Similarity Analysis of Feature Ranking Techniques on Imbalanced DNA Microarray Datasets

David Dittman*, Taghi Khoshgoftaar*, Randall Wald*, and Amri Napolitano*
*Florida Atlantic University, Boca Raton, FL 33431
Email: ddittman@fau.edu; khoshgof@fau.edu; rwald1@fau.edu; amrifau@gmail.com

***Abstract*—DNA microarrays are a modern advancement in the analysis of genetic data. This technology allows a researcher to test samples for thousands of genes simultaneously. However, once the samples in the DNA microarrays have been tested, the researcher must then search through the data collected and identify genes important to their problem. A possible solution to this issue is the data mining pre-processing technique called feature selection. Feature (gene) selection takes the original set of features (in the case of DNA microarrays, gene probes) and chooses an optimal subset to perform analysis from. Ideally, the reduced subset only contains the most important features as determined by the feature selection technique (or set of feature selection techniques), which allows for further research in the discovered genes. However in the case of using multiple feature selection techniques, the set of techniques must be diverse in order to reduce redundancy among the chosen features. Another benefit of increasing diversity is that any features chosen across a diverse set of feature selection techniques will have more importance than those chosen by a single technique or a set of related ones. Therefore, it would be useful to know how similar the feature selection techniques are to each other. In this study we perform an analysis of eighteen feature selection techniques across nine imbalanced DNA microarray datasets and using four feature subset sizes. Our results found that one should not use Gini Index and Probability Ratio together or the Kolmogorov-Smirnov statistic and Geometric Mean together at any feature subset size in order to minimize redundancy, and that the members of the first of these pairs (along with the pair of ReliefF and ReliefF-W) are very dissimilar to all rankers outside their own cluster. We also found that Chi-Squared, Information Gain, and Symmetric Uncertainty form a cluster of similarity, as do Chi-Squared, Deviance, F-Measure, and Mutual Information.**

***Keywords*-Similarity; DNA microarray; feature selection;**

I. Introduction

One of the many goals of the fields of genetics and bioinformatics is being able to take the data that lies within the genetic structure of an organism and use this to determine or discover important genes or biomarkers and use them to identify various traits that are not immediately apparent from the phenotypical features of said organism. While acquiring the data is not particularly difficult (recent advances in chemistry and technology allow researchers to test a sample for thousands of genes simultaneously), working with the large amount of data that results from the acquisition process can be a difficult task. A number of possible solutions to the problem of performing classification with the large amounts of data can be found within the domain of data mining.

Perhaps one of the most dangerous issues when it comes to large amounts of data is high dimensionality. High dimensionality occurs when there are a large number of attributes (or features) that are associated with each instance. This can cause a number of problems for data mining and bioinformatics experiments including: reduced performance, large computational time, and the use of features that may be either redundant or irrelevant to the problem being studied. One of the more effective methods of handling high dimensionality is the data mining pre-processing technique known as feature selection. Feature (gene) selection takes the set of features and chooses an optimum subset of these features and uses only these features for subsequent analysis. Surprisingly, despite the loss of data, feature selection can lead to more accurate and efficient classifiers [1].

An example of a dataset which can benefit greatly from the use of these techniques is the DNA microarray or gene expression profile dataset. The DNA microarray allows the researcher to test a sample for thousands of genes simultaneously. Due to the high dimensionality of these datasets, feature selection can be an effective tool. Feature selection will allow us to sort through the features and determine which are the most important, letting us choose the optimum subset for analysis. In addition to choosing the optimum feature subset, the result of feature selection can be used by itself as a topic of research as the features chosen are determined to be the most important.

When choosing a set of feature selection techniques it is essential to determine just how similar or different the techniques are to each other. Diversity between feature selection techniques can be very beneficial to experiments in genetics and bioinformatics, especially in areas such as biomarker identification or confirming the importance of a particular gene. If the techniques are diverse, then there is more reason to have confidence in any genes found by all techniques, while it is unsurprising (and not confidence-enhancing) when similar techniques all select the same genes. Additionally, diversity is an important factor when using ensemble feature selection: it allows for smaller ensembles to nonetheless have the breadth necessary to maximize performance. Ensemble techniques are important because they can both improve classification performance and stability of the chosen features [2].

This paper is a study of eighteen feature selection tech-

978-1-4673-2560-8/12/$31.00 ©2012 IEEE

niques (eleven of which were designed and implemented by our research group) and how similar they are to each other. In addition to the eighteen feature selection techniques, we also use nine imbalanced datasets and four feature subset sizes. Our study found that if diversity is the goal then one should not use Gini Index and Probability Ratio together or the Kolmogorov-Smirnov statistic and Geometric Mean together at any feature subset size. When looking at dissimilar feature selection techniques, the first of these pairs along with the pair of ReliefF and ReliefF-W are dissimilar to all other rankers. In addition to these results, we found that Chi-Squared, Information Gain, and Symmetric Uncertainty form a cluster, as do Chi-Squared, Deviance, F-Measure, and Mutual Information.

The remainder of the paper is organized as follows. Section II contains some related works relevant to our study. Section III outlines our methods when performing this study. Section IV contains the results of our experiment. Lastly, Section V contains our conclusions and suggestions for future work.

II. Related Works

Due to the high dimensionality of DNA microarray datasets, it has become necessary to include dimensionality reduction techniques. A study performed by Inza et al. [3] found that classification performed on reduced feature subsets derived from the original DNA microarray datasets outperformed classification using the whole feature set in a majority of cases and that feature selection drastically reduced computation time. However, the large degree of high dimensionality causes a number of techniques to be infeasible in terms of computational time. For example, if one wanted to choose the best pair of features from a dataset of 15,154 features it would require evaluating 114,814,281 pairs. If it takes 0.1 seconds to evaluate each pair it will take 132.9 days of continuous computation to evaluate them all [4]. Therefore, a majority of the work on feature selection has been using filter-based feature rankers. There are a number of reasons why these techniques are ideal for this problem, including: the output of the techniques (a ranked list of features) is intuitive and easy to understand; the ranking of genes makes it easy for researchers to further validate the results through laboratory techniques; and the relatively small computational time when compared to other types of feature selection techniques (filter-based subset evaluation, wrapper, etc.) [5].

In a number of cases, DNA microarray datasets have been known to exhibit class imbalance (number of instances in each class are not equal). In the case of DNA microarrays the class of interest may have very few instances attributed to its class. This can cause a number of problems when it comes to performing analysis on the data. A study performed by Blagus et al. [6] found that when one performs analysis on imbalanced datasets a bias is created toward the majority class (not the class of interest). While there are cases where the class imbalance can be reduced by certain methods (i.e. undersampling, oversampling, etc.), the larger levels of imbalance require more complex methods in order to alleviate the issue. Another study conducted by Al-Shahib et al. [7] states that there are two main reasons for the poor performance of the minority class in bioinformatics. The first is that the main goal of a number of machine learning algorithms focus on accuracy and therefore will optimize the prediction for the majority class even if it will fail for the minority class. The second reason is that due to the small number of features in the class of interest, outliers are given more weight and can sometimes distort the findings from the real cause.

While there has been very little research regarding similarity between rankers there have been studies in the related topic of stability in recent years. Since one of the more common methods of measuring the stability of a feature ranker involves calculating the similarity of different feature subsets [8], these techniques can be applied toward the study of similarity. There are a number of stability measures used to compare feature subsets. Kalousis et al. [9] used different measures of correlation to measure the stability of the feature ranker. Lustgarten et al. [10] devised a new stability measure called Adjusted Stability Measure that can be applied to a set of feature subsets and determine the cardinality of the subsets. The adjusted part of Adjusted Stability Measure comes from measures used to adjust for the chances of similarity occuring based on chance or random feature selection.

In order to better understand our methods, some background information on our chosen method of measuring similarity is necessary. In 2007, Kuncheva et al. [8] devised a framework to study the stability of feature selection methods by calculating the similarity of multiple feature subsets derived from the same technique using randomly selected instance subsets of the original dataset as the input. This study defined the term consistency index. The consistency index is a measure of similarity between two different feature subsets. They devised this measure as a way to choose the best set of features for an experiment. As this method derives the commonality between the two feature subsets we use it to determine the similarity of the feature selection techniques.

III. Methods

Our methods can be described in three aspects: Datasets, Feature Selection, and Similarity Measure. Section III-A contains details on the datasets used in this study. Section III-B describes the eighteen feature selection techniques that are being analyzed for similarity. Lastly, Section III-C outlines the stability measure used to compare the feature subsets.

Table I
DETAILS OF THE DATASETS

Name	Total # of Instances	Percent Minority Instances	# of Attributes
ECML_Pancreas [1]	90	8.89%	27680
lung-Michigan [11]	96	10.42%	7130
lung-cancer [1]	181	17.13%	12534
lung50k [12]	400	17.50%	54614
Lymphoma [1]	96	23.96%	4027
acute-lymphoblastic-leukemia [13]	327	24.16%	12559
ovarian_mat [14]	66	24.24%	6001
lymphoma_mat [14]	77	24.68%	7130
Brain_Tumor [13]	90	25.56%	27680

A. Datasets

In this experiment we choose a set of nine datasets that were acquired from various real world bioinformatics, genetics, and medical projects. All nine datasets show a large degree of class imbalance (the majority class vastly outunumbers the minority class). The reason behind choosing these imbalanced datasets is that frequently DNA microarray experiments can have very few samples or instances in the minority class despite the minority class being the class of interest. For the purposes of simplicity, we will describe the class with the most instances as the majority class and the class with the least amount of instances as the minority class, regardless of the actual class label of the two classes. Table I describes the nine datasets in terms of the charactistics as well as the references of their previous use.

B. Feature Selection

Feature selection is a necessary part of the analysis of DNA microarray data. Feature selection is a process which chooses an optimal subset of features to be used in later analysis, rather than analyzing the entire dataset. What is interesting is that despite the loss of data, feature selection can be useful in creating efficient and accurate classifiers. In our study we used eighteen filter-based feature rankers on each of the nine datasets. Therefore a total of (9 datasets $\times$ 18 feature rankers) = 162 different rankings were computed.

The feature rankers (filters) chosen can be placed into three categories: commonly used filter-based feature selection techniques, a rarely used filter technique called Signal to Noise, and threshold-based feature selection techniques (TBFS) that were developed by our research team. For convenience, we collectively refer to the commonly used filters and Signal to Noise as "non-TBFS" techniques.

1) Non-TBFS Feature Selection Techniques: Seven of the techniques (Chi-Squared, Information Gain, Gain Ratio, Symmetric Uncertainty, ReliefF, ReliefF-W, and S2N) are of the non-TBFS category, and are implemented in the open-source Weka machine learning toolkit [15]. The Chi-Squared (CS) test compares the observed distribution of class-feature value pairs to the distribution predicted by a chi-squared random distribution, and those features which are distinct from this null distribution are preferred. Information Gain (IG) is based on the entropy inherent in the class-value distribution, and selects features which reduce this entropy when the instances are divided up based on their values. Gain Ratio (GR) is a modification of IG which takes into account the inherent entropy of the feature values to reduce the problem of features with many values having artificially-high IG scores. Symmetric Uncertainty (SU) also modifies IG by taking into account the feature's inherent entropy, but it does this by dividing the IG by the sum of the feature and class's independent entropies. ReliefF (RF) decides the relevance of features by seeing how much they vary when taking a randomly-selected instance and comparing its values with those of its nearest hit (instance in same class) and nearest misses (instances in different classes). ReliefF-W (RFW) is a variant of ReliefF which applies distance-based weights to the nearest neighbors prior to adjusting the feature scores. Finally, Signal-to-Noise finds the signal-to-noise ratio of the feature, which is the ratio of the difference of the mean values for each class to the sum of the standard deviations for each class.

2) Threshold-Based Feature Selection Techniques: Eleven of the techniques used in this work (F-Measure, Odds Ratio, Power, Probability Ratio, Gini Index, Mutual Information, Kolmogorov-Smirnov statistic, Deviance, Geometric Mean, Area Under the ROC Curve, and Area Under the Precision Recall Curve) fall under the category of TBFS techniques. These feature ranking techniques were proposed and implemented recently by our research group [12]. In TBFS, each attribute is evaluated against the class, independent of all other features in the dataset. After normalizing each attribute to have a range between 0 and 1, simple classifiers are built for each threshold value $t \in [0, 1]$ according to two different classification rules (e.g., whether instances with values above the threshold are considered positive or negative class examples). The normalized values are treated as posterior probabilities: however, no real classifiers are being built. Instead, these ersatz posterior probabilities are used to calculate various classifier performance metrics, and the results of these metrics are the quality of the feature being examined.

C. Similarity Measure

We decided to use consistency index [8] because it takes into consideration bias due to chance. First, we assume that a given dataset has n features. Let T_i and T_j be subsets of features, where $|T_i| = |T_j| = k$. The consistency index [8] is obtained as follows:

$$I_C(T_i, T_j) = \frac{dn - k^2}{k(n-k)}, \tag{1}$$

where d is the cardinality of the intersection between subsets T_i and T_j, and $-1 < I_C(T_i, T_j) \leq +1$. The greater the consistency index, the more similar the subsets are. However, in order to determine the similarity of the subsets we must

determine which values of k (the feature subset size of both subsets) to use. In this case, four subset sizes are chosen per feature ranking. The sizes of the four subsets are as follows: 50, 75, 100, and 200. We limited our investigation to four subset sizes due to space limitations, and chose these four specifically because they are appropriate according to previous research [14].

IV. Results

This study is an analysis of eighteen filter-based feature selection techniques across nine imbalanced datasets. Table II contains the pairwise comparisons of all eighteen feature selection methods for feature subset sizes 50 through 200. Each entry in the table is the average similarity across all nine datasets and across all of the assigned feature subsets. Any pairwise combination which results in a measurement over 0.8 are in boldface, and any under 0.25 are in italics. The reason for the cutoff points at similarity values of 0.8 and 0.25 is to ensure that the pairs reported are either truly similar or dissimilar respectively.

Looking at the results from Table II, the combinations can be combined into clusters of similar techniques. The clusters {PR and GI} and {KS and GM} are most notable, as the former pair are particularly distinct from the other rankers, and the latter pair has the highest overall similarity. Outside of those two clusters, there are two other clusters found within this range. The first is {CS, F, Dev, and MI}. It should be noted that the similarity between CS and MI is less than 0.8, but due to how close the similarity score is (0.77673), and because of MI's similarity to F and Dev (0.8323 and 0.8002, respectively), it was still included in the cluster. The last cluster is {CS, SU, and IG}. While this cluster also includes CS the other two techniques were different enough from the {CS, F, Dev, and MI} cluster to necessitate their own cluster: although SU and IG only having a similarity of 0.7845 with one another, each has > 0.8 similarity with CS. As noted, KS and GM has the highest level of similarity with a similarity of 0.9352. In terms of dissimilarity, we find that RF and RFW are dissimilar to most of the other rankers except each other. In addition, the {KS and GM} and {PR and GI} clusters are extremely dissimilar from each other: each member of one of those clusters has less than 0.25 similarity with the members of the other cluster. The {PR and GI} cluster is also somewhat dissimilar from most of the remaining rankers, with the exceptions of GR and POW.

It is instructive to look more closely at the {PR and GI} and {KS and GM} clusters. When considering {PR and GI}, we find that not only are these two rankers consistently together but they are only similar to each other. We believe this pattern holds because the two feature techniques both seek to find the threshold which maximizes the precision. In particular, GI = $\min_{t\in[0,1]}[2PRE(t)(1 - PRE(t)) + 2NPV(t)(1 - NPV(t))]$ and PR = $\max_{t\in[0,1]} \frac{TPR(t)}{FPR(t)}$, where PRE, NPV, TPR, and FPR are the precision, negative predictive value (the percentage of instances predicted to be negative that are actually negative), true positive rate, and false positive rate respectively. For {KS and GM}, we note that these two techniques score the largest similarity and are the only pair of techniques whose similarity is above 0.9. We believe this is due to the fact that both KS and GM perform optimally when the True Positive Rate is maximized (KS = $\max_{t\in[0,1]} |TPR(t) - FPR(t)|$ and GM = $\max_{t\in[0,1]} \sqrt{TPR(t) \times TNR(t)}$, where TNR is the true negative rate). When we look at the patterns of dissimilarity, we find that the members of the {PR and GI} cluster, in addition to those of the {RF and RFW} cluster, tend to be dissimilar to all rankers not within their own cluster. It should be noted that RF and RFW don't pass the 0.8 threshold to be counted as similar to one another, but both due to their being the only ReliefF-based rankers used in this paper and due to their extreme dissimilarity to the other rankers, we consider them as a cluster.

V. Conclusions

When deciding upon a set of feature selection techniques, diversity is an important factor to consider during the process. If the set of feature selection techniques are diverse, then any features (in the case of DNA microarrays, gene probes) found across all of the techniques is given more importance than features chosen by only a single feature selections or from a set of very similar techniques. In addition, diversity is an essential aspect in the creation of stable ensemble feature selection techniques which can be used to build high-performing classifiers.

Our results found that in order to maximize diversity, there are certain high-similarity clusters of feature selection techniques that should not be used in conjunction with one another. First of all, two clusters stand out: {PR and GI} and {KS and GM}. The {PR and GI} cluster is particularly dissimilar to all other rankers, and the {KS and GM} cluster shows the highest similarity for any pair of rankers. The cluster of {RF and RFW} was also notable, because although the two rankers did not have an impressive similarity to each other, they were dissimilar to all other rankers. In addition, the {CS, IG, and SU} and {CS, Dev, F, MI} clusters showed high within-cluster similarity.

Future work on the similarity of feature selection techniques in DNA microarray datasets includes performing experiments on more datasets with varying levels of class balance. This will allow us to further validate our findings. Another possibility is to select a specific goal (i.e. biomarker identification) and only perform the analysis on datasets built for this goal. This will allow us to derive more specific cases of feature selection similarity.

References

[1] J. Van Hulse, T. M. Khoshgoftaar, A. Napolitano, and R. Wald, "Feature selection with high dimentional imbalanced

Table II
SIMILARITY MEASURE: 50–200 FEATURES

	CS	GR	IG	RF	RFW	SU	F	OR	Pow	PR	GI	MI	KS	Dev	GM	ROC	PRC	S2N
CS		0.5203	**0.8002**	0.3802	0.3132	**0.8199**	**0.8091**	0.6598	0.6152	0.3524	0.3611	0.7767	0.6556	**0.8331**	0.6431	0.6335	0.7314	0.5317
GR			0.4626	0.2789	*0.2446*	0.6095	0.4692	0.5221	0.5437	0.5643	0.5685	0.4467	0.3792	0.4884	0.3661	0.3807	0.4650	0.3568
IG				0.3584	0.2924	0.7845	0.7456	0.6371	0.5173	0.3010	0.3134	0.7881	0.7506	0.7286	0.7188	0.7067	0.6796	0.5796
RF					0.6172	0.3664	0.3979	0.3569	0.3613	0.2526	0.2683	0.3751	0.3383	0.3937	0.3408	0.3545	0.4111	0.3783
RFW						0.3068	0.3188	0.2904	0.2955	*0.2251*	*0.2259*	0.2920	0.2747	0.3130	0.2775	0.2833	0.3296	0.3205
SU							0.7453	0.6978	0.6325	0.3990	0.4108	0.7510	0.6365	0.7798	0.6121	0.6062	0.6937	0.5392
F								0.6199	0.5471	0.3074	0.3172	**0.8323**	0.7292	**0.8420**	0.7263	0.7002	0.7276	0.5472
OR									0.6524	0.4216	0.4308	0.6460	0.5339	0.6630	0.5101	0.5469	0.6323	0.5008
Pow										0.5404	0.5474	0.5107	0.4105	0.6108	0.4039	0.4597	0.6564	0.4185
PR											**0.8690**	0.2766	*0.2392*	0.3451	*0.2381*	0.2788	0.4075	0.2891
GI												0.2882	0.2507	0.3728	*0.2495*	0.2859	0.4187	0.3005
MI													0.7872	**0.8002**	0.7533	0.7055	0.6891	0.5563
KS														0.6620	**0.9352**	0.7720	0.6151	0.5615
Dev															0.6491	0.6467	0.7202	0.5314
GM																0.7741	0.6085	0.5522
ROC																	0.6887	0.6134
PRC																		0.5922
S2N																		

data," in *Proceedings of the 9th IEEE International Conference on Data Mining - Workshops (ICDM'09)*. Miami, FL: IEEE Computer Society, December 2009, pp. 507–514.

[2] Y. Saeys, T. Abeel, and Y. Peer, "Robust feature selection using ensemble feature selection techniques," in *ECML PKDD '08: Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases - Part II*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 313–325.

[3] I. Inza, P. Larraaga, R. Blanco, and A. J. Cerrolaza, "Filter versus wrapper gene selection approaches in dna microarray domains," *Artificial Intelligence in Medicine*, vol. 31, no. 2, pp. 91 – 103, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0933365704000193

[4] R. Somorjai, B. Dolenko, and R. Baumgartner, "Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions," *Bioinformatics*, vol. 19, no. 12, pp. 1484–1491, 2003. [Online]. Available: http://bioinformatics.oxfordjournals.org/content/19/12/1484.abstract

[5] Y. Saeys, I. Inza, and P. Larraaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007. [Online]. Available: http://bioinformatics.oxfordjournals.org/content/23/19/2507.abstract

[6] R. Blagus and L. Lusa, "Class prediction for high-dimensional class-imbalanced data," *BMC Bioinformatics*, pp. 523–539, 2010. [Online]. Available: http://www.biomedcentral.com/1471-2105/11/523

[7] A. Al-Shahib, R. Breitling, and D. Gilbert, "Feature selection and the class imbalance problem in predicting protein function from sequence," *Applied Bioinformatics*, vol. 4, no. 3, pp. 195–203, 2005. [Online]. Available: http://www.ingentaconnect.com/content/adis/abi/2005/00000004/00000003/art00004

[8] L. I. Kuncheva, "A stability index for feature selection," in *Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications*. Anaheim, CA, USA: ACTA Press, 2007, pp. 390–395. [Online]. Available: http://portal.acm.org/citation.cfm?id=1295303.1295370

[9] A. Kalousis, J. Prados, and M. Hilario, "Stability of feature selection algorithms: a study on high-dimensional spaces," *Knowledge and Information Systems*, vol. 12, no. 1, pp. 95–116, Dec. 2006. [Online]. Available: http://www.springerlink.com/index/10.1007/s10115-006-0040-8

[10] J. L. Lustgarten, V. Gopalakrishnan, and S. Visweswaran, "Measuring stability of feature selection in biomedical datasets," in *AMIA 2009 Symposium Proceedings*, 2009, pp. 406–410.

[11] D. G. Beer, S. L. Kardia, C.-C. Huang, T. J. Giordano, A. M. Levin, D. E. Misek, L. Lin, G. Chen, T. G. Gharib, D. G. Thomas, M. L. Lizyness, R. Kuick, S. Hayasaka, J. M. Taylor, M. D. Iannettoni, M. B. Orringer, and S. Hanash, "Gene-expression profiles predict survival of patients with lung adenocarcinoma," *Nature Medicine*, vol. 8, pp. 816–824, 2002. [Online]. Available: http://dx.doi.org/10.1038/nm733

[12] D. Dittman, T. Khoshgoftaar, R. Wald, and J. Van Hulse, "Comparative analysis of dna microarray data through the use of feature selection techniques," in *Proceedings of the Ninth IEEE International Conference on Machine Learning and Applications (ICMLA)*. ICMLA, 2010, pp. 147 –152.

[13] J. Van Hulse, T. M. Khoshgoftaar, A. Napolitano, and R. Wald, "A comparative evaluation of feature ranking methods for high dimensional bioinformatics data," in *Proceedings of the IEEE International Conference on Information Reuse and Integration - IRI'11*, August 3-5, 2011, pp. 315–320.

[14] X. Chen and M. Wasikowski, "Fast: a roc-based feature selection metric for small samples and imbalanced data classification problems," in *Proceedings of the 14th ACM SIGKDD Int'l Conf. Knowldege Discovery and Data Mining (KDD '08)*. New York, NY: ACM, 2008, pp. 124–132.

[15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.