

A Hybrid Simplified Swarm Optimization Method for Imbalanced Data Feature Selection

Wei Chang Yeh, Yu Tai Yang, Chyh Ming Lai

National Tsing Hua University, Taiwan

ABSTRACT

In recent years, feature selection has become an important field in data mining and been widely used in numerous regions. The purpose of feature selection is to search an optimal subset of features from existing data to maximize the accuracy. However, there are still few studies investigating the impact of data imbalance, the existence of underrepresented categories of data, on feature selection problem. Therefore, the aim of this study is to provide a feature selection method for increasing classifying high-dimensional imbalanced data accuracy. In this study, we proposed a hybrid method which can spot a better optimal features subset. In the proposed method, information gain as a filter selects the most informative features from the original dataset. The imbalance of the dataset with selected features is justified by using Synthetic minority over-sampling technique. Then, simplified swarm optimization is implemented as feature search engine to guide the search for an optimal feature subset. Finally, support vector machine serve as a classifier to evaluate the performance of the proposed method. To evaluate the performance of proposed algorithm, we apply our algorithm in four benchmark datasets and compare the results with existing algorithm. The results show that our algorithm has a better performance than its competitor.

Keywords: Data Mining, Feature Selection, Imbalanced Data, Soft Computing, Simplified Swarm Optimization, Support Vector Machine

1. INTRODUCTION

Feature selection has become an important field in data mining, and been widely used in numerous regions, including text categorization, image retrieval, and genomic analysis (Liu & Yu, 2005). Feature selection is the process of selecting a subset of features (variables) from data. The reduction of features can increase the accuracy of machine learning result and reduce the time of building model. In addition, feature selection can reduce the risk of overfitting, which is more common on high-dimensional data sets (Maldonado, Weber, & Famili, 2014).

The feature selection has three kinds of algorithms: filter, wrapper, and embedded method. Filter method use pre-defined metric to evaluate the goodness of features before using classifier. The common filter method includes χ^2 static, Information Gain (Y. Yang & Pedersen, 1997), and Relief (Kira & Rendell, 1992). Filter method has lower computation time but has poorer performance than other two methods. Wrapper method generates subsets and judges the subsets by the performance of implying classifier. The embedded method finds the subset during the procedure of classifier. The wrapper and embedded method usually have better performance but are more computational incentive than filter method (Maldonado et al., 2014).

For the wrapper method, many methods have been proposed to evaluate a better optimal features subset. In recent years, soft computing algorithm has been implied as searching for wrapper method. The randomness of these stochastic algorithm can reduce the sensitive to the dataset (Al-Ani, Alsukker, & Khushaba, 2013). Commonly used method such as particle swarm optimization (PSO) (Chuang, Chang, Tu, & Yang, 2008), genetic algorithm (GA) (C.-H. Yang, Chuang, & Yang, 2010), artificial bee colony (ABC) (Schiezero & Pedrini, 2013), and simulated annealing (SA) (Lin, Lee, Chen, & Tseng, 2008). Among all these method, PSO and GA are the most commonly used method, as well as PSO has better algorithmic efficiency and effortless to implement than GA (Al-Obeidat, Belacel, Carretero, & Mahanti, 2011). However, PSO has its disadvantages. It may drop to the local optimal and is inadequate for discrete problem (Liang, Qin, Suganthan, & Baskar, 2006). To overcome these disadvantages, Yeh proposed the simplified swarm optimization (SSO) (Yeh, 2009). It modified the update process of variable in PSO and has few parameters to tune. The SSO is also perform well in feature selection problem (Yeh, Chang, & Chiu, 2011). Thus, in this study we use SSO as the wrapper method in our hybrid algorithm.

The feature selection technique can also be implied in imbalanced data, which is also a crucial issue in recent years (He & Garcia, 2009). Imbalanced data is that one of the class of data set has relatively few instances, which calls minority class, is underrepresented. In many cases, the minority class is the major target of the data, like cancer diagnosis (Mazurowski et al., 2008), the patient with malignant tumour is way less than patient with benign tumour. Other applications, such as fraud detection (Anil Kumar & Ravi, 2008), helicopter fault monitoring (Japkowicz, Myers, & Gluck, 1995), have a valuable minority class as well. The issue of imbalanced data is that the normal classification method may misjudge the minority class. For example, if we have 95 instances majority class, the class holds the most instances, and 5 instances of minority class in the dataset. If the classifier identifies all class as the majority class will have 95% accuracy, which normally considered good enough. However, this result can't reflect that the minority class is 100% misclassified, as well as the minority class are way more important than the majority class. Therefore, how to handle the biased dataset and classify minority class as accurate as possible are the main challenge in class imbalanced problem.

The class imbalanced feature selection problem have been studied in many fields (Villar, Fernández, Carrasco, & Herrera, 2012). However, there are still relatively few studies investigating the impact of data imbalance on high-dimensional feature selection problem (Maldonado et al., 2014). Hence, we proposed a hybrid algorithm for high-dimensional class imbalanced feature selection problem. The algorithm combine IG and simplified swarm optimization (SSO) (Yeh, 2013), an soft computing algorithm, as filter and wrapper method to explore the optimal subset. The resampling technique is implied to reduce class imbalanced problem. We use the support vector machine (SVM) as the classifier.

The paper is organized as follows: section 2 we provide a brief review of related work associate to our study. In section 3 we describe the algorithm proposed. The experiment result is showed in section 4, and a conclusion is given in section 5.

2. RELATED WORK 1000-2000

2.1 Imbalanced Data

To overcome the underrepresented of minority class, we can change the weight of different class. For example, let the weight of minority class be W_+ , weight of majority class be W_- , and $W_+ > W_-$. That makes the classifier be more sensitive to the misclassification of minority class. Another way is the resampling technique, including oversampling and undersampling. The former duplicate the minority class while the latter delete the majority class from dataset, which may cause a loss of information (Van Hulse, Khoshgoftaar, Napolitano, & Wald, 2009).

The above method can help balancing the data set and increasing the accuracy of classifying class imbalanced dataset. However, due to the monotonic of minority class and there are no new instances is added, these method still may cause overfitting (Van Hulse et al., 2009). Thus, the synthetic minority over-sampling technique (SMOTE) (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) is proposed to overcome above problem. SMOTE can generate new instances from existed data, which makes the decision regions less specific and preventing from overfitting (Han, Wang, & Mao, 2005). The procedure of SMOTE is described as follows: First find each minority instances k nearest minority neighbourhood (k usually set to 5). Then randomly pick one of k neighbourhood, generate new instances randomly between them. This process repeat depends on the portion we want for the new artificial instance. In most case, the portion is set to 200% (Barua, Islam, Yao, & Murase, 2014).

For the class imbalanced problem, we usually discuss the binary (two-class) problem. The minority class usually denotes as positive, while majority class denotes as negative. The confusion matrix is shown in Figure 1. As we described in section 1, the performance of imbalanced data can't be easily presented by accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad (1)$$

where TP and TN denote the number of true positive and true negative instances, while FP and FN denote the number of false positive and false negative instances. Therefore, numerous evaluation metric are introduced. In this study, we use geometric mean (g-mean) as the performance measurement metric:

$$\text{G-mean} = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \quad (2)$$

		Actual Value	
		Minority (positive)	Majority (negative)
Prediction (classification) Result	Minority (positive)	True Positive	False Positive
	Minority (positive)	False Negative	True Negative

Figure 1: Confusion matrix for binary class problem

Other common metric includes F-measure (also called F-score) and Area under curve (AUC)(Sokolova, Japkowicz, & Szpakowicz, 2006).

2.2 Information Gain

Information Gain is a filter feature selection method. It scores the informational entropy of features and determines the importance of these features. Informational entropy is theoretically the number of bits of data it would take to encode a given piece of information (Y. Yang & Pedersen, 1997). The more space of a piece of information takes to encode, the more entropy it has (Van Hulse et al., 2009). To explain the above description we can see an example, a sequential data can be easily transfer to a smaller archive file using compression algorithm, while a totally random data, which has maximum entropy, cannot be compressed.

For classification, the information of instances belong to which classes is the data we want to describe/compressed. If all instances belong to one classes, the compress rate is huge (all instances is in the first class). But if all instances are randomly separate to each classes, the space we need to record the information is huge, which means the entropy of this situation is high. The equation calculating entropy is described as follows:

$$H(T) = -\sum_{i=1}^q \left(\frac{n_i}{n} \right) \times \log \left(\frac{n_i}{n} \right) \quad (3)$$

where training dataset T has $n=|T|$ instances and qclasses. The i^{th} class has n_i instances. Now we want to know the entropy related to each attribute. For attribute a which has v distinct values $a_j(j=1,2,\dots,v)$, the entropy is calculated by summing the entropy for each a_j :

$$H(T|a) = \sum_{j=1}^v \left(\frac{|a_j|}{n} \right) \times H(T|a = a_j) \quad (4)$$

where $|a_j|$ denotes the number of instances has the same attribute a value a_j . If attribute a is highly related to the classes belong, the entropy of $H(T|a)$ will be low, otherwise the entropy will close to $H(T)$. We get the reduction of entropy, the value we measure information gain for each attribute (IG value), as shown in Eq. 5:

$$IG(a) = H(T) - H(T|a) \quad (5)$$

The more the entropy decrease, the more significant feature x is for prediction (Van Hulse et al., 2009). The IG value we get in Eq. 5 is between 0 and 1. We can filter the features by a pre-determined number of features to keep, or by setting a threshold value and reserve the features that has higher IG value.

2.3 Simplified Swarm Optimization (SSO)

Simplified swarm optimization is a soft computing algorithm proposed by Yeh (Yeh, 2009). It is evolved from the PSO algorithm and is developed to overcome the drawback of PSO in discrete problem. SSO has been implied in many fields, including network intrusion detection (Chung & Wahid, 2012) and disassemble sequencing problem (Yeh, 2012). The main idea of SSO is generate a set of solutions and randomly update them by each solution's history best solution (pbest) and global best solution (gbest).

Let $X_i^t = \{x_{i1}^t, x_{i2}^t, \dots, x_{im}^t\}$ denotes the i^{th} solution in t^{th} iteration, where each solution has m variables and x_{ij}^t is the j^{th} variable of i^{th} solution in t^{th} iteration. The x_{ij}^t is updated by generating a random variable $\rho_{ij}^t \in (0,1)$, then update the variable:

$$x_{ij}^t = \begin{cases} x_{ij}^{t-1}, & \text{if } \rho_{ij}^t \in (0, c_w) \\ p_{ij}^{t-1}, & \text{if } \rho_{ij}^t \in (c_w, c_p) \\ g_j^{t-1}, & \text{if } \rho_{ij}^t \in (c_p, c_g) \\ x_{new}, & \text{if } \rho_{ij}^t \in (c_g, 1) \end{cases} \quad (6)$$

where c_w, c_p, c_g is the pre-defined parameter for update that $0 < c_w < c_p < c_g < 1$. The $p_{ij}^{(t-1)}$ and $g_j^{(t-1)}$ denote the pbest of solution i and gbest of variable j in $t-1^{\text{th}}$ iteration. The x_{new} is a new random value in the feasible field of variable j .

The procedure of SSO is as follows: First generate a set of solutions randomly and calculate the fitness function of each solution. The pbest is initially set as these solutions and gbest is the solution of pbest that has the best fitness value. For each iteration, all the variable in each solution is updated by Eq. 6. Then we calculate the fitness function of updated solutions. If the fitness value of solution is higher than pbest, replace the pbest by present solution. If the new pbest has higher fitness value than gbest, replace the gbest by pbest. The procedure is repeated until we reach the threshold or reach the maximum iterations.

2.4 Support Vector Machine (SVM)

Support vector machine is a supervised learning method and can be use as classifier in machine learning. SVM determine an optimal hyperplane or a set of hyperplanes $f(A) = w^T \times A + b$ that separating different classes as wide as possible. In this study, we only discuss the binary linear SVM.

For binary problem, we denote $A_i \in R^m$ as the vector of attributes of instance ($i=1, \dots, n$) i and y_i as the class label of instance i where $y_i = \{-1, 1\}$, $i=1, \dots, n$. The goal isto find the optimal hyperplane that separate different classes. The formulation is stated as follows:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i \times (w^T \times A_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, n \end{aligned} \quad (7)$$

where ξ_i is the slake variable and C is the parameter for the penalty function of training error (Vapnik & Vapnik, 1998).

As far as we know, there are no wrapper method investigating imbalanced data problem in high-dimensional dataset using SVM. Villar et al. (Villar et al., 2012) proposed a genetic algorithm based feature selection method for class imbalanced problem but with low-dimensional datasets. Maldonado et al. (Maldonado et al., 2014) also proposed an embedded method using backward elimination. The method is for class imbalanced problem in high-dimensional dataset, which is same as our target, and got good prediction result with fewer features.

3. PROPOSED METHOD

3.1 Encoding

The encoding technique in our SSO wrapper method is quite elementary. The main idea is whether we pick the feature or not. We define there are m variables, the number of total features, in each solution. Each variable is a binary number (0, 1) that represents we pick the features or not. Figure 2 represents an example of a solution, which picks the 1, 2, 4 features and the dataset has 5 features in total.

3.2 Fitness Function

The fitness function is calculated depending on the solution subset classification result using SVM. In feature selection problems, if we have a set of subsets with similar accuracy, then the subset that contains fewer features is considered a better subset. And as we mentioned in section 2.1, the accuracy can't represent the performance of imbalanced data completely. Therefore, we refer to previous studies (C.-H. Yang et al., 2010) and set the fitness function by g-mean:

$$fit(X_i) = \delta \times g\text{-mean} + (1 - \delta) \times \frac{T - S}{T} \quad (8)$$

where T and S denote the number of total features and selected features. The δ is a pre-defined parameter between 0 and 1. The fitness function is a bi-objective function which considers both the number of selected features and the g-mean.

3.3 The Proposed Ig-Sso

The proposed hybrid method for high-dimensional class imbalanced feature selection problems has three phases. First we use SMOTE pre-processing the data. It can generate artificial minority class in order to balance the dataset. Next we imply the information gain to the dataset. The elimination of insignificant features can help reduce the computation complexity for the following stage. Finally we use SSO as the wrapper method and SVM as the classifier to investigate the optimal features subset.

In the second phase of information gain, the selection of threshold is a challenging work. If the threshold is too strict, we may eliminate valuable features. If the threshold is too loose, the uninformative features will increase the computational complexity for the next phase. Therefore, instead of setting a threshold, we select the top 60 features with the highest IG value for the next step (C.-H. Yang et al., 2010). The procedure of our proposed method is presented as follows:

- Step 1: Input the dataset D . Set parameters $c_w, c_p, c_g, \text{rnk}, \text{pop}, \text{iter}$.
- Step 2: Use SMOTE to add artificial instances set SM . $D \leftarrow D \cup SM$
- Step 3: Calculate the IG value and hold the top rnk features. $D \leftarrow D_{IG(\text{rnk})}$
- Step 4: Generate X_i ($i = 1, \dots, \text{rnk}$), let $t = 1$.
- Step 5: Let $i = 1$.
- Step 6: Let $j = 1$.
- Step 7: Update x_{ij} by Eq. 6.
- Step 8: If $j < \text{rnk}$ then $j = j + 1$ and go to Step 7.
- Step 9: Calculate $fit(X_i)$ and update $pbest$ and $gbest$.
- Step 10: If $i < \text{pop}$ then $i = i + 1$ and go to Step 6.
- Step 11: If $t < \text{iter}$ then $t = t + 1$ and go to Step 5.
- Step 12: Return the g-mean of $gbest$.

4. EXPERIMENT RESULT

To evaluate the performance of our proposed method (IG-SSO), we test four benchmark problems, which have been used for feature selection problems (K. Yang, Cai, Li, & Lin, 2006), and compare the result with the method proposed by

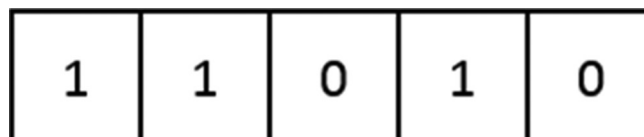


Figure 2: The example of encoding

Maldonado (Maldonado et al., 2014). They are a family of class imbalanced high-dimensional method using SVM. We evaluate the performance the feature selection method by g-mean.

The information of benchmark datasets are shown in Table 1. Since we are studying binary problem, we modified these datasets to binary problem. In the CAR dataset we set the kidney (11 instances) class as the minority class, and the rest classes as the majority class. The other datasets are set as above, where GLIOMA dataset we use cancer oligodendrogliomas (7 instances) class, LUNG2 datasets we use small-cell lung carcinomas (20 instances) class, and SRBCT datasets we use Burkitt lymphoma (11 instances) class as the minority class (Maldonado et al., 2014).

Considering all these datasets have relatively few instances, we built the evaluation model using leave-one-out cross validation (LOOCV). The percentage of SMOTE to generate is 200%. For the information gain, we select the top 60 features ($rnk = 60$) for the succeeding phase. The penalty function parameter C of SVM is set to 1. And for the SSO algorithm, $c_w = 0.1$, $c_p = 0.4$, $c_g = 0.9$, $\delta = 0.95$. The number of solutions (pop) is set to 30, the number of iterations (iter) is set to 100, and the IG-SSO is repeated 10 times.

The result are shown in Table 2 and the best result is marked in bold. The HO-BFE_{bi} and BFE-SVM_{bi} are two method proposed in Maldonado's research. Since the number of features is not the main part of Maldonado's study, some of the results are not shown in this table. As we can see in Table 2, the IG-SSO has the best g-mean value among all result. The classification result reach 100% accuracy in LUNG2 and SRBCT datasets. In LUNG2 datasets, the HO-BFE_{bi} has lower number of selected features, but other three datasets the IG-SSO pick fewer features for classification and got better g-mean.

For the resampling technique, the result shows that SMOTE do increase the performance of classification in IG-SSO, all datasets performed better with SMOTE except LUNG2 dataset which already reach 100% g-mean. However, it seems the resampling technique has low influence to the HO-BFE_{bi} and BFE-SVM_{bi} method and cause a decrease in g-mean.

5. CONCLUSION

The imbalance data seriously affect the outcome of feature selection and classification. In this study, we proposed a hybrid algorithm IG-SSO for the high-dimensional feature selection problem with imbalanced dataset. The proposed method used resampling technique SMOTE to deal with class imbalanced problem, and combined filter method, information gain, and wrapper method, SSO, for the feature selection. To the best of our knowledge, it is the first SVM based wrapper method facing class imbalanced high-dimensional problem. The performance of proposed method is compared with previous research (Maldonado et al., 2014) and shows that our method can spot a better optimal features subset and thus increase the accuracy of classifying high-dimensional imbalanced data problem.

Table 1: Descriptions of datasets

Name	Features	Instances	Minority	%Minority
CAR	9182	174	11	6.3
GLIOMA	4433	50	7	14.0
LUNG2	3312	203	20	9.8
SRBCT	2308	83	11	13.3

Table 2: Average g-mean, in percentage, and the number of features selected for the dataset

	CAR		GLIOMA		LUNG2		SRBCT	
	gmean	#	gmean	#	gmean	#	gmean	#
No SMOTE								
HO-BFE _{bi}	93.1	50	81.7		100.0	10	96.5	100
BFE-SVM _{bi}	92.1		80.8	50	100.0		99.9	
IG-SSO	93.7	30.7	98.0	33.7	100.0	33.0	100.0	26.7
SMOTE								
HO-BFE _{bi}	92.6		74.8		98.3		95.8	50
BFE-SVM _{bi}	92.6	20	72.7	250	99.2	10	99.2	
IG-SSO	95.3	32.7	99.2	28.7	100.0	27.7	100.0	31.0

Although our experiment shows a promising result for our proposed method, there are some work left for the future researchers. The number of features selected by information gain is determined by researchers and may vary by case. For the future investigation, researchers can focus on the filter method phase and construct a threshold that vary the number of features selected, or use other filter method for the pre-selected phase. We expected our method can be applied to more scenario and help improving feature selection problem.

REFERENCE

- Al-Ani, A., Alsukker, A., & Khushaba, R. N. (2013). Feature subset selection using differential evolution and a wheel based search strategy. *Swarm and Evolutionary Computation*, 9, 15-26.
- Al-Obeidat, F., Belacel, N., Carretero, J. A., & Mahanti, P. (2011). An evolutionary framework using particle swarm optimization for classification method PROAFTN. *Applied Soft Computing*, 11 (8), 4971-4980.
- Anil Kumar, D., & Ravi, V. (2008). Predicting credit card customer churn in banks using data mining. *International Journal of Data Analysis Techniques and Strategies*, 1 (1), 4-28.
- Barua, S., Islam, M. M., Yao, X., & Murase, K. (2014). MWMOTE--majority weighted minority oversampling technique for imbalanced data set learning. *Knowledge and Data Engineering, IEEE Transactions on*, 26 (2), 405-425.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 321-357.
- Chuang, L.-Y., Chang, H.-W., Tu, C.-J., & Yang, C.-H. (2008). Improved binary PSO for feature selection using gene expression data. *Computational Biology and Chemistry*, 32 (1), 29-38.
- Chung, Y. Y., & Wahid, N. (2012). A hybrid network intrusion detection system using simplified swarm optimization (SSO). *Applied Soft Computing*, 12 (9), 3014-3022.
- Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning *Advances in intelligent computing* (pp. 878-887): Springer.
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *Knowledge and Data Engineering, IEEE Transactions on*, 21 (9), 1263-1284.
- Japkowicz, N., Myers, C., & Gluck, M. (1995). A novelty detection approach to classification. Paper presented at the IJCAI.
- Kira, K., & Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. Paper presented at the AAAI.
- Liang, J. J., Qin, A. K., Suganthan, P. N., & Baskar, S. (2006). Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. *Evolutionary Computation, IEEE Transactions on*, 10 (3), 281-295.
- Lin, S.-W., Lee, Z.-J., Chen, S.-C., & Tseng, T.-Y. (2008). Parameter determination of support vector machine and feature selection using simulated annealing approach. *Applied Soft Computing*, 8 (4), 1505-1512.
- Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *Knowledge and Data Engineering, IEEE Transactions on*, 17 (4), 491-502.
- Maldonado, S., Weber, R., & Famili, F. (2014). Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines. *Information Sciences*, 286, 228-246.
- Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A., & Tourassi, G. D. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural networks*, 21 (2), 427-436.
- Schiezaro, M., & Pedrini, H. (2013). Data feature selection based on Artificial Bee Colony algorithm. *EURASIP Journal on Image and Video Processing*, 2013 (1), 1-8.
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation *AI 2006: Advances in Artificial Intelligence* (pp. 1015-1021): Springer.
- Van Hulse, J., Khoshgoftaar, T. M., Napolitano, A., & Wald, R. (2009). Feature selection with high-dimensional imbalanced data. Paper presented at the IEEE International Conference on Data Mining Workshops.
- Vapnik, V. N., & Vapnik, V. (1998). *Statistical learning theory*: Wiley New York.
- Villar, P., Fernández, A., Carrasco, R. A., & Herrera, F. (2012). Feature selection and granularity learning in genetic fuzzy rule-based classification systems for highly imbalanced data-sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 20 (03), 369-397.
- Yang, C.-H., Chuang, L.-Y., & Yang, C. H. (2010). IG-GA: a hybrid filter/wrapper method for feature selection of microarray data. *Journal of Medical and Biological Engineering*, 30 (1), 23-28.
- Yang, K., Cai, Z., Li, J., & Lin, G. (2006). A stable gene selection in microarray data analysis. *BMC bioinformatics*, 7 (1), 228.
- Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. Paper presented at the ICML.
- Yeh, W.-C. (2009). A two-stage discrete particle swarm optimization for the problem of multiple multi-level redundancy allocation in series systems. *Expert Systems with Applications*, 36 (5), 9192-9200.
- Yeh, W.-C. (2012). Optimization of the disassembly sequencing problem on the basis of self-adaptive simplified swarm optimization. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 42 (1), 250-261.
- Yeh, W.-C. (2013). New parameter-free simplified swarm optimization for artificial neural network training and its application in the prediction of time series. *Neural Networks and Learning Systems, IEEE Transactions on*, 24 (4), 661-665.
- Yeh, W.-C., Chang, W.-W., & Chiu, C.-W. (2011). A simplified swarm optimization for discovering the classification rule using microarray data of breast cancer. *International Journal of Innovative Computing, Information and Control*, 7 (5), 2235-2246.