f) \geq R$ where R denotes a threshold indicating a significant association between a feature and the class variable.

Definition 3 (Pruned Feature Correlation Network). A pruned feature correlation network is a feature correlation network constructed from a subset of relevant features.

Our implementation of the FSFCN method for datasets with real-valued features and categorical class variables uses pruned feature correlation networks which are constructed without explicitly stating the threshold T. This means

that the algorithm for constructing pruned correlation networks has only one parameter R separating relevant from irrelevant features. Additionally, the algorithm uses the Spearman correlation coefficient to determine correlations among relevant features (the C_f measure), while correlations between relevant features and the class variable are quantified by their mutual information (the C_c measure). The mutual information between a real-valued feature f and the categorical class variable c, denoted by $I(f, c)$, can be approximated by

$$I(f,c) \approx \sum_{y \in c} \sum_{x \in f'} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right),$$

where f' is the set of discrete values obtained by a discretization of f, $p(x, y)$ is the joint probability distribution function of f' and c, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of f' and c, respectively. $I(f, c)$ equal to 0 means that f and c are totally unrelated. A larger value of $I(f, c)$ implies a stronger association between f and c.

The algorithm for constructing pruned correlation network consists of the following steps (see Algorithm 1):

1. The subset of relevant features F_r is determined using the mutual information measure. Then, the nodes of the network are created such that each node corresponds to one feature from F_r.
2. For each pair of relevant features f_i and f_j, the algorithm forms a list L, where elements are tuples in the form (f_i, f_j, S_{ij}), where S_{ij} denotes the value of the Spearman correlation coefficient between features f_i and f_j.
3. L is sorted by the third component (S_{ij}) in decreasing order, i.e. the first element of the sorted list is the pair of features exhibiting the highest correlation, while the last element is the pair of features with the lowest correlation.
4. In the last step, the algorithm forms the links of the network by iterating through the sorted list L beginning from the first element. Let $e_k = (f_i, f_j, S_{ij})$ denotes the element processed in the k-th iteration. The algorithm forms a link l_{ij} connecting f_i and f_j with weight S_{ij}. If the addition of l_{ij} results in a connected graph (i.e., a graph that has exactly one connected component or, equivalently, a graph in which there is path between each pair of nodes) then the algorithm stops, otherwise it goes to the next element in the sorted list and repeats the same procedure. In other words, the algorithm iteratively builds the network by connecting features having the highest correlation until the network becomes a connected graph. Consequently, the weight of the last added link determines the value of the threshold T.

The basic idea of the FSFCN method is to cluster a pruned feature correlation network in order to obtain cohesive groups of relevant features such that correlations between features within a group are stronger than correlations between features belonging to different groups. The FSFCN method leans on community detection techniques to identify clusters in feature correlation networks. The

Algorithm 1: Construction of pruned feature correlation networks

input : D, R
D – a dataset of instances with real-valued features $F = \{f_1, f_2, \ldots, f_k\}$ and a categorical class variable c
R – the threshold separating relevant from irrelevant features

output: $G = (V, E)$ – the pruned feature correlation network of D

// determine relevant features and form nodes in G
$F_r :=$ empty set of relevant features
foreach $f \in F$ **do**
  $m :=$ the value of the mutual information of f and c
  if $m \geq R$ **then**
    $F_r := F_r \cup \{f\}$
  end
end
$V := F_r$

// compute the Spearman correlation for each pair of relevant features
$L :=$ empty list of tuples (f_i, f_j, S_{ij})
foreach $(f_i, f_j) \in F_r \times F_r, i \neq j$ **do**
  $s :=$ the value of the Spearman correlation for f_i and f_j
  $L := L + (f_i, f_j, s)$
end
$L :=$ sort L in non-increasing order of the Spearman correlation

// form links
$i := 1$, $cont := \top$
while $cont$ **do**
  $s :=$ the first component of $L[i]$
  $d :=$ the second component of $L[i]$
  $E := E \cup \big\{\{s, d\}\big\}$
  $i := i + 1$
  $cont :=$ G is not a connected graph
end

development of community detection techniques started with Newman and Girvan [18] who introduced a measure called modularity to estimate the quality of a partition of a network into communities. The main idea behind the modularity measure is that a subgraph can be considered a community if the actual number of links connecting nodes within the subgraph is significantly higher than the expected number of links with respect to some null random graph model. In the case of weighted networks, modularity accumulates differences between the total weight of links within a community and the mathematical expectation of the previous quantity with respect to a random network having the same degree and link weight distribution [17].

Definition 4 (Modularity). For weighted networks modularity Q is defined as

$$Q = \sum_{c=1}^{n_c} \left[\frac{W_c}{W} - \left(\frac{S_c}{2W} \right)^2 \right],$$

where n_c is the number of communities in the network, W_c is the sum of weights of intra-community links in c, S_c is the total weight of links incident to nodes in c, and W is the total weight of links in the network.

Four widely used community detection algorithms provided by the iGraph library [5] are employed to detect non-overlapping communities in feature correlation networks:

1. The Greedy Modularity Optimization (GMO) algorithm [4]. This algorithm relies on a greedy hierarchical agglomeration strategy to maximize modularity. The algorithm starts with the partitioning in which each node is assigned to a singleton cluster. In each iteration of the algorithm, the variation of modularity obtained by merging any two communities is computed. The merge operation that maximally increases (or minimally decreases) modularity is chosen and the merge of corresponding clusters is performed.
2. The Louvain algorithm [1]. This method is an improvement of the previous method. The algorithm uses a greedy multi-resolution strategy to maximize modularity starting from the partition in which all nodes are put in different communities. When modularity is optimized locally by moving nodes to neighboring clusters, the algorithm creates a network of communities and then repeats the same procedure on that network until a maximum of modularity is obtained.
3. The Walktrap algorithm [19]. This algorithm relies on a node distance measure reflecting probability that a random walker moves from one node to another node in exactly k steps (k is the only parameter of the algorithm having default value $k = 4$). The clustering dendrogram is constructed by Ward's agglomerative clustering technique and the partition which maximizes modularity is taken as the output of the algorithm.
4. The Infomap algorithm [23]. This method reveals communities by optimally compressing descriptions of information flows on the network. The algorithm uses a greedy strategy to minimize the map equation which reflects the expected description length of a random walk on a partitioned network.

Each of used community detection algorithms defines one concrete implementation instance (i.e. one variant) of the FSFCN method.

The final step in the FSFCN method is the selection of features according to obtained community partitions in pruned feature correlation networks. The main idea is to select one or more features within each community such that:

1. selected features have a strong association with the class variable, and
2. any two selected features belonging to the same community are not directly connected.

Algorithm 2: The FSFCN algorithm

input : D, R, CDA
D – a dataset of instances with real-valued features $F = \{f_1, f_2, \ldots, f_k\}$ and a categorical class variable c
R – the threshold separating relevant from irrelevant features
CDA – a community detection algorithm

output: S – the set of selected features

// form the pruned feature correlation network corresponding to D
$G :=$ Algorithm1(D, R)

$C :=$ the set of clusters in G obtained by CDA

$S :=$ empty set
foreach $c \in C$ **do**
  $(V_q, E_q) :=$ subgraph of G induced by nodes in c
  while $V_q \neq$ *empty set* **do**
    // determine feature having the highest mutual information
    // with the class variable
    $f := \text{argmax}_{x \in V_q}\ C_c(x)$

    // remove f and its neighbors from (V_q, E_q)
    $V_r := \{a \in V_q\ :\ \{f, a\} \in E_q\} \cup \{f\}$
    $E_r := \{\{a, b\} \in E_q\ : a \in V_r \vee b \in V_r\}$
    $V_q := V_q \setminus V_r$
    $E_q := E_q \setminus E_r$

    // add f to the set of selected features
    $S := S \cup \{f\}$
  end
end

The procedure for forming the set of selected features is described in Algorithm 2.

After the pruned correlation network is constructed and clustered, the FSFCN method forms subgraphs of the network corresponding to identified communities where one subgraph is induced by nodes belonging to one community. For each of community subgraphs the following operations are performed:

1. A feature having the highest association with the class variable is identified and put in the set of selected features. Then, it is removed from the community subgraph together with its neighbors.
2. The previous step is repeated while the community subgraph is not empty.

In other words, for each of identified communities the FSFCN method selects one or more features which represent the whole community. The method also takes into account the size of communities – for larger communities a higher number of features is selected. Also, when a feature is added to the set of selected features its

neighbors are removed from the community subgraph which implies that the set of selected features will not contain features having a high mutual correlation (otherwise, such two features would be directly connected in the community subgraph).

4 Experiments and Results

The experimental evaluation of the FSFCN feature selection method was performed on a dataset with 120 plasma signaling protein features related to the diagnosis of Alzheimer's disease [21]. The class variable indicates whether a patient was diagnosed with Alzheimer's or not. The total number of instances in the dataset is equal to 176 where 64 data instances correspond to patients diagnosed with Alzheimer's.

We performed feature selection using 4 variants of the FSFCN method. Each of those variants relies on a different community detection technique to cluster pruned feature correlation networks obtained at the threshold R equal to 0.05. The variants of the method are denoted by:

1. FG – the FSFCN method with the Fast greedy modularity optimization community detection algorithm,
2. LV – the FSFCN method with the Louvain algorithm,
3. WT – the FSFCN method with the Walktrap algorithm, and
4. IM – the FSFCN method with the Infomap algorithm.

Using the WEKA machine learning workbench [28, 8] we trained 7 different classifiers on datasets containing features selected by different variants of the FSFCN method. The examined classification models are denoted by:

1. RF – the random forest classifier,
2. J48 – the C4.5 decision tree classifier,
3. LMT – the logistic model tree classifier,
4. JRIP – the RIPPER rule induction classifier,
5. LOGR – the logistic regression classifier,
6. SMO – the Support Vector Machine classifier, and
7. NB – the Naive Bayes classifier.

The classifiers were trained and evaluated using the 10-fold cross-validation procedure with the default WEKA values for the parameters of classification algorithms. We used the classification accuracy measure (the fraction of correctly classified data instances) to compare the performance of classifiers. The classification accuracy of classifiers trained after feature selection by different variants of the FSFCN method was also compared to the accuracy of the same classifiers trained on the full dataset (the original dataset without any feature selection) and the dataset containing features selected by the CFS method [10] provided by WEKA.

The pruned feature correlation network of the dataset contains 35 nodes which means that 35 out of 120 features exhibit significant association with the

class variable in terms of mutual information. Those 35 nodes representing relevant features are connected by 161 links which implies that a randomly selected relevant feature has a significant correlation with 9.2 other relevant features on average. The maximal and the minimal absolute value of link weights are 0.72 and 0.32, respectively, which means that there are moderate to strong Spearman correlations among relevant features.

The results of community detection on the pruned feature correlation network are summarized in Table 1. To compare obtained community partitions we computed the Rand index [20] for each pair of them. The FG and LV methods identified exactly the same communities: the Rand index for the community partitions obtained by FG and LV is equal to 1. The WT method identified a partition with a higher number of communities and a lower value of modularity compared to FG/LV. The Rand index between partitions obtained by WT and FG/LV is equal to 0.79 which indicates that those two partitions are highly similar. Finally, it can be seen that the IM method failed to identify communities in the network, i.e. this method identified one community encompassing all nodes in the network. Consequently, the features selected by this method can be seen as features selected from the pruned correlation network without the clustering step.

Table 1. The results of community detection on the pruned feature correlation network. NC – the number of identified communities, Q – the value of the modularity measure, S – the vector giving the size of identified communities.

Method	NC	Q	S
FG	4	0.275	(14, 8, 7, 6)
LV	4	0.275	(14, 8, 7, 6)
WT	7	0.218	(13, 8, 5, 5, 2, 1, 1)
IM	1	0	(35)

The features selected by different variants of the FSFCN method are shown in Table 2. FS and LV selected the same features since community partitions obtained by the corresponding community detection algorithms are equal. It can be observed that each of the FSFCN variants drastically reduced the dimensionality of the dataset – the number of selected features varies from 7 to 12. On the other hand, the CFS method implemented in WEKA selected 25 features.

The accuracy of trained classifiers are shown in Table 3. It can be observed that classifiers trained without feature selection tend to exhibit the worst performance. The classifiers trained on the dataset containing features selected by the WEKA CFS method are always better than the classifiers trained on the full dataset. On the other hand, the classifiers trained on the datasets containing features selected by FG/LV and WT show a better performance compared to the classifiers trained on the full dataset except in one case. Namely, the accu-

Table 2. The features selected by four different variants of the FSFCN method. Feature ranks are determined according to the mutual information with the class variable.

FG/LV		WT		IM	
	Rank		Rank		Rank
IL-1a	1	IL-1a	1	IL-1a	1
IL-8	2	TNF-a	3	PDGF-BB	7
TNF-a	3	GCSF	6	sTNF RI	12
PDGF-BB	7	PDGF-BB	7	Eotaxin	15
sTNF RI	12	sTNF RI	12	MCP-2	17
VEGF-B	14	Eotaxin	15	IGFBP-2	23
Eotaxin	15	SCF	16	TPO	31
MIP-1d	19	MIP-1d	19		
IGFBP-2	23	CTACK	22		
		IGFBP-2	23		
		BTC	30		
		TPO	31		

racy of the LMT classifier trained on the full dataset is equal to the accuracy of the same classifier trained after feature selection based on the FG/LV and WT methods. Consequently, we can say feature selection based on properly clustered feature correlation networks does not decrease the performance of all examined classifiers while drastically reducing the dimensionality of the dataset.

Table 3. The classification accuracy of examined classifiers. The column FULL corresponds to classifiers trained without feature selection, while the column WEKA-CFS corresponds to classifiers trained on the dataset containing features selected by the CFS feature selection method implemented in WEKA. One star indicates the worst performance, while two stars indicate the best performance.

	FULL	WEKA-CFS	FG/LV	WT	IM
RF	0.82	0.85**	0.82	0.85**	0.79*
J48	0.74*	0.77	0.77	0.81**	0.74*
LMT	0.84	0.85**	0.84	0.84	0.83*
JRIP	0.72*	0.81**	0.79	0.78	0.75
LOGR	0.73*	0.81	0.85**	0.85**	0.84
SMO	0.82*	0.83	0.84	0.86**	0.85
NB	0.78*	0.84	0.88**	0.88**	0.84

The next important result that can be observed in Table 3 is that the IM variant of the FSFCN method exhibits the worst performance compared to other three FSFCN variants. The IM variant is actually equivalent to the FSFCN

method without clustering since IM identified exactly one cluster encompassing features in the network. Therefore, we can conclude that clustering of feature correlation networks enables a better selection of relevant features.

The best performing classifier trained without feature selection is LMT achieving accuracy of 0.84, the best classifier trained after the WEKA CFS feature selection are RF and LMT achieving accuracy of 0.85. On the other hand, the classifier with the highest accuracy is NB trained after features are selected by three different variants of the FSFCN method. Finally, the classifiers trained after the WT variant of the FSFCN method tend to exhibit the best overall performance. It can be observed that this feature selection method outperforms other feature selection methods in case of 5 out of 7 classifiers. It is also important to emphasize that this variant of the FSFCN method drastically improves the performance the J48, LOGR and NB in comparison with the classifiers trained on the full dataset.

5 Conclusions and Future Work

In this paper we presented a novel method for feature selection based on feature correlation networks. Feature correlation networks are weighted graphs showing the strongest correlations between features. The main idea of the approach is to cluster feature correlation networks using community detection techniques in order to identify groups of features such that correlations between features within a group tend to be stronger than correlations between features belonging to different groups. Then, one or more features representing each group is selected considering their correlations with the class variable, the size of groups and connections within them.

The experimental evaluation of four variants of the method, each of them relying on a different community detection technique, was conducted on a highly dimensional dataset (120 features) related to the diagnosis of Alzheimer's disease. More specifically, we compared the accuracy of 7 different classifiers trained without feature selection, with feature selection by the CFS method implemented in WEKA, and with feature selection performed by the variants of our method. The obtained results show that the variant of our method which employs the Walktrap community detection algorithm exhibits the best overall performance compared to the alternatives. Additionally, our results indicate that clustering of feature correlation networks is a necessary step to obtain relevant sets of features for classification purposes.

The main task in our future work will be to perform a more comprehensive evaluation of our approach considering high dimensional datasets from various domains. The evaluation will also include a statistically robust comparison with other representative graph-based and clustering-based feature selection methods. It is also possible to experiment with additional variants of the method taking into account other correlation measures and community detection algorithms (including also detection of overlapping communities [27]). Finally, in this paper we focused on feature selection in the context of classification. In our future

work we will also focus on adaptations of the method for clustering problems. Currently, the selection of features representing feature clusters is guided by the mutual information between a feature and the class variable. We plan to examine different network centrality measures instead of the mutual information in order to be able to apply the method on uncategorized data and investigate its performance in this setting.

Acknowledgments. This work is supported by the bilateral project "Intelligent computer techniques for improving medical detection, analysis and explanation of human cognition and behavior disorders" between the Ministry of Education, Science and Technological Development of the Republic of Serbia and the Slovenian Research Agency. M. Savić, V. Kurbalija and M. Ivanović also thank the Ministry of Education, Science and Technological Development of the Republic of Serbia for additional support through project no. OI174023, "Intelligent techniques and their integration into wide-spectrum decision support."

References

1. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008(10), P10008 (2008)
2. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U.: Complex networks: Structure and dynamics. Physics Reports 424(4-5), 175–308 (2006)
3. Butterworth, R., Piatetsky-Shapiro, G., Simovici, D.A.: On feature selection through clustering. In: Proceedings of the Fifth IEEE International Conference on Data Mining. pp. 581–584. ICDM '05, IEEE Computer Society, Washington, DC, USA (2005)
4. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Physical Review E 70, 066111 (Dec 2004)
5. Csardi, G., Nepusz, T.: The igraph software package for complex network research. InterJournal Complex Systems p. 1695 (2006)
6. Duch, W.: Filter Methods, pp. 89–117. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)
7. Fortunato, S.: Community detection in graphs. Physics Reports 486(35), 75 – 174 (2010)
8. Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I.H., Trigg, L.: Weka-A Machine Learning Workbench for Data Mining, pp. 1269–1277. Springer US, Boston, MA (2010)
9. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157–1182 (Mar 2003)
10. Hall, M.A.: Correlation-based Feature Subset Selection for Machine Learning. Ph.D. thesis, University of Waikato, Hamilton, New Zealand (1998)
11. Horvath, S.: Correlation and Gene Co-Expression Networks, pp. 91–121. Springer New York, New York, NY (2011)
12. Kononenko, I.: Estimating attributes: Analysis and extensions of RELIEF, pp. 171–182. Springer Berlin Heidelberg, Berlin, Heidelberg (1994)

13. Krier, C., Franois, D., Rossi, F., Verleysen, M.: Feature clustering and mutual information for the selection of variables in spectral data. In: Proc European Symposium on Artificial Neural Networks Advances in Computational Intelligence and Learning. pp. 157–162 (2007)
14. Lal, T.N., Chapelle, O., Weston, J., Elisseeff, A.: Embedded Methods, pp. 137–165. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)
15. Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H.: Feature selection: A data perspective. arXiv preprint arXiv:1601.07996 (2016)
16. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45(2), 167–256 (2003)
17. Newman, M.E.J.: Analysis of weighted networks. Physical Review E 70, 056131 (Nov 2004)
18. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E 69, 026113 (Feb 2004)
19. Pons, P., Latapy, M.: Computing communities in large networks using random walks. Journal of Graph Algorithms and Applications 10(2), 191–218 (2006)
20. Rand, W.M.: Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association 66(336), 846–850 (Dec 1971)
21. Ray, S., Britschgi, M., Herbert, C., Takeda-Uchimura, Y., Boxer, A., Blennow, K., Friedman, L., Galasko, D., Jutel, M., Karydas, A., Kaye, J., Leszek, J., Miller, B., Minthon, L., Quinn, J., Rabinovici, G., Robinson, W., Sabbagh, M., So, Y., Sparks, D., Tabaton, M., Tinklenberg, J., Yesavage, J., Tibshirani, R., Wyss-Coray, T.: Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. Nature Medicine 13(11), 1359–1362 (11 2007)
22. Robnik-Šikonja, M., Kononenko, I.: Theoretical and empirical analysis of relieff and rrelieff. Machine Learning 53(1), 23–69 (2003)
23. Rosvall, M., Bergstrom, C.T.: Maps of information flow reveal community structure in complex networks. Proceedings of the National Academy of Sciences of the United States of America 105(4), 1118–1123 (2007)
24. Sánchez-Maroño, N., Alonso-Betanzos, A., Tombilla-Sanromán, M.: Filter Methods for Feature Selection – A Comparative Study, pp. 178–187. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
25. Song, Q., Ni, J., Wang, G.: A fast clustering-based feature subset selection algorithm for high-dimensional data. IEEE Transactions on Knowledge and Data Engineering 25(1), 1–14 (Jan 2013)
26. Van Dijck, G., Van Hulle, M.M.: Speeding Up the Wrapper Feature Subset Selection in Regression by Mutual Information Relevance and Redundancy Analysis, pp. 31–40. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)
27. Wang, M., Yang, S., Wu, L.: Improved community mining method based on LFM and EAGLE. Computer Science and Information Systems 13(2), 515–530 (2016)
28. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2005)
29. Yu, L., Liu, H.: Feature selection for high-dimensional data: A fast correlation-based filter solution. In: Fawcett, T., Mishra, N. (eds.) Proceedings of the 20th International Conference on Machine Learning (ICML-03). pp. 856–863 (2003)
30. Zhang, Z., Hancock, E.R.: A Graph-Based Approach to Feature Selection, pp. 205–214. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
31. Zhao, Z., Liu, H.: Searching for interacting features. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence. pp. 1156–1161. IJCAI'07, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2007)