Afr. J. Lib. Arch. & Inf. Sc. Vol. 29, No. 2 (2019) 103-113

Investigating Optimal Feature Selection Method to Improve the Performance of Amharic Text Document Classification

Tamir Anteneh Alemu and
Alemu Kumilachew Tegegnie
Faculty of Computing, Bahir Dar institute of Technology (BiT),
Bahir Dar University, Bahir Dar, Ethiopia,
tamirat.1216@gmail.com
alemupilatose@gmail.com

Abstract

*Feature selection is one of the famous solutions to reduce high dimensionality problem of text categorisation. In text categorisation, selection of good features (terms) plays a crucial role in improving accuracy, effectiveness and computational efficiency. Due to the nature of the language, Amharic documents suffered from high dimensionality feature space that degrades the performance of the classifier and increases the computational cost. This paper investigates optimal feature selection methods for Amharic Text Document Categorisation among various feature selection techniques such as Term Frequency*Inverse Document Frequency (tf*idf), Information Gain (IG), Mutual Information (MI), Chi-Square (-X^2), and Term Strength (TS) using Support Vector Machine (SVM) classifiers. Experimentations carried out based on the collected datasets showed that X^2 and IG method performed consistently well on Amharic document Texts among other methods. Using both methods, the SVM classifier showed a significant improvement of the classification accuracy and computational efficiency.*

Keywords: Feature selection, Amharic, SVM, Classification

Introduction

Amharic is the working language of the Federal Government of Ethiopia. It is the native language of people living in the north central part of Ethiopia. The language is also spoken as a second language in many parts of the country. Significant number of immigrants in the Middle East, Asia, Western Europe and North America also speak Amharic (Solomon and Menzel, 2007). Amharic is written from left to right similar to English unlike other Semitic languages, such as Arabic and Hebrew. Amharic language has its own writing system that uses the Ge'ez alphabet. The Amharic writing system consists of a core of thirty three characters each of which occur in basic form and in six other forms called orders (Alemu, 2010).

Due to the advancement of technology, there are numerous electronic documents produced and stored in Amharic. As most of information is stored in the form of texts, text mining has gained paramount importance. With the high availability of information from diverse sources, the task of automatic categorisation of documents has become a vital method for managing, organising vast amount of information and knowledge discovery. Text classification is the task of assigning predefined categories to documents. With the increasing availability of text documents in electronic form, it is of great importance to label the contents with a predefined set of thematic categories in an automatic way, what is also known as automated text categorisation. In the last decades, a growing number of advanced machine learning algorithms have been developed to address this challenging task by formulating it as a classification problem (Yiming and Pedersen, 2015; Bo, 2016; Thorsten, 1998). Commonly, an automatic text classifier is built with a learning process from a set of pre-labelled documents.

Documents need to be represented in a way that is suitable for a general learning process. The most widely used representation is "the bag of words": a document is represented by a vector of features, each of which corresponds to a term or a phrase in a vocabulary collected from a particular dataset. The value of each feature element represents the importance of the term in the document according to a specific feature measurement (Schütze, Hull, and Pedersen. (1995). A big challenge in text categorisation is the learning from high dimensional data. On one hand, tens and hundreds of thousands terms in a document may lead to a high computational burden for the learning process. On the other hand, some irrelevant and redundant features may hurt predictive performance of classifiers for text categorisation. To avoid the issue of the "curse of dimensionality" and to speed up the learning process, it is necessary to perform feature reduction to reduce the size of features. A common feature reduction approach for text categorisation is feature selection. Feature selection is the process of selecting a specific subset of the terms occurring in the training set and using only this subset as features in the classification algorithm. The feature selection process takes place before the training of the classifier and serves two main purposes. First, it makes training and applying a classifier more efficient by decreasing the size of the effective vocabulary. Second, feature selection often increases classification accuracy by eliminating noise features (Monica and Yiming, 2002; Honavar, 1998).

In the last decades, a number of feature selection methods have been proposed, which can be usually categorised into the following five types of approach: each of which a term – goodness criterion threshold – has been taken to achieve a desired degree of term elimination from the full vocabulary of a document corpus. These methods are: document frequency (DF), information gain (IG), mutual information (MI), X^2 statistic (CHI), term strength (TS) (Yiming and Pedersen, 2015; and Jan, 2016). The purpose of feature selection is to improve the performance of the text classification task by reducing high dimensionality of feature space and identifying a subset of the most useful features from the original entire set of features. Hence, it aimed at making text document classifiers more efficient and accurate. It also provides a way of reducing computation time, improving prediction performance, and a better understanding of the data.

The main focus of this research is to investigate an optimal feature selection method for large Amharic document classification. While feature selection in text categorisation is considered, it is only two feature selection techniques that have been tried for Amharic text. i.e. tf*idf and DF (Zelalem, 2001; Surafel, 2003; Yohannes, 2007; Worku, 2009; Alemu, 2010). However, Automatic text categorisation through many feature selection methods has never been carried out for large Amharic text categorisation problems. The main objective of this study was to investigate the optimal feature selection method using SVM that can improve the performance of the classifier in the process of Amharic text document classification. The study also examined the feature selection methods used in document classification, tested the classifier using testing data set, evaluated and selected an optimal feature selection method that shows high performance in classification accuracy. Thus, we sought answers to the following questions with empirical evidence:

- To what extent can feature selection improve the accuracy of a classifier?
- How much of the document vocabulary can be reduced without losing useful information in category prediction?
- Which feature selection methods are both computationally scalable and high-performing across classifiers and collections?

Literature Review

Feature selection for text classification is a well-studied problem in England and China; its goals are improving classification effectiveness, computational efficiency, or both. Aggressive reduction of the feature space has been repeatedly shown to lead to little accuracy loss, and to a performance gain in many cases. Lewis and Ringuette (1994) used an information gain measure to aggressively reduce the document vocabulary in a naïve Bayes model and a decision-tree approach to binary classification. In this research, the authors present empirical results on the performance of a Bayesian classier and a decision tree learning algorithm on two text categorisation datasets. The results showed that both algorithms achieve reasonable performance and allow controlled trade-offs between false positives and false negatives. Wiener, Pedersen and Weigend (1995) used mutual information and X^2 statistic to select features for input to neural networks. The results indicate that term selection and modified latent semantic indexing (LSI) representations lead to similar topic spotting performance, and that this performance is equal to or better than other published results on the same corpus. In the test of the Expert Network method on CACM documents, for example, an 87% removal of unique words reduced the vocabulary of documents from 8,002 distinct words to 1,045 words, which resulted in a 63% time savings and a 74% memory savings in the computation of category ranking, with a 10% precision improvement on average over not using word removal. Yang (1999) and Schutze et al. (1995) used principal component analysis to find orthogonal dimensions in the vector space of documents. Yang and Wilbur (1995) used document clustering techniques to estimate probabilistic term strength and used to reduce the variables in linear regression and nearest neighbour classification. Moulinier et al. (1996) used an inductive learning algorithm to obtain features in disjunctive normal form of news story categorisation. Lang (1995) used a minimum description length principle to select terms for Netnews categorisation. The results showed that a learning algorithm based on the Minimum Description Length (MDL) principle was able to raise the percentage of interesting articles to be shown to users from 14% to 52% on average. Asker, Argar, Gamback, Asfeha and Habte (2009) made experiments to investigate the effect of operations such as stemming and part-of-speech tagging using Self-organizing Maps (SOM) for classifying Amharic web news. In their findings, the best accuracy was achieved using the full text as representation. Sahlemariam, M., Yacob, D. and Libsie, M. (2009) proposed a framework that automatically categorises Amharic documents into predefined categories using concepts. The finding in this research shows that the use of concepts for an Amharic document categorises results in 92.9% accuracy. Abate and Assabie (2014) proposed a supervised data-driven experimental approach to develop Amharic morphological analyser. The researchers use a memory-based supervised machine learning method which extrapolates new unseen classes based on previous examples in memory. As the result in this study showed that the performance of the model is evaluated using 10-fold cross-validation with IB1 and IGtree algorithms resulting in the overall accuracy of 93.6% and 82.3% respectively.

Haliu and Assabie (2016) presents a system that categorises Amharic documents based on the frequency of item sets obtained after analysing the morphology of the language. The researchers selected seven categories into which a given document is to be classified. In this research work, the task of categorisation is achieved by employing an extended version of *a priori* algorithm to implement the proposed system. The results in their research work show that the proposed item-sets-based method has superior performance over other methods tested so far for Amharic text categorisation.

In recent years, text categorisation has been studied in other languages such as Amharic. As several researchers studied (Zelalem, 2001; Surafel, 2003; Yohannes, 2007; Worku, 2009; Alemu, 2010), the problem is attempted with different machine learning and statistical methods but with a common feature selection method – document frequency method (DF). As far as the knowledge of the researchers, other feature selection methods have never been used for the dimensionality reduction of the feature space in Amharic text documents, which is the concern of this paper. Hence, the results of classification accuracy showed minimal similarity with texts which used other feature selection methods in categorisation despite the challenges of the languages and machine learning methods. Only one research (Alemu, 2010) showed better result in hierarchical classification since dimensionality of the feature space is reduced along with the hierarchy, and the classifier is focused on features that are relevant to the classification problem at hand.

Research Methodology

Data Source and Data Collection Methods

The data used in this study were collected from different sources such as news agencies, corporations and sites which are assumed to produce and store large amount of Amharic text documents. To collect electronic Amharic text data, Ethiopia

Television (EBC), Ethiopia Radio, Ethiopian News Agency (ENA), Walta TV, Amhara Radio, sport media agencies/enterprises have been consulted and about 52,300 data for the study have been collected. The data from TV and radio stations were collected by tracing manually their websites after they published it on their websites for further access once the news is being broadcasted to the public.

Implementation Tools

LibSVM$^{\text{multiclass}}$ is a tool used for experimentation. The study was performed using eight document classes, i.e. Economy, Business, Agriculture, Sport, Politics, Culture and Tourism, Weather and Climate, and Entertainment. After the data pre-processing was finalised and relevant features generated, the data was classified into training and testing data where these data were prepared using Python 3.7 according to the format that the LibSVM$^{\text{multiclass}}$ tool required.

Experimental Setup and Evaluation Procedures

The experiment followed five major activities: pre-processing, feature selection, input preparation, building the classifier, testing and evaluation. The first step, document preprocessing procedures, included the following tasks. These are:

- Data pre-processing (data cleaning, normalization, stop word removal, stemming and exception handling, and term weighting);
- Transform (prepare) the data to the format of an LibSVM package;
- Systematically try a few kernels and parameters and select the one which performs best; and
- Using the best parameter to train the whole training set and at the end test the classifier.

For stemming Amharic words, the researchers used successor variety stemming algorithm (method). This is because it has the advantage of avoiding the need of affix removal rules that are based on the morphological structure of a language. Once document pre-processing activities has been applied, the number of document features has been reduced. As far as the knowledge of the researchers, there is no general list of stop words for the Amharic language. But stop words in Amharic have the following properties:

- They are non-informative words if they are used alone.
- They occur frequently in documents.
- There are important for the structure of the language not important for the semantics purpose.
- Most of the time they can be adjectives, pronouns, articles.
- General words for the language are not domain specific.

The second step, feature selection was tried to apply feature selection algorithms over the 38,500 documents. From these, different methods generate different numbers of features. Of these features, input data was prepared as a third step. In fourth and fifth steps, the classifier was built using training data and tested for its performance using a test data respectively. Finally, evaluation and analysis was made using the five feature selection method to select the one that showed optimal performance of the classifier. To evaluate the classification performance, the researchers used accuracy, precision and recall as metrics. The accuracy metric is widely used in machine learning fields, which indicates the overall classification performance. The precision is the percentage of documents that are correctly classified as positive out of all the documents that are classified as positive, and the recall is the percentage of documents that are correctly classified as positive out of all the documents that are actually positive. The metrics of accuracy, precision and recall are defined as:

$$\text{Accuracy} = \frac{\begin{array}{c}\text{The number of correctly classified documents}\\ \text{determined by the classifier}\end{array}}{\begin{array}{c}\text{The number of expected text documents}\\ \text{stored in the database}\end{array}} = (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (3)$$

Where TP denotes the number of true positive, FP denotes the number of false positive, and FN denotes the number of false negative. These two metrics have an inverse relationship between each other. In other words, increasing the precision is at the cost of reducing the recall, and vice versa. Among those measures that attempt to combine precision and recall as one single measure, the F1 measure is one of the most popular, which is defined by

$$\text{F1} = \frac{2 * \text{precission} * \text{recall}}{\text{precision} + \text{Recall}} \tag{4}$$

The metrics of precision, recall and F1 measure are originally defined for binary class. For multi-class classification, we followed several other studies (Yiming, 2015; J.R., 1986) in which binary classifiers are built for each individual class and a global F1 measure is obtained by averaging the F1 measure of each class weighted by the class prior.

Types of Feature Selection Methods

Feature selection methods are used to remove trivial terms and reduce high dimension of feature set to optimise the categorisation efficiency and effectiveness. In the following section, we briefly describe the following feature selection methods: six methods are included in this study, each of which uses a term – goodness criterion threshold – to achieve a desired degree of term elimination from the full vocabulary of a document corpus. These criteria are: document frequency (DF), information gain (IG), mutual information (MI), X^2 statistic (CHI), tf*idf and term strength (TS).

Document Frequency:- Document frequency (DF) is the number of documents in which a term occurs. The researchers computed the document frequency for each unique term in the training corpus and removed from the feature space those terms whose document frequency was less than some predetermined threshold. The basic assumption is that rare terms are either non-informative for category prediction or not influential in global performance. In either case, removal of rare terms reduces the dimensionality of the feature space. Improvement in categorisation accuracy is also possible in rare terms which happen to be noise terms. DF thresholding is the simplest technique for vocabulary reduction. It easily scales to very large corpora, with a computational complexity approximately linear in the number of training documents. However, it is usually considered an ad hoc approach to improve efficiency, not a principled criterion for selecting predictive features. Also, DF is typically not used for aggressive term removal because of a widely perceived assumption in information retrieval. That is, low-DF terms are assumed to be relatively informative and therefore should not be removed aggressively.

Information Gain (IG): Information gain is frequently employed as a term – goodness criterion – in the field of machine learning (Liu, and Setiono, 1995; Yang, and Honavar, 1998). It measures the number of bits to find information obtained for category prediction by knowing the presence or absence of a term in a document. Information Gain (IG) measures the number of bits of information obtained for category prediction and by knowing presence or absence of a term in a document (Schutze et al., 1995). The idea behind IG is to select features that reveal the most related information about the classes. IG reaches its maximum value if a term is an ideal indicator for class association, i.e., if the term is present in a document if and only if the document belongs to the respective class. The IG method fails to identify discriminatory features, particularly when they are distributed over multiple classes (Song, Liu, and Yang, Song, 2005; Jing, Huang, and Shi, 2002).

Mutual Information (MI): Mutual information is a criterion commonly used in statistical language modelling of word associations and related applications. MI measures how much information presence or absence and term contribution to make the correct categorisation decision on a category.

Chi-Squared: Chi-Square ($-X^2$) is a statistical feature selection method (Asker et. al., 2009). $-X^2$ is used to measure the association between a term and category in text categorisation. It also used to test whether the occurrence of a specific term and the occurrence of a specific category are independent. Thus, we estimate the quantity for each term and we rank them by their score. If a term is close to more categories, then the score of that term is higher. High scores on -X^2 indicate that the null hypothesis of independence should be rejected and thus that the occurrence of the term and category are dependent. If they are dependent, then we select the feature for the text categorisation. Feature Selection via chi square (X^2) test is another very commonly used method (Koller and Sahami, 1996). Chi-squared attribute evaluation evaluates the worth of a feature by computing the value of the chi-squared statistic with respect to the class. Where O_{ij} is the observed frequency and E_{ij} is the expected (theoretical) frequency, asserted by the null hypothesis. The greater the value of X^2, the greater the evidence against the hypothesis H_0 is.

Term Strength (TS): This method estimates term importance based on how commonly a term is likely to appear in "closely related" documents (Lam et al., 1999). It uses a training set of documents to derive document pairs whose similarity (measured using the cosine value of the two document vectors) is above a threshold. Term strength is then computed, based on the estimated conditional probability that a term occurs in the second half of a pair of related documents given that it occurs in the first half.

Term Frequency-Inverse Document Frequency (tf*idf): This method is commonly used technique for term weighting in the field of text classification (Jing et al., 2002). It determines the relative frequency of terms in a specific document through an inverse proportion of the term over the entire document corpus (Sahlemariam, Libsie and Yacob, 2009). The tf*idf weight is composed by two conditions: the first condition computes the normalised term frequency tf, and the second condition is the inverse document frequency (idf). The term frequency (tf) measures the number of times a term occurs in a document, and it is used to calculate the describing ability of the term.

General Feature Selection Structure

From most of the feature selection algorithms, the following general architecture can be arrived with four basic steps:

Subset Generation: Subset generation is a search procedure; it generates subsets of features for evaluation. The total number of candidate subsets is 2N, where N is the number of features in the original data set, which makes exhaustive search through the feature space infeasible with even moderate N. Non-deterministic search like evolutionary search is often used to build the subsets (Mukras, Wiratunga, Lothian, Chakraborti and Harper (2007). It is also possible to use heuristic search methods. There are two main families of these methods: forward addition (Qu, Wang, and Zou, 2008) (starting with an empty subset, we add features after features by local search) or backward elimination (the opposite).

Subset Evaluation: Each subset generated by the generation procedure needs to be evaluated by a certain evaluation criterion and compared with the previous best subset with respect to this criterion. If it is found to be better, then it replaces the previous best subset. A simple method for evaluating a subset is to consider the performance of the classifier algorithm when it runs with that subset. The method is classified as a wrapper, because in this case, the classifier algorithm is wrapped in the loop. In contrast, filter methods do not rely on the classifier algorithm, but use other criteria based on correlation notions (Liu and Setiono, 1995).

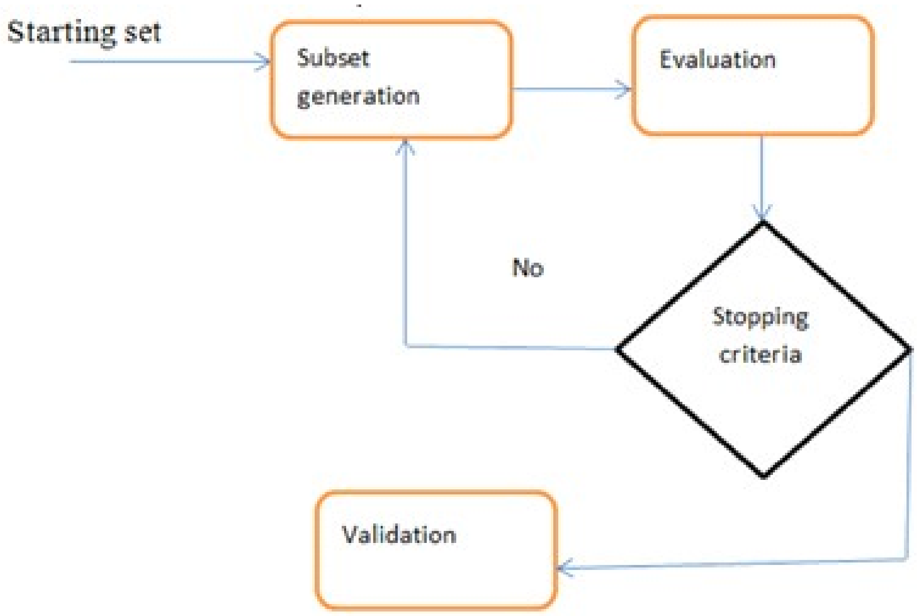

Fig. 1: General feature selection structure

Stopping Criteria: Without a suitable stopping criterion, the feature selection process may run exhaustively before it stops. A feature selection process may stop under one of the following reasonable criteria: (1) a predefined number of features are selected, (2) a predefined number of iterations are reached, (3) in case the addition (or deletion) of a feature fails to produce a better subset, (4) an optimal subset according to the evaluation criterion is obtained.

Validation: The selected best feature subset needs to be validated by carrying out different tests on both the selected subset and the original set and comparing the results using artificial data sets and/ or real-world data sets. A feature selection process may stop under one of the following reasonable criteria (Mcallum, and Nigam, 1998).

i. A pre-defined number of features are selected

ii. A pre-defined number of iterations are reached

iii. In case the addition or deletion of a feature fails to produce a better subset

iv. Obtained an optimal subset according to the evaluation criterion.

Discussion and Results

For this study, about 52,300 documents were used and processed according to text processing techniques. After pre-processing, about 38,500 documents were generated. Among these documents, feature selection techniques were applied. Table1 showed the result generated by respective feature selection techniques.

Table 1: Generated features as the result of different feature selection techniques

Methods	Cutting threshold (%)	Generated features
Tf*idf	>=0.5	50,000
DF	>=0.5	60,000
X2	>=0.5	37,000
IG	>=0.5	35,000
MI	>=0.5	75,000
TS	>=0.5	25,000

During the experimentation, SVM classifier is built based on the hierarchical approach (Alemu, 2010). The hierarchical classification approach considers the structural relationship among a given category. In such a hierarchical structure, document types become more specific as we go down in the hierarchy. The generated features are set to be used for training and testing data sets. The classifier is built using the document features selected by the respective feature selection technique as shown in Table 1. Although there are many Machine learning algorithms used to predict the features of documents in the test data set, such as Support Vector Machine (SVM), Decision Tree, Naïve Bayes and Artificial Neural Network used for hierarchical text classification, SVM was selected for this study. This is due to its capability of providing a number of benefits as compared to other algorithms (Alemu, 2010). Classifying documents for training and testing is taken based on 70/30 principle (Sebastiani, 2002). As the result, Table 2 below showed that the performance of the classifier based on document features selected by various feature selection techniques.

Table 2: Performance of the classifier based on different feature selection methods

Feature selection Techniques	Accuracy in %	Recall in %	Precision in %
IG	83.6	79.5	64.2
Tf*idf	70.5	72	57.5
X^2	87.33	82	67
MI	68	71	54.6
DF	64.5	73.6	51
TS	75	69	55

As shown in the Table 2 above, IG and X^2 showed highest performance for accuracy and both precision and recall. . In terms of recall, this research achieved a very good result in IG and X^2. But precision is somewhat lower compared to the recall value. This is because of the trade-off between precision and recall and small number of classes.

In the experimentation, we also tried to investigate the performance of the classifier for a collection with various number of features selected in these techniques. Figure 2 shows that both IG and X^2 increase the performance of the classifier as the number of documents and document features increases. This shows that both IG and X^2 are working well when we have more documents and more document features.

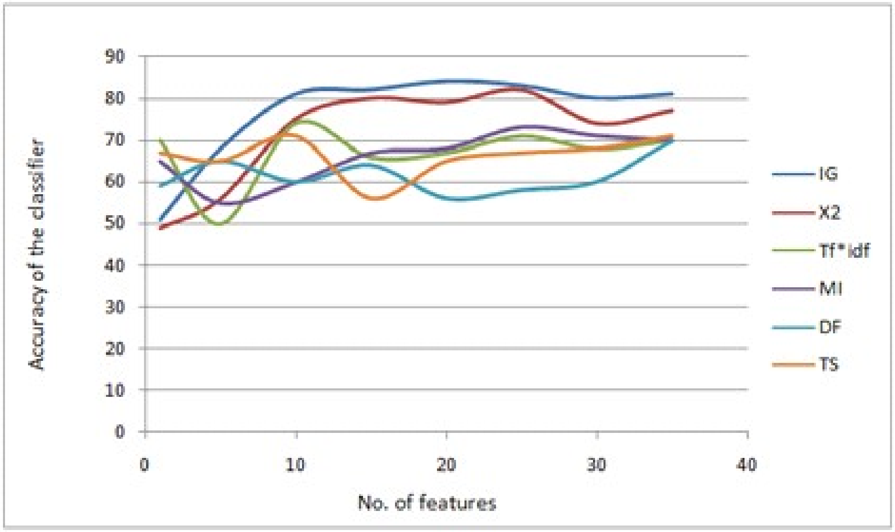

Fig 2(a): Performance of the classifiers based on the five feature selection methods

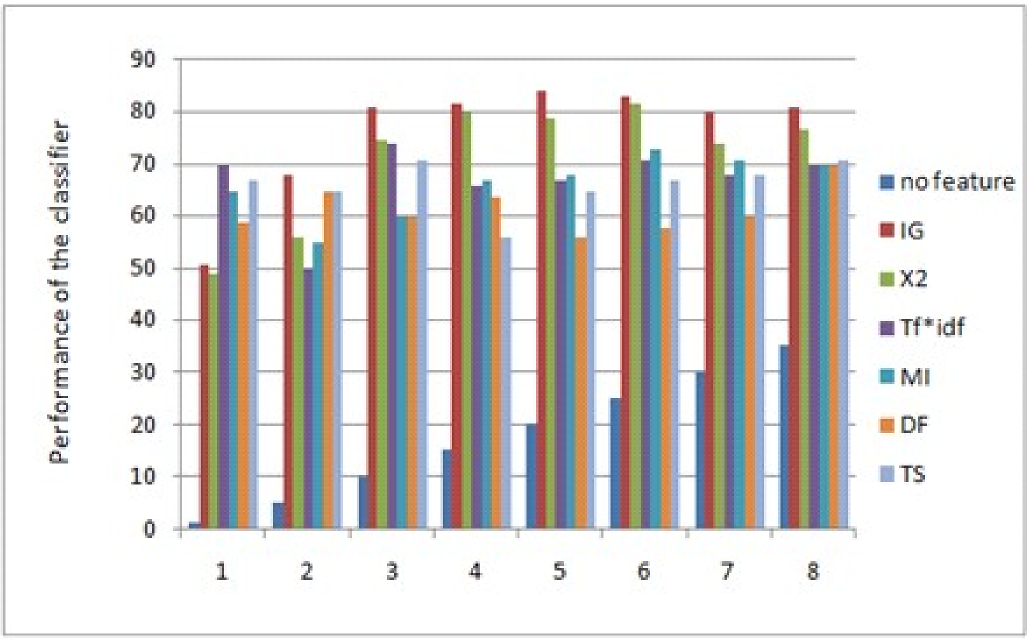

Fig 2(b): Performance of the classifiers based on the five feature selection methods

As shown in Figures 1 and 2 above, as the number of document and document features increases, the classification accuracy based IG and X^2 increases.

Conclusion and Future Work

In text categorisation, high dimensionality of feature space is about critical issues. These issues are resolved by using various feature selection approaches, which increases the efficiency of text categorisation. In this paper, we reported an experimental evaluation on the most widely used text feature selection methods using SVM classifiers, to categorise text documents. In the experiments, X^2 and IG method performed consistently well on all the other methods using the datasets with SVM classifier. The evaluation demonstrated that the X^2 method has tremendous influence on improving the categorisation accuracy. In the future, it is also intended to work on the computational complexity of various feature selection methods (FSM) using different classifiers.

References

Abate, M. and Assabie, Y. (2014). Development of Amharic Morphological Analyzer Using Memory-Based Learning. In *International Conference on Natural Language Processing* (pp. 1-13). Springer, Cham.

Abate, S. T and Menzel, W. (2007). Syllable-based Speech Recognition for Amharic. In Proceedings of the 2007 Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources (pp. 33-40). Association for Computational Linguistics.

Alemu, K. T. (2010). Hierarchical Amharic News Text Classification (Doctoral dissertation, Addis Ababa University).

Asker, L., Argaw, A. A., Gambäck, B., Asfeha, S. E., and Habte, L. N. (2009). Classifying Amharic webnews. Information Retrieval, 12(3), 416-435.

Haliu, A. and Assabie, Y. (2016). Item sets-based Amharic Document Categorisation Using an Extended a Priori Algorithm. Lecture Notes in Computer Science, Volume 9561, 317-326.

Jing, L. P., Huang, H. K., and Shi, H. B. (2002, November). Improved Feature Selection Approach TFIDF in Text Mining. In Proceedings. International Conference on Machine Learning and Cybernetics (Vol. 2, pp. 944-946). IEEE.

Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning With Many Relevant Features. In European Conference On Machine Learning (pp. 137-142). Springer, Berlin, Heidelberg.

Kelemework, W. (2013). Automatic Amharic Text News Classification: A Neural Networks Approach. Ethiopian Journal of Science and Technology, 6(2), 127-137.

Koller, D., and Sahami, M. (1996). Toward Optimal Feature Selection. Stanford InfoLab.

Lam, W., Ruiz, M., and Srinivasan, P. (1999). Automatic Text Categorization And Its Application to Text Retrieval. IEEE Transactions on Knowledge and Data engineering, 11(6), 865-879.

Lang, K. (1995). Newsweeder: Learning to Filter Netnews. In Machine Learning Proceedings 1995 (pp. 331-339). Morgan Kaufmann.

Lewis, D. D., and Ringuette, M. (1994). A Comparison of Two Learning Algorithms for Text Categorization. In Third Annual Symposium on Document Analysis and Information Retrieval Vol.33 pp.81-93).

Liu, H., and Setiono, R. (1995). Chi2: Feature Selection and Discretization of Numeric Attributes. In Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence (pp. 388-391). IEEE.

McCallum, A., and Nigam, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. In AAAI-98 workshop on learning for text categorization (Vol. 752, No. 1, pp. 41-48).

Moulinier, I., Ganascia, J. G., and Raškinis, G. (1996). Text Categorization: A Symbolic Approach. In Fifth Annual Symposium on Document Analysis and Information Retrieval, April 15-17, 1996, Las Vegas, Nevada: proceedings. Las Vegas: University of Nevada, Las Vegas, 1996.

Mukras, R., Wiratunga, N., Lothian, R., Chakraborti, S., and Harper, D. (2007). Information Gain Feature Selection for Ordinal Text Classification Using Probability Re-Distribution. In Proceedings of the Textlink workshop at IJCAI (Vol. 7, p. 16).

Qu, S., Wang, S., and Zou, Y. (2008). Improvement of Text Feature Selection Method based on tfidf. In 2008 International Seminar on Future Information Technology and Management Engineering (pp. 79-81). IEEE.

Quinlan, J. R. (1986). Induction of Decision Trees. Machine learning, 1(1), 81-106.

Rogati, M., and Yang, Y. (2002). High-performing Feature Selection for Text Classification. In Proceedings of the Eleventh International Conference on Information and Knowledge Management (pp. 659-661). ACM.

Sahlemariam, M., Libsie, M., and Yacob, D. (2009). Concept-Based Automatic Amharic Document Categorization. AMCIS 2009 Proceedings, 116.

Schütze, H., Hull, D. A., and Pedersen, J. O. (1995). A Comparison of Classifiers and Document Representations for the Routing Problem. In Annual ACM Conference on Research and Development in Information Retrieval-ACM SIGIR.

Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47.

Sintayehu, Z. (2001). Automatic Classification of Amharic News Items: The Case of Ethiopian News Agency. Master of Science Thesis, School of Information Studies for Africa, Addis Ababa University, Addis Ababa, Ethiopia.

Song, F., Liu, S., and Yang, J. (2005). A Comparative Study on Text Representation Schemes in Text Categorization. Pattern Analysis and Applications, 8(1-2), 199-209.

Tang, B., Kay, S., and He, H. (2016). Toward Optimal Feature Selection in Naive Bayes for Text Categorization. IEEE Transactions on Knowledge and Data Engineering, 28(9), 2508-2521.

Teklu, S. (2012). Automatic Categorization of Amharic News Text: A Machine Learning Approach. LAP Lambert Academic Publishing.

Wiener, E., Pedersen, J. O., and Weigend, A. S. (1995, April). A Neural Network Approach to Topic Spotting. In Proceedings of SDAIR-95, 4th Annual Symposium on Document Analysis and Information Retrieval (Vol. 317, p. 332).

Yang, J., and Honavar, V. (1998). Feature Subset Selection Using a Genetic Algorithm. In Feature Extraction, Construction and Selection (pp. 117-136). Springer, Boston, MA.

Yang, Y. (1999). An Evaluation of Statistical Approaches to Text Categorization. Information retrieval, 1(1-2), 69-90.

Yang, Y., and Pedersen, J. O. (1997). A Comparative Study on Feature Selection in Text Categorization. In International Conference On Machine Learning (Vol. 97, No. 412-420, p. 35).

Yang, Y., and Wilbur, J. (1996). Using Corpus Statistics to Remove Redundant Words in Text Categorization. Journal of the American Society for Information Science, 47(5), 357-369.

Yohannes, A. (2007). Automatic Amharic News Text Classification Using Support Vector Machine Approach. Master of Science Thesis, School of Information Studies for Africa, Addis Ababa University, Addis Ababa, Ethiopia.

Tamir Anteneh Alemu obtained his M.sc degree in Information Science from Addis Ababa University, Ethiopia and B.sc degree in Information Technology from Bahir Dar University, Ethiopia. He is currently working as a lecturer at the Faculty of Computing, Bahir Dar University, Institute of Technology.

Alemu Kumilachew Tegegne obtained his M.sc degree in Information Science from Addis Ababa University and B.sc degree in Information Technology from Jimma University, Ethiopia. He is currently working as a lecturer at the Faculty of Computing, Bahir Dar University, Institute of Technology (BiT).