International Journal of Hybrid Intelligent Systems 10 (2013) 33–41
DOI 10.3233/HIS-130163
IOS Press

Hybrid classifiers based on semantic data subspaces for two-level text categorization

Nandita Tripathi[a,*], Michael Oakes[a] and Stefan Wermter[b]
[a] *University of Sunderland, Sunderland, UK*
[b] *University of Hamburg, Hamburg, Germany*

Abstract. Many organizations are nowadays keeping their data in the form of multi-level categories for easier manageability. An example of this is the Reuters Corpus which has news items categorized in a hierarchy of up to five levels. The volume and diversity of documents available in such category hierarchies is also increasing daily. As such, it becomes difficult for a traditional classifier to efficiently handle multi-level categorization of such a varied document space. In this paper, we present hybrid classifiers involving various two-classifier and four-classifier combinations for two-level text categorization. We show that the classification accuracy of the hybrid combination is better than the classification accuracies of all the corresponding single classifiers. The constituent classifiers of the hybrid combination operate on different subspaces obtained by semantic separation of data. Our experiments show that dividing a document space into different semantic subspaces increases the efficiency of such hybrid classifier combinations. We further show that hierarchies with a larger number of categories at the first level benefit more from this general hybrid architecture.

Keywords: Hybrid classifiers, text classification, multi-level categorization, semantic subspace learning, maximum significance value

1. Introduction

Documents today are often maintained in category hierarchies rather than a flat categorization system. As the volume and diversity of documents grow, so do the size and complexity of the corresponding category hierarchies. Traditional methods of text categorization proceed by flattening these hierarchies but this leads to the curse of dimensionality [8] which degrades the performance of many classifiers. To be able to access such documents in real time, we need fast automatic methods to navigate these hierarchies. As data gets progressively sparser at deeper levels of a hierarchy, it is more appropriate to concentrate on the first few topic levels for categorization. Instead of using a classification algorithm at the first level, methods which can directly point to a relevant main topic should be explored. Such methods will help in increasing the search/categorization speeds. Subspace learning is an area popular with non-text domains such as pattern recognition to increase speed and accuracy. In this paper we present hybrid classifiers based on semantic data subspaces as a means to improve two-level categorization of text documents.

Classifier combinations are increasingly being used nowadays to improve learning performances. Instead of trying to improve performance of individual classifiers, the emphasis is now on various methods of combining simple classifiers. Diverse classifiers are being combined in an effort to improve classification performances [6]. Hybrid classifiers resulting from the combination of such diverse classifiers are being applied in various problem domains. Some examples of these are the Global-Local Hybrid Ensemble [3], hybrid pattern classifier for medical applications [15] and a hybrid classifier for spoken language classification [20].

The vast data space in today's world is divided into many subspaces which are quite different from each

*Corresponding author: Nandita Tripathi, University of Sunderland, Sunderland, UK. Tel.: +44 191 515 3631; Fax: +44 191 515 2781; E-mail: Nandita.Tripathi@research.sunderland.ac.uk.

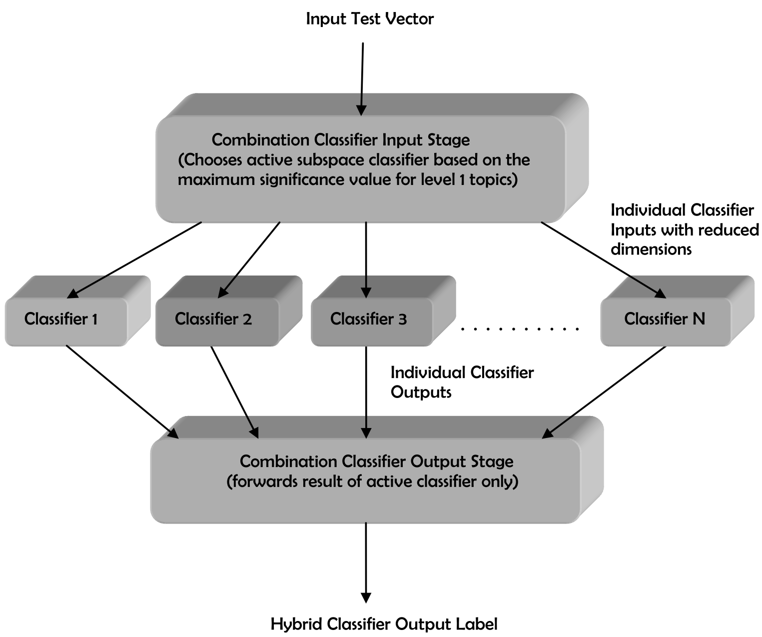

Fig. 1. Hybrid classifier architecture for subspace learning.

other, e.g. medicine and politics. These subspaces are often subdivided into further categories. Therefore, we need methods which can detect categories within these subspaces [14]. Hybrid classifiers can be effectively applied to subspace analysis. Since each subspace can be viewed as an independent dataset, different classifiers can be used to process different subspaces. Instead of using the complete set of full space feature dimensions, classifier performances can be boosted by using only a subset of the dimensions. The method of choosing an appropriate reduced set of dimensions is an active research area [10]. In the Random Subspace Method (RSM) [24], classifiers were trained on randomly chosen subspaces of the original input space and the outputs of the models were then combined. However, a random selection of features does not guarantee that the selected inputs have necessary distinguishing information. To address this problem, several variations of RSM have been proposed by various researchers such as Relevant Random Feature Subspaces for Co-training (Rel-RASCO) [25], Not-so-Random Subspace Method (NsRSM) [17] and Local Random Subspace Method [23].

In the real world, documents can be divided into major semantic subspaces with each subspace having its own unique characteristics. The above research does not take this division into account. In this paper, we describe a hybrid parallel architecture (Fig. 1) which takes advantage of the different semantic subspaces existing in the data. We use this architecture to show how the performance of various basic classifiers can be improved by combining them with classifiers of other types. We test the hybrid combinations of classifiers using the conditional significance vector representation [18] which is a variation of the semantic significance vector [21,22] to incorporate semantic information in the document vectors. The conditional significance vector enhances the distinction between subtopics within a given main topic. The region of the test data is determined by the maximum significance value [18] which is evaluated in O(k) time where k is the number of level 1 topics and thus can be very effective where time is critical for returning search results.

2. Methodology and architecture

Initially, we used two text datasets for our experiments – the Reuters Headlines dataset and the Reuters Full Text dataset – both drawn from the Reuters Corpus [19]. The Reuters Corpus is a well-known test bench for text categorization experiments. It has a hierarchical organization with four major groups which is well suited to test our hybrid architecture. Ten thousand Reuters Headlines along with their topic codes were extracted from the Reuters Corpus to constitute our first dataset. Some examples of Reuters Headlines are:

"Questar signs pact to buy oil, gas reserves."
"Ugandan rebels abduct 300 civilians, army says."
"Estonian president faces reelection challenge."

For the second dataset, we extracted ten thousand Reuters Full Text items which included both headline as well as body text for each news item. The news items in both cases were chosen so that there was no overlap at the first level categorization. Each news item belonged to only one level 1 category. At the second level, since most news items had multiple level 2 subtopic categorizations, the first subtopic was taken as the assigned subtopic. Thus, each news item (for both the headlines as well as the full text datasets) had two labels associated with it – the main topic (Level 1) label and the subtopic (Level 2) label. The news items were then preprocessed to separate hyphenated words. Dictionaries with term frequencies were generated using the Text to Matrix Generator (TMG) [4] toolbox. These were then used to generate the Full Significance Vector [18] and the Conditional Significance Vector [18] for each document. The two datasets were then randomized and divided into training sets of 9000 documents and corresponding test sets of 1000 documents.

The Waikato Environment for Knowledge Analysis (WEKA) [16] is a machine learning workbench with a number of learning algorithms of different types. We used two Bayesian algorithms (Naïve Bayes and BayesNet), two tree-based algorithms (J48 and Random Forest), one rule-based algorithm (PART) and one neural network (Multilayer Perceptron) as our test algorithms. These algorithms act as representatives of different classes of learning. We combined each algorithm with algorithms from other types in various new hybrid architectures in order to test a variety of learning algorithms. Classification accuracy, which is a comparison of the predicted class to the actual class, was recorded for each experimental run.

In order to scale up the number of categories, we also conducted similar experiments on a part of the Large Scale Hierarchical Text Collection (LSHTC) [11] challenge dataset provided on the LSHTC website (http://lshtc.iit.demokritos.gr). This challenge was part of the ECIR 2010 conference. The data was formed by crawling the web pages found in the ODP directory located at www.dmoz.org and the data was presented in the form of content vectors.

3. Experimental data generation

3.1. Text data preprocessing

From the Reuters Corpus, ten thousand Reuters Headlines and ten thousand Reuters Full Text items were used in these experiments. The level 1 categorization of the Reuters Corpus divides the data into four main topics, namely Corporate/Industrial (CCAT), Economics (ECAT), Government/Social (GCAT) and Markets (MCAT). Level 2 categorization further divides these into subtopics e.g. C11 (Strategy), E21 (Government Finance), GVIO (War) and M14 (Commodity Markets), which are subtopics of CCAT, ECAT, GCAT and MCAT respectively. A total of 50 subtopics were included in these experiments. Since all the news items had multiple subtopic assignments, e.g. C11/C15/C18 (Strategy/Performance/Ownership changes), only the first subtopic e.g. C11 (Strategy) was taken as the assigned subtopic. Our assumption here is that the first subtopic used to tag a particular Reuters news item is the one which is most relevant to it. The text data was then processed in two ways to generate data vectors in two different formats.

The LSHTC dataset has a multi-level categorization scheme and each document is tagged with the lowest level category code. The categorization scheme is provided in a separate file. These topic codes have numerical values. We pre-processed this data to replace the tagged code with the corresponding top 2 level category codes. The level 1 numeric codes were replaced by the letters A–J and level 2 numeric codes were replaced with the alphanumeric codes such as A01-A19 for the subtopics of A, B01-B36 for the subtopics of B, etc. The LSHTC data used in our experiments had 10 main and 158 subtopics as opposed to the 4 main and 50 subtopics in the Reuters datasets.

3.2. Semantic significance vector generation

We use a vector representation which represents the significance of the data and weighs different words according to their significance for different topics. Significance vectors [21,22] are determined based on the frequency of a word in different semantic categories. A modification of the significance vector called the semantic vector uses normalized frequencies where each word w is represented by a vector $(c_1, c_2, \dots, c_n)$ where c_i represents a certain semantic category and n is the total number of categories. A value $v(w, c_i)$ is calculated for each element of the semantic vector as follows:

$$v(w, c_i) = \frac{\text{Normalised Frequency of } w \text{ in } c_i}{\sum_k \text{Normalised Frequency of } w \text{ in } c_k}$$

where $k \in \{1 \dots n\}$.

For each document, the document semantic vector is obtained by summing the semantic vectors for each word in the document and dividing by the total number of words in the document. Henceforth it is simply referred to as the *significance vector*. The TMG Toolbox [4] was used to generate the term frequencies for each word in each news document. Two separate word vectors were generated for level 1 and level 2 categories and then concatenated. The final word vector consisted of 54 columns (for 4 main topics and 50 subtopics) for the two Reuters Corpus datasets and 168 columns (for 10 main topics and 158 subtopics) for the LSHTC dataset. While calculating the *significance vector* entries for each word, its occurrence in all subtopics of all main topics was taken into account. The vector generated this way was called the *Full Significance Vector*. We also generated the *Conditional Significance Vector* [18] where a word's occurrence in all subtopics *of only a particular main topic* is taken into account while calculating the word significance vector entries.

3.3. *Data vector sets generation*

As will be described below, two of these different vector representations (Full Significance Vector and Conditional Significance Vector) were generated for our experiments. The Conditional Significance Vectors were processed further to generate four main category-wise data vector sets for Reuters and ten main category-wise datasets for LSHTC.

3.3.1. *Full significance vector*

Here, the document vectors were generated from the full significance word vectors as explained in Section 3.2. For each Reuters Full Significance document vector the first four columns, representing four main topics – CCAT, ECAT, GCAT and MCAT, were ignored leaving a vector with 50 columns representing 50 subtopics. Both Reuters datasets were split into 9000 training/1000 test vectors. Similarly, for the LSHTC dataset the first ten columns representing the ten main topics were ignored leaving a vector with 158 columns representing 158 subtopics. The LSHTC data was split into 4000 training/463 test vectors.

3.3.2. *Category-wise conditional significance vectors*

Here, the conditional significance word vectors (see Section 3.2) were used to generate the document vectors and a train/test split of 90/10 was taken generating 9000 training and 1000 test vectors for each Reuters dataset. The training set was then divided into four sets according to the main topic labels. For each of these sets, only the relevant subtopic vector entries (e.g. C11, C12, etc. for CCAT; E11, E12, etc. for ECAT) for each main topic were retained. Thus the CCAT category training dataset had 18 columns for 18 subtopics of CCAT. Similarly the ECAT training dataset had 8 columns, the GCAT training dataset had 20 columns and the MCAT training dataset had 4 columns. These four training sets were then used to train the four parallel classifiers of the Reuters hybrid classifier. For LSHTC, the training set was divided into ten sets according to the main topic labels and for each of these sets, only the relevant subtopic vector entries for each main topic were retained. These ten training sets were then used to train the ten parallel classifiers of the LSHTC hybrid classifier. The main category of a test data vector was determined by the maximum significance vector entry for the first four columns representing the four main categories in the Reuters Corpus and the first ten columns representing the ten main categories of the LSHTC corpus. After this, the entries corresponding to the subtopics of this predicted main topic were extracted along with the *actual* subtopic label and given to the classifier trained for this predicted main category.

3.3.3. *Upper limit for hybrid classifier accuracy*

For the Reuters Headlines dataset, the accuracy of choosing the correct main topic by selecting the maximum significance level 1 entry was found to be 96.80% for the 1000 test vectors, i.e. 968 vectors were assigned the correct trained classifiers whereas 3.20% or 32 vectors were assigned to a wrong classifier – resulting in a wrong classification decision for all these 32 vectors. Hence, the upper limit for classification accuracy is 96.80% for our hybrid parallel classifier for the Reuters Headlines dataset. Similarly, this upper limit was found to be 82.50% for the hybrid parallel classifier for the Reuters Full Text dataset. The upper limit for the hybrid parallel classifier for the LSHTC dataset was found to be 85.31%.

3.4. *Classification algorithms*

Six classification algorithms were tested with our datasets, namely Random Forest, J48 (C4.5), the Multilayer Perceptron, Naïve Bayes, BayesNet and PART. Random Forests [12] are a combination of tree predic-

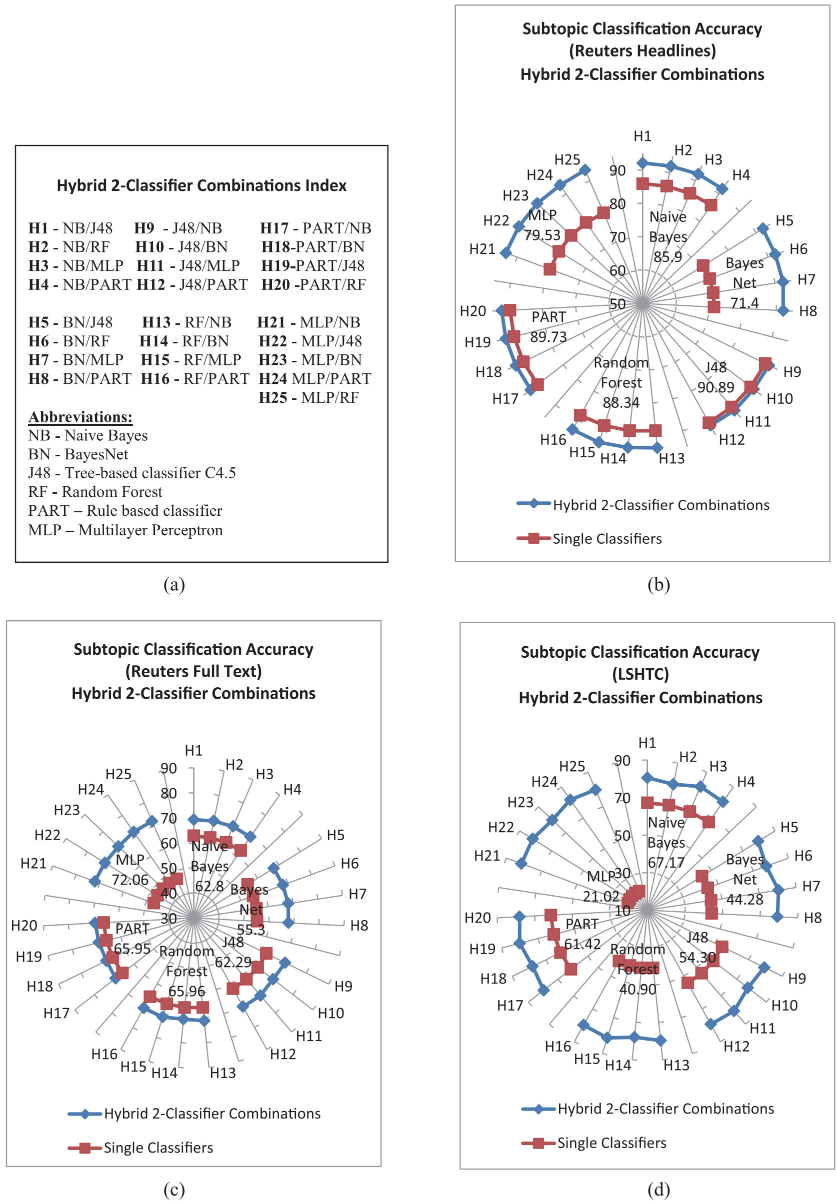

Fig. 2. (a): Hybrid 2-classifier combinations index; (b): Subtopic classification accuracy (reuters headlines) hybrid 2-classifier combinations; (c): Subtopic classification accuracy (reuters full text) hybrid 2-classifier combinations; 2(d): Subtopic classification accuracy (LSHTC) hybrid 2-classifier combinations. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130163)

tors so that each tree depends on the values of a random vector sampled independently. C4.5 [9] is an inductive tree algorithm with two pruning methods: subtree replacement and subtree raising. The Multilayer Perceptron [2] is a neural network which uses backpropagation for training. Naive Bayes [13] is the simplest form of Bayesian Network, in which all attributes are independent given the value of the class variable. BayesNet [7] implements Bayes Network learning using various search algorithms and quality measures. A PART [5] decision list uses C4.5 decision trees to generate rules.

4. Results and analysis

We tested various hybrid 2-classifier and 4-classifier combinations. For the hybrid 2-classifier combina-

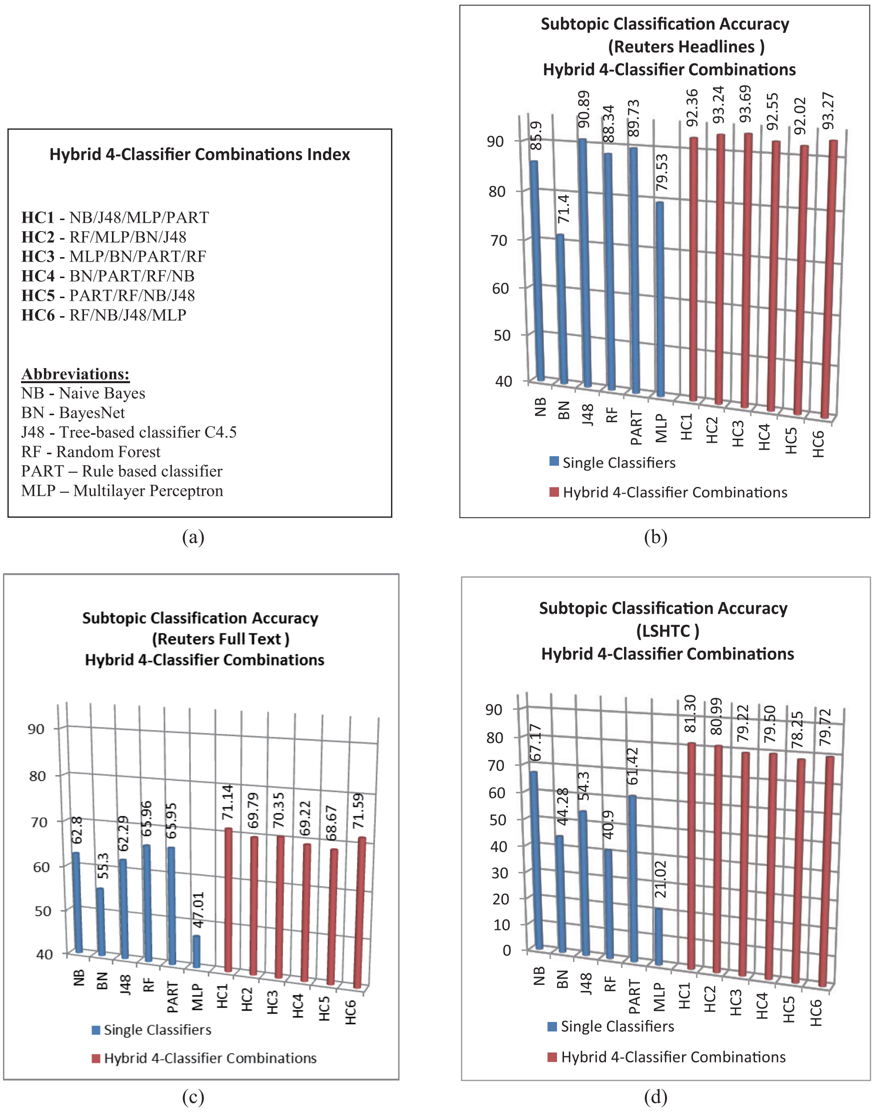

Fig. 3. (a): Hybrid 4-classifier combinations index; (b): Subtopic classification accuracy (reuters headlines) hybrid 4-classifier combinations; (c): Subtopic classification accuracy (reuters full text) hybrid 4-classifier combinations; (d): Subtopic classification accuracy (LSHTC) hybrid 4-classifier combinations. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/HIS-130163)

tions, a classifier of one type was combined with classifiers of other types in a large variety of combinations. The performance of each single classifier on the full data was compared with the performance of the hybrid 2-classifier combinations in which this particular classifier also participated. For the single classifier experiments, the Full Significance Vector representation was used whereas for the hybrid classifier experiments, the category-wise separated Conditional Significance Vector representation was used. The standard text classification train/test split [1] of 90:10 was taken in all cases for the Reuters datasets. The numerical values in all the figures represent the average of 10 runs with different parameter values.

Figure 2(a) shows the various hybrid 2-classifier combinations used in our experiments. In all combinations, it was observed that hybrid 2-classifier combinations performed better than the single basic classifier. Figures 2(b) and Fig. 2(c) show the subtopic classification accuracy of the hybrid 2-classifier combinations along with the subtopic classification accuracy of single basic classifiers for the Reuters Headlines and the Reuters Full Text datasets. Both datasets follow a similar pattern where all the hybrid classifiers perform bet-

An Example of Reuters Full Text

Headline:

Nicaraguan parliament rejects election postponement

Body Text:

Nicaragua's National Assembly on Monday rejected a proposed election reform that would have postponed the Oct. 20 presidential elections by two weeks.
In a letter sent to legislators, Vice Secretary of the National Assembly Jaime Bonilla said under the constitution the reform could not be passed by the current parliament and needed to wait until the next legislature is formed.
The rejected bill contained six election reforms including the postponement of the election till Nov. 3, 1996.
A group of presidential candidates for small parties led by Alfredo Cesar of the UNO 96 Alliance had proposed postponing the elections after the government failed to give the parties their campaign funding by the legal deadline.
The parties received the first installment of their campaign money two weeks after the start of the election campaign.
The smaller parties argued the delay put them at a disadvantage and had been engineered by the Liberal Alliance, the right wing party headed by Arnoldo Aleman that is ahead in the polls.
This was a political game. Here are the interests of the Liberal Alliance who don't want anyone else to be able to lift their head to threaten them - one of the bill's sponsors, conservative presidential candidate, Miriam Arguello, told reporters.
But the reforms were opposed by powerful groups in Nicaragua including President Violeta Chamorro, the two leading political parties, the Catholic Church and most international observer groups, who argued postponing the election date would hurt the vote's credibility.

Full Text = Headline + Body Text

Fig. 4. An example of reuters full text.

ter than any of the single classifiers. In both cases, this was statistically significant (Wilcoxon Signed Rank $h = 1$, $p =$ 1.304e-05). Numerically, the classification accuracy values for the Reuters Headlines are better than those of Reuters Full Text.

The single classifier performances also show a similar pattern for both datasets. In the tree-based classifiers, J48 performs better than Random Forest for Reuters Headlines and vice versa for Reuters Full Text. In Figs 2(b), (c) and (d) the hybrid classifier data points immediately above a particular single classifier show the 2-classifier combinations which include that single classifier e.g. the hybrid classifier data points H1-H4 which are above the single classifier Naïve Bayes show the 2-classifier combinations which include Naïve Bayes. As can be seen in the figures all the hybrid 2-classifier combinations perform better than the corresponding single classifiers.

Figure 2(d) shows the subtopic classification accuracy of the hybrid 2-classifier combinations along with the subtopic classification accuracy of the single basic classifiers for the LSHTC dataset. In this case too, all the hybrid classifiers performed better than any of the single classifiers. This result is statistically significant (Wilcoxon Signed Rank $h = 1$, $p =$ 1.2290e-005).

The single classifier performances also show a similar pattern to the Reuters datasets. In the tree-based classifiers, J48 performs better than Random Forest as with Reuters Headlines. The improvement in the performance obtained by using hybrid classifiers is much more marked with the LSHTC dataset than for both the Reuters Headlines and the Reuters Full Text dataset. For example, the hybrid 2-classifier combinations containing J48 show an improvement of about 1.5% for Reuters Headlines, 8% for Reuters Full Text and 25% for LSHTC over the corresponding baseline single J48 classifiers. Similarly, the hybrid 2-classifier combinations containing Naïve Bayes show an improvement of about 6% for Reuters Headlines, 6% for Reuters Full Text and 13% for LSHTC over the corresponding baseline single Naïve Bayes classifiers. The BayesNet classifier similarly shows an improvement of about 21% for Reuters Headlines, 12% for Reuters Full Text and 35% for the LSHTC dataset on hybridization.

As the LSHTC dataset has 10 main topics, the LSHTC hybrid classifier has 10 basic classifiers to deal with these 10 subspaces (main topics). The total number of subtopics in the LSHTC dataset is 158. The average number of subtopics to be distinguished by each subspace classifier is therefore 15.8 (158/10). The average vector length handled by a subspace classifier is also 15.8 in this case. The baseline single LSHTC classifiers, however, has to distinguish between all the 158 subtopics and deal with a vector length of 158. The combined effect of high dimensions and high number of categories considerably reduces the classification accuracies of the baseline single LSHTC classifiers. Thus the gap between the classification accuracies of the hybrid classifiers and the baseline single classifiers is very large for the LSHTC dataset. The Reuters datasets, on the other hand, has 4 main topics and 50 subtopics. The Reuters hybrid classifier thus has 4 basic classifiers for the 4 subspaces (main topics). The average number of subtopics to be distinguished by each Reuters subspace classifier is 12.5 (50/4). The average vector length is also 12.5 here. Thus the baseline single Reuters classifiers has to distinguish between 50 subtopics with a vector length of 50. While the complexity required to be handled by the subspace classifiers is similar for both the LSHTC and Reuters hybrid classifiers (subtopics and vector lengths of 15.8 vs. 12.5), the complexity to be handled by the baseline single classifiers is very different (subtopics and vector lengths of 158 for LSHTC vs. 50 for Reuters). The subtopic classification accuracies obtained with baseline single classifiers for LSHTC are much less than those obtained with the baseline single classifiers for Reuters. This is the cause of the greater improvement observed by the LSHTC dataset with the use of hybrid classifiers. Thus increasing the number of level 1 topics (subspaces) causes an increase in the number of subspace classifiers employed by a hybrid classifier thereby causing an increased improvement in the subtopic classification performance. Therefore, the effectiveness of hybrid classifiers increases with an increasing number of categories.

Figure 3(a) gives the details of the various hybrid 4-classifier combinations used in our experiments. Figure 3(b) and (c) show the subtopic classification accuracy of these hybrid 4-classifier combinations along with the subtopic classification accuracy of single basic classifiers for the Reuters Headlines and the Reuters Full Text datasets. Here again, both datasets follow a similar pattern where all the hybrid classifiers perform better than any of the single classifiers. Once again, this is statistically significant for both headlines and full text (Wilcoxon Signed Rank $h = 1, p = 0.03125$). In this case too, the numerical classification accuracy values for the Reuters Headlines are better than those of Reuters Full Text.

Figure 3(d) shows the subtopic classification accuracy of the hybrid 4-classifier combinations along with the subtopic classification accuracy of the single basic classifiers for the LSHTC dataset. Here, too, all the hybrid classifiers perform better than any of the single classifiers. These results follow a pattern similar to that of the Reuters datasets for the 4-classifier combinations. This is again statistically significant (Wilcoxon Signed Rank $h = 1, p = 0.03125$).

5. Conclusion

Our experiments highlight the fact that Reuters Headlines perform better than Reuters Full Text for the purpose of news categorization. This finding is consistent across all types of experiments – single classifiers, hybrid 2-classifier combinations as well as hybrid 4-classifier combinations. This can be attributed to the fact that Reuters Full Text contains a lot of text which is just introduced to make reading interesting. From a text processing point of view, this acts as noise which interferes with the relevant words (see example in Fig. 4). On the other hand, Reuters Headlines provide a concise summary of the news article which improves classification accuracy.

Our results also show that combining a basic classifier in parallel with classifiers of other types in a hybrid combination improves the classification accuracy of the corresponding basic classifier where the data is divided into distinct semantic categories. They also show that combining various types of classifiers in a hybrid combination results in a classification accuracy better than that of all the constituent single classifiers. The experiments confirm the fact that the maximum significance value is very effective in detecting the relevant subspace of a test document and that training different classifiers on different subsets of the original data enhances overall classification accuracy. The datasets with a larger number of categories benefit more from this architecture. This result is particularly encouraging for real-world applications where the number of categories would be much larger than the number present in these experimental datasets.

References

[1] Andrew K. McCallum MALLET: A Machine Learning for Language Toolkit. http://mallet.cs.umass.edu. 2002.

[2] B.Verma, Fast training of multilayer perceptrons, In IEEE Transactions on Neural Networks, Vol 8, Issue 6, Nov 1997, pp. 1314–1320.

[3] D. Baumgartner and G. Serpen, Comparative performance evaluation of global-local hybrid ensemble, *In International Journal of Hybrid Intelligent Systems* **8** (2011), 59–70.

[4] D. Zeimpekis and E. Gallopoulos, TMG: A MATLAB Toolbox for Generating Term Document Matrices from Text Collections, in: *Book Chapter in Grouping Multidimensional Data: Recent Advances in Clustering*, J. Kogan and C. Nicholas, eds, Springer, 2005.

[5] E. Frank and I.H. Witten, Generating Accurate Rule Sets Without Global Optimization, in: *Machine Learning: Proceedings of the Fifteenth International Conference*, J. Shavlik, ed., Morgan Kaufmann Publishers, 1998.

[6] F. Bernardini, M. Monard and R. Prati, Constructing ensembles of symbolic classifiers, *In International Journal of Hybrid Intelligent Systems* **3**(Issue 3) (August 2006), 159–167.

[7] F. Pernkopf, Discriminative learning of Bayesian network classifiers, In Proceedings of the 25th IASTED International Multi-Conference: Artificial intelligence and applications, 2007, pp. 422–427.

[8] J.H. Friedman, On Bias, Variance, 0/1 – Loss, and the Curse-of-Dimensionality, *In Data Mining and Knowledge Discovery* **1**(1) (1997), 55–77.

[9] J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA. 1993.

[10] K.R. Varshney and A.S. Willsky, Learning dimensionality-reduced classifiers for information fusion, In Proceedings of the 12th International Conference on Information Fusion, Seattle, Washington, July 2009, pp. 1881–1888.

[11] Kosmopoulos, Gaussier, Paliouras, and Aseervatham, 2010, The ECIR 2010 Large Scale Hierarchical Classification Workshop, *In SIGIR Forum* **44**(1) (June 2010), 23–32.

[12] L. Breiman, Random Forests, *In Machine Learning* **45**(1) (Oct. 2001), 5–32.

[13] L. Jiang, D. Wang, Z. Cai and X. Yan, Survey of Improving Naïve Bayes for Classification, Proceedings of the 3rd International Conference on Advanced Data Mining and Applications (ADMA '07), 2007, pp. 134–145.

[14] L. Parsons, E. Haque and H. Liu, Subspace Clustering for High Dimensional Data: A Review, *In ACM SIGKDD Explorations Newsletter* **6**(1) (2004), 90–105.

[15] M. Frize, N. Yiu and S. Weyand, Effectiveness of a hybrid pattern classifier for medical applications, *In International Journal of Hybrid Intelligent Systems* **8** (2011), 71–79.

[16] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. Witten, The WEKA Data Mining Software: An Update, *In ACM SIGKDD Explorations Newsletter* **11**(1) (July 2009), 10–18.

[17] N. Garcia-Pedrajas and D. Ortiz-Boyer, Boosting Random Subspace Method, *Neural Networks* **21** (2008), 1344–1362.

[18] N. Tripathi, S. Wermter, C. Hung and M. Oakes, Semantic Subspace Learning with Conditional Significance Vectors, Proceedings of the IEEE International Joint Conference on Neural Networks, Barcelona July 2010, pp. 3670–3677.

[19] T. Rose, M. Stevenson and M. Whitehead, The Reuters Corpus Volume 1 – from Yesterday's News to Tomorrow's Language Resources, In Proceedings of the Third International Conference on Language Resources and Evaluation (LREC-02), 2002, pp. 827–833.

[20] S. Garfield, S. Wermter and S. Devlin, Spoken language classification using hybrid classifier combination, *In International Journal of Hybrid Intelligent Systems* **2**(Issue 1) (January 2005), 13–33.

[21] S. Wermter, Hybrid Connectionist Natural Language Processing, Chapman and Hall, 1995.

[22] S. Wermter, C. Panchev and G. Arevian, Hybrid Neural Plausibility Networks for News Agents, In Proceedings of the Sixteenth National Conference on Artificial Intelligence, 1999, pp. 93–98.

[23] S.B. Kotsiantis, Local Random Subspace Method for constructing multiple decision stumps, International Conference on Information and Financial Engineering, 2009, pp. 125–129.

[24] Tin Kam Ho, The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(Issue 8) (Aug 1998), 832–844.

[25] Y. Yaslan and Z. Cataltepe, Co-training with relevant random subspaces, *Neurocomputing* **73** (2010), 1652–1661 (Elsevier).