International Journal of Computational Linguistics and Applications
vol. 7, no. 1, 2016, pp. 101–119
Received 19/11/2015, accepted 06/03/2016, final 21/06/2016
ISSN 0976-0962, http://ijcla.bahripublications.com

LDA-based Topic Modelling in Text Sentiment Classification: An Empirical Analysis

AYTUĞ ONAN[1], SERDAR KORUKOĞLU[2], AND HASAN BULUT[2]

[1] *Celal Bayar University, Turkey*
[2] *Ege University, Turkey*

ABSTRACT

Sentiment analysis is the process of identifying the subjective information in the source materials towards an entity. It is a subfield of text and web mining. Web is a rich and progressively expanding source of information. Sentiment analysis can be modelled as a text classification problem. Text classification suffers from the high dimensional feature space and feature sparsity problems. The use of conventional representation schemes to represent text documents can be extremely costly especially for the large text collections. In this regard, data reduction techniques are viable tools in representing document collections. Latent Dirichlet allocation (LDA) is a popular generative probabilistic model to represent collections of discrete data. In this regard, this paper examines the performance of LDA in text sentiment classification. In the empirical analysis, five classification algorithms (Naïve Bayes, support vector machines, logistic regression, radial basis function network and K-nearest neighbor algorithms) and five ensemble methods (Bagging, AdaBoost, Random Subspace, voting and stacking) are evaluated on four sentiment datasets.

KEYWORDS: topic modelling, text classification, ensemble learning.

1 Introduction

Topic modelling is an important research direction in machine learning, natural language processing and information retrieval. Topic modelling is the process of identifying the underlying semantic structure of a document with the use of a hierarchical Bayesian analysis on the collection of documents [1]. The Web is an important source of information. The amount of information that is digitized and stored has been progressively expanding [2]. However, the progressively expanding digital content is unstructured. Topic modelling algorithms enable to automatically organize, understand, search and summarize large and unstructured collections of documents [2]. Topic modelling methods have been successfully applied in several applications, such as automatic document indexing, document classification, topic discovery, entities relationship discovery and temporal topic trends and community discovery [3].

In topic modelling, documents are represented with a mixture of topics, topics are represented with a probability distribution over words and the documents are represented by a probability distribution over topics [4]. Probabilistic topic modelling methods can be broadly assigned into five main categories as basic, inter-document correlated, intra-document correlated, temporal and supervised probabilistic directed topic models [3]. Though there are many methods of topic modelling in the literature, the seminal works on the area are latent semantic analysis, probabilistic latent semantic analysis and latent Dirichlet allocation [5]. Latent semantic analysis is a method of natural language processing in which statistical and mathematical computations are done to extract and represent the contextual usage meaning of the words in a large collection of text corpus [6]. Latent semantic analysis (LSA) utilizes singular value decomposition to reduce the dimensionality of TF-IDF scheme. LSA can capture synonyms of words, but it is difficult to determine the number of topics in LSA [7]. Probabilistic latent semantic analysis (PLSA) is a probabilistic topic model in which each word in the document are modelled as a sample from a mixture models [7]. In PLSA, there is no a probabilistic model at the document-level [7, 8]. Latent Dirichlet allocation (LDA) is a generative probabilistic method to model collections of discrete data, such as a text corpus [8]. LDA can provide a full generative model and can handle long-length documents [7].

Sentiment analysis (also known as opinion mining) is a subfield of natural language processing, text mining and web mining which aims to extract subjective information in the source materials towards an entity. Since it can be extremely useful to identify opinions/sentiments towards a particular event or entity, sentiment analysis is an important research direction. The process of identifying the sentimental orientation of a text document can be modelled as a text classification problem. Text classification is a field with high dimensional feature space problem and the determination of an appropriate representation of text documents can be extremely important in the performance of text classifiers [9, 10]. TF-IDF is a one of the most widely utilized text representation methods, yet it suffers from high dimensionality problem [11]. Besides, TF-IDF scheme indicates little information about the inter-document and intra-document statistical structure [7]. Hence, other reduction methods, such as latent Dirichlet allocation can be utilized.

As mentioned in advance, the high dimensional feature space and feature sparsity are two major problems encountered in conventional representation schemes of text mining applications. Topic modelling is the process of identifying latent topics in document. The utilization of latent topics extracted from text documents as the features instead of a large number of words can overcome the curse of dimensionality problem and improve the predictive performance of text classifiers.

In this paper, we have examined the performance of latent Dirichlet allocation based feature representation in text sentiment classification. In the empirical analysis, Naïve Bayes, support vector machines, logistic regression, radial basis function network and K-nearest neighbor algorithms are utilized as the weak learners and Bagging, AdaBoost, Random Subspace, Voting and Stacking methods are utilized as the ensemble learning methods. In summary, the experimental analysis seeks answer to the following research questions:

(1) Can ensemble learning methods be utilized to enhance the predictive performance of classifiers when latent Dirichlet allocation method is used to represent text collections?

(2) Is there an optimal number of latent topics in LDA-based representation of text documents in text sentiment classification?

(3) Which configuration of classification algorithms, ensemble learning methods, number of latent topics yield promising results for text sentiment classification?

The rest of this paper is organized as follows. Section 2 briefly reviews the literature on latent Dirichlet allocation. Section 3 presents latent Dirichlet allocation method. Section 4 briefly describes the classification algorithms, Section 5 briefly describes the ensemble methods. Section 6 presents the experimental results and discussion. Finally, Section 7 presents the concluding remarks.

2 Literature Review

Latent Dirichlet allocation can be used as a feature representation method in conjunction with machine learning algorithms to classify text documents. The related work on LDA-based feature representation is briefly presented here. Tian et al. [12] utilized latent Dirichlet allocation method to index and analyze the source code documents as a mixture of probabilistic topics. Based on this representation, software systems in open-source repositories are automatically categorized. Taşçı and Güngör [13] examined the performance of latent Dirichlet allocation based representation in text categorization. In order to deal with high dimensionality problem of text mining problems, feature selection methods, such as information gain, chi-square statistics and document frequency threshold are taken into consideration. The performance of LDA is compared to these feature selection methods. In the comparative evaluation, TF-IDF scheme is utilized to weight the terms and support vector machines are utilized as the base learners. Ramage et al. [14] presented a labelled latent Dirichlet allocation model which incorporates labels and topic priors to learn word-tag correspondences. The experimental results indicate that labelled LDA method can yield better performance than support vector machines classifier owing to its explicit modelling of the importance of each label in the document.

Hong and Davison [15] utilized two topic modelling methods (the author-topic model and latent Dirichlet allocation) to predict popular Twitter messages and to classify Twitter users and corresponding messages into topical categories.

Liu et al. [16] empirically evaluated the performance of vector space model, latent semantic indexing and latent Dirichlet allocation methods on text classification. In LDA-based text classification, latent Dirichlet allocation method was utilized to represent the text documents. The experimental results indicated that the use of LDA in conjunction to support vector machines yields better performance than the other compared configurations.

For each document *w* in a corpus *D*:

1. Choose $N \sim$ Poisson (ξ).
2. Choose $\Theta \sim$ Dir (α).
3. For each of the *N* words w_n:
    a. Choose a topic $z_n \sim$ Multinomial (Θ).

Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

Fig. 1. The generative process of LDA [8]

3 Latent Dirichlet Allocation

The latent Dirichlet allocation model (LDA) is a generative probabilistic topic model where each document is represented as a random mixture of latent topics and each topic is represented as a distribution over fixed set of words [8]. LDA aims to identify the underlying latent topic structure based on the observed data. In LDA, the words of each document are the observed data. For each document in the corpus, the words are generated in a two-staged procedure. First, a distribution over topics is randomly chosen. Based on this distribution, a topic from the distribution over topics is randomly chosen for each word of the document [2]. In LDA, a word is a discrete data from a vocabulary indexed by {1, …, *V*}, a sequence of *N* words $w=(w_1, w_2, ..., w_n)$ and a corpus is a collection of *M* documents denoted by $D=\{\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_M\}$. The generative process of LDA can be summarized in Figure 1.

The process of LDA can be modelled by a three-level Bayesian graphical model, where random variables are represented by nodes and possible dependencies between the variables are represented by edges, as depicted in Figure 2. In this representation, α refers to Dirichlet parameter, Θ refers to document-level topic variables, *z* refers to per-word topic assignment, *w* refers to the observed word and β refers to the topics. As it can be observed from the three-layered representation, α and β parameters are sampled once while generating the corpus, document-level topic variables are sampled for each document and word-level variables are sampled for each word of the document [8].

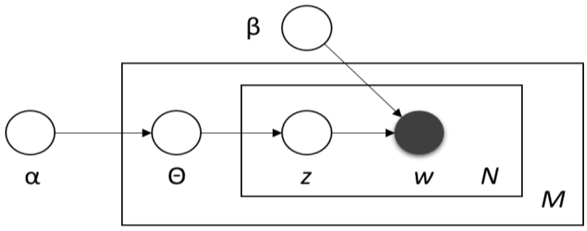

Fig. 2. The graphical representation of LDA [8]

The generative process of LDA indicates a joint distribution over random variables. The probability density function of a k-dimensional Dirichlet random variable is computed by Equation 1, the joint distribution of a topic mixture is determined by Equation 2 and the probability of a corpus is computed as given by Equation 3 [8]:

$$p(\Theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \Theta_1^{\alpha_1 - 1} \ldots \Theta_k^{\alpha_k - 1} \quad (1)$$

$$p(\Theta, z, w|\alpha, \beta) = p(\Theta|\alpha) \prod_{n=1}^{N} p(z_n|\Theta)\, p(w_n|z_n, \beta) \quad (2)$$

$$p(D|\alpha, \beta) = \prod_{d=1}^{M} \int p(\Theta_d|\alpha) \big(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\Theta_d) p(w_{dn}|z_{dn}, \beta) \big) d\Theta_d \quad (3)$$

For a particular document, the computation of the posterior distribution of the hidden variables is an essential inferential task in LDA. The exact inference of the posterior distribution of the hidden variables can be an intractable problem. Hence, approximation algorithms, such as Laplace approximation, variational approximation, Gibbs sampling, Markov chain Monte Carlo have been widely utilized in conjunction with LDA [8, 17].

4 Classification Algorithms

This section briefly presents the machine learning classifiers utilized in the experimental analysis as the weak learners.

Naïve Bayes algorithm (NB) is a statistical classifier which is based on Bayes' theorem. Naïve Bayes algorithm is a simple, computationally efficient classification algorithm with high predictive performance owing to its independence assumptions. It can yield better predictive performance than other learning algorithms with more sophisticated structures and less computational efficiencies [18]. Naïve Bayes algorithm is widely employed in several areas, including text classification with comparable results to decision trees and artificial neural networks [19].

Logistic regression (LR) is a linear classifier. Logistic regression uses a linear function of a set of predictor variables to model the probability of some event's occurring [20]. Linear regression can yield good results. However, the membership values generated by linear regression cannot be always in [0-1] range, which is not an appropriate range for probabilities. In logistic regression, a linear model is constructed on the transformed target variable whilst eliminating the mentioned problems.

K-nearest neighbour algorithm (KNN) is an instance based classifier. In the algorithm, the classification model is constructed based on the similarity of instances to *k* closest training instances [21].

Support vector machines (SVM) are classification algorithms that can be used to classify both linear and non-linear data [22]. In SVM, the original data set is transformed into a higher dimension by a non-linear matching. SVM intends to identify an optimal decision boundary that can be used to classify different classes.

Radial basis function networks (RBF) has a feedforward artificial neural network architecture with radial basis functions as activation functions. It contains a radial basis layer and a linear layer. Radial basis function networks can be utilized in classification owing to their simple yet efficient characteristics [23].

5 Ensemble Learning Methods

This section briefly describes the ensemble methods utilized in the experimental evaluations.

Bagging (Bootstrap aggregating) is a popular ensemble learning method which aims to enhance the predictive performance by combining classifiers trained on different training sets obtained from the original training set by sampling with replacement [24]. In Bagging, the sizes of each sample in the new training sets are kept identical to the size of the original training set. The use of sampling scheme provides diversity that is necessary for efficiency of ensemble classification. To combine the outputs of weak learners, bagging algorithm generally utilizes majority voting or weighted majority voting scheme.

Boosting algorithm is an ensemble learning method which aims to enhance the predictive performance of weak learning algorithms by training the algorithms recursively on different sampling distributions. In this scheme, the training sets of each classifier is modified so that classifiers can focus on incorrectly classified instances. AdaBoost algorithm [25] is a widely utilized boosting algorithm owing to its speed, robustness and simplicity. AdaBoost algorithm aims to focus on difficult data points to enhance the predictive performance of weak learning algorithms. In the algorithm, a weight value is assigned to each instance of the training set. During the iterations of the algorithm, the weight values of misclassified instances are increased and the weight values of correctly classified instances are decreased. In this way, difficult data points are dedicated more iterations [26].

Random Subspace [27] is an ensemble learning method where multiple classifiers are trained on randomly selected feature subspaces. In Random Subspace algorithm, classifiers are trained on different samples on feature space. The method intends to eliminate the overfitting problem while providing high predictive performance.

Stacking (also known as Stacked generalization) [28] is an ensemble learning method which contains a two-levelled structure with multiple weak learners. In Stacking, a meta-learning algorithm is utilized to combine the outcomes of individual weak learners (the base-level classification algorithms).

Voting is the simplest form of combining the outputs of base classification algorithms. There are several different combination rules to combine the individual classification algorithms. The voting schemes include minimum probability, maximum probability, majority voting, product of probability and average of probabilities.

6 Experimental Results

6.1 Datasets

In order to examine the performance of LDA-based feature representation in text sentiment classification, we have used four public sentiment analysis datasets from several domains. In Table 1, the basic descriptive information regarding the text collections utilized in the experimental analysis is presented. The number of features listed in Table 1 corresponds to the number of terms extracted when vector space model is utilized to represent text documents [29]. In this study, text collections are modelled with the use of latent Dirichlet allocation (LDA) method to eliminate the high dimensionality problem. Text collections are represented by latent topics. In order to examine the performance of different number of features, features ranging from 50 to 200 are taken into consideration.

Table 1. Descriptive information for the datasets [29]

Dataset	Number of documents	Number of features	Number of classes
Multi-Domain Sentiment	8000	13360	2
Review-Polarity	2000	15698	2
Irish-Sentiment	1660	8659	3
Reviews	4069	22927	5

6.2 Evaluation Measures

In order to evaluate the performance of classification algorithms, classification accuracy (ACC) and F-measure metrics are utilized. Classification accuracy (ACC) is the proportion of true positives and true negatives obtained by the classification algorithm over the total number of instances as given by Equation 4:

$$ACC = \frac{TN + TP}{TP + FP + FN + TN} \tag{4}$$

where *TN* denotes number of true negatives, *TP* denotes number of true positives, *FP* denotes number of false positives and *FN* denotes number of false negatives.

Precision (PRE) is the proportion of the true positives against the true positives and false positives as given by Equation 5:

$$PRE = \frac{TP}{TP + FP} \tag{5}$$

Recall (REC) is the proportion of the true positives against the true positives and false negatives as given by Equation 6:

$$REC = \frac{TP}{TP + FN} \tag{6}$$

F-measure takes values between 0 and 1. It is the harmonic mean of precision and recall as determined by Equation 7:

$$F - \text{measure} = \frac{2 \times PRE \times REC}{PRE + REC} \tag{7}$$

6.3 Experimental Procedure

In the experimental analysis, 10-fold cross validation is used. In this scheme, the original data set is randomly divided into 10 equal-sized subsamples. For each time, a single subsample is used for validation and the other nine subsamples are utilized for training. The process is repeated ten times and average results are reported. In the experimental analysis, classification algorithms and ensemble learning methods are performed by WEKA 3.7.11, which is an open source Java software for machine learning research. The default parameters of the toolkit are employed. In the experimental analysis, Naïve Bayes, support vector machines, logistic regression, radial basis function network and K-nearest neighbor algorithms are utilized as the weak learners and Bagging, AdaBoost, Random Subspace, Voting and Stacking methods are utilized as the ensemble learning methods. Voting and Stacking ensembles are generated by the utilization of five weak learning algorithms.

6.4 Results and Discussion

In Table 2 and Table 3, classification accuracies and F-measure results obtained by the classification algorithms and ensemble learning methods on LDA-based sentiment analysis datasets are presented, respectively. In the tables, the highest results for each dataset are indicated by using boldface and underline, the second highest results are indicated by using only boldface and the third highest results are indicated by using boldface and italics. In Table 2 and Table 3, the average results for different number of latent topics in LDA-based representation of text documents are presented.

Table 2. Classification accuracies of sentiment analysis datasets

	Irish Sentiment	**Reviews**	**Multi-Domain Sentiment**	**Review Polarity**
NB	56.52	86.48	67.56	65.41
SVM	**64.85**	**92.97**	**73.40**	**77.21**
LR	60.44	89.90	72.10	76.51
KNN	55.95	86.93	60.53	65.25
RBF	58.72	89.54	64.06	67.70
AdaBoost+NB	56.52	86.54	67.99	65.41
AdaBoost+SVM	63.66	92.67	72.88	75.85
AdaBoost+LR	62.84	92.15	72.09	75.82
AdaBoost+KNN	55.95	86.93	60.53	65.25
AdaBoost+RBF	61.76	91.34	67.54	69.39
Bagging+NB	57.86	87.96	67.70	65.70
Bagging+SVM	***64.43***	92.88	***73.26***	***76.84***
Bagging+LR	59.75	89.78	72.04	76.20
Bagging+KNN	56.85	87.32	61.69	65.94
Bagging+RBF	59.80	90.62	67.11	69.41
Random Subspace+NB	54.85	84.31	67.39	65.05
Random Subspace+SVM	61.82	90.31	72.21	75.15
Random Subspace+LR	52.16	83.81	70.95	74.73
Random Subspace+KNN	58.36	89.00	64.38	68.37
Random Subspace+RBF	60.35	89.70	66.97	69.63
Voting (Average of Probabilities)	63.96	***92.89***	71.98	74.85
Voting (Product of Probabilities)	54.67	88.86	69.66	74.54
Voting (Majority Voting)	63.69	92.69	71.82	74.69
Voting (Minumum Probability)	54.67	88.86	69.66	74.54
Voting (Maximum Probability)	60.27	89.93	72.04	75.34
Stacking	**64.60**	**93.03**	**73.30**	**77.07**

Table 3. F-measure values of sentiment analysis datasets

	Irish Sentiment	**Reviews**	**Multi-Domain Sentiment**	**Review Polarity**
NB	0.56	0.49	0.67	0.61
SVM	**0.67**	**0.78**	**0.74**	**0.78**
LR	0.66	0.49	***0.72***	**0.77**

KNN	0.58	0.70	0.59	0.63
RBF	0.60	0.66	0.63	0.66
AdaBoost+NB	0.56	0.49	0.67	0.61
AdaBoost+SVM	***0.66***	**0.79**	**0.73**	***0.76***
AdaBoost+LR	***0.66***	0.75	***0.72***	***0.76***
AdaBoost+KNN	0.58	0.70	0.59	0.63
AdaBoost+RBF	0.63	0.74	0.67	0.69
Bagging+NB	0.58	0.54	0.68	0.61
Bagging+SVM	**0.67**	**0.78**	**0.74**	**0.77**
Bagging+LR	***0.66***	0.48	***0.72***	**0.77**
Bagging+KNN	0.59	0.71	0.60	0.64
Bagging+RBF	0.60	0.74	0.67	0.69
Random Subspace+NB	0.53	0.43	0.67	0.59
Random Subspace+SVM	***0.66***	0.55	**0.73**	***0.76***
Random Subspace+LR	0.62	0.32	***0.72***	***0.76***
Random Subspace+KNN	0.61	0.68	0.65	0.68
Random Subspace+RBF	0.63	0.47	0.67	0.69
Voting (Average of Probabilities)	**0.72**	0.71	**0.74**	**0.78**
Voting (Product of Probabilities)	**0.72**	0.71	**0.74**	**0.78**
Voting (Majority Voting)	***0.66***	***0.76***	***0.72***	0.74
Voting (Minumum Probability)	**0.72**	0.71	**0.74**	**0.78**
Voting (Maximum Probability)	***0.66***	0.68	**0.73**	**0.77**
Stacking	**0.67**	**0.79**	**0.73**	**0.77**

The first concern of the study is the performance of ensemble learning methods and classification algorithms in LDA-based representation of text documents in text sentiment classification. As it can be observed from the results summarized in Table 2, support vector machines yield the highest predictive performance for Irish-sentiment, multi-domain sentiment and review-polarity dataset. For Reviews dataset, Stacking method yields the highest predictive performance. Bagging ensemble of support vector machines also obtain promising results. Though ensemble learning is a promising research direction to enhance the predictive performance of classification algorithms, the performance improvement obtained by the ensemble learning for LDA-based sentiment classification is not significant for the datasets utilized in the experimental analysis. In general, support vector machines and Stacking ensemble yields the highest results. Regarding the performance of classifiers in terms of F-measure values listed in Table 3, ensemble learning methods yield generally better (higher) F-measure values compared to the weak learning algorithms. Among all the configurations of experimental analysis, the highest F-measure value (0.79) is obtained by AdaBoost ensemble of support vector machines and Stacking ensemble. Similarly, the highest classification accuracy (93.03) is obtained by Stacking ensemble.

In Figure 3 and Figure 4, we have summarized the main experimental findings obtained from the empirical analysis. Figure 3 and Figure 4 presents the main effects plots for classification accuracy and F-measure, respectively. Datasets, number of latent topics and classifiers are the three main effects of the empirical analysis. As it can be observed from Figure 3, the highest predictive performance is obtained by Reviews dataset, whereas the lowest predictive performance is obtained by Irish-sentiment dataset. As mentioned in advance, the experimental analysis aims to identify whether is there an optimal number of latent topics in LDA-based representation of text documents in text classification. As it can be observed from Figure 3, the change of predictive performance in terms of different number of latent topics exhibits a relatively flat line. Hence, the performance variations among different number of topics are not significant in terms of classification accuracies.

In Figure 4, the main effects plot for F-measure is given. The results obtained for F-measure are not exhibit the same patterns as the results obtained for classification accuracies. The highest (the best) F-measure values are obtained by Review-polarity dataset. However, the F-measure values obtained for Reviews dataset are also relatively low. In terms of number of latent topics, the highest F-measure values are obtained for 150 topics. Though the classification accuracies obtained by voting and stacking ensembles are not very high, F-measure values obtained by voting and stacking ensembles are high. To further evaluate the experimental results, we performed general linear model analysis to perform factorial ANOVA test in Minitab statistical program. The results for ANOVA test of overall results obtained by algorithms and methods are summarized in Table 4, where DF, SS, MS, F and P denote degrees of freedom, adjusted sum of squares, adjusted mean square, F-statistics and probability value, respectively. There are statistically meaningful differences between the results of compared datasets, the results of compared number of latent topics and the results of compared classifiers at 95% confidence level. Besides, the p-values ($p<0.001$) indicate that the three main factors (datasets, number of latent topics and classifiers) have statistically significant effect on the experimental results.

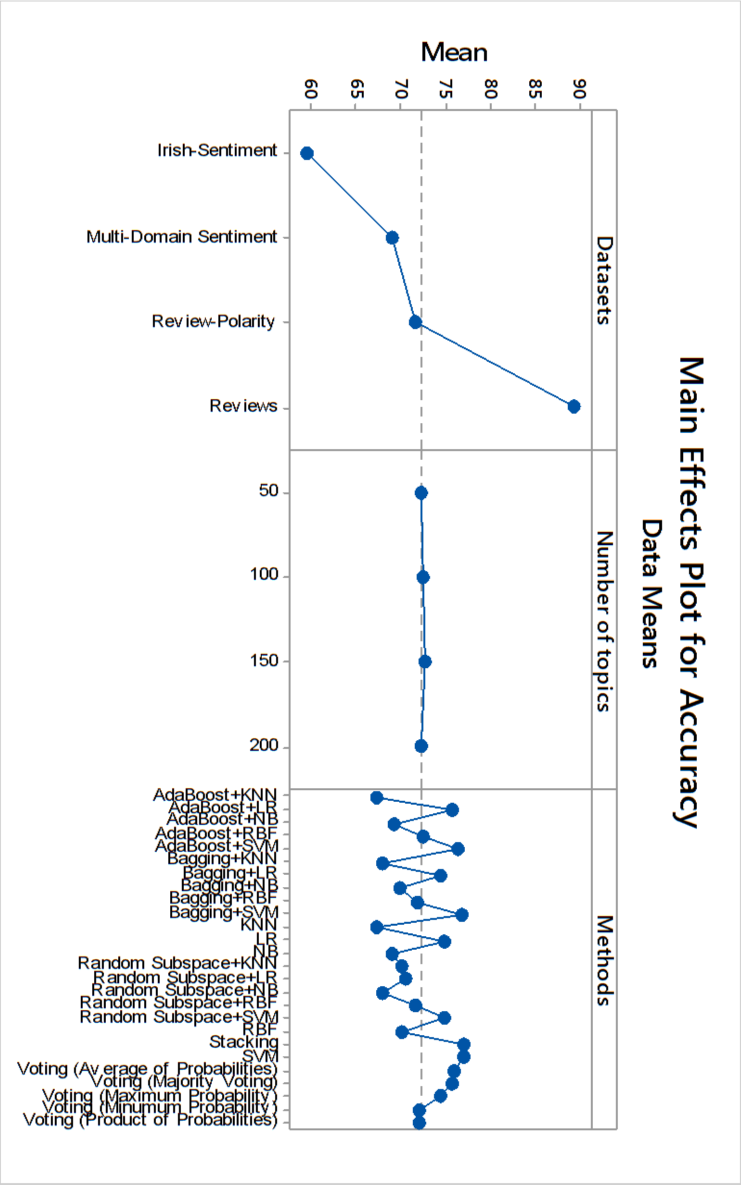

Fig. 3. Main effects plot for accuracy

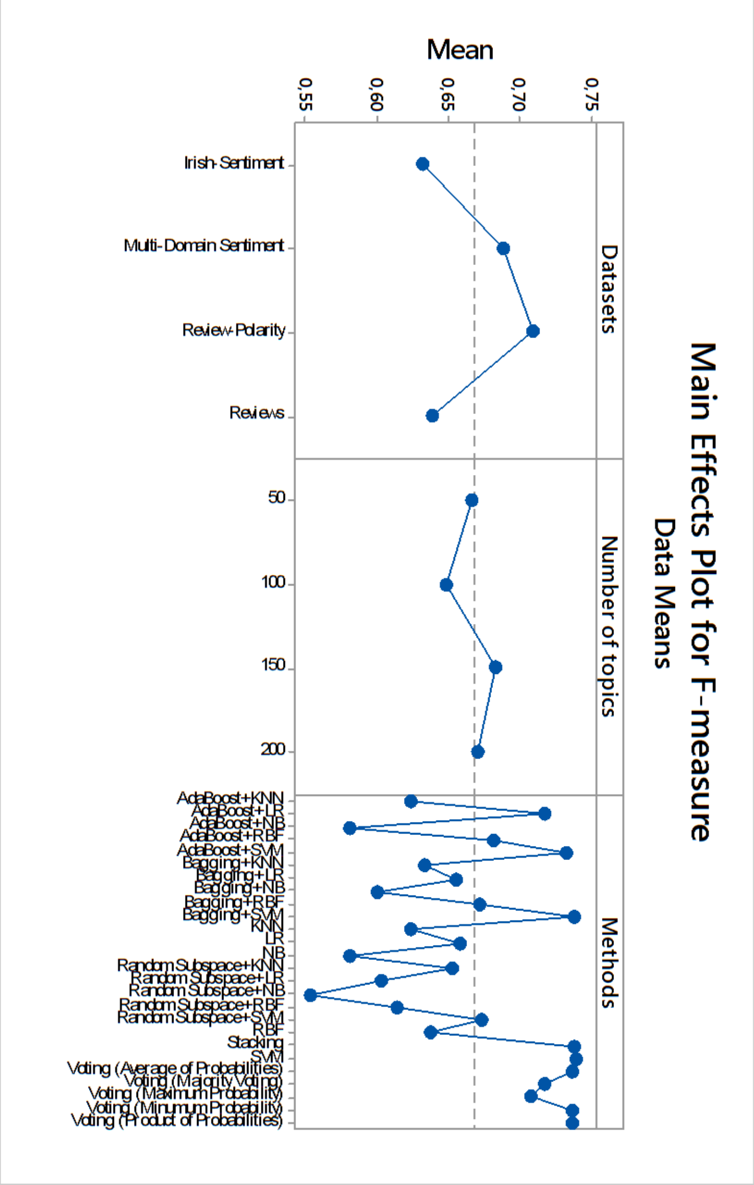

Fig. 4. Main effects plot for F-measure

Table 4. ANOVA test results

Source (Accuracy Values)	**DF**	**SS**	**MS**	**F**	**P**
Datasets	3	49294.9	16431.6	21064.72	0.000
Number of topics	3	13.9	4.6	5.94	0.001
Classifiers	25	4367.2	174.7	223.94	0.000
Datasets*Number of topics	9	64.2	7.1	9.15	0.000
Datasets*Classifiers	75	1485.1	19.8	25.38	0.000
Number of topics*Classifiers	75	131.1	1.7	2.24	0.000
Error	225	175.5	0.8		
Total	415				
Source (F-measure Values)	**DF**	**SS**	**MS**	**F**	**P**
Datasets	3	0.44508	0.148361	132.18	0.000
Number of topics	3	0.06337	0.021124	18.82	0.000
Classifiers	25	1.39521	0.055808	49.72	0.000
Datasets*Number of topics	9	0.16076	0.017863	15.91	0.000
Datasets*Classifiers	75	1.35645	0.018068	16.11	0.000
Number of topics*Classifiers	75	0.09636	0.001285	1.14	0.225
Error	225	0.25255	0.001122		
Total	415				

Among the compared p-values listed in Table 4, number of topics-classifiers interaction has a p value of 0.225. This indicates that regarding the performance of classification algorithms in terms of F-measure values, there is no statistically meaningful differences based on the different number of latent topics.

7 Conclusion

Ensemble learning is a promising research direction which aims to integrate the predictions of multiple learning algorithms to construct a robust classification model with better predictive performance. Sentiment analysis is the process of extracting and identifying subjective information in source materials which may serve potentially useful information to decision support systems and decision makers. In this paper, we have examined the predictive performance of classification algorithms (Naïve Bayes, support vector machines, logistic regression, radial basis function network and K-nearest neighbor algorithms) and ensemble learning methods (Bagging, AdaBoost, Random Subspace, voting and stacking) for text sentiment classification when LDA-based representation is utilized. LDA can be used as a viable method to represent text collections in a compact yet efficient way. Ensemble learning can be used to enhance the predictive performance of classification algorithms. Though the use of ensemble learning in conjunction to classification algorithms do not yield significant performance improvements for the datasets utilized in the experimental analysis, further research may yield more promising results by improving LDA-based representation or ensemble learning methods.

References

1. Blei, D.M., Lafferty, J.D.: Topic models. In: Srivastava, A., Sahami, M. (eds.) Text Mining: Classification, Clustering, and Applications. pp. 71-93. Chapman & Hall/CRC, London (2009)
2. Blei, D.M.: Probabilistic topic models. Communications of the ACM 55(4), 77-84 (2012)
3. Daud, A., Li, J., Zhou, L., Muhammad, F.: Knowledge discovery through directed probabilistic topic models: a survey. Frontiers of Computer Science in China 4(2), 280-301 (2010)
4. Steyvers, M., Griffiths, T.: Probabilistic topic models. In: Landauer, T., Mcnamara, D., Dennis, S., Kintsch, W. (eds.) Latent Semantic Analysis: a road to meaning. pp. 2-15. Laurence Erlbaum, New Jersey (2007)
5. Alghamdi, R., Alfalqi, K.: A survey of topic modeling in text mining. International Journal of Advanced Computer Science and Applications 6(1), 147-153 (2015)
6. Landauer, T.K., Laham, D., Rehder, B., Schreiner, M.E.: How well can passage meaning be derived without using word order? A comparison of latent semantic analysis and humans. In: Proceedings of the 19th Annual Meeting of the Cognitive Science Society, pp. 412-417. Erlbaum, New Jersey (1997)
7. Lee, S., Baker, J., Song, J., Wetherbe, J.C.: An empirical comparison of four text mining methods. In: Proceedings of the 43rd Hawaii International Conference on System Sciences, pp. 1-10. IEEE, New York (2010)
8. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research 3, 993-1022 (2003)
9. Joachims, T.: Learning to classify text using support vector machines. Springer, New York (2002)
10. Aggarwal, C.C., Zhai, C.X.: A survey of text classification algorithms. In: Aggarwal, C.C., Zhai, C.X. (eds.) Data Mining and Knowledge Discovery Handbook. pp. 77-128. Springer, Berlin (2012)
11. Zhang, W., Yoshida, T., Tang, X.: A comparative study of TF-IDF, LSI and multi-words for text classification. Expert Systems with Applications 38, 2758-2765 (2011)

12. Tian, K., Revelle, M., Poshyvanyk, D.: Using Latent Dirichlet Allocation for automatic categorization of software. In: Proceedings of the 6th International Working Conference on Mining Software Repositories. pp. 163-166. IEEE, New York (2009)
13. Tasci, S., Gungor, T.: LDA-based keyword selection in text categorization. In: Proceedings of the 24th International Symposium on Computer and Information Sciences. pp. 230-235. IEEE, New York (2009)
14. Ramage, D., Hall, D., Nallapati, R., Manning, C.D.: Labeled LDA: a supervised topic model for credit attribution in multi-label corpora. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. pp. 248-256. ACL, Stroudsburg (2009)
15. Hong, L., Davison, B.D.: Empirical study of topic modeling in Twitter. In: Proceedings of the First Workshop on Social Media Analytics. pp. 80-88. ACM, New York (2010)
16. Zelong, L., Maozhen, L., Yang, L., Ponraj, M.: Performance evaluation of latent dirichlet allocation in text mining. In: Proceedings of Eight International Conference on Fuzzy Systems and Knowledge Discovery. pp. 2695-2698. IEEE, New York (2011)
17. Jordan, M.: Learning in Graphical Models. MIT Press, Cambridge (1999)
18. Shmueli, G., Patel, N.R., Bruce, P.C.: Data mining for business intelligence: concepts, techniques and applications in Microsoft Office Excel with XLMiner. John Wiley & Sons, New York (2010)
19. Han, J., Kamber, M.: Data mining concepts and techniques. Morgan Kaufmann Publishers, San Francisco (2006)
20. Kantardzic, M.: Data mining: concepts, models, methods, and algorithms. Wiley-IEEE Press, New York (2011)
21. Aha, D.W., Kibler, D., Albert, M.K.: Instance-based learning algorithm. Machine Learning 6, 37-66 (1991)
22. Vapnik, V.: The nature of statistical learning theory. Springer, New York (1995)
23. Bors, A.G.: Introduction of the radial basis function networks. Online Symposium for Electronic Engineers 1, 1-7 (2001)
24. Breiman, L.: Bagging predictors. Machine Learning 4, 123-140 (1996)
25. Rokach, L.: Ensemble-based classifiers. Artificial Intelligence Review 33, 1-39 (2010)
26. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: Proceedings of the Thirteenth International Conference on Machine Learning. pp. 1-8. IEEE, New York (1996)
27. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(8), 832-844 (1998)
28. Wolpert, D.H.: Stacked generalization. Neural Networks 5(2), 241-259 (1992)

29. Rossi, R.G., Maraccini, R.M., Rezende, S.O.: Benchmarking text collections for classification and clustering tasks. Technical Report, University of Sao Paulo (2013)

AYTUĞ ONAN
DEPARTMENT OF COMPUTER ENGINEERING,
FACULTY OF ENGINEERING,
CELAL BAYAR UNIVERSITY, MANISA, 45140, TURKEY
E-MAIL: <AYTUG.ONAN@CBU.EDU.TR>

SERDAR KORUKOĞLU
FACULTY OF ENGINEERING,
DEPARTMENT OF COMPUTER ENGINEERING,
EGE UNIVERSITY,
IZMIR, 35100, TURKEY
E-MAIL: <SERDAR.KORUKOGLU@EGE.EDU.TR>

HASAN BULUT
FACULTY OF ENGINEERING,
DEPARTMENT OF COMPUTER ENGINEERING,
EGE UNIVERSITY,
IZMIR, 35100, TURKEY
E-MAIL: <HASAN.BULUT@EGE.EDU.TR>