The Influence of Parameter Choice on the Performance of SVM RBF Classifiers for Argumentative Zoning

Renny Pradina Kusumawardani*, Bambang Riyanto Trilaksono, Masayu Leylia Khodra****
* Department of Information System, Faculty of Information Technology, Institut Teknologi Sepuluh Nopember
** School of Electrical Engineering and Informatics, Institut Teknologi Bandung

Keywords:
automatic classification
Argumentative Zoning
Support Vector Machine
RBF kernel
best parameters
grid search
feature selection

ABSTRACT

Today scientists are inundated by the plethora of works that may or may not be relevant to their research interests and needs. One of the concepts proposed for overcoming this problem is Argumentative Zoning. It classifies information within scientific papers by assigning labels to sentences according to their rhetorical role. In a previous work, Support Vector Machines with RBF kernels are shown to give the best results when compared to other methods for the task of Argumentative Zoning. This paper aims to investigate the influence of further treatment on their performance, namely, using grid search for finding optimal parameter values, both under and without feature selection. Experiment results show that feature selection generally gives higher mean accuracy value when compared with when all features are used, averaging at 65.36% with a maximum value of 68.10%. However, when feature selection is not performed, using parameter values from the grid search, accuracy may reach 73.95%, although the average performance is lower. These results indicate that for the task of Argumentative Zoning with SVM RBF classifiers, good selection of parameter values is more important when compared feature set optimization.

Corresponding Author:

Renny Pradina Kusumawardani,
Departement of Information System, Faculty of Information Technology,
Institut Teknologi Sepuluh Nopember,
Jalan Raya Kampus ITS, Gedung Sistem Informasi, Sukolilo, Surabaya, Indonesia.
Email: renny@is.its.ac.id

1. INTRODUCTION

The vast availability of electronic documents nowadays has made it more difficult to find relevant information. There is a multitude of techniques invented and developed to overcome this problem, the most commonly used being the internet search engine. However, such solutions usually applies only to the general domain, and may not be entirely suitable for specific purposes, such as for scientific texts.

Due to the specific nature of scientists' information needs, there are specific important aspects in scientific texts; for example, attribution of parts of the scientific works and the aim of a particular paper. A system which is able to correctly recognize these aspects automatically will be very important in assisting the retrieval of relevant information. Argumentative Zoning [1] is a concept which classifies information within scientific papers by labeling sentences in the text based on their rhetorical roles. One of the basic premises of Argumentative Zoning is that such assignment can be done based on the shallow characteristics of the sentence. This enables the classification to be performed automatically using machine learning methods.

Various algorithms may be used for doing the actual classification task. In [2] it has been argued that SVMs are suitable for the task of Argumentative Zoning. In particular, SVM RBF deliver better performance when compared to a SVM of linear kernel, and also when compared to the baseline classifier, Naïve Bayes.

However, an important question remains. In Argumentative Zoning, various sets of attributes have already been proposed as features; for example by [1], [3], and [4]. The question of what set of parameters is most beneficial to the task of Argumentative Zoning has already been discussed in [5]. These works have given us a good idea on the set of features which would deliver good results. Yet we still do not know, whether given the choice of the algorithm for building the classifier, the treatment for optimizing the set of features will actually give the best performance.

According to Hsu, et. al. [3], the performance of SVM classifiers is highly dependent on the choice for their parameter values. As these values practically characterize the classifier, this paper aims to answer the question on the relationship between a specific classifier, the SVM RBF, and the optimized feature set by comparing the performance of the SVM RBF under combinations of feature selection and parameter values. Specifically, we will compare the performance of the SVM RBF classifier under systematic parameter search both when feature selection is performed and without feature selection.

2. RELATED WORKS

Argumentative Zoning is first proposed by Teufel [1] as a concept for classifying information within scientific texts. It assigns labels to each sentence in the text based on their rhetorical role. In [1], each sentence is assigned one of the seven categories enlisted in Table 1.

Table 1. Annotation schema in Argumentative Zoning 7 [1]

Category	Sentence Description
BACKGROUND	Describes background knowledge that is generally accepted in the field.
OTHER	Contains description of other work or knowledge claim that is not claimed by the current paper. Both knowledge claims held either by other authors or by the author of the current paper in a previous paper fall into this category.
OWN	Covers own work; including method, results, conclusions, proposal for future works, or other aspects in the current work.
AIM	States the specific research goal or hypothesis of the paper.
TEXTUAL	The sole purpose is to give information on the structure of the paper
CONTRAST	Contrasts or compares the current work with other works, or points out weaknesses of other works.
BASIS	Describes other works which provide the basis for the current work.

a. Features used in Argumentative Zoning

One of the basic premises of Argumentative Zoning is that the classification can be performed by shallow characteristics of the texts that would enable these to be captured easily as features and subsequently used for automatic classification using machine learning techniques.

Table 2 presents the features originally proposed by Teufel [1]. Other works have proposed variations of this set, for example Teufel and Moens [4] and Merity, et. al. [5]. Table 3 summarizes these variations. Of these sets, Merity et. al. achieved the best performance with an overall accuracy of 70.25%.

Given the variety of features used by researchers, it is interesting to investigate the influence of the features on the performance of the classifier. This is the object of discussion in [6].

b. SVM RBF Parameter Optimization

The work of Kusumawardani [2] explores the use of SVM classifier for the task of Argumentative Zoning, comparing SVM classifiers with RBF and linear kernels and also with the baseline of Naïve Bayes. The results shows that for this task using SVM classifier with RBF kernel gives better performance when compared to the other two classifiers. However, the use of SVM RBF classifiers requires selecting parameter values, as can be seen in the mathematical formulation for SVM RBF in Equation 1.

$$K(x_i, x_j) = \exp(-\gamma \parallel x_i - x_j \parallel^2), \quad \gamma > 0 \qquad \text{(Eq. 1)}$$

x_i, x_j : feature vectors of a pair of data instance

$K(x_i, x_j)$: kernel value for instance pairs x_i and x_j

γ : parameter value for the RBF

In most cases, it is unnecessarily difficult or even impossible to try to find a decision boundary which would cleanly separates between classes. A solution to this problem is suggested by Cortes and Vapnik [7] by incorporating an upper bound on the number of allowable training error. This approach is called the *soft-margin* SVM. This requires the selection of the constant C, which will be the upper bound of error. Therefore, there are two parameters that needs to be optimized, the γ parameter of Equation 1 and the constant C.

Table 2. Teufel's features [1]

Feature	**Description**
Cont-1	Whether the sentence contains significant terms, as determined by the tf/idf measure
Cont-2	Whether the sentence contains words that are also a part of the title of the paper
Loc	Sentence position relative to a defined segmentation
Struct-1	Position of the sentence within the section
Struct-2	Relative position of the sentence within the paragraph
Struct-3	Type of the title of the section in which the sentence is located
Length	Whether the sentence contains more words than a defined threshold
Syn-1	The voice of the first finite verb in the sentence
Syn-2	The tense of the first finite verb in the sentence
Syn-3	Whether the first finite verb of the sentence is modified by modal auxiliary
Cit-1	Whether the sentence contain a citation of an author in the reference list, either formally (which enable direct linking to a specific reference in the reference list) or only mentioning the name
Cit-2	Whether the citation in the sentence cites a work of one of the authors of the current paper
Cit-3	Relative location of the citation in the sentence
Formu	Type of the first formulaic expression that is found in the sentence
Ag-1	Type of the first agent found in the sentence
Ag-2	Type of the first action found in the sentence

Table 3. Summary of research settings of previous works

	Teufel [1]	**Teufel and Moen [4]**	**Merity et al. [5]**	
			1	**2**
Task	Argumentative Zoning 7			
Corpus	Teufel [1] corpus			
Classifier	Naïve Bayes	Naïve Bayes	Naïve Bayes	Maximum Enthropy
Features	Teufel [1]	Teufel [1] + history	Teufel [1] – verb syntax	Teufel [1] – verb syntax

3. RESEARCH METHOD

In order to observe the influence of the choice of parameter values and feature selection on the performance of the classifier, in this research we use the black-box method as described by Kohavi [8]. Figure 1 illustrates the set-up we use for applying this method for our case.

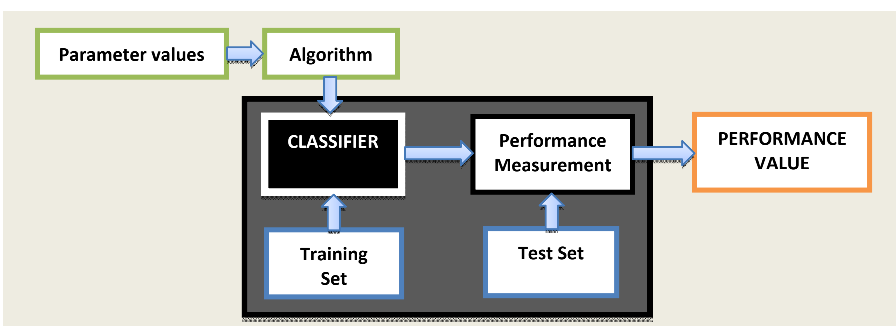


Figure 1. Experiment Setting

In this paper, we use SVM RBF classifiers, adjusting two parameter values; the γ of the RBF kernel and the soft-margin SVM parameter C. The implementation used is the LibSVM by Chang and Lin [9].

The performance metric used here is the accuracy of the classifier. Accuracy values are reported from experiments using 5-fold cross-validation [8]. This method is applied on a corpus consisting of 9200 sentences, each labeled according to the seven-category scheme of Teufel [1]. The sentences are from 60 papers on the computational linguistic domain, a subset of those used by Teufel in her corpus. A further description of this corpus can be found in [2].

From each sentence in the corpus, features listed in Table 4 are extracted. These features are selected by combining Teufel's and Merity et. al.'s features. However, some features are omitted, either because our experiments show them to be ineffectual, or because they require extensive time and computational resources. Furthermore, Merity et.al. obtain better accuracy than Teufel when syntactic features are not included, thus we obmit these as well. In this experiment we also use the unigram features which are absent from both Teufel and Merity et. al. A more thorough discussion on this can be found in [2] and [6].

Table 4. Features extracted from sentences in this research

Feature Names				
Cont-1	Struct-1	Loc	Cit-1	Cit-3
Cont-2	Struct-2	Length	Cit-2	Unigram

The Unigram feature codes whether a word of the corpus exists in the sentence.
For the meaning of other feature names, please refer to Table 2.

a. Observing the Influence of Parameter Values and Feature Selection

The work described in this paper arises from the need to find the best parameter values for the SVM classifiers in [2]. For this purpose, we use the grid search as they are computationally simple and efficient [3]. Moreover, some researches also indicate that the parameters found using the grid often results in better classifier performance when compared to other complicated, more sophisticated methods; for example as shown by Boardman and Trappenberg [10].

In this research, we use a procedure similar to Hsu et. al. [3], in which at first a coarse grid is used. A finer grid is subsequently selected by choosing the range which gives the best performance at the previous iteration. The procedure stops when using finer grids does not result in improvement when compared to the coarser grid of the previous iteration.

4. RESULTS AND ANALYSIS

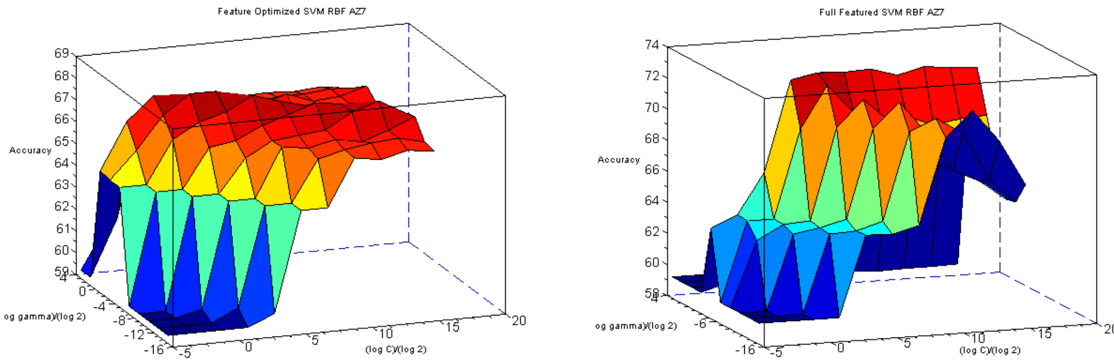
Using the grid search procedure of Hsu et. al. [3], the best grids found in our experiment are as shown in Table 5. These grids result in 190 performance data points when all features are used and 182 performance data points when feature selection is performed.

Table 5. The best grids found in the experiment

With All Features						
Iteration	**exp(C)**			**exp(γ)**		
	Lower Value	**Upper Value**	**Step**	**Lower Value**	**Upper Value**	**Step**
Iteration 1	-5	15	2	-15	3	2
Iteration 2	0	8	1	-6	-4	1
Iteration 3	2.5	7.5	0.5	-5.5	-4.5	0.5
Iteration 4	2.75	3.25	0.25	-5.25	-4.75	0.25
With Feature Optimization [6]						
Iteration	**exp(C)**			**exp(γ)**		
	Lower Value	**Upper Value**	**Step**	**Lower Value**	**Upper Value**	**Step**
Iteration 1	-5	15	2	-15	3	2
Iteration 2	3.5	6.5	1	-6.5	-3.5	1
Iteration 3	4.75	5.25	0.5	-5.25	-4.75	0.5

At first glance, there seems to be little regularity shown by these performance points. However, on a closer inspection, it is apparent that with full feature set the performance of the SVM RBF classifier is somewhat more fluctuated; achieving high peaks at some points, but falling sharply at parameter values adjacent to those peaks. The performance given by optimized feature set seems to be more stable; however, it is interesting to note that the maximum accuracy given by the optimized feature set, which is 68.10%, is considerably lower than the maximum performance with full feature set at 73.95% accuracy. Therefore, in order to obtain more insight into the behaviour of SVM RBF classifier, we plot the accuracy values obtained from the experiments against the parameter field. The result is shown in Figure 2a. for the SVM RBF with feature optimization and in Figure 2b. for when all features are included.

Comparing these two figures, we see that the surface plot for accuracy with feature selection in Fig.3 starts at a steeper slope, and then when it reaches fairly high accuracy values, it settles. No sharp increase or decrease is subsequently observed. This is in contrast with when all features are used. Without feature selection, the accuracy surface starts rising gradually, reaching low values at most points. However, the accuracy rises steeply until it reaches some peaks. In order to get a more accurate idea of this behavior, we plot the histograms and calculated some descriptive statistics for the accuracy values produced in both cases. These are shown in Figure 3.



a. With feature selection   b. All Features are included

Figure 2. Plot of accuracy versus log C/log 2 and log γ/log 2 for SVM RBF

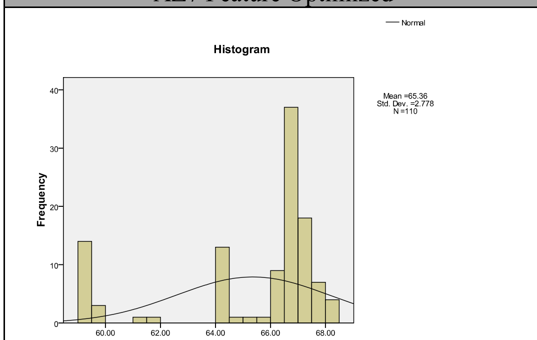
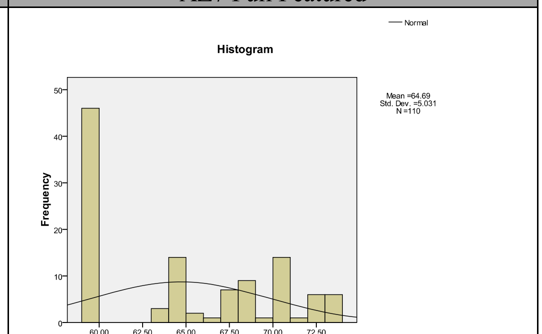
AZ7 Feature Optimized		AZ7 Full Featured	
			
Min.	59.37	Min.	59.37
Max.	68.10	Max.	73.95
Mean	65.3637	Mean	64.6919
Std. Dev.	2.77752	Std. Dev.	5.03097

Figure 3. Histograms and statistics on the accuracy of the SVM RBF classifier

Figure 3 shows that the mean accuracy of the SVM RBF on the task of Argumentative Zoning is higher when an optimized set of features is used. The histogram shows that on similar parameter fields, the accuracy of the classifier with feature optimization tends to higher values. This is in contrast with when all features are used; most parameter combinations results in lower accuracy. However, the maximum accuracy of 73.95% with all features is considerably higher than the maximum for the optimized set of only 68.10%.

What these seem to indicate is that using SVM RBF, it is more beneficial to include the full feature set, as it gives the SVM RBF the opportunity to reach a high performance. This is in agreement with the argument of Joachim [11] that the SVM has the ability to process large dimensions of features. However, we must be cautious with respect to the selection of parameter values, as incorrect selection of these values may result in low performance. This is as indicated by the standard deviation in both cases; the higher value of 5.03 in the case of all features compared to 2.7 for optimized feature set shows that with respect of parameter values, the performance given by all features is less stable. In conclusion, even though without feature selection the classifier may achieve higher performance, it is important that in such case parameter values are selected carefully.

5. CONCLUSION

From the discussion in the previous sections, it can be concluded that for the SVM RBF, omitting the process of feature selection opens the possibility to the classifier to achieve higher performance than otherwise. However, due to the fluctuant nature of the performance of SVM RBF in such setting, parameter values of the classifier must be carefully chosen.

In this work, we have explored the use of the grid approach for selecting parameter values of the SVM RBF, as this method has been shown to give good or better results when compared to other approaches. Naturally, the maximum found is a local maximum with respect to the selection of initial grid values. In future work, it is interesting to explore this issue further, either through more principled selection of initial values or the exploration of other parameter search methods.

REFERENCES

[1] S. Teufel. “Argumentative Zoning: Information Extraction from Scientific Text”. PhD. dissertation, University of Edinburgh, Edinburgh, 1999.
[2] R. P. Kusumawardani. “Automatic Sentence Classification for Argumentative Zoning using Support Vector Machines”. Master’s thesis, Institut Teknologi Bandung, Bandung, Indonesia. 2010.
[3] C. W. Hsu, C. C. Chang, C. J. Lin. *A Practical Guide to Support Vector Classification*. 2008.
[4] S. Teufel, and M. Moens, “Summarising Scientific Articles – Experiments with Relevance and Rhetorical Status”. *Computational Linguistics*. 2002; 28(4):409–445.
[5] S. Merity, T. Murphy, J. R. Curran. “Accurate Argumentative Zoning with Maximum Entropy Models”. *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries, ACL-IJCNLP 2009*. 2009; pp. 19-26.
[6] R. P. Kusumawardani, B. R. Trilaksono, M. L. Khodra. “Feature Selection for Automatic Classification of Sentences in Scientific Texts”. *Proceedings of the 2010 National Conference in Intelligent Computation Systems*. 2010.
[7] C. Cortes and V. Vapnik. “Support-Vector Networks” in *Machine Learning*, vol. 20, number 3. Springer Netherlands, 1995, pp 273 - 297.
[8] R. Kohavi. “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection”, in *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995.
[9] C. C. Chang and C. J. Lin. “LIBSVM: A Library for Support Vector Machines”. 2001.
[10] M. Boardman and T. Trappenberg. “A Heuristic for Free Parameter Optimization with Support Vector Machines”. *2006 International Joint Conference on Neural Networks*. 2006.
[11] T. Joachims. “Text Categorization with Support Vector Machines: Learning with Many Relevant Features”. *Proceedings of the European Conference on Machine Learning (ECML)*, Springer, 1998.

BIOGRAPHY OF AUTHORS

	Renny Pradina Kusumawardani graduated from the Department of Electrical Engineering, Institut Teknologi Bandung (ITB) in 2004. Her Master’s degree is also from ITB, where she graduated in 2010 *cum laude*, ranking first in the Informatics program. Currently she is an Associate Professor at the Department of Information Systems, Institut Teknologi Sepuluh Nopember (ITS) Surabaya, Indonesia. She has authored several conference paper, both at the national and international level, one of which won her a Best Paper Award in 2010.
	Prof. Bambang Riyanto Trilaksono received his bachelor degree in Electrical Engineering from Institut Teknologi Bandung (ITB) in 1986, and his Master’s and Doctoral Degree from Waseda University, Japan, in 1991 and 1994, respectively. He is currently a Professor at the School of Electrical Engineering and Informatics, ITB. His research interests include robust control, intelligent control and intelligent systems, multi agent systems, control applications, robotics, and machine learning, in which he has undertaken several high-profile projects such as the development of aircraft simulation and in ground flight control test for the N250 aircraft. He has published over 200 papers in journals and conferences and received several awards including the Indonesian Toray Science Foundation Research Award in 2004. He is currently the Director of Advanced Robotics Research Laboratory, ITB. He is the chief editor of ITB Journal of Science, ITB Journal of Engineering Science and ITB Journal of Information and Communication Technology, while also serving as the editorial board member for several other international journals. He is a member of the IEEE and ACPA, and a research fellow of the University of New South Wales, Australia.

Dr. Masayu Leylia Khodra received her bachelor degree in Informatics from Institut Teknologi Bandung (ITB), where she also obtained her Master's degree in 2006 and Doctoral degree in 2012, both *cum laude*. She is one of the scientists at the forefront of Natural Language Processing in Indonesia. She has published more than thirty conference papers, most of which are international conferences, and authored five papers published in both national and international journals.