JMLR: Workshop and Conference Proceedings 10: 4-13 The Fourth Workshop on Feature Selection in Data Mining

Feature Selection: An Ever Evolving Frontier in Data Mining

Huan Liu HUAN.LIU@ASU.EDU
Computer Science and Engineering,
Arizona State University, USA

Hiroshi Motoda MOTODA@AR.SANKEN.OSAKA-U.AC.JP
Asian Office of Aerospace Research and Development,
Air Force Office of Scientific Research,
US Air Force Research Laboratory, JAPAN
and Institute of Scientific Research,
Osaka University , JAPAN

Rudy Setiono RUDYS@COMP.NUS.EDU.SG
School of Computing,
National University of Singapore, SINGAPORE

Zheng Zhao ZHAOZHENG@ASU.EDU
Computer Science and Engineering,
Arizona State University, USA

Editor: Neil Lawrence

Abstract

The rapid advance of computer technologies in data processing, collection, and storage has provided unparalleled opportunities to expand capabilities in production, services, communications, and research. However, immense quantities of high-dimensional data renew the challenges to the state-of-the-art data mining techniques. Feature selection is an effective technique for dimension reduction and an essential step in successful data mining applications. It is a research area of great practical significance and has been developed and evolved to answer the challenges due to data of increasingly high dimensionality. Its direct benefits include: building simpler and more comprehensible models, improving data mining performance, and helping prepare, clean, and understand data. We first briefly introduce the key components of feature selection, and review its developments with the growth of data mining. We then overview FSDM and the papers of FSDM10, which showcases of a vibrant research field of some contemporary interests, new applications, and ongoing research efforts. We then examine nascent demands in data-intensive applications and identify some potential lines of research that require multidisciplinary efforts.

Keywords: Feature Selection, Feature Extraction, Dimension Reduction, Data Mining

1. An Introduction to Feature Selection

Data mining is a multidisciplinary effort to extract nuggets of knowledge from data. The proliferation of large data sets within many domains poses unprecedented challenges to data mining (Han and Kamber, 2001). Not only are data sets getting larger, but new types of data become prevalent, such as data streams on the Web, microarrays in genomics and proteomics, and networks in social computing and system biology. Researchers are realizing that in order to achieve successful data mining, feature selection is an indispensable component (Liu and Motoda, 1998; Guyon and Elisseeff, 2003; Liu and Motoda, 2007). It is a process of selecting a subset of original features according to certain criteria, and an important and frequently used technique in data mining for dimension reduction. It reduces the number of features, removes irrelevant, redundant, or noisy features, and brings about palpable effects for applications: speeding up a data mining algorithm, improving learning accuracy, and leading to better model comprehensibility. Various studies show that some features can be removed without performance deterioration (Ng, 2004; Donoho, 2006). Feature selection has been an active field of research for decades in data mining, and has been widely applied to many fields such as genomic analysis (Inza et al., 2004), text mining (Forman, 2003), image retrieval (Gonzalez and Woods, 1993; Swets and Weng, 1995), intrusion detection (Lee et al., 2000), to name a few. As new applications emerge in recent years, many challenges arise requiring novel theories and methods addressing high-dimensional and complex data. Feature selection for data of ultrahigh dimensionality (Fan et al., 2009), steam data (Glocer et al., 2005), multi-task data (Liu et al., 2009; G. Obozinski and Jordan, 2006), and multi-source data (Zhao et al., 2008, 2010a) are among emerging research topics of pressing needs.

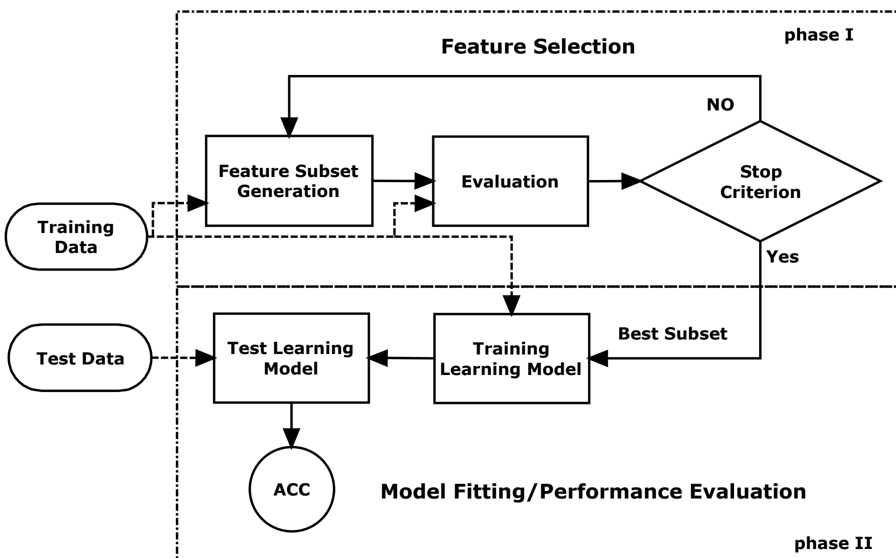

Figure 1: A unified view of a feature selection process

Figure 1 presents a unified view for a feature selection process. A typical feature selection process contains two phases: feature selection, and model fitting and performance evaluation. The feature selection phase contains three steps: (1) generating a candidate set containing a subset of the original features via certain research strategies; (2) evaluating the candidate set and estimating the utility of the features in the candidate set. Based on the evaluation, some features in the candidate set may be discarded or added to the selected feature set according to their relevance; and (3) determining whether the current set of selected features are good enough using certain stopping criterion. If it is, a feature selection algorithm will return the set of selected features, otherwise, it iterates until the stopping criterion is met. In the process of generating the candidate set and evaluating it, a feature selection algorithm may use the information from the training data, current selected features, target learning model, and given prior knowledge (Helleputte and Dupont, 2009) to guide their search and evaluation. Once a set of features is selected, it can be used to filter the training and test data for model fitting and prediction. The performance achieved by a particular learning model on the test data can also be used to as an indicator for evaluating the effectiveness of the feature selection algorithm for that learning model.

In the process of feature selection, the training data can be either labeled, unlabeled or partially labeled, leading to the development of **supervised**, **unsupervised** and **semi-supervised** feature selection algorithms. In the evaluation process, a supervised feature selection algorithm (Sikonja and Kononenko, 2003; Weston et al., 2003; Song et al., 2007; Zhang et al., 2008) determines features' relevance by evaluating their correlation with the class or their utility for achieving accurate predication, and without labels, an unsupervised feature selection algorithm may exploit data variance or data distribution in its evaluation of features' relevance (Dash and Liu, 2000; Dy and Brodley, 2004; He et al., 2005). A semi-supervised feature selection algorithm (Zhao and Liu, 2007c; Xu et al., 2009) uses a small amount of labeled data as additional information to improve unsupervised feature selection.

Depending on how and when the utility of selected features is evaluated, different strategies can be adopted, which broadly fall into three categories: **filter**, **wrapper** and **embedded** models. To evaluate the utility of features, in the evaluation step, feature selection algorithms of filter model rely on analyzing the general characteristics of data and evaluating features without involving any learning algorithm. On the other hand, feature selection algorithms of wapper model require a predetermined learning algorithm and use its performance on the provided features in the evaluation step to identify relevant feature. Algorithms of the embedded model, e.g., C4.5 (Quinlan, 1993), LARS (Efron et al., 2004), 1-norm support vector machine (Zhu et al., 2003), and sparse logistic regression (Cawley et al., 2007), incorporate feature selection as a part of the model fitting/training process, and features' utility is obtained based on analyzing their utility for optimizing the objective function of the learning model. Compared to the wrapper and embedded models, algorithms of the filter model are independent of any learning model, therefore do not have bias associated with any learning models, one advantage of the filter model. Another advantage of the filter model is that it allows the algorithms to have very simple structure, which usually employs a straightforward search strategy, such as backward elimination or forward selection, and a feature evaluation criterion designed according to certain criterion. The benefit of the simple structure is two-folds. First, it is easy to design, and after it is implemented, it is also easy to understand for other researchers. This actually explains why most feature selection algorithms are of the filter model. And in real world applications, many most frequently used feature selection algorithms are also filters. Second, since the structure of the algorithms is simple, they are usually very fast. On the other hand, researcher also recognized that compared to the filter model, feature selection algorithms of the wrapper and embedded models can usually select features that result in higher learning performance for a particular learning model, which is used in the feature selection process. Comparing with the wrapper model, feature selection algorithms of embedded model are usually more efficient, since they look into the structure of the involved learning model and use its properties to guide feature evaluation and search. In recent years, the embedded model is gaining increasing interests in feature selection research due to its superior performance. Currently, most embedded feature selection algorithms are designed by applying L_0 norm (Weston et al., 2003; Huang et al., 2008) or L_1 norm (Liu et al., 2009; Zhu et al., 2003; Zhao et al., 2010b) as a constraint to existing learning models to achieve a sparse solution. When the constraint is of L_1 norm form, and the original problem is convex, existing optimization techniques can be applied to obtain the unique global optimal solution for the regularized problem in a very efficient way (Liu et al., 2009).

Feature selection algorithms with the filter and embedded models may return either a subset of selected features or the weights (measuring features' relevance) of all features. According to the type of the output, feature selection algorithms can be divided into either **feature weighting** algorithms or **subset selection** algorithms. Feature selection algorithms of the wrapper model usually return feature subsets, therefore are subset selection algorithms. To the best of our knowledge, currently, most feature selection algorithms are designed to handle learning tasks with **single data source**. Researchers have started exploring the capability of using multiple auxiliary data and prior knowledge sources for **multi-source** feature selection (Zhao and Liu, 2008) to effectively enhance the reliability of relevance estimation (Lu et al., 2005; Zhao et al., 2008, 2010a).

Given a rich literature exists for feature selection research, a systematical summarization and comparison studies are of necessity to facilitate the research and application of feature selection techniques. Recently, there have been many surveys published to serve this purpose. A comprehensive surveys of existing feature selection techniques and a general framework for their unification can be found in (Liu and Yu, 2005). Guyon and Elisseeff (2003) reviewed feature selection algorithms from statistical learning point of view. In (Saeys et al., 2007), the authors provided a good survey for applying feature selection techniques in bioinformatics. In (Inza et al., 2004), the authors reviewed and compared the filter with the wrapper model for feature selection. In (Ma and Huang, 2008), the authors explored the representative feature selection approaches based on sparse regularization, which is a branch of embedded feature selection techniques. Representative feature selection algorithms are also empirically evaluated in (Liu et al., 2002; Li et al., 2004; Sun et al., 2005; Lai et al., 2006; Ma, 2006; Swartz et al., 2008; Murie et al., 2009) under different problem settings and from different perspectives. We refer readers to these survey works to obtain comprehensive understanding on feature selection research.

2. Toward Cross-Discipline Collaboration in Feature Selection Research

Knowledge discovery and data mining (KDD) is a multidisciplinary effort. Researchers and practitioners in multiple disciplines and various IT sectors confront similar issues in feature selection, and there is a pressing need for continuous exchange and discussion of challenges and ideas, exploring new methodologies and innovative approaches. The international workshop on Feature Selection in Data Mining (FSDM) serves as a platform to further the cross-discipline, collaborative effort in feature selection research. FSDM 2005[1]

1. http://enpub.fulton.asu.edu/workshop/

and 2006[2] were held with the SIAM Conference on Data Mining (SDM) 2005 and 2006, respectively. FSDM 2008[3] was held with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD) 2008. And FSDM 2010[4] is the fourth workshop of this series, and is held at the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2010. This collection consists of one keynote and 8 peer-reviewed papers, among which, there are three on exploring novel problems in feature selection research; four on developing new feature selection algorithms or improving existing ones; two on designing effective algorithms to solve real-world problems. Below we give an overview on the papers of FSDM 2010.

Two novel feature selection research problems are investigated. In the keynote paper (Chawla, 2010), the author studies the interesting research problem of detecting feature dependence, which is also the topic of (Salehi et al., 2010). Both works are based on the techniques related to association rule mining. A concept that is closely related to feature dependence is feature interaction, in which a set of features cooperate with each other to define the target concept. The problem of feature interaction is studied in (Jakulin and Bratko, 2004; Zhao and Liu, 2007b). Besides detecting feature dependence, the problem of feature extraction for heterogeneous data with ontology information is also studied in (Gorodetsky and Samoylov, 2010). It is an interesting feature extraction problem related to information fusion and multi-source feature selection (Zhao and Liu, 2008).

The filter, wrapper, and embedded models are the major models used in feature selection for algorithm design. In (Esseghir, 2010), an interesting hybrid approach is proposed to combine the wrapper with the filter model through a so-called greedy randomized adaptive search procedure (GRASP). The advantage of the method is that it can inherit the strength of both models to improve the performance of feature selection. In (Jaiantilal and Grudic, 2010), a new feature selection algorithm based on the embedded model is proposed. The intrinsic point of the paper is to develop a random sampling framework, which can effectively estimate feature weights for weighted L_1 penalty based sparse learning models (Zou, 2006). The pairwise sample similarity is an important way to depict the relationships among samples, and has been widely used in designing feature selection algorithms (Zhao and Liu, 2007a). Improving the quality of similarity measurements is beneficial to the feature selection algorithms by taking sample similarity as their input. In (Zagoruiko et al., 2010), the authors propose to apply FRiS-function to improve similarity estimation. And in (Xie et al., 2010), the authors proposed to construct continuous variables from categorical features to achieve better similarity estimation.

Text mining is an important research area, where feature selection is widely applied for dimension reduction. In (Singh et al., 2010), the authors develop a new feature evaluation criterion for text mining based on Gini coefficient of inequality. Their empirical study shows that the proposed criterion significantly improve the learning performance compared to several existing criteria in feature selection, including mutual information, information gain and chi-square statistic. Besides text mining, feature extraction for outlier detection is also studied. In (Nguyen and Gopalkrishnan, 2010), the authors propose to use weight adjusted scatter matrices in feature extraction to address the class unbalance issue in outlier

2. http://enpub.fulton.asu.edu/workshop/2006/
3. http://www.psb.ugent.be/ yvsae/fsdm08/index.html
4. http://featureselection.asu.edu/fsdm10/index.html

detection, and empirical results show that the proposed method can bring about nontrivial improvement over the existing algorithms.

3. Advancing Feature Selection Research

The current development in scientific research will lead to the prevalence of ultrahigh dimensional data generated from the high-throughput techniques (Fan et al., 2009) and the availability of many useful knowledge sources resulting from collective work of cutting-edge research. Hence one important research topic in feature selection is to develop computational theories that help scientists to keep up with the rapid advance of new technologies on data collection and processing. We also notice that there is a chasm between symbolic learning and statistical learning that prevents scientists from taking advantage of data and knowledge in a seamless way. Symbolic learning works well with knowledge and statistical learning works with data. Explanation-based learning is one such example that would provide an efficient way to bridge this gap. The technique of explanation-based feature selection will enable us to use the accumulated domain knowledge to help narrow down the search space and explain the learning results by providing reasons why certain features are relevant. Below are our conjectures about some interesting research topics in feature selection of potential impact in the near future.

Feature selection for ultrahigh dimensional data: selecting features on data sets with millions of features (Fan et al., 2009). As high-throughput techniques keep evolving, many contemporary research projects in scientific discovery generate data with ultrahigh dimensionality. For instance, the next-generation sequencing techniques in genetics analysis can generate data with several giga features on one run. Computation inherent in existing methods makes them hard to directly handle data of such high dimensionality, which raises the simultaneous challenges of computational power, statistical accuracy, and algorithmic stability. To address these challenges, researchers need to develop efficient approaches for fast relevance estimation and dimension reduction. Prior knowledge can play an important role in this study, for example, by providing effective ways to partition original feature space to subspaces, which leads to significant reduction on search space and allows the application of highly efficient parallel techniques.

Knowledge oriented sparse learning: fitting sparse learning models via utilizing multiple types of knowledge. This direction extends multi-source feature selection (Zhao and Liu, 2008). Sparse learning allows joint model fitting and features selection. Given multiple types of knowledge, researchers need to study how to use knowledge to guide inference for improving learning performance, such as the prediction accuracy, and model interpretability. For instance, in microarray analysis, given gene regulatory network and gene ontology annotation, it is interesting to study how to simultaneously infer with both types of knowledge, for example, via network dynamic analysis or function concordance analysis, to build accurate prediction models based on a compact set of genes. One direct benefit of utilizing existing knowledge in inference is that it can significantly increase the reliability of the relevance estimation (Zhao et al., 2010a). Another benefit of using knowledge is that it may reduce cost by requiring fewer samples for model fitting.

Explanation-based feature selection (EBFS): feature selection via explaining training samples using concepts generalized from existing features and knowledge. In many real-world applications, the same phenomenon might be caused by disparate reasons. For example, in a cancer study, a certain phenotype may be related to mutations of either genes A or gene B in the same functional module M. And both gene A and gene B can cause the defect of M. Existing feature selection algorithm based on checking feature/class correlation may not work in this situation, due to the inconsistent (variable) expression pattern of gene A and gene B across the cancerous samples[5]. The generalization step in EBFS can effectively screen this variation by forming high-level concepts via using the ontology information obtained from annotation databases, such as GO. Another advantage of EBFS is that it can generate sensible explanations to show why the selected features are related. EBFS is related to the research of explanation-based learning (EBL) and relational learning.

Feature selection remains and will continue to be an active field that is incessantly rejuvenating itself to answer new challenges.

Acknowledgments

This work is, in part, supported by NSF Grant (0812551).

References

G. C. Cawley, N. L. C. Talbot, and M. Girolami. Sparse multinomial logistic regression via bayesian l1 regularisation. In *NIPS*, 2007.

Sanjay Chawla. Feature selection, association rules network and theory building. In *The 4th Workshop on Feature Selection in Data Mining*, 2010.

M. Dash and H. Liu. Feature selection for clustering. In *Proceedings of 4th Pacific Asia Conference on Knowledge Discovery and Data Mining, 2000*. Springer-Verlag, 2000.

D. Donoho. Formost large underdetermined systems of linear equations, the minimal l1-norm solution is also the sparsest solution. *Comm. Pure Appl. Math.*, 59:907–934, 2006.

Jennifer G. Dy and Carla E. Brodley. Feature selection for unsupervised learning. *J. Mach. Learn. Res.*, 5:845–889, 2004. ISSN 1533-7928.

B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32:407–49, 2004.

M. A. Esseghir. Effective wrapper-filter hybridization through grasp schemata. In *The 4th Workshop on Feature Selection in Data Mining*, 2010.

Jianqing Fan, Richard Samworth, and Yichao Wu. Ultrahigh dimensional feature selection: Beyond the linear model. *Journal of Machine Learning Research*, 10:2013–2038, 2009.

George Forman. An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, 3:1289–1305, 2003.

B. Taskar G. Obozinski and M. I. Jordan. Multi-task feature selection. Technical report, Statistics Department, UC Berkeley, 2006.

5. For a cancerous sample, either gene A or gene B has abnormal expression, but not both.

K. Glocer, D. Eads, and J. Theiler. Online feature selection for pixel classification. In *Proceedings of the 22nd international conference on Machine learning (ICML)*, 2005.

R. Gonzalez and R. Woods. *Digital Image Processing*. Addison-Wesley, 2nd edition, 1993.

V. Gorodetsky and V. Samoylov. Feature extraction for machine learning: Logic-probabilistic approach. In *The 4th Workshop on Feature Selection in Data Mining*, 2010.

I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufman, 2001.

X. He, D. Cai, and P. Niyogi. Laplacian score for feature selection. *Advances in Neural Information Processing Systems 18*, Cambridge, MA, 2005. MIT Press.

Thibault Helleputte and Pierre Dupont. Partially supervised feature selection with regularized linear models. In *ICML*, 2009.

Kaizhu Huang, Irwin King, and Michael R. Lyu. Direct zero-norm optimization for feature selection. In *Proceeding of The 8th IEEE International Conference on Data Mining*, 2008.

I. Inza, P. Larranaga, R. Blanco, and A. Cerrolaza. Filter versus wrapper gene selection approaches in dna microarray domains. *Artificial Intelligence in Medicine*, 31:91–103,2004.

A. Jaiantilal and G. Grudic. Increasing feature selection accuracy for l1 regularized linear models in large datasets. In *the 4th Workshop on Feature Selection in Data Mining*, 2010.

A. Jakulin and I. Bratko. Testing the significance of attribute interactions. In *ICML*, 2004.

Carmen Lai, Marcel J T Reinders, Laura J van't Veer, and Lodewyk F A Wessels. A comparison of univariate and multivariate gene selection techniques for classification of cancer datasets. *BMC Bioinformatics*, 7:235, 2006.

W. Lee, S. J. Stolfo, and K. W. Mok. Adaptive intrusion detection: A data mining approach. *AI Review*, 14(6):533 – 567, 2000.

Tao Li, Chengliang Zhang, and Mitsunori Ogihara. A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, 20(15):2429–2437, 2004.

H. Liu and H. Motoda. *Feature Selection for Knowledge Discovery and Data Mining*. Boston: Kluwer Academic Publishers, 1998. ISBN 0-7923-8198-X.

H. Liu and H. Motoda, editors. *Computational Methods of Feature Selection*. Chapman and Hall/CRC Press, 2007.

H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Trans. on Knowledge and Data Engineering*, 17(3):1–12, 2005.

Huiqing Liu, Jinyan Li, and Limsoon Wong. A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. *Genome Inform*, 13:51–60, 2002.

J. Liu, S. Ji, and J. Ye. Multi-task feature learning via efficient l2,1-norm minimization. In *the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, 2009.

J. Lu, G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz, and T. R. Golub. Microrna expression profiles classify human cancers. *Nature*, 435:834–838, 2005.

S. Ma. Empirical study of supervised gene screening. *BMC Bioinformatics*, 7:537, 2006.

Shuangge Ma and Jian Huang. Penalized feature selection and classification in bioinformatics. *Brief Bioinform*, 9(5):392–403, Sep 2008.

Carl Murie, Owen Woody, Anna Lee, and Robert Nadon. Comparison of small n statistical tests of differential expression applied to microarrays. *BMC Bioinformatics*, 10:45, 2009.

A. Y. Ng. Feature selection, l1 vs. l2 regularization, and rotational invariance. In *the 21st international conference on Machine learning*. ACM Press, 2004.

H. V. Nguyen and V. Gopalkrishnan. Feature extraction for outlier detection in high-dimensional spaces. In *The 4th Workshop on Feature Selection in Data Mining*, 2010.

J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

Yvan Saeys, Iaki Inza, and Pedro Larraaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, 2007.

Elham Salehi, Jayashree Nyayachavadi, and Robin Gras. A statistical implicative analysis based algorithm and mmpc algorithm for detecting multiple dependencies. In *The 4th Workshop on Feature Selection in Data Mining*, 2010.

M. R. Sikonja and I. Kononenko. Theoretical and empirical analysis of Relief and ReliefF. *Machine Learning*, 53:23–69, 2003.

Sanasam Ranbir Singh, Hema A. Murthy, and Timothy A. Gonsalves. Feature selection for text classification based on gini coecient of inequality. In *The 4th Workshop on Feature Selection in Data Mining*, 2010.

L. Song, A. Smola, A. Gretton, K. Borgwardt, and J. Bedo. Supervised feature selection via dependence estimation. In *International Conference on Machine Learning*, 2007.

Y. Sun, C. F. Babbs, and E. J. Delp. A comparison of feature selection methods for the detection of breast cancers in mammograms: adaptive sequential floating search vs. genetic algorithm. *Conf Proc IEEE Eng Med Biol Soc*, 6:6532–6535, 2005.

Michael D Swartz, Robert K Yu, and Sanjay Shete. Finding factors influencing risk: Comparing bayesian stochastic search and standard variable selection methods applied to logistic regression models of cases and controls. *Stat Med*, 27(29):6158–6174, Dec 2008.

D. L. Swets and J. J. Weng. Efficient content-based image retrieval using automatic feature selection. In *IEEE International Symposium On Computer Vision*, pages 85–90, 1995.

J. Weston, A. Elisseff, B. Schoelkopf, and M. Tipping. Use of the zero norm with linear models and kernel methods. *Journal of Machine Learning Research*, 3:1439–1461, 2003.

Jierui Xie, Boleslaw Szymanski, and Mohammed J. Zaki. Learning dissimilarities for categorical symbols. In *The 4th Workshop on Feature Selection in Data Mining*, 2010.

Zenglin Xu, Rong Jin, Jieping Ye, Michael R. Lyu, and Irwin King. Discriminative semi-supervised feature selection via manifold regularization. In *IJCAI' 09: Proceedings of the 21th International Joint Conference on Artificial Intelligence*, 2009.

Nikolai G. Zagoruiko, Irina A. Borisova, Vladimir V. Duybanov, and Olga A. Kutnenko. Attribute selection based on fris-compactness. In *The 4th Workshop on Feature Selection in Data Mining*, 2010.

Yi Zhang, Chris Ding, and Tao Li. Gene selection algorithm by combining relieff and mrmr. *BMC Genomics*, 9:S27, 2008.

Z. Zhao and H. Liu. Spectral feature selection for supervised and unsupervised learning. In *International Conference on Machine Learning (ICML)*, 2007

Z. Zhao, J. Wang, H. Liu, J. Ye, and Y. Chang. Identifying biologically relevant genes via multiple heterogeneous data sources. In *The Fourteenth ACM SIGKDD International Conference On Knowledge Discovery and Data Mining*, 2008.

Zheng Zhao and Huan Liu. Searching for interacting features. In *International Joint Conference on AI (IJCAI)*, 2007.

Zheng Zhao and Huan Liu. Semi-supervised feature selection via spectral analysis. In *Proceedings of SIAM International Conference on Data Mining*, 2007.

Zheng Zhao and Huan Liu. Multi-source feature selection via geometry-dependent covariance analysis. In *Journal of Machine Learning Research, Workshop and Conference Proceedings Volume 4: New challenges for feature selection in data mining and knowledge discovery*, volume 4, pages 36–47, 2008.

Zheng Zhao, Jiangxin Wang, Shashvata Sharma, Nitin Agarwal, Huan Liu, and Yung Chang. An integrative approach to identifying biologically relevant genes. In *Proceedings of SIAM International Conference on Data Mining (SDM)*, 2010.

Zheng Zhao, Lei Wang, and Huan Liu. Efficient spectral feature selection with minimum redundancy. In *Proceedings of the Twenty-4th AAAI Conference on Artificial Intelligence (AAAI)*, 2010.

Ji Zhu, Saharon Rosset, Trevor Hastie, and Rob Tibshirani. 1-norm support vector machines. In *Advances in Neural Information Processing Systems 16*, 2003.

Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101 (12):1418–1429, 2006.