Enhancing Semi-Supevised Text Classification using Document Summaries

Esaú Villatoro-Tello[2], Emmanuel Anguiano[1], Manuel Montes-y-Gómez[1], Luis Villaseñor-Pineda[1], and Gabriela Ramírez-de-la-Rosa[2]

[1] Language Technologies Lab., Computational Sciences Department, Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE), México.
{eanguiano,mmontesg,villasen}@ccc.inaoep.mx
[2] Language and Reasoning Research Group, Information Technologies Dept., Universidad Autónoma Metropolitana (UAM) Unidad Cuajimalpa, México
{evillatoro,gramirez}@correo.cua.uam.mx

Abstract. The vast amount of electronic documents available on the Internet demands for automatic tools that help people finding, organizing and easily accessing to all this information. Although current text classification methods have alleviated some of the above problems, such strategies depend on having a large and reliable set of labeled data. In order to overcome such limitation, this work proposes an alternative approach for *semi-supervised* text classification, which is based on a new strategy for diminishing the sensitivity to the noise contained on labeled data by means of automatic *text summarization*. Experimental results showed that our proposed approach outperforms traditional semi-supervised text classification techniques; additionally, our results also indicate that our approach is suitable for learning from only one labeled example per category.

Keywords: Text Classification, Text Summarization, Semi-Supervised Learning, Self-training, Feature Selection

1 Introduction

Nowadays there are millions of digital texts available on the Internet, which rapidly increase. This situation has produced a growing need for tools that help users to find, organize, and analyze all these resources in short periods of time. Particularly, Text Classification (TC)[1], *i.e.* the task of automatic assignment of free text documents to one or more predefined categories or topics, has emerged as a very important component in many information management tasks.

Traditionally, TC is approached by means of supervised techniques, *i.e.,* automatically constructing a classifier from a large set of labeled (*i.e.,* already categorized) documents. In addition, a common practice considered in the TC pipeline is applying feature selection methods, which use statistical information from the training set in order to identify those attributes that better describe the documents among different categories [2]. However, under this paradigm, a major problem is the high cost involved in collecting enough labeled data for building an effective classification model as well as for performing an appropriate feature selection. For example, in classification tasks such as deceptive opinion identification [3] and author profiling [4] it is very difficult to effectively collect or even validate enough labeled data.

In order to overcome the above mentioned problems, semi-supervised learning techniques have gained popularity among the scientific community. For instance, the *self-training* algorithm represents a learning strategy that aims at building an accurate classification model through iteratively increasing the training set by means of labeling and selecting the most confident recently classified instances from the unlabeled data. Some research works have demonstrated the pertinence of the *self-training* method in some TC tasks [5,6,7,8]. However, some of the difficulties faced by these approaches is the sensitivity of the algorithm to the noise contained in the labeled data as well as how to effectively select useful new labeled instances in every iteration.

In this paper we propose a modification to the self-training approach for diminishing the sensitivity of the learning algorithm to the noise contained on the labeled data, which consists in using document summaries. This idea was motivated by the work described in [9,10], where the benefits of using document summaries in a supervised TC approach are described. Contrastingly, we evaluate the pertinence of text summaries in a semi-supervised TC scenario, *i.e.,* when very few labeled data are available for building the classification model. As an additional contribution, we propose a new strategy for performing the instance selection process within the self-training algorithm, which helps in the process of preserving high homogeneity values among classes. To sum up, we aim to determine the usefulness of summarization as a noise filtering technique for improving *semi-supervised* text classification. Our main hypothesis establishes that if we incorporate a highly-confident summary, instead of a full document, to the set of labeled data during the training process of a TC model, we will be able to avoid dramatic changes on the behavior of our self-training algorithm, leading to a correct learning of the true target function.

Particularly, we are interested in evaluating the proposed method under exceptional circumstances, *i.e.*, having only one labeled document. Under this scenario, known as *one-shot learning* in the pattern recognition field [11]; our results indicate that having just one labeled summary per category provides enough information to our text classification system, outperforming the traditional semi-supervised configuration (*i.e.*, using full documents). Furthermore, obtained results also demonstrate that our proposed method for selecting highly-confident labeled documents tends to get better performance across iterations when short summaries are added to the set of labeled data rather than full documents.

The rest of this document is organized as follows. Section 2 presents some related work concerning the use of *text summarization* in the task of TC. Section 3 describes our proposed method; particularly it details the automatic *text summarization* process as well as our proposed self-training method. Then, Section 4 describes the used datasets, the experimental configuration and shows the results achieved by the proposed approach as well as some baseline results corresponding to the application of traditional text classification techniques under a semi-supervised paradigm. Finally, Section 5 depicts our conclusions and some future work ideas.

2 Related work

Although there are several works that proposed solving the problem of TC using different strategies of Text Summarization (TS), to the best of our knowledge there is no prior work that has explored the importance of using automatic generated summaries as noise filtering strategies under a *semi-supervised* text classification configuration, *i.e.*, having very few labeled data for training a classifier.

Ideally, a summary contains only the most relevant information from a document and given that such summary represents a significantly shorter document than the original, there are several approaches that prefer using summaries instead of full documents for improving the performance of a supervised TC system. For example in [12], authors proposed a new form of weighting terms by taking into account their frequency and their position within documents; whereas in [13], it is considered a weighting scheme that rewards terms from those phrases selected by a summarization method. Similar methods are described in [14], where relevant sentences are used for training a supervised classifier.

Works described in [15,10] explicitly proposed using summaries as a feature selection strategy. Authors applied different summarization techniques and compared their achieved TC results against those obtained when a statistical feature selection techniques are employed, *e.g.* information gain. Both papers conclude that using document summaries as a feature selection method, represents a competitive strategy against traditional statistical techniques. As a consequence of summarizing documents prior to a TC process, the dimensionality required for representing such documents is considerably reduced. The advantages obtained from the dimensionality reduction has been widely discussed in [13] and [9].

As we mentioned before, to the best of our knowledge there is no prior work related to the use of TS under a semi-supervised paradigm, hence we describe only those that we consider are the more related to the proposed approach. In [5], authors introduce a self-trainig algorithm that employs an ensemble of classifiers, namely *Ordered Classification*. This method allows selecting highly confident documents (in every iteration) to be included in the training data. In [7] authors describe a method for discovering the constant common knowledge in both, training and test sets by means of semi-supervised strategies. One of the closest works is the proposed by [6] which proposes a semi-supervised method for TC, which considers the extraction of unlabeled examples from the Web and the application of a enriched self-training approach for the construction of the classification model. And finally, in [8] authors propose a novel self-training approach for sentiment classification, their method uses multiple feature subspace-based classifiers for exploring a set of good features for better classification decision and to select the informative samples for automatically labeling data. Although these are related works, all of them employ complete documents during training and classification stages. In this work, we incorporate the advantages of text summaries as a noise filtering strategy for improving semi-supervised text classification. In addition, we propose a novel strategy for selecting high-confident instances during self-training, which allows to obtain high homogeneity among classes. Following sections describe into detail the proposed approach.

3 Proposed method

Our proposed method represents a modification of the traditional self-training algorithm. This algorithm assumes that, at the beginning, there are very few labeled data (D_L) and a very large set of unlabeled data (D_U). The goal is to obtain and use relevant information, extracted from D_U in order to improve the initial classifier (Φ_0) which was trained over D_L. To obtain such information, Φ_0 is used to classify the elements of D_U, then, by means of a specific selection criteria, some elements of D_U are considered for augmenting the set D_L. Once D_L has been updated, a new training process is performed to construct the classifier Φ_1. As expected, self-training algorithm represents an iterative process that is repeated until some stop criteria is reached.

A general view of the proposed self-training method is shown in Figure 1. Generally speaking, our proposed method starts by automatically constructing summaries from the set of labeled data D_L (*Text Summarization* module). Then, while some stop criteria is not achieved, such summaries are employed to construct a classification model Φ_0 (*Classifier Construction* module), which is used to classify all unlabeled data D_U. Next, we evaluate and preserve, through our *Instance Selection* module, those documents that represent the most confident labeled instances. The exact same number of documents for each category c_j are preserved, and selected documents are removed from D_U. The following step consists in creating their respective summaries and incorporating them into the original training data set. Then, we retrain our classifier to create the Φ_i classification model.

The *instance selection* module represents an important contribution since contrary to the traditional self-training algorithm, we perform this step separately from the classification stage *i.e.,* the classifier's confidence is not considered for this process. By means of this we favour high homogeneity among classes and avoid the bias introduced by a classifier trained with very few labeled data. Following sections describe in detail each one of the main modules from our proposed method (see Figure 1).

3.1 Text summarization module

For the implementation of the TS module, we used an unsupervised strategy that has demonstrated being able to construct high quality summaries, particularly we employed a graph based technique as explained in [9].

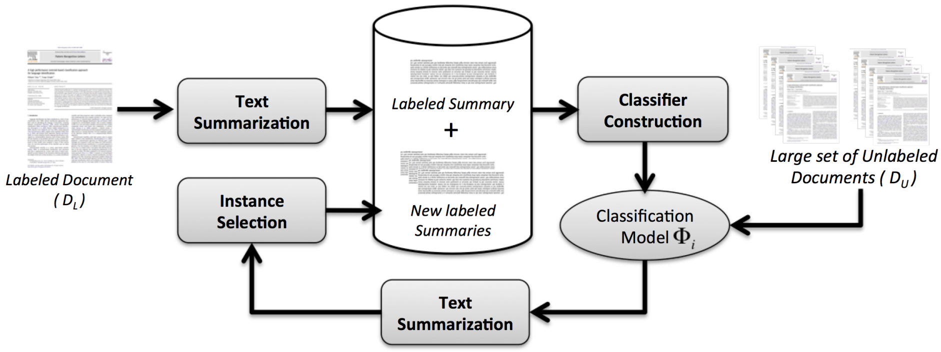

Fig. 1. General Architecture of the proposed self-training text classification method.

The underlying idea of this method is that every sentence can be represented as the vertex of some graph. By means of similarity measures it is possible to assign an specific rank to each sentence (*i.e.*, an importance value to each element on the graph). Finally, assuming an ideal ranking, this strategy preserve the top n better ranked sentences to construct the final summary.

Similarly to [9], we employed a measure of ranking, originally proposed for Web pages which is called *Hyperlinked Induced Topic Search* [16] to determine the associated value of ranking for each sentence from a document. First, the similarity of each sentence against the rest is computed in order to assign a weight value to all edges[3]. Once all edges have their corresponding weight, two different coefficients are iteratively computed as follows:

$$HITS_A(V_i) = \sum_{V_j \in In(V_i)} w_{ji} HITS_H(V_j) \quad (1)$$

$$HITS_H(V_i) = \sum_{V_j \in Out(V_i)} w_{ij} HITS_A(V_j) \quad (2)$$

Where V_i represent a vertex on the graph $G = (V, E)$; $In(V_i)$ represent the number of incoming links to V_i, whereas $Out(Vi)$ represent the out-coming links from vertex V_i. Consequently, $HITS_A$ represent the "authority value" (vertex with a large number of incoming links), while $HITS_H$ the "hub value" (vertex with a large number of outgoing links).

Once the HITS value of each vertex has been computed, the assigned value represents the importance of each sentence within the document. From here we preserve only the top n most relevant sentences to construct the final summary.

[3] Normally the direction of the edges is determined by the order of the sentences in the original document.

For our experiments, n is defined in function of the length of the document[4], *i.e.*, is defined dynamically rather than being a fixed number.

3.2 Instance selection module

The most natural form of selecting new instances to be added on each iteration to D_L is through the classification confidence degree assigned by the classification model. Nevertheless, this criterion might not be preferable since it depends directly from the classifier's quality, which in turn depends on the quantity and quality of the available training data. As we mentioned in Section 1, we are considering a scenario where very few labeled data are available (*e.g.,* one-shot learning), thus, traditional instance selection criteria are not suitable for this type of situations[5].

To overcome the above problem, we propose a novel method for assigning a confidence value to recently classified documents, which is independent from the employed learning algorithm as well as independent from the quantity and the quality of the labeled data. Our proposed instance selection criteria allows to preserve high homogeneity values among classes, and represents a distance based approach, which is computed as follows:

$$dist(d_U, C_j) = \frac{\sum_{d_L \in C_j} dist(d_U, d_L)}{|C_j|} \tag{3}$$

where d_L represents the summary of a labeled document such that $d_L \in D_L$. Hence, after applying Formula 3, those documents from D_U that were assigned to the label c_j and with the minor average distance to the corresponding class C_j are considered as highly confident.

During our experimental phase, we defined $dist(d_U, C_j)$ by means of an euclidean distance. Then, we preserve the top k documents most similar to the class c_j to include them into the labeled set of summaries.

3.3 Classification module

As explained before, determining the confidence degree of the classified instances does not depend on any particular algorithm, thus, any learning algorithm can be employed in the classification module.

In particular, we use the Support Vector Machine (SVM) method given that is especially suited to work with datasets with high dimensionality. For performing our experiments we employed the SVM implementation included in the Weka[6] toolkit with the default parameters.

[4] The parameter that defines the length of a summary is also known as the *compression rate parameter*, and represents a number that indicates the percentage of the information that we are requiring to preserve from the original document.

[5] One disadvantage of self-training is that mistakes reinforce/strengthen themselves; it is well known that accuracies lower than random at the beginning tend to conduct to worst results in subsequent iterations.

[6] `http://www.cs.waikato.ac.nz/ml/weka/`

Table 1. Statistics from the employed document collection: number of documents per category, average documents size (in tokens) and average vocabulary size.

Categories names	***Training Documents***			***Test Documents***		
	Num. docs	*Docs size*	*Vocab. size*	*Num. docs*	*Docs size*	*Vocab. size*
earn	2701	49.99	31.00	1040	45.01	26.42
acq	1515	74.74	50.36	661	71.25	48.16
trade	241	121.78	81.46	72	125.12	83.01
crude	231	110.57	71.87	112	102.21	66.66
money-fx	191	97.46	65.14	76	95.55	65.92
interest	171	88.43	57.35	73	91.43	59.04
ship	98	82.56	59.01	32	79.25	57.71
grain	41	115.89	75.40	9	85.66	50.33
Total:8	*5189*	*67.36*	*43.80*	*2075*	*63.42*	*40.66*

4 Experimental results

4.1 Data set

For validating our hypothesis, we performed experiments with the R8 data set. This collection is formed by the eight largest categories from the Reuters-21578 corpus, which documents belong to only one class. In order to know a more detailed description of this data set refer to [17]. Table 1 shows some basic statistics (*e.g.*, number of training/test documents, vocabulary size, etc.) from the employed data set. As can be seen, it is a highly unbalanced data set.

For performing our experiments we randomly select *one* document from the training set as the labeled training document, *i.e.*, our algorithm always begin iterating with only one single document and the rest are considered as the unlabeled data. It is important to mention that our method does not require knowing if the dataset is unbalanced or not, since given the nature of the proposed algorithm, this will converge to an accurate classification model.

4.2 Method configuration

As mentioned in Section 3, our method requires an user defined parameter k, which represents the number of documents to be added for each category into the labeled data. For our experimental phase we considered adding $k = 1$ and $k = 5$ documents in every iteration. It is worth to remember that at the beginning, the self-training algorithm starts by having only one labeled document, replicating a *one-shot learning* scenario. Our reported experiments represent the average performance obtained when randomly varying five times the initial labeled document. As stop-criteria we defined the following: *i)* when a top of 20 iterations is reached, and *ii)* when there are no more unlabeled documents in D_U. Finally, it is worth mentioning that for all the experiments, documents were represented following the traditional *vector model* from the Information Retrieval field, specifically a *tf* (term-frequency) weighting scheme.

4.3 Baseline definition

Since we aim at demonstrating that using summaries allows to improve the performance of the proposed self-training method, we defined as our *baseline* approach the exact same algorithm but using complete documents instead of summaries. In other words, this baseline depicts the traditional self-training algorithm for text classification without summarizing the considered documents. Similarly, documents were represented by means of a vector model using a *tf* weighting scheme.

As an additional baseline, we consider the result obtained by the classifier when training only with a single labeled data. In our experiment this result corresponds to iteration number 0. This baseline aims at demonstrating that using summaries for training a TC model allows to obtain, from the beginning, better classification results.

4.4 Evaluation metrics

The effectiveness of the proposed method was measured by means of the macro-averaged F_1 evaluation measure. Using this type of measure is very useful since it allows obtaining a confident perspective of the system's performance, particularly for cases where classes are highly unbalanced.

4.5 Results

Figure 2 shows the obtained results for the performed experiments having different compression rate summaries (30%, 50% and 70%). The graph on the left represents the behaviour of the proposed method when one document ($k = 1$) is added in every iteration, whereas the graph on the right depicts the results when five documents ($k = 5$) are added in every iteration. As we mentioned, all the experiments started with only one document per class on the D_L set.

It is important to remark that our proposal of using summaries for training rather than full documents allows to improve the performance of the classification system from iteration 0, *i.e.*, when our proposed method still has not started iterating yet. Notice that for the baseline configuration, *i.e.,* when training with a full document the $F_1 = .34$ at iteration 0, and for the same case we get a $F_1 = .43$ when using a summary of 30% compression rate, which confirms the ability of *text summarization* as feature selection method.

These results also indicate that the baseline configuration hardly improves its performance even when more documents are added to the training data. Particularly, it goes from $F_1 = .34$ on iteration 0 to $F_1 = .38$ on iteration 20 when $k = 1$, and from $F_1 = .34$ on iteration 0 to $F_1 = .37$ on iteration 20 when

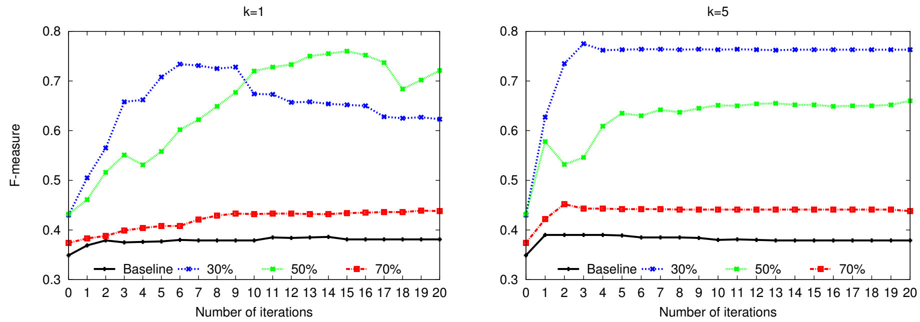

Fig. 2. Obtained results on the R8 collection. On the left 1 document is added every iteration ($k = 1$), whilst on the right 5 documents are added each iteration ($k = 5$).

$k = 5$. On the contrary, our proposed method obtains a significant improvement as more iterations are performed. Particularly, when we add only 1 document per iteration per category (Figure 2, $k = 1$), and using summaries that represent only 30% of the size of the original document, our system goes from a $F_1 = .43$ at iteration 0 to $F_1 = .65$ at iteration 3. A similar situation occurs for the case of adding 5 documents on each iteration (Figure 2, $k = 5$), where using summaries of 30% size allow to our method to go from $F_1 = .43$ at iteration 0 to $F_1 = .73$ at iteration 3. This behavior indicates that our *instance selection* criteria, in combination with the use of summaries allows to maintain high homogeneity among classes, reduces the noise contained in labeled documents and converges to the true target function.

Finally, in order to validate the importance of the compression rate parameter from the *text summarization* module, we performed a series of experiments varying the size of the produced summary across the 20 iterations that are executed by the self-training algorithm. Figure 3 show the statistical variance of the F-score for $k = 1$ and $k = 5$, respectively.

In general, we can observe that the compression rate is an important parameter of the proposed method, however, it is clear that using summaries is in fact a better strategy than using full documents (*i.e.,* the baseline configuration). In addition, it is possible to notice that using summaries of 30%-50% compression rate allow obtaining the best results.

4.6 Discussion

What to summarize: With the intention of evaluating the pertinence of using summaries in both training and test phases, we carried out some experiments considering a supervised classification scenario. Table 2 shows the results from these experiments. The conclusion is clear: summarizing both training and test documents (*Sum-Sum*) worsen the classifier performance; on the contrary, the best performance was obtained when only training documents were summarized

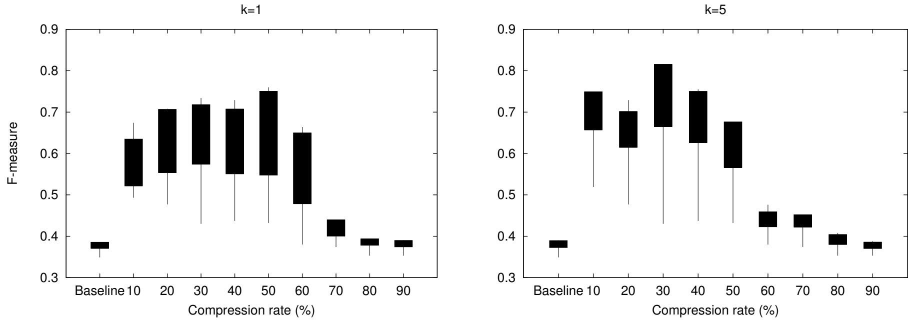

Fig. 3. Variance of the F-score performance for the proposed method when $k = 1$ and $k = 5$. Reported results are for 20 iterations at different compression rate values (10%-90%); first column depicts the baseline performance.

(*Sum-Doc*). These results supported the design of the proposed system architecture, where summaries were employed only on the training (see Figure 1).

Table 2. Results of a supervised TC method varying the TS module location. Notice that when no summarization is applied neither to training or test documents, the obtained performance is $F = 0.84$

Compression Rate	***Method Configuration***		
	Sum-Sum	*Doc-Sum*	*Sum-Doc*
30%	0.87	0.77	**0.89**
50%	0.86	0.82	**0.87**
70%	0.86	0.83	**0.86**

Some learned lessons: Despite obtained results, experiments indicate that the success of the proposed method depends to some extent on the following factors: *i*) the quality of the initial labeled data, *i.e.*, the initial labeled documents must be representative of their categories; *ii*) the number of unlabeled data: for those categories having a small number of examples in the unlabeled set, the selection of confident instances at every iteration becomes into a very hard task; *iii*) a good selection of the compression rate parameter: our method showed outstanding results in news classification when using compression rates from 10 to 60%, however, for noisier documents, such as social media texts, the compression rate definition could be complex and critical.

5 Conclusions

In this paper we have proposed a modification to the *self-training* algorithm for improving text classification when very few labeled documents are available. Our proposal considers using text summaries instead of full documents as a strategy for diminishing the sensitivity of the learning algorithm to the noise contained on the labeled data. Additionally, as a second contribution of this work, we have proposed a novel criterion for selecting highly confident elements to be included in the set of labeled data. The proposed criteria is performed separately from the classification process, which makes it independent from the learning algorithm, allowing to preserve high homogeneity values among labeled documents.

The performed experiments showed that our proposed algorithm is able to incorporate information from the unlabeled data for improving the performance of the classifier. By means of using automatic text summaries we are able to discard noisy information during the *self-training* process. Particularly, experimental results showed that shorter summaries are the best choice (30% to 50% compression rate). Further more, we demonstrated that the proposed approach is very suitable for collections with very few labeled data. Particularly, we evaluated our proposed method under a *one-shot learning* scenario, *i.e.,* having only one labeled document. Obtained results are promising and represent an initial effort towards the problem of one-shot text classification.

As future work we are interested in evaluating different unsupervised summarization techniques [18], aiming at determining the sensitivity of the proposed method towards the TS module. We are also interested in evaluating the performance of the proposed method using other semi-supervised strategies, *e.g.*, co-training and multi-view approaches, as well as in other larger datasets. Additionally, we intent to determine the pertinence of the proposed algorithm for solving non-thematic text classification tasks, such as author profiling problems (*e.g.,* age, gender, and personality recognition), where not enough/reliable labeled data are available.

Acknowledgments. This work was partially funded by CONACyT, project number 247870 and 258588. We appreciate the support provided by the Thematic Networks program (Language Technologies Thematic Network projects 260178 and 271622). We thank to UAM Cuajimalpa and SNI for their support.

References

1. F. Sebastiani, "Machine learning in automated text categorization," *ACM computing surveys (CSUR)*, vol. 34, no. 1, pp. 1–47, 2002.
2. Y. Villuendas-Rey and M. M. Garcia-Lorenzo, "Attribute and case selection for nn classifier through rough sets and naturally inspired algorithms," *Computación y Sistemas*, vol. 18, no. 2, pp. 295–311, 2014.
3. D. H. Fusilier, M. M. y Gómez, P. Rosso, and R. G. Cabrera, "Detecting positive and negative deceptive opinions using pu-learning," *Information Processing & Management*, vol. 51, no. 4, pp. 433 – 443, 2015.

4. A. P. López-Monroy, M. Montes-y Gómez, H. J. Escalante, L. Villaseñor-Pineda, and E. Stamatatos, "Discriminative subprofile-specific representations for author profiling in social media," *Knowledge-Based Systems*, vol. 89, pp. 134–147, 2015.
5. T. Solorio, "Using unlabeled data to improve classifier accuracy," *M. Sc. Degree Thesis, Computer Science Department, Inaoe, Mexico*, 2002.
6. R. Guzmán-Cabrera, M. Montes-y Gómez, P. Rosso, and L. Villaseñor-Pineda, "Using the web as corpus for self-training text categorization," *Information Retrieval*, vol. 12, no. 3, pp. 400–415, 2009.
7. Y. Zheng, S. Teng, Z. Liu, and M. Sun, "Text classification based on transfer learning and self-training," in *2008 Fourth International Conference on Natural Computation*, vol. 3, pp. 363–367, Oct 2008.
8. W. Gao, S. Li, Y. Xue, M. Wang, and G. Zhou, *Chinese Lexical Semantics: 15th Workshop, CLSW 2014, Macao, China, June 9–12, 2014, Revised Selected Papers*, ch. Semi-supervised Sentiment Classification with Self-training on Feature Subspaces, pp. 231–239. Cham: Springer International Publishing, 2014.
9. R. Mihalcea and S. Hassan, "Using the essence of texts to improve document classification," in *Proceedings of the Recent Advances in Natural Language Processing (RANLP-05)*, 2005.
10. E. Anguiano-Hernández, L. Villaseñor-Pineda, M. Montes-y Gómez, and P. Rosso, "Summarization as feature selection for document categorization on small datasets," in *Advances in Natural Language Processing*, pp. 39–44, Springer, 2010.
11. L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 594–611, April 2006.
12. S. J. Ker and J.-N. Chen, "A text categorization based on summarization technique," in *Proceedings of the ACL-2000 workshop on Recent advances in natural language processing and information retrieval: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 11*, pp. 79–83, Association for Computational Linguistics, 2000.
13. Y. Ko, J. Park, and J. Seo, "Automatic text categorization using the importance of sentences," in *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, pp. 1–7, Association for Computational Linguistics, 2002.
14. J. Xiao-Yu, F. Xiao-Zhong, W. Zhi-Fei, and J. Ke-Liang, "Improving the performance of text categorization using automatic summarization," in *Computer Modeling and Simulation, 2009. ICCMS'09. International Conference on*, pp. 347–351, IEEE, 2009.
15. A. Kolcz, V. Prabakarmurthi, and J. Kalita, "Summarization as feature selection for text categorization," in *Proceedings of the tenth international conference on Information and knowledge management*, pp. 365–370, ACM, 2001.
16. J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.
17. A. M. d. J. C. Cachopo, *Improving methods for single-label text categorization.* PhD thesis, Universidade Técnica de Lisboa, 2007.
18. M. Litvak and N. Vanetik, "Multi-document summarization using tensor decomposition," *Computación y Sistemas*, vol. 18, no. 3, pp. 581–589, 2014.