Statistical Validation of ACO-KNN Algorithm for Sentiment Analysis

Siti Rohaidah Ahmad[1], Azuraliza Abu Bakar[2], Mohd Ridzwan Yaakub[2], Nurhafizah Moziyana Mohd Yusop[1]
[1]*Department of Computer Science, Faculty of Defence Science & Technology, Universiti Pertahanan Nasional Malaysia, 57000, Kuala Lumpur, Malaysia*
[2]*Data Mining and Optimization Research Group, Centre for Artificial Intelligence Technology, Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, 46000, Bangi, Selangor, Malaysia.*
sitirohaidah@upnm.edu.my

***Abstract*—This research paper aims to propose a hybrid of ant colony optimization (ACO) and k-nearest neighbour (KNN) algorithms as feature selections for selecting and choosing relevant features from customer review datasets. Information gain (IG), genetic algorithm (GA), and rough set attribute reduction (RSAR) were used as baseline algorithms in a performance comparison with the proposed algorithm. This paper will also discuss the significance test, which was used to evaluate the performance differences between the ACO-KNN, the IG-GA, and the IG-RSAR algorithms. The dependency relation algorithm was used to identify actual features commented by customers by linking the dependency relation between product feature and sentiment words in customers sentences. This study evaluated the performance of the ACO-KNN algorithm using precision, recall, and F-score, which was validated using the parametric statistical significance tests. The evaluation process has statistically proven that this ACO-KNN algorithm has been significantly improved compared to the baseline algorithms. In addition, the experimental results have proven that the ACO-KNN can be used as a feature selection technique in sentiment analysis to obtain quality, optimal feature subset that can represent the actual data in customer review data.**

***Index Terms*—Feature Selection; Sentiment Analysis; Statistical Analysis; Ant Colony Optimization.**

I. INTRODUCTION

Sentiment analysis (SA) is often used to mine customers' sentiments by examining written texts [1]. The main challenge with sentiment analysis is the large-sized customer comments dataset, which may contain irrelevant and overlapping features [2] – [4]. Feature selection (FS) is the main step in SA that selects the subset feature from the real features without altering the original data content [5]. This process also involves selecting and evaluating the optimum subset based on the evaluation criteria [6]. Researches by [5, 7] suggested using the ant colony optimization (ACO) and k-nearest neighbour (KNN) as feature selections to select text features from a dataset. An experiment by Aghdam et al. [7] had shown that a hybrid ACO-KNN, when used as a text feature selection, was able to select the relevant features, thus improving the performance. Therefore, this study has applied ACO-KNN as the text feature selection to select and choose relevant features from customer review datasets.

In this study, the performance of the ACO-KNN was tested using three performance metrics; the precision, recall, and F-score. More importantly, the results were validated using the statistical significance test. This paper will subsequently discuss the testing and validating process. The proposed ACO-KNN algorithm was compared with baseline algorithms, namely the IG, and GA that were applied by Abbasi et al. [9], as well as IG combined with the RSAR technique, as used in [10]. Customer review datasets on five different types of electronic products from the Amazon website were used as the experimental data.

The remainder of this paper is organised as follows: Section 2 will discuss the feature selection in sentiment analysis. Next, Section 3 will outline the statistical analysis, and Section 4 will describe the experimental set-up. Then, Section 5 will present the results and discussion. Lastly, Section 6 will conclude this work.

II. FEATURE SELECTION IN SENTIMENT ANALYSIS

Several methods can be employed to perform sentiment analysis. This section will discuss related work on feature selection in sentiment analysis.

Ahmad et al. [5] reported that feature selection techniques in sentiment analysis can be divided into two categories: feature selection techniques based on natural language processing, and based on modern methods. There are three types of feature selection techniques based on modern methods, namely, filter techniques, wrapping techniques, and hybrid techniques. Feature selection techniques play a major role in improving the performance of sentiment classification. Sentiment analysis techniques are based on the machine learning approach due to the large-sized features. Several studies on sentiment analysis have combined filter techniques with metaheuristic techniques to overcome the weaknesses of each technique. For example, Abbasi et al. [9], and Agarwal and Mittal [10] applied the IG technique to identify important features in sentiment classification. According to Agarwal and Mittal [10], IG was used to determine the reduction of uncertainty in identifying the feature class properties when the value of the feature has been identified. The most important features are selected to reduce the size of the feature vector to obtain a better classification. Agarwal and Mittal [10] claimed that IG is a filter technique that can determine the importance of these features in a document. Nonetheless, its weakness is that the threshold value must be set in advance. In addition, this technique does not take into account the surplus between features, and there is an absence of communication between features [9, 10]. In their study, Abbasi et al. [9] combined the IG filter technique with the Entropy Weight Genetic Algorithm (EWGA) metaheuristic technique. The resulting combination of filters managed to increase the performance of sentiment classification and obtained an optimal feature subset. However, their study was focused on the document level, which only considered the entire document as either positive or negative. The disadvantage of this technique is that the contents of the document are not thoroughly filtered, and the focus is only on one product and not multiple products [1]. In their study, Agarwal and Mittal [10] had combined the IG technique with the RSAR. The RSAR technique was implemented to reduce the number of irrelevant and excessive features, as well as noise. RSAR has the advantage of taking into account the dependent nature of the combinations of features [12]. However, the RSAR has two drawbacks: a) Obtaining an optimum feature subset, and this is a non-deterministic polynomial-time hard (NP-hard), therefore, the metaheuristic algorithm was proposed to overcome this problem [10]; and b) Feature selection process is time-consuming [9, 12, 13]. Selecting a feature subset is a non-deterministic polynomial problem that requires an efficient algorithm, such as a metaheuristic algorithm to solve feature selection problems [15] – [18]. Metaheuristic techniques, such as the ACO and GA have been used by Aghdam et al. [7], and Basiri and Nemati [19] as a feature selection technique for text classification. Conversely, GA is widely used for text classification in sentiment analysis, such as by Abbasi et al. [9], Zhu et al. [20], and Kalaivani and Shunmuganathan [21] as a feature selection technique. Meanwhile, Liu et al. [22] proposed a multi-swarm particle swarm optimization (MPSO) algorithm as a feature selection technique to choose emotional features found in course reviews. The PSO technique has been identified as being used only in the study of Chinese [23] and Arabic text classifications [24]. Findings by Aghdam et al. [7] showed that ACO was able to obtain the optimal feature subset compared to GA in text classification. The authors [7] found that the ACO has several advantages: a) it can produce a rapid convergence process; b) efficient at solving problem space, and c) it can efficiently obtain a minimum feature subset. Meanwhile, the GA was found to be inefficient at controlling a lot of features, which makes it difficult to obtain the optimal feature subset. A combination of the ACO-KNN in a study by Aghdam et al. [7] has helped obtain the optimal feature subset, and improve the performance of text classification. The advantages of the ACO-KNN are seen as being potentially able to solve feature selection problems in sentiment classification. Therefore, it is proposed that the ACO-KNN is used as a feature selection technique in this study.

A. Ant Colony Optimization

During early 1999, the ACO, which is a metaheuristic approach, was proposed as a way to solve hard combinatorial optimization problems [25]. The ACO algorithm [26], [27] is based on the behaviour of ants that interact using a chemical medium called pheromone. This chemical leaves a trace on the soil to mark their route. An ant uses this marker to find its way back to its nest. Additionally, other ants can use the marker to identify the best route to a food source. Numerous studies have applied the ACO as a feature selection technique. For instance, [28] used the ACO as a subset search procedure for a voice clarification process. In addition, [29] used it in a face identification system. Previous studies [7], [19] used the ACO to select features in text form in text classification processes, which were applied to the Reuters-12578 dataset. Kabir et al. [30] also applied the ACO as a feature selection technique, whereby a neural network (NN) assessed the feature subset derived from the ACO. The subset assessment was based on the classification accuracy percentage achieved by the NN on the testing dataset. Meanwhile, Aghdam et al. [31] combined the ACO with a Bayesian classification as the feature selection technique and applied on the Post-synaptic dataset. The results of that study showed that the combination of ACO and Bayesian classification was effective, and resulted in high classification accuracy compared to other techniques. Several other studies have also implemented ACO as the feature selection technique as it is more advantageous compared to other techniques [18 – 21].

B. K-Nearest Neighbour

The k-nearest neighbour is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions) [36]. The KNN algorithm has been used in statistical estimation, pattern recognition, and other processes since the 1970s. There are two types of KNN algorithm [36], [37]:

1) Structureless nearest neighbour (NN) techniques, where the distance from all training points to the test point is evaluated, and the point with the shortest distance is called the nearest neighbour [37].
2) Structure-based NN techniques, which are based on data structures, such as the orthogonal structure tree (OST), Ball Tree, and nearest feature line (NFL) [37].

The advantages of KNN are its simplicity, easy to implement, and it is effective with large training dataset. These advantages allow the KNN to be integrated or hybridized with the ACO.

C. ACO-KNN Feature Selection Technique for Mining Relations

The idea of using the ACO as the feature selection technique was derived from [7], and this technique was adjusted to fit the type of dataset used in this study. The feature selection problem is depicted in Figure 1, where each node represents a feature, while the edges between the nodes are the paths to the next node. The search process for the minimum feature subset normally starts with the movement of ants through the nodes that are present in the graph. Each node that the ants pass needs to be assessed by a subset assessment method, such as an entropy-based measure [38], and a mutual information evaluation function (MILF)[28].

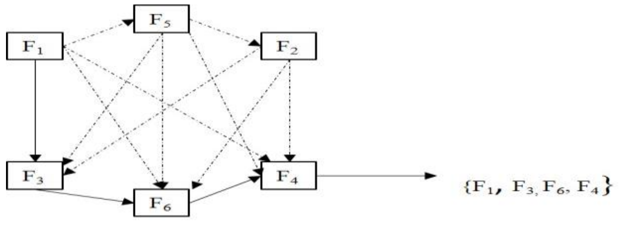

Figure 1: Graphical illustration of the movement of ants during feature selection

Based on the previous example, an ant at F_1 has many routes (dotted lines) to choose to travel to the next node. If the ant selected F_3 as its next node based on the probabilistic transition rule formula, then, F_6 and F_4 will be chosen as the nodes through which it will continue its journey. Thus, the selected nodes would become subset A $\{F_1, F_3, F_6, F_4\}$. If subset A matches the stopping criterion that has been set for traversing the search space, then the search will end. An example of a stopping criterion is that when subset A has achieved a high classification value, the ants need to finish the search process when the optimum feature subset has been acquired. If the stopping criterion is not met, the pheromone will be updated, a new set of ants will be constructed, and the process of searching for features will be repeated again.

D. The Process of Feature Selection

It is important to identify the actual features commented by the customers prior to the process of identifying their relationships because each sentence might contain more than one feature. The ACO-KNN approach was used to produce an optimum feature set. Features that have been selected by the ACO and KNN for each sentence were compared with the optimum feature set. If a selected feature is present in the optimum set, then the next process would be to identify the type of the relationship based on Table 1.

Words that were related to a feature were compared with a list of positive or negative sentiment words. This step was conducted to categorise the relationship between the feature and the sentiment word into either the positive or the negative sentiment group.

E. Identifying The Actual Features in the Customer's Review Sentences

To identify the actual feature that was commented by users, a method proposed by [39] was used to identify the sentiment word that can be connected to it. A sentiment word can describe elements of positive sentiments, such as good, excellent, and amazing. Meanwhile, a sentiment word can also portray negative sentiments, such as bad and worst. This study applied a technique by [39] to determine the actual feature as commented by customers. The dependency relation algorithm [39] was used to extract the features and identify the sentiment words related to them. Somprasertsri et al. [39] suggested six types of relationships, as listed in the following Table 1.

Table 1
Types of Relationships [39]

Relationship	Description
Child	Product feature depends on the sentiment word
Grandchild	Product feature depends on the word that depends on the sentiment word
Sibling	Both the sentiment word and the product feature depend on the same word.
Parent	Sentiment word depends on the product feature
Grandparent	Sentiment word depends on the word that depends on the product feature.
Indirect	None of the above relationships

Figure 2 illustrates the dependency relation that describes the relationship between a feature and a sentiment word.

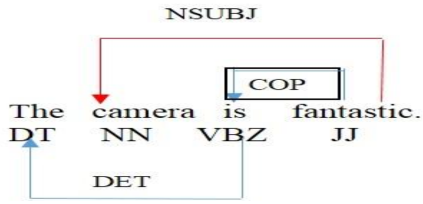

Figure 2: Dependency relation

Table 2
Sample of a Sentence

Sentence 1= "The camera is fantastic."	Dependency Relations
Feature = 'camera' Sentiment word = 'fantastic'	NN→JJ

In Table 2, sentence 1 has the 'child'-type relationship. The product feature is the 'child'. This relationship explains that the product feature is the subject or object of the verb. The sentiment word is the verb or a complement of the copular verb. The 'child' relationship means that the product feature depends on the sentiment word.

For example, typed dependency relation: -

{det(camera-2, The-1, nsubj(fantastic-4, camera-2, cop(fantastic-4, is-3)}

The phrase "*camera*" is the product feature. It is also a noun phrase that represents the subject. The word "*fantastic*" is the sentiment word, which is also a complement of the copular verb.

F. Sentiment Classification

Each comment in the customer review dataset contains a range of information; features, sentiment words as well as sentiment strength, which can be either positive or negative.

The proposed ACO-KNN algorithm was used to determine the optimal feature subset. Results from the ACO-KNN were used as input for the algorithm proposed by [39] to detect the sentiment words connected to the features found in customers' sentences. The results of this process were categorised as {feature, sentiment word, type of sentiment}. The outputs were manually compared with the customer data review set.

When the review process was completed, the next process was the evaluation. This process was conducted to determine the performance of the ACO-KNN as a feature selection algorithm when choosing the actual features commented by the customers. Thus, in this evaluation process, precision, recall, and F-score were utilised to measure its effectiveness in identifying the relations between the features and the sentiment words. The precision, recall, and F-score were calculated using the following formulas [40]:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

$$\text{F-score} = \frac{(2*(Precision*Recall))}{Precision+Recall)} \tag{3}$$

Where TP (true positive) is the number of true relations between features and sentiment words. False positive (FP) is the number of unidentifiable relations; feature or sentiment word. False negative (FN) is the number of relations that the algorithm has failed to determine between the feature and the sentiment word. If no relationship can be determined, then, it is possible that the feature did not match the subset feature in the sentence. In addition, it is also possible that the sentiment word has been implicitly described. The evaluation results are listed in Table 4.

III. Statistical Validation

In this study, statistical validation in the form of a significance test was used to evaluate the mean difference between two experimental results for algorithm performance. The two methods that can be used to test significance level are the parametric and non-parametric tests.

1) Non-parametric tests, such as the Wilcoxon signed-rank test can be used to test the significance level of a model if the experiment has a distribution that is abnormal or has a small number of sample testing [41].
2) Parametric test methods, such as the t-test and z-test can be conducted when the results of two experiments are normal and have a number of experiments with at least 30 sets of testing [42].

In this study, the parametric test (t-test) was applied because this experiment has more than 30 samples. The performance of the algorithm was considered significant if the ACO-KNN meet two criteria; (1) the value of p must be less than the significance level ($p < 0.05$), and (2) the mean value of the proposed algorithm must be greater than that of the baseline algorithms. The mean value in this study refers to the larger average value for the group of data set that used the proposed algorithm, compared to the average value for the group of data set for baseline algorithm. The average values for each group were obtained from the t-test. Additionally, one of the essential parts in data mining research is to statistically validate the experimental results. The differences between the algorithms can show whether the proposed algorithm was significant or not. In this study, the t-test was used to determine the significance of the two algorithms. The statistical validation approaches in this study included; (1) the significance test, and (2) the mean value of the algorithm.

IV. Experimental Setup

The proposed algorithm was tested on five benchmark customer review datasets on electronic products collected by Hu and Liu [43] from the global retailer, Amazon (www.amazon.com; see Table 3). These datasets have already been manually annotated by [43]. In order to evaluate the proposed algorithm according to a data mining approach, the dataset was divided into a group of training data. This dataset was divided into 10 fractions of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%. Each fraction has a different set of data, which did not overlap. The value of 10 was given because it has been comprehensively tested, and it was indicated that this value is the most appropriate representation in generating expected errors [44].

Abbasi et al. [9] and Zhu et al. [20] were the earliest studies that applied the metaheuristic technique. Abbasi et al. [9] combined the GA algorithm with the IG filter algorithm, while Zhu et al. [20] used only GA as the feature selection algorithm. Thus, the study by Abbasi et al. [9] was chosen as the baseline algorithm. Furthermore, another baseline algorithm, which was a combination of IG algorithm and the RSAR algorithm by Agarwal & Mittal [10], was also chosen. This is because the RSAR has proven its efficiency as a feature selection algorithm, capable of removing irrelevant features and noise, and yet, it has difficulty in obtaining optimal feature subset. Since Abbasi et al. [9], and Agarwal and Mittal [10] used different datasets compared to the dataset in this study, the basic comparison technique was reevaluated using the dataset in this study. This means that the proposed ACO-KNN algorithm, and the baseline algorithms (IG-GA and IG-RSAR) must be applied 10 times for each dataset with different data percentages.

Table 3
Summary of Tagged Features for Each Dataset

	Apex	Canon	Creative	Nokia	Nikon
Number of sentences	739	597	1716	546	346
Number of manually tagged features	110	100	180	109	74

V. Experiment and Result

Table 4 shows the comparative results between ACO-KNN, IG-GA, and IG-RSAR in terms of the average value of precision (P), recall (R), and F-score for every dataset. The results of the significance test are shown in Table 5 (ACO-KNN & IG-GA), and Table 7 (ACO-KNN & IG-RSAR). The p value represents the t-test, where the p value of the ACO-KNN should be less than 0.05 to make it statistically significant compared to the IG-GA and IG-RSAR. The p value for the data on Canon was 0.1107, which was greater than $p = 0.05$. These results indicated that ACO-KNN and IG-GA were not statistically significant. Similarly, the p value for the data on Creative was 0.2118, which was greater than $p = 0.05$. Nonetheless, the data on Canon in Table 6 showed that the average mean value for ACO-KNN was 83.1% higher than the average mean value for the IG-GA algorithm. Correspondingly, the average value for ACO-KNN was 85.4% higher than the IG-GA algorithm for the data on Creative. However, Table 5 shows that the mean values for the ACO-KNN algorithm were higher than the IG-GA algorithm for all datasets.

Table 4
The Comparative Results (Precision, Recall, and F-score (FS)) for ACO-KNN, IG-GA, and IG-RSAR Algorithms

Dataset	ACO-KNN			IG-GA			IG-RSAR		
	P	R	FS	P	R	FS	P	R	FS
Nikon	89.2	92.7	90.7	74.1	76	74.1	72.3	84	77.3
Nokia	81.3	84.6	82.7	73.3	61.6	65.9	62.8	61	61.6
Apex	71.5	71.8	71.5	63	60.5	61.6	62.4	60.5	60.6
Canon	80.6	86	83.1	78.8	83.4	80.5	76.2	85.3	80.2
Creative	84.8	86	85.4	82.6	84.5	83.5	78.3	80.2	79.2

Table 5
The t-Test for ACO-KNN and IG-GA

Dataset	p value	Significant
Nikon	0.0489	+
Nokia	0.0002	+
Apex	0.0012	+
Canon	0.1107	-
Creative	0.2118	-

Table 6
The Mean Values for ACO-KNN and IG-GA

	ACO-KNN	IG-GA
Nikon	81.6	74
Nokia	82.7	65.9
Apex	71.5	61.6
Canon	83.1	80.6
Creative	85.4	83.5

Table 7 shows that the ACO-KNN was statistically improved when its significance results in most datasets were less than the significance level, which was $p < 0.05$, compared to the IG-RSAR. Meanwhile, Table 8 shows that the mean values for ACO-KNN were higher than for the IG-RSAR algorithm for all datasets.

Table 7
The t-test for ACO-KNN and IG-RSAR

Dataset	p value	Significant
Nikon	0.0195	+
Nokia	0.0027	+
Apex	0.0083	+
Canon	0.0085	+
Creative	0.0494	+

Table 8
The mean values for ACO-KNN and IG-RSAR

	ACO-KNN	IG-RSAR
Nikon	81.6	77.3
Nokia	82.7	61.6
Apex	71.5	60.6
Canon	83	80
Creative	85.4	79.2

These experimental results showed that the ACO-KNN algorithm was significant as a feature selection technique for selecting relevant and optimum features.

VI. CONCLUSION

This study has evaluated the performance of the proposed ACO-KNN algorithm using precision, recall, and F-score. Moreover, the statistical tests included a significance test, and the mean values for these algorithms were used to validate the performance of the ACO-KNN, IG-GA, and IG-RSAR algorithms. These evaluations have statistically proven that the ACO-KNN algorithm had surpassed the IG-GA and the IG-RSAR in all performance metrics with significant differences. This statistical analysis has proven that the ACO-KNN was effective as a feature selection technique in sentiment analysis, for choosing relevant feature subsets, and for representing the actual data. Appropriate choice of feature selection technique can improve the performance of sentiment classification, as well as be helpful in determining the actual feature commented by the customers.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Ministry of Higher Education and Universiti Pertahanan Nasional Malaysia for supporting this research project.

REFERENCES

[1] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[2] H. Arafat, R. M.Elawady, S. Barakat, and N. M.Elrashidy, "Different Feature Selection for Sentiment Classification," *International Journal of Information Science and Intelligent System*, vol. 1, no. 3, pp. 137–150, 2014.

[3] A. Abbasi, S. France, Z. Zhang, and H. Chen, "Selecting Attributes for Sentiment Classification Using Feature Relation Networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 3, pp. 447–462, 2011.

[4] Vinodhini G. and R. Chandrasekaran, "Effect of Feature Reduction in Sentiment Analysis of Online Reviews," *International Journal of Advance in Computer Engineering & Technology*, vol. 2, no. 6, pp. 2165–2172, 2013.

[5] S. R. Ahmad, A. Abu Bakar, and M. R. Yaakub, "Metaheuristic Algorithms for Feature Selection in Sentiment Analysis: A Review," in *Science and Information Conference (SAI)*, 2015, pp. 222–226.

[6] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, 2005.

[7] M. H. Aghdam, N. Ghasem-Aghaee, and M. E. Basiri, "Text Feature Selection using Ant Colony Optimization," *Journal Expert Systems with Applications: An International Journal*, vol. 36, no. 3, pp. 6843–6853, 2009.

[8] A. A. B. M. R. Y. S. R. Ahmad, "Metaheuristic algorithms for feature selection in sentiment analysis," in *Science and Information Conference (SAI), 2015*, 2015, pp. 222–226.

[9] A. Abbasi, H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web forums," *ACM Transactions on Information Systems*, vol. 26, no. 3, pp. 1–34, 2008.

[10] B. Agarwal and N. Mittal, "Sentiment Classification using Rough Set based Hybrid Feature Selection," in *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2013)*, 2013, no. June, pp. 115–119.

[11] Y. Saeys, I. Inza, and P. Larrañaga, "A Review of Feature Selection Techniques in Bioinformatics," *Bioinformatics (Oxford, England)*, vol. 23, no. 19, pp. 2507–2517, 2007.

[12] R. Jensen and Q. Shen, "Fuzzy-Rough Sets Assisted Attribute Selection," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, 2007.

[13] R. Jensen and Q. Shen, "Fuzzy-rough attribute reduction with application to web catagorization," *In the Transaction on Fuzzy Sets and System*, vol. 141, no. 3, pp. 469–485, 2004.

[14] R. Jensen and Q. Shen, "New Approaches to Fuzzy-Rough Feature Selection," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, 2009.

[15] R. Kohavi and G. H. John, "Wrappers for Feature Subset Selection," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, 1997.

[16] M. Mafarja and D. Eleyan, "Ant Colony Optimization based Feature Selection in Rough Set Theory," *International Journal of Computer Science and Electronics Engineering (IJCSEE)*, vol. 1, no. 2, pp. 244–247, 2013.

[17] A. Unler and A. Murat, "A Discrete Particle Swarm Optimization Method for Feature Selection in Binary Classification Problems," *European Journal of Operational Research*, vol. 206, no. 3. pp. 528–539, 2010.

[18] S. C. Yusta, "Different Metaheuristic Strategies to Solve the Feature Selection Problem," *Pattern Recognition Letters*, vol. 30, no. 5. pp. 525–534, 2009.

[19] M. E. Basiri and S. Nemati, "A Novel Hybrid ACO-GA Algorithm for Text Feature Selection," *2009 IEEE Congress on Evolutionary Computation*, 2009.

[20] J. Zhu, H. Wang, and J. T. Mao, "Sentiment Classification using Genetic Algorithm and Conditional Random Field," in *Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on*, 2010, pp. 193–196.

[21] P. Kalaivani and K. L. Shunmuganathan, "Feature Reduction Based on Genetic Algorithm and Hybrid Model for Opinion Mining," *Scientific Programming*, p. 15, 2015.

[22] Z. Liu, S. Liu, L. Liu, J. Sun, X. Peng, and T. Wang, "Sentiment recognition of online course reviews using multi-swarm optimization-based selected features," *Neurocomputing*, vol. 185, pp. 11–20, Apr. 2016.

[23] Y. Jin, W. Xiong, and C. Wang, "Feature Selection for Chinese Text Categorization Based on Improved Particle Swarm Optimization," in *Natural Language Processing and Knowledge Engineering (NLP-KE)*, 2010, pp. 1–6.

[24] H. K. Chantar and D. W. Corne, "Feature Subset Selection for Arabic Document Categorization using BPSO-KNN," in *Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress*, 2011, pp. 546–551.

[25] M. Dorigo and G. Di Caro, "Ant colony optimization: A new metaheuristic," *Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on*, vol. 2, pp. 1470–1477, 1999.

[26] M. Dorigo and T. Stützle, "Ant Colony Optimization," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28–39, 2004.
[27] M. Dorigo and T. Stützle, "Ant Colony Optimization: Overview and Recent Advances," in *Handbook of Metaheuristics*, vol. 146, 2010, pp. 227–263.
[28] A. Al-Ani, "Ant Colony Optimization for Feature Subset Selection," *Society*, vol. 4, no. February, pp. 35–38, 2005.
[29] H. R. Kanan and K. Faez, "An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 716–725, 2008.
[30] M. M. Kabir, M. Shahjahan, and K. Murase, "A new hybrid ant colony optimization algorithm for feature selection," *Expert Systems with Applications*, vol. 39, no. 3. pp. 3747–3763, 2012.
[31] M. H. Aghdam, T. Jafar, A. R. Naghsh-Nilchi, and M. E. Basiri, "Combination of Ant Colony Optimization and Bayesian Classification for Feature Selection in a Bioinformatics Dataset," *Journal of Computer Science & Systems Biology*, vol. 2, no. 3, pp. 186–199, 2009.
[32] M. E. Basiri, N. Ghasem-Aghaee, and M. H. Aghdam, "Using Ant Colony Optimization-Based Selected Features for Predicting Post-synaptic Activity in Proteins," in *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2008, pp. 12–23.
[33] S. Nemati, M. E. Basiri, N. Ghasem-Aghaee, and M. H. Aghdam, "A novel ACO–GA hybrid algorithm for feature selection in protein function prediction," *Expert Systems with Applications*, vol. 36, no. 10. pp. 12086–12094, 2009.
[34] E. Sarac and Se. A. Ozel, "An ant colony optimization based feature selection for web page classification," *The Scientific World Journal*, vol. 2014, 2014.
[35] H. S. Alghamdi, H. L. Tang, and S. Alshomrani, "Hybrid ACO and TOFA Feature Selection Approach for Text Classification," in *WCCI 2012 IEEE World Congress on Computational Intelligence*, 2012, vol. 1–6, pp. 1–6.
[36] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," *Procedia Technology*, vol. 10, pp. 85–94, 2013.
[37] N. Bhatia and Vandana, "Survey of Nearest Neighbor Techniques," *International Journal of Computer Science and Information Security*, vol. 8, no. 2, p. 4, 2010.
[38] R. Jensen, "Combining rough and fuzzy sets for feature selection," Edinburgh University, 2005.
[39] G. Somprasertsri and P. Lalitrojwong, "Mining Features-Opinion in Online Customer Reviews for Opinion Summarization," *Journal of Universal Computer Science*, vol. 16, no. 6, pp. 938–955, 2010.
[40] M. R. Yaakub, Y. Li, A. Algarni, and B. Peng, "Integration of Opinion into Customer Analysis Model," in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 2012, pp. 164–168.
[41] M. F. Triola, *Elementary Statistics*, 10, illust ed. Pearson, 2007.
[42] J. Demšar, "Statistical Comparisons of Classifiers over Multiple Data Sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.
[43] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining KDD 04*, pp. 168–177, 2004.
[44] I. H. Written and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edi. Elsevier Science, 2005.