

Improving Image steganalysis performance using a graph-based feature selection method

Amir Nouri¹, Alimohammad Nazari²

¹ Islamic Azad University, Ashtian Branch, Faculty of Computer Engineering, Ashtian, Iran
amirnouiri0016@gmail.com

² Department of Mathematics, Arak university, Arak, Iran, P.O. Box, 38156-8-8349
a-nazari@araku.ac.ir

Abstract

Steganalysis is the skill of discovering the use of steganography algorithms within an image with low or no information regarding the steganography algorithm or/and its parameters. The high-dimensionality of image data with small number of samples has presented a difficult challenge for the steganalysis task. Several methods have been presented to improve the steganalysis performance by feature selection. Feature selection, also known as variable selection, is one of the fundamental problems in the fields of machine learning, pattern recognition and statistics. The aim of feature selection is to reduce the dimensionality of image data in order to enhance the accuracy of Steganalysis task. In this paper, we have proposed a new graph-based blind steganalysis method for detecting stego images from the cover images in JPEG images using a feature selection technique based on community detection. The experimental results show that the proposed approach is easy to be employed for steganalysis purposes. Moreover, performance of proposed method is better than several recent and well-known feature selection-based Image steganalysis methods.

Keywords: Image steganalysis; Feature selection; Graph clustering; Feature clustering.

1. Introduction

Data hiding is a collection of techniques to embed secret data into digital media. These techniques can be used in many different application scenarios, such as secret communications, copyright protection or authentication of digital contents, among others. Images are the most common carriers for data hiding because of their widespread use in the Internet.

Within data hiding, steganography is a major branch whose goal is to secretly communicate data, making it undetectable for an attacker. On the other hand, steganalysis is another branch whose goal is to detect messages previously hidden using steganography.

One of the most important recent discussions in information forensics is related to the improvement of steganalysis performance and protection of concealed

information. On the other hand, it is important to transfer the most important information through a safe way somehow; it cannot be attacked, detected and accessed using steganography techniques. Several methods have been presented to improve the steganalysis performance by increasing feature space to be blind steganalysis that take a long time to steganalysis.

In recent years, with advancement of science and technology, images have grown hugely and now include large number of features. Accordingly machine learning methods often deal with samples consisting of thousands of features. In high-dimensional image, typically many features are irrelevant and/or redundant for a given learning task, having harmful consequences in terms of performance and/or computational cost [1-3].

Moreover, a large number of features require a large storage space. To deal with such datasets, several dimensionality reduction methods have been proposed in literature with the goals of reducing the computational cost and improving the general abilities of the learning models [4, 5]. Clearly, computation time to build models with smaller number of features will be lower than large ones. Moreover, low-dimensional representation of the problem reduces the risk of “over fitting”. Furthermore, dimensionality reduction methods’ provide us a way to better understanding of the data in machine learning or pattern recognition applications.

Many of the image steganalysis methods in the state-of-the-art use feature based steganalysis and machine learning classification. In order to apply this methodology, the steganalysis needs to extract a set of features from a training data set and train a classifier. Then, the classifier is tested using a testing data set and, if the results are satisfactory, the classifier is considered successful.

Since the presentation of a method to handle both irrelevant and redundant features in an acceptable time is an important issue, a major purpose of the current study is to attempt to select a high quality feature subset within a reasonable time.

In this paper, we present novel graph-based feature selection methods using feature clustering. In this paper, a graph-based algorithm is applied on the given data to obtain the reduced features. However, the performance of the features totally depends on the clustering algorithm and also the features may not themselves be the most discriminative features. In the present work, graph clustering algorithm is used to feature selection. Moreover, the methods use criteria to analyze the relevance and redundancy of the features which are used as graph representation to guide the search process.

The rest of the paper is organized as follows: Section 2 gives a brief review of previous works. Section 3 presents the preliminary concepts of feature selection, Section 4 presents the proposed feature selection method based on graph theoretic approach. Section 5 reports the experimental results on well-known medical dataset. Finally, Section 6 presents the conclusion.

2. Related Work

The data set obtained after feature extraction depends on many factors, such as the steganographic algorithm used for hiding data into the cover source, the algorithm used for feature extraction or

the properties of the cover source in different aspects (e.g. size, noise and hardware used for acquisition) If similar cover source is used, the feature extraction process provides data sets with similar representation and, therefore, the machine learning tools work properly and the classification results are satisfactory. However, if different cover source is used, the data sets obtained by feature extraction are also different, producing a degradation of the sification results. Machine learning [6] literature refers to this problem as domain adaptation, whereas the term used to refer to this situation in steganalysis is cover source mismatch (CSM).

Several approaches to deal with the CSM problem have been proposed in the recent years. In the BOSS competition [7], the BOSSrank database (which suffers from CSM) had to be used as a testing set. Some participants of the competition tried to include the testing set images in the training set. This idea was called “training on a contaminated database”. This approach consists in applying denoising algorithms to estimate the cover sources of the testing set and using these estimated covers to generate new stego samples, by embedding new information into them. After that these new estimated cover and stego samples are included in the training set.

In 2012, a solution based on training a classifier with a huge variety of images was proposed [8]. This approach consists in applying machine learning to millions of images. Due to the high time and memory requests, this step is performed using on-line classifiers. Later on, in

2013, the use of rich features in universal steganalysis was analyzed [9]. Since rich features are not sensitive enough for their application in universal steganalysis, the authors apply linear projections informed by embedding methods and an anomaly detector. This approach tries to make these projections sensitive to stego content and, at the same time, insensitive to cover variation.

In 2014, different methods to deal with CSM were presented [10] show the possibility of centering features when there is a shift in the cover sources, by subtracting an estimated centroid of the cover features.

In recent years, a considerable number of research studies have adopted dimension reduction as a pre-analysis processing to separate the irrelevant and unimportant features from the relevant and important features. This process can be classified into feature selection methods and feature extraction methods. The former are techniques of selecting a possible features set from the whole set of candidate features. The latter are techniques that have been used to extract many features from the original data (Image) in order to generate dataset. Furthermore, The feature selection techniques are a subset of the more general field of feature extraction.

Based on whether an image includes hidden message or not, images can be categorized into the image with no hidden message called cover-image, and the image with a message hidden called stego-image. Steganalysis can thus be concentrated as a pattern recognition procedure to decide which class a test image belongs to. The important issue for steganalysis just similar to pattern recognition is feature extraction. The features must be sensitive to the data hiding procedure. That is, the features must be somewhat dissimilar for the cover image and for the stego-image. The superior the difference, the better the features are. The features must be as common as possible, i.e., they are effective to overall different kinds of images and different data hiding systems.

Unlike feature extraction, feature selection techniques have been applied to data sets with identified features. These techniques try to identify the important features and remove irrelevant or redundant features from the whole features set. The feature selection process looks for and selects the optimal subset of feature that can effect on the whole contents of the dataset with the minimum error and information loss.

3. Feature selection

During the last decade, the motivation for applying feature selection (FS) techniques has shifted from being an optional subject to becoming a real prerequisite for model building. The main reason for this change is the high-dimensional nature of many modeling tasks. The selection of features and the removal or reduction of redundant

information unrelated to the classification task on hand will not only reduce the complexity of the problem and improve the efficiency of the processing [11]. Feature selection helps to improve classification performance (accuracy, etc.) and also to obtain more interpretable classifiers or to detect outliers

Dimensionality reduction techniques can be mainly classified into feature extraction and feature selection methods [4, 12]. In feature extraction approach, features are projected into a new space with lower dimensionality. Examples of feature extraction technique include Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Singular Value Decomposition (SVD), to name a few. On the other hand, the feature selection approach aims to select a small subset of features that minimize redundancy and maximize relevance to the target (i.e. class label). Feature selection has been established as an important technique in many practical applications such as text processing [13-15], face recognition [16, 17], image retrieval [18, 19], medical diagnosis [20, 21], knowledge-based authentication [22] and Image steganalysis [23]. Feature selection has been a fertile field of research and development since 1970s in statistical pattern recognition, machine learning, data mining, and there have been a number of attempts to review the feature selection methods [24-26].

feature selection methods that can be classified into four categories including filter, wrapper, embedded, and hybrid approaches. Moreover, graph based feature selection methods are also reviewed.

The filter approach relies on the characteristics of the learning data and selects a subset of features without involving any learning model. Thus, the methods in this approach are typically fast. The filter-based feature selection methods can be classified into univariate and multivariate methods. In the univariate methods, informativeness of each feature is evaluated individually, according to a specific criterion, such as the Information gain [27], gain ratio [6], term variance (TV) [28], Laplacian score (LS) [29], and Fisher score (FS) [30]. This means that each feature is considered separately, thereby ignoring feature dependencies may lead to reduce classification performance compared to other types of feature selection methods. On the other hand, the multivariate approach, evaluates the relevance of the features considering how they function as a group, taking into account their dependencies [31, 32].

In contrast, the wrapper approach requires one predetermined mining algorithm and uses its performance to evaluate and determine which features are selected. The wrapper approach applies a specific learning model in the feature selection process to evaluate a subset of selected features iteratively, and then the accuracy of the learning model is used to guide the search process. [25, 26]. The wrapper based feature selection methods apply a learning

algorithm to evaluate the quality of feature subsets in the search space iteratively. These methods can effectively identify and remove irrelevant and redundant features. [25]. Due to the frequent use of the learning algorithm in the search process, this model requires high computational time, especially for high-dimensional datasets [3, 20].

In practice, the filter methods have much lower computational complexity than the wrappers; meanwhile, they achieve comparable classification accuracy for most classifiers. Thus, the filter methods are very popular to high-dimension data set.

The hybrid approach attempts to take advantage of the filter and wrapper approaches. It is often found that, hybrid technique capable of locating a good solution, while a single technique often traps into an immature solution [2, 33]. Hybrid methods use the ranking information obtained using filter methods to guide the search in the optimization algorithms used by wrapper methods..

In addition, recently a new category of feature selection approach has been proposed. That is to say, the embedded approach in which FS process is integrated with the classifier construction. The performance of this approach is similar to wrapper approach, since the main concern is to the interaction between the feature selection and classification [34, 35].

Moreover, according to whether the class labels of training data are available, feature selection algorithms can be roughly grouped into two families, i.e., supervised feature selection and unsupervised feature selection.

4. Proposed method

Recently, the graph-based methods, such as spectral embedding[36], spectral clustering [37], and semi-supervised learning [38], have played an important role in machine learning due to their ability to encode similarity relationships among data. The best known methods are the Fisher score [30] and Laplacian score[29], both of them are belong to general graph-based feature selection methods.

In this section a novel method is described which can efficiently and effectively deal with both the irrelevant and redundant features. The proposed method consists of three steps including: (1) graph representation of the problem space, (2) graph clustering for clustering the original features, and (3) select the best final features from each cluster by applying the term variance.

In the first step feature set is represented as a weighted graph in which each node in the graph denotes a feature and each edge weight indicates similarity value between its corresponding features. In the second step, the features are divided into several clusters using a specific community detection method. The goal of clustering features is to group most correlated features into the same

cluster. Finally, in the third step a novel algorithm based on node centrality concepts is proposed to select the most informative features from each cluster.

4.1 Graph representation

A preliminary step for all graph-based methods is to represent training data with an undirected graph. For this purpose, the feature set is mapped into its equivalent graph $G = (F, E, w_F)$, where $F = \{F_1, F_2, \dots, F_n\}$ is a set of original features, $E = \{(F_i, F_j) : F_i, F_j \in F\}$ denotes the edges of graph and w_{ij} indicates similarity between two features F_i and F_j connected by the edge (F_i, F_j) . Different measures for computing vertex similarities (i.e. edge weights) leads to different performances on the graph-based feature selection methods. In this work, we have used well-known Pearson product-moment correlation coefficient [3] to measure similarity between different features of a given training set.

4.2 Feature clustering

Feature clustering is an efficient approach for dimensionality reduction [39, 40]. The main idea of feature clustering is to group original features into different clusters based on their similarities; thus, the features in the same clusters are similar to each other. In this paper community detection is used to feature clustering. Community detection is often divided into groups or communities with dense connections within communities and sparse connections between communities. Detection of community structures in the weighted complex networks is significant to understand the network structures and analysis of the network properties. In recent years, community detection has been in the center of attention due to its wide use in data mining, information retrieval and social network analysis. [41, 42]. Classical clustering approach, k-means, has been shown to be very efficient to detect communities in networks. However, k-means is quite sensitive to the initial centroids or seeds, especially when it is used to detect communities [41, 43].

In this work, we have used the Louvain community detection algorithm [44] to identify the feature clusters. The Louvain Method for community detection is a method to extract communities from large networks. [44]. This method outperforms other methods in terms of computation time, which allows us to analyze networks of unprecedented size.

In the Louvain Method of community detection, first small communities are found by optimizing modularity locally on all nodes, then each small community is grouped into one node and the first step is repeated.

In order to maximize this value efficiently, the Louvain Method has two phases that are repeated iteratively. First, each node in the network is assigned to its own community. Then for each node i , the change in modularity is calculated for removing i from its own community and moving it into the community of each neighbor j of i . This value is easily calculated by:

$$\Delta Q = \frac{\sum_C + k_i^C}{2m} - \left(\frac{\sum_C + k_i}{2m} \right)^2 - \left[\frac{\sum_C}{2m} - \left(\frac{\sum_C}{2m} \right)^2 - \left(\frac{k_i}{2m} \right) \right] \quad (1)$$

where \sum_C is sum of the edges weights' inside C , $\sum_{\hat{C}}$ is sum of the edges weights' incident to nodes in C , k_i is sum of edges weights' incident to node i , k_i^C is sum of the edges weights' from i to nodes in C , and m is sum of the all edges weights' in the network. The second step simply makes a new network consist of nodes that are those communities previously found. Then the process iterates until a significant improvement of the network modularity is obtained. The method is also implemented in several software's of network analysis, including NetworkX [45] and Gephi [46].

4.3 Select final feature set

The main purpose of this step is to identify relevant and influential features from each cluster. To this end, term variance is utilized to identify final feature set. Term variance of feature can be calculated as follows:

$$TV(F_i) = \frac{1}{|S|} \sum_{j=1}^{|S|} (A_{ij} - \bar{A}_i)^2 \quad (2)$$

where A_{ij} indicates the value of feature F_i for the pattern j , and $|S|$ is the total number of patterns. After calculating the efficient value of features, some feature with efficient value less than δ parameter are removed and reminder feature are select as final feature set.

5. Experimental results

The experiments have been run on a machine with a 3.2 GHz CPU and 2GB of RAM. Moreover, the proposed method was compared to the well-known and state-of-the-art unsupervised filter-based methods including: Laplacian Score (LS) [29], Relevance-redundancy feature selection (RRFS) [31], Fisher Score (F-Score) [30], Minimal-redundancy-maximal-relevance (mRMR) [47].

Three sets of experiments were carried out in steganalysis. In experimental studies, the breaking out steganography system [48] (BOSS), version 1.01 grey scale image databases is employed that the rate of hidden text embedding is 0.4 per pixel. This database contains 10,000 cover images and 10,000 stego images. Likewise, the number of classes in our experiments is two.

In this paper, both the subtractive pixel adjacency model (SPAM) method and CC-PEV are employed to extract the features for steganalysis. SPAM has 686 features and one

class feature, and CC-PEV, 548 features and one class feature.

To show the generality of the proposed method, the classification prediction capability of the selected features was tested using several well-known classical classifiers:

Support Vector Machine (SVM) is supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. The goal of SVM is to maximize margin between data samples.

Decision Tree (DT) a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility.

Naïve Bayes (NB) is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set.

K-Nearest Neighbor (KNN) is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition.

In this section performance of the proposed method was evaluated over different classifiers. Tables 1 and 2 show the average classification accuracy (in %) of the proposed method with ten independent runs over the SVM, DT, NB and KNN classifiers respectively. Moreover the results are compared with those of the unsupervised filter methods including LV, RRFS, FS and mRMR.

Table 1: Accuracy in SPAM dataset

	LS	RRFS	FS	mRMR	Proposed method
SVM	57.32	59.11	56.32	58.73	60.14
DT	56.38	58.19	56.38	57.69	61.19
NB	58.13	58.48	56.78	55.18	60.78
KNN	57.89	57.91	57.12	58.19	61.69

Table 2: Accuracy in CC-PEV dataset

	LS	RRFS	FS	mRMR	Proposed method
SVM	67.78	69.17	69.38	67.91	71.44
DT	66.27	68.87	68.49	68.09	71.54
NB	68.49	68.78	66.59	65.13	70.36
KNN	67.18	67.29	67.26	68.19	71.03

From the results it can be observed that in most cases the proposed method obtained the highest classification

accuracy compared to those of unsupervised feature selection methods.

Moreover, several experiments were conducted to compare the accuracy of the proposed method with the other feature selection methods based on the different numbers of selected features. Figs. 1 and 2 plot the classification accuracy (average over 10 independent runs) curves of SVM and DT classifiers on SPAM and CC-PEV datasets, respectively. In these plots, the x-axis denotes the subset of selected features, while the y-axis is the average classification accuracy. Figs. 1 and 2 illustrates that the performance of the proposed method is superior to the performances of all methods for different numbers of selected features when the SVM classifier are used in the experiments.

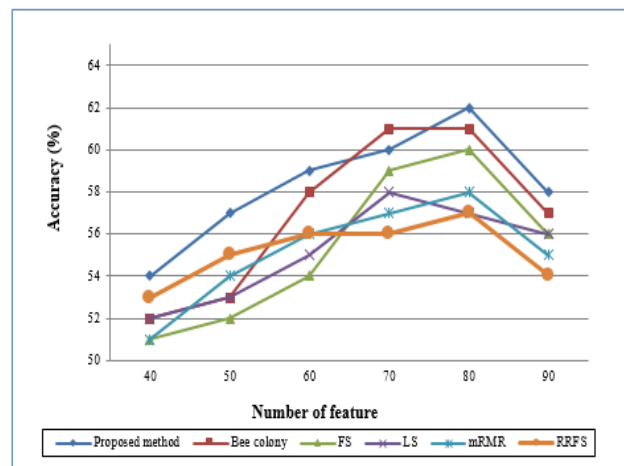


Fig 1: Accuracy in SPAM dataset with number of different feature

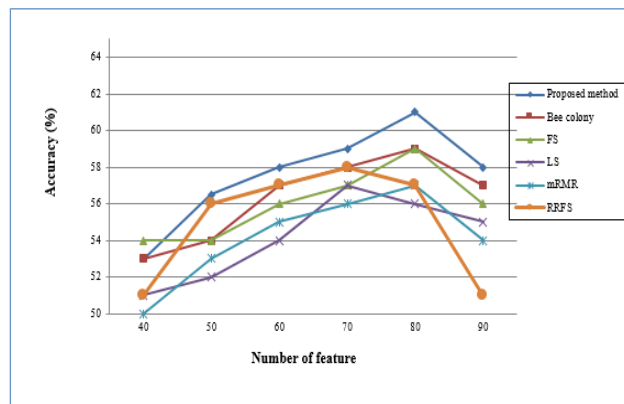


Fig 2: Accuracy in CC-PEV dataset with number of different feature

6. Conclusions

In Image steganalysis problems, an image usually involves a large number of features, often including relevant, irrelevant and redundant features. However,

irrelevant and redundant features are not useful for steganalysis and they may even reduce the performance due to the large search space. In this paper, novel graph-based-feature selection methods were proposed based on the feature clustering by analyzing the relevance and redundancy of features. The proposed methods combined the efficiency of the filter model with the advantages of the graph representation. Moreover, a term variance criteria is used to consider the dependencies between subsets of features which enhanced the quality of the found solution.

To show the usefulness of the proposed algorithm, and compare with well-known feature selection methods, three sets of experiments were carried out using both SPAM and CC-PEV data set of features in Image steganalysis. Experimental results show that proposed method is significantly superior to the existing methods over different classifiers and datasets.

References

- [1] J. M. Cadenas, M. C. Garrido, and R. Martínez, "Feature subset selection Filter-Wrapper based on low quality data," *Expert Systems with Applications*, vol. 40, pp. 6241-6252, 2013.
- [2] A. Unler, A. Murat, and R. B. Chinnam, "mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification," *Information Sciences*, vol. 181, pp. 4625-4641, 2011.
- [3] Md. MonirulKabir, Md. Shahjahan, and K. Murase., "A new local search based hybrid genetic algorithm for feature selection," *Neurocomputing*, vol. 74, pp. 2914-2928, 2011.
- [4] Y. Liu and Y. F. Zheng, "FS_SFS: A novel feature selection method for support vector machines," *Pattern Recognition*, vol. 39, pp. 1333-1345, 2006.
- [5] Md. Monirul Kabir, Md. Monirul Islam, and K. Murase., "A new wrapper feature selection approach using neural network," *Neurocomputing*, vol. 73, pp. Pages 3273-3283, 2010.
- [6] T. M. Mitchell, "Machine Learning," McGraw-Hill, New York., 1997.
- [7] "Filler, T., Pevný, T., Bas, P., 2010. Break our steganographic system (BOSS). <http://exile.felk.cvut.cz/boss/>, (accessed 21.09.12)."
- [8] "Lubenko, I., Ker, A.D., 2012. Going from small to large data in steganalysis. In: Media Lubenko, I., Ker, A.D., 2012. Going from small to large data in steganalysis. In: Media national Society for Optical Engineering, vol. 8303, pp. 0M01-0M10.."
- [9] "Pevný, T., Ker, A.D. 2013. The Challenges of Rich Features in Universal Steganalysis. In: Proceedings of SPIE - The International Society for Optical Engineering 8665, 86650M-86650M-15.."
- [10] "Ker, A.D., Pevný, T., 2014. A mishmash of methods for mitigating the model mismatch mess. In: Proceedings of SPIE - The International Society for Optical Engineering 9028, 90280I-90280I-15."
- [11] B. Bonev, F. Escolano, D. Giorgi, and S. Biasotti, "Information-theoretic selection of high-dimensional spectral features for structural recognition," *Computer Vision and Image Understanding*, vol. 117, pp. 214-228, 2013.
- [12] Ahmed K. Farahat, Ali Ghodsi, and M. S. Kamel, "Efficient greedy feature selection for unsupervised learning," *Knowledge and Information Systems* vol. 35, pp. 285-310, 2013.
- [13] M. H. Aghdam, N. Ghasem-Aghaee, and M. E. Basiri, "Text feature selection using ant colony optimization," *Expert Systems with Applications*, vol. 36, pp. 6843-6853, 2009.
- [14] P. Shamsinejadbabki and M. Saraee, "A new unsupervised feature selection method for text clustering based on genetic algorithms," *Journal of Intelligent Information Systems*, vol. 38, pp. 669-684, 2011.
- [15] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowledge-Based Systems*, vol. 24, pp. 1024-1032, 2011.
- [16] L. D. Vignolo, D. H. Milone, and J. Scharcanski, "Feature selection for face recognition based on multi-objective evolutionary wrappers," *Expert Systems with Applications*, vol. 40, pp. 5077-5084, 2013.
- [17] L. Zini, N. Noceti, G. Fusco, and F. Odone, "Structured multi-class feature selection with an application to face recognition," *Pattern Recognition Letters*, 2014.
- [18] S. F. da Silva, M. X. Ribeiro, J. d. E. S. Batista Neto, C. Traina-Jr, and A. J. M. Traina, "Improving the ranking quality of medical image retrieval using a genetic feature selection method," *Decision Support Systems*, vol. 51, pp. 810-820, 2011.
- [19] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "A simultaneous feature adaptation and feature selection method for content-based image retrieval systems," *Knowledge-Based Systems*, vol. 39, pp. 85-94, 2013.
- [20] H. H. Inbarani, A. T. Azar, and G. Jothi, "Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis," *Comput Methods Programs Biomed*, vol. 113, pp. 175-85, 2014.
- [21] P. Jaganathan and R. Kuppuchamy, "A threshold fuzzy entropy based feature selection for medical database classification," *Comput Biol Med*, vol. 43, pp. 2222-9, Dec 2013.
- [22] Y. Chen and D. Liginlal, "A maximum entropy approach to feature selection in knowledge-based authentication," *Decision Support Systems*, vol. 46, pp. 388-398, 2008.
- [23] "J. Fridrich, "Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes," in *Information Hiding*, 2005.."
- [24] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, pp. 2507-17, Oct 1 2007.
- [25] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, pp. 16-28, 2014.
- [26] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, vol. 17, pp. 491 - 502 2005.
- [27] Lei Yu and H. Liu., "Feature selection for high-dimensional data: A fast correlation-based filter solution," In: Proceedings of the 20th International Conference on Machine Learning, pp. pp. 856-863, 2003.

- [28] S. Theodoridis and C. Koutroumbas, "Pattern Recognition, 4th Edn," Elsevier Inc, 2009.
- [29] Xiaofei He, Deng Cai, and P. Niyogi, "Laplacian Score for Feature Selection," *Adv. Neural Inf. Process. Syst.*, vol. 18, pp. 507-514, 2005.
- [30] Quanquan Gu, Zhenhui Li, and J. Han, "Generalized Fisher Score for Feature Selection," In: *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, 2011.
- [31] Artur J. Ferreira and M. A. T. Figueiredo, "An unsupervised approach to feature discretization and selection," *Pattern Recognition*, vol. 45, pp. 3048–3060, 2012.
- [32] S. Tabakhi, P. Moradi, and F. Akhlaghian, "An unsupervised feature selection algorithm based on ant colony optimization," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 112-123, 2014.
- [33] Hui-Huang Hsu, Cheng-Wei Hsieh, and M.-D. Lu, "Hybrid feature selection by combining filters and wrappers," *Expert Systems with Applications*, vol. 38, pp. 8144–8150, 2011.
- [34] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *Neural Networks, IEEE Transactions on*, vol. 5, pp. pp. 537 - 550 1994.
- [35] R. Archibald and G. Fann, "Feature Selection and Classification of Hyperspectral Images With Support Vector Machines," *Geoscience and Remote Sensing Letters, IEEE*, vol. 4, pp. pp. 674 - 677 2007.
- [36] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Neural Inform. Process. Systems 1*, pp. 585-592, 2002.
- [37] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 888–905, 2000.
- [38] F. Chung, "Spectral Graph Theory," In: *Regional Conference Series in Mathematics American Mathematical Society*, vol. 92, pp. 1-212, 1997.
- [39] Jung-Yi Jiang, Ren-Jia Liou, and S.-J. Lee, "A Fuzzy Self-Constructing Feature Clustering Algorithm for Text Classification," *IEEE Transactions Knowledge and Data Engineering*, vol. 23, pp. 335 - 349 2011.
- [40] Xi Zhao, W. Deng., and Y. Sh, "Feature Selection with Attributes Clustering by Maximal Information Coefficient " *Procedia Computer Science*, vol. 17, 2013.
- [41] Chuan Shi, Yanan Cai , Di Fu , Yuxiao Dong , and B. Wu, "A link clustering based overlapping community detection algorithm," *Data & Knowledge Engineering*, vol. 87, pp. Pages 394–404, 2013.
- [42] Yakun Li , Hongzhi Wang, Jianzhong Li , and H. Gao, "Efficient community detection with additive constrains on large networks," *Knowledge-Based Systems*, vol. 52, pp. Pages 268–278, 2013.
- [43] Yawen Jiang , Caiyan Jia, and J. Yu, "An efficient community detection method based on rank centrality," *Physica A*, vol. 392, pp. Pages 2182–2194, 2013.
- [44] J. G. V. Blondel, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 10008, pp. pp. 1–12, 2008.
- [45] A.A. Hagberg, D.A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in: *Proceedings of the 7th Python in Science Conference (SciPy2008)*, Pasadena, CA, USA, pp. pp. 11–15, 2008.
- [46] M. Bastian, et al., "Gephi: an open source software for exploring and manipulating networks.," In: *International AAAI Conference on Weblogs and Social Media*, North America, 2009.
- [47] Hanchuan Peng, Fuhui Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226 - 1238 2005.
- [48] "Bas, P. Filler, T. Pevn' {y}, T., 2011. Information hiding. In: *13th International Workshop. Springer-Verlag, New York..*"