An Ensemble Classification Algorithm for Short Text Data Stream with Concept Drifts

Gang Sun, Zhongxin Wang*, Zhengqi Ding, and Jia Zhao

***Abstract*—With the rapid development of information technology and instant messaging, many short text data streams are continuously generated. Due to short texts' inherent features such as sparse features, weak concept description, and concept drifts, most of the existing data stream classification methods are difficult to adapt and decrease performance. In this paper, we propose a new short text data stream classification algorithm to handle the problem. Specifically, we first select the distinguishing features from short texts to form the feature space and use the similar features in the feature space to expand. Then, a concept drift detection method based on topic distribution is used to detect and adapt to the concept drifts in this short text data stream. Finally, the ensemble classification model is updated in accordance with whether concept drift happening or not. The experimental results show that the classification performance of the short text data stream algorithm proposed in the paper is better than other algorithms.**

***Index Terms*—Short text data stream, ensemble classification algorithm, feature extension, concept drift**

I. Introduce

With the rapid development of information technology and instant messaging, network users and servers generate a large amount of short text data at all times, such as Weibo, online comments, instant messages. Short text data streams play a vital role in user intent understanding, question-answering systems, and intelligent information retrieval. Short text data streams have three characteristics: sparsity, weak feature expression, and concept drifts. Due to the three characteristics mentioned above, it is not easy to apply traditional data stream classification methods to short text data streams. Traditional algorithms pay rare attention to the characteristics of the short text data stream and the concept drift hidden in the data stream at the same time [1-4]. Therefore, in accordance with the features of short text data streams, this paper proposes a new short text data stream classification algorithm to handle the corresponding problems. First, the algorithm selects the distinguishing features from short texts to form the feature space and use the similar features in the feature space to expand. Then, in order to effectively detect and adapt to concept drifts in the short text data stream, this paper uses a concept drift detection method based on topic distribution. It can determine whether the concept drift occurs in the new data block, through accumulating the semantic distance between each short text in the new and present data block. Finally, according to whether concept drift happening or not, update the ensemble classification model will be updated.

The main contributions of this paper are as follows:

1) First of all, this paper proposes a short text feature extension method based on the features of short texts. This method selects the distinguishing features from the short texts themselves to form the feature space and uses the similar features to expand the space.

2) Secondly, this paper proposes a concept drift detection method based on topic distribution, so as to detect the concept drifts hidden in the short text data stream. By accumulating the semantic distance between each short text in the new and present data block, it is determined whether the concept drift occurs in the new data block.

3) Finally, compared with most of existing data stream classification methods and short text classification methods, the ensemble classification model proposed in this paper can effectively detect the concept drifts hidden in the short text data stream and has better classification performance than other classification models.

II. Related Works

Short text feature extension methods are mainly divided into two categories: the first category is an extension method based on its resources, and the second one is an extension method based on external resources. Based on its resources, the feature extension method mainly uses the text's context or semantics to perform feature extension on short texts. Yuan et al. [5] improved the naive Bayes algorithm to adapt and deal with short text's sparse feature. Wang et al. [6] proposed a new short text method based on semantic clustering and convolutional neural networks. Gao et al. [7] used a structured sparse representation classifier to classify short text data effectively. Haddoud et al. [8] proposed more than 80 metrics for the weighting problem of vocabulary labeling. Bicalho et al. [9] presented a topic classification model, which uses the words "co-occurrence" and "vectors" to create an important document in the original document. Doulamis et al. [10] used a fuzzy time feature sequence to detect similar and related words in Weibo events. Flisar et al. [11] showed an improving short text classification using information from DBpedia ontology.

Manuscript received January 21, 2021; revised April 6, 2021. The work was supported by the National Natural Science Foundation of China (61906044), Big Data and Intelligent Computing Innovation Team of Fuyang Normal University (XDHXTD201703), Fuyang Humanities and Social Science Research Special Project (FYSK2019QD10), and the Natural Science Foundation of the Anhui Higher Education Institutions of China (KJ2019A0532, KJ2019A0542 and KJ2020ZD48).

Gang Sun is an associate professor of Fuyang Normal University, Fuyang 236037, China (e-mail: sungang@fynu.edu.cn).

Zhongxin Wang is a lecturer of Fuyang Normal University, Fuyang 236037, China (corresponding author phone: 86-558-2591393; e-mail: wzx@fynu.edu.cn).

Zhengqi Ding is an assistant experimenter of Fuyang Normal University, Fuyang 236037, China (e-mail: 201009001@fynu.edu.cn).

Jia Zhao is an associate professor of Fuyang Normal University, Fuyang 236037, China (e-mail: 201410003@fynu.edu.cn).

At present, there are few researches on short text data stream classification. Bouaziz et al. [12] introduced the IGLM model, which improves data stream classification by continuously updating the classifier. Ren et al. [13] proposed hierarchical multi-label short text data stream classification, which uses a block-based structure optimization strategy to classify short texts. Li et al. [14] proposed an incremental ensemble model to adapt to the short text data stream, which compensates for data sparsity by introducing more semantic contextual information hidden in a short text.

There have been also some researches on concept drift detection methods. Gmama et al. [15] proposed a method to detect the change of sample probability distribution. Kurlej et al. [16] proposed an active learning method, which detects concept drift by calculating the closest distance. Sun et al. [17] came up with the financial distress concept drift method, which creates a dynamic FDP model to select and predict instances. Hegedus et al. [18] presented an adaptive GoLF method, which mainly includes age-based drift processing and concept drift detection. Li et al. [19] showed an incremental EDTC method, which uses three random features to define three tangent points to complete the tree. Loeffel et al. [20] proposed an online algorithm that deals with various types of concept drifts and did not use a fixed threshold. Chen et al. [21] proposed to use a genetic algorithm to obtain fuzzy concept drift mode. Mirza et al. [22] proposed the ESOS-ELM method, which can handle unbalanced data streams. Sethi and Kantardzic et al. [23] proposed the MD3 method, a drift detection algorithm for unmarked stream data. Tennant et al. [24] used the minimum cluster neighbor method to calculate statistical data. Silva et al. [25] proposed the FEAC-Stream algorithm, which uses k-means clustering and automatically estimates the value of k from the stream. Deng et al. [26] proposed a concept drift detection method based on rough set and granular computing.

The algorithms mentioned above pay rare attention to the characteristics of the short text data stream, and the concept drifts hidden in the short text data stream at the same time. Therefore, it is not easy to meet the short text data stream's classification requirements with concept drifts.

III. Short Text Data Stream Ensemble Classification Algorithm

A. Problem Definition

The classification of short text data streams is defined as follows. Given a short text data stream D, the data stream D is divided into N data blocks of the same size, so D can be expressed as $D=\{D_1,D_2,D_3,\ldots,D_N\}(N\rightarrow\infty)$; Each short text data block Di is defined as $Di=\{d_1,d_2,d_3,\ldots,d_M\}(M=|Di|)$; each d_i can be defined as $d_i=\{(v_j,y_j)|v_j\epsilon E, y_j\epsilon Y\}$, where E represents the feature space of the data block, and Y represents the class label set. The goal of short text data stream classification is to train a dynamic classifier: $E\rightarrow Y$, which maps the feature vector to a set of labels. This classifier automatically adapts to changes in short text data streams and detects concept drifts in the short text data stream. Figure 1 shows the basic framework of the proposed algorithm in this paper.

B. Feature Space

Distinguishing features are chosen to form the feature space F, and are tried to be contained in each short text appear in the feature space, so as to ensure that each short text is directly related to the feature space. In addition, the selected features should be shared by many texts in a category, which can ensure the balance of feature distribution in the text and avoid sparseness. For this reason, this paper focuses on the difference in the number of texts in each category and the relevance between the feature and the category when making feature selection. The features that have a large contribution to the category are selected and used to represent the corresponding category. Then the selected features are combined to form a feature space except the repetitive features. The feature space will cover the text of each category to the greatest extent, and the selected features are also better to the classification of short texts.

This paper selects the same ratio of feature words for each category. The advantage of the method is that when the number of categories A and B is extremely unbalanced, the small category is not easily covered by the large category. Besides, this paper uses the intra-category dispersion CD_i to represent the distribution of features in a category, and the calculation is shown in formula (1).

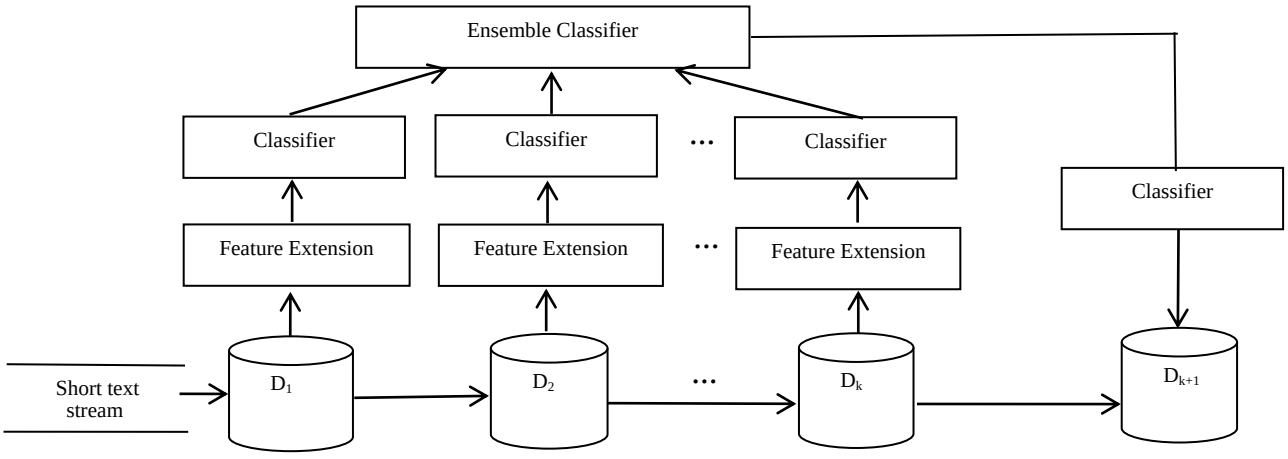

Fig. 1. The framework of data streams classification.

$$CD_i(f, C_i) = \frac{\sqrt{\sum_{j=1}^{n} tf_{ij}(f) - \overline{tf_i(f)}/n}}{\frac{n}{\sqrt{n-1}}\overline{tf_i(f)}} \quad (1)$$

where n represents the total number of documents in the category Ci, $tf_{ij}(f)$ indicates the number of occurrences of the feature f in the j-th document in the category Ci, and $\overline{tf_i(f)}$ refers to the average number of occurrences of the feature f in all documents in the category Ci. The smaller the intra-category dispersion CD_i of feature f, the more balance its distribution within the category, and the better the ability to distinguish category. Here, the feature CD_i in each category is calculated and sort from largest to smallest. For each category, the paper selects the top N features according to the ratio R, then combines all the selected non-repetitive features, and takes the features to form a feature space.

C. Feature Extension

When the feature f in the short text d is in the feature space F, the feature f' in F that has the greatest correlation with f is added to the short text. When the feature f is not in the feature space F, this paper uses the feature f'' contained in the short text that has the greatest correlation with the feature f to expand the short text. In the past, the main method of calculating the correlation between features was to use mutual information. Mutual information can measure the degree of dependence between variables, but it is not good for measuring the degree of correlation between features. It is very sensitive to inaccuracies caused by sparse data. The sparsity problem may cause the mutual information value between features to be negative, which will be adverse to later application and processing. Therefore, this paper uses the improved correlation calculation formula $RMI(x,y)$ based on mutual information to calculate the correlation degree, which weakens the influence of data sparseness on the correlation degree between features. The calculation is shown in formula (2):

$$RMI(x,y) = log\left(\frac{P(x,y)}{P(x)\times P(y)}\right) / log\left(\frac{2}{P(x)+P(y)}\right) \quad (2)$$

Where $P(x,y)$ represents the probability that x and y appear in the data set at the same time, $P(x)$ represents the probability that x appears in the data set, and $P(y)$ represents the probability that y appears in the data set.

D. Concept Drift Detection

In the short text data stream, the topic's potential drift over time is called concept drift, which seriously affects the classification performance. It has become an urgent problem to adapt to the concept drift in the short text data stream. Thus, this paper proposes a concept drift detection method based on topic distribution. We accumulate the semantic distance between each short text in the new and current data block to determine whether the concept drift occurs in the new data block. The calculation is shown in formula (3):

$$dist(D'_{i+1}, D'_i) = 1/|D_{i+1}| \sum_{j=1}^{|D_{i+1}|} dist(d'_j, D'_i) \quad (3)$$

To calculate the semantic distance $dist(d'_j, D'_i)$ in formula (3), according to the class distribution, the current data block is first divided into class clusters $\{L_c\}$, where L_c represents the short text set of class lable c; then the distance between the short text and all class clusters $\{L_c\}$ is calculated. The semantic distance $dist(d'_j, D'_i)$ between the short text and the current data block is the value of the smallest semantic distance to a certain class cluster. The calculation is shown in formula (4):

$$dist(d'_j, D'_i) = arg\ min_{c\in\{L_c\}}\ dist(d'_j, L_c) \quad (4)$$

Given a threshold μ, it can be judged whether a concept drift occurs. If $dist(D'_{i+1}, D'_i)$ is greater than μ, we consider that the concept drift occurs in the new data block.

E. Construction and Update of Ensemble Classification Model

K base classifiers are constructed on K expanded data blocks, respectively. SVM is used as the base classifier because it is widely used for text classification. When the new data block D_{new} arrives, we first perform feature extension on the short text to obtain the expanded short text data block D'_{new}. The semantic distance is calculated between the new data block D'_{new} and the data block D'_k that constructs the base classifier C_k. When the data block D'_{new} has a concept drift relative to each data block in the ensemble model, if the number of base classifiers in the ensemble model EC is less than K, the classifier C_{new} is added to the ensemble model EC; if it is equal, the classifier C_{new} is replaced the oldest base classifier in the ensemble model EC; otherwise, the classifier C_{new} is used to replace the base classifier constructed from the data block with the smallest semantic distance from the data block D'_{new}.

Our algorithm

Input: short text data stream D, the number of base classifiers K, the threshold of concept drift detection μ;

Output: an ensemble classification model EC;

For D_i in D do:

Performing feature extension on the short text data block D_i to obtain the expanded short text data block D'_i;

Constructing a base classifier Ci in the short text data block D'_i;

The weight of the base classifier C_i is the classification accuracy of the base classifier C_i on the data block D'_i;

If (i>K):

Calculating the semantic distance between the short text data block D'_{new} and D'_k;

Determining whether concept drift occurs according to the threshold of concept drift detection μ;

if (concept drift occurs):

Constructing a base classifier C_{new} in the data block D'_{new} and updateing the ensemble classification model EC;

End for.

IV. Experimental Results and Analysis

This section mainly compares the experimental results of the algorithm proposed in this paper with the benchmark algorithm. First, the short text data sets and the benchmark algorithms used in the experiment are given [27-31], and then the evaluation indexes and parameter settings of the experiment are explained. Finally, the effectiveness of the algorithm proposed in this paper in concept drift detection and classification performance is verified.

A. Experimental Data Sets and Evaluation Indexes

Table 1 shows the detailed information of the experimental data sets:

Snippets data set: Snippets is the search result of putting a preset expression on the Internet search engine, which has 8 categories and 12340 short texts.

News data set: News comes from TagMyNews data set, which has 7 categories. Each news contains a short headline, description and a date, and other information. In the experiment, we only select descriptions as a data set, and there are 32604 short texts.

Tweets data set: With the help of the Twitter keyword tracking API, the data is obtained from Twitter from November to December 2012, which has 4 categories. In the experiment, we only used the data from December 2012, and there are 803613 short texts.

This paper presents 4 classical data stream classification algorithms, 4 short text data classification algorithms, and 4 concept drift detection algorithms to verify the effectiveness of the algorithm proposed in this paper in classification and concept drift detection. Table 2 gives detailed information on the benchmark algorithms.

B. Parameter Setting

On the *Snippets* and *News* data sets, the data block size is set to 100. The *Tweets* data set is too large, so the data block size is set to 1000. In concept drift detection, we set a threshold, $\mu = 1 - 0.5/C$, where C represents the number of categories. When using an ensemble framework, too many or too few buffer data blocks are not appropriate. This paper selects 6 buffer data blocks to construct an ensemble classifier. Finally, we choose the *SVM* from libsvm as the base classifier, and the corresponding parameters are *C-SVC* and linear kernel function.

C. Parameter R

The feature selection ratio R is an essential parameter in the algorithm, which directly affects the feature space's dimension. If the selection ratio is small, the dimension of the feature space is small. Otherwise, the dimension of the feature space is enormous. Figure 2 shows this method's variation curve with the feature selection ratio on the data sets *Snippets*, *News*, and *Tweets*. It can be seen from Figure 2 that as the ratio of feature selection increases, the classification accuracy increases first and then stabilizes. This is because the selected features are highly representative. After reaching a certain accuracy, as the selection ratio decreases, the feature dimension becomes smaller. The sparsity of short texts causes a downward trend in classification accuracy. In the experiment, the feature selection ratio R is set to 1/60.

D. Concept Drift Detection

To verify the effectiveness of the concept drift detection method in this paper, three statistical evaluation indexes are used. False alarms: the error probability of concept drift detection; Missing: the probability that there is concept drift but not detected; Delay: the average number of short texts that delay detection of concept drift when concept drift occurs. Table 3 shows the evaluation results of the concept drift detection method proposed in this paper and the benchmark concept detection method. The experimental results show that this paper's concept drift detection method can effectively detect concept drift.

TABLE 1
Data Sets.

Data set	Field	Quantity	Total
Snippets	Business	1500	12340
	Computer	1420	
	Culture-arts-ent	2210	
	Education-science	2660	
	Engineering	370	
	Health	1180	
	Politics-socienty	1500	
	Sport	1420	
News	Sport	8190	32604
	Business	5367	
	U.S.	4783	
	Health	1851	
	Sci&tech	2872	
	World	6255	
	Enterainment	3286	
Tweets	Arsenal	276744	803613
	Blackfriday	34481	
	Chelsea	340194	
	Smartphone	152194	

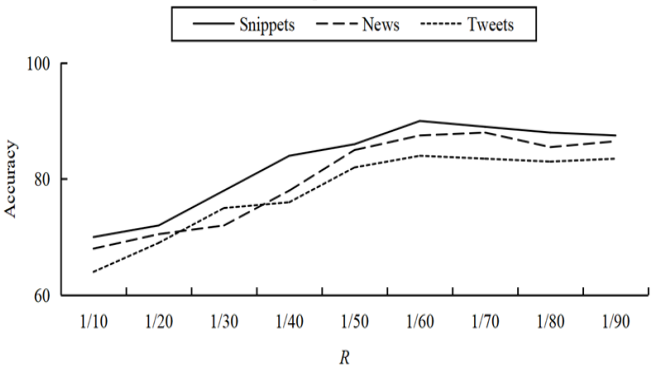

Fig. 2. The curve of classification accuracy with parameter *R*.

TABLE 2
BENCHMARK ALGORITHMS.

Category	Method	Description
Data Stream Classification	Naive Bayes	Data stream classification method based on single-sample Bayesian model
	Spegasos	Spegasos method based on random variables
	KNN+PAW+ADWIN	K-nearest neighbor method based on probability approximation window and adaptive sliding window
Short Text Data Classification	S-SVM	Support vector machine method based on self-information
	MaxEnt	Classification method based on maximum entropy
	INPath	Classification method based on semi-supervised learning
Concept Drift Detection	EDDM	Early concept drift detection method
	ADWINChangeDetector	Concept drift detection method based on adaptive sliding window
	HDDM-A-Test	On-line concept drift detection method based on Hoeffding boundary
	HDDM-W-Test	On-line concept drift detection method based on McDiarmid boundary

TABLE 3
STATISTICS OF CONCEPT DRIFT DETECTION IN DATA SETS.

Measure	EDDM	ADWIN	HDDM-A-Test	HDDM-W-Test	Ours
			Snippets		
False Alarms	4.16	13.64	8.53	5.42	1.78
Missing	16.52	14.53	7.52	9.26	7.24
Delay	28.34	23.42	9.84	11.73	5.92
			News		
False Alarms	6.98	12.35	7.59	7.64	6.75
Missing	36.42	28.54	16.36	16.23	6.42
Delay	28.56	20.36	9.87	10.21	3.14
			Tweets		
False Alarms	5.38	9.82	7.85	8.14	4.92
Missing	32.41	18.93	34.36	32.52	9.64
Delay	67.52	62.46	82.24	78.43	1.86

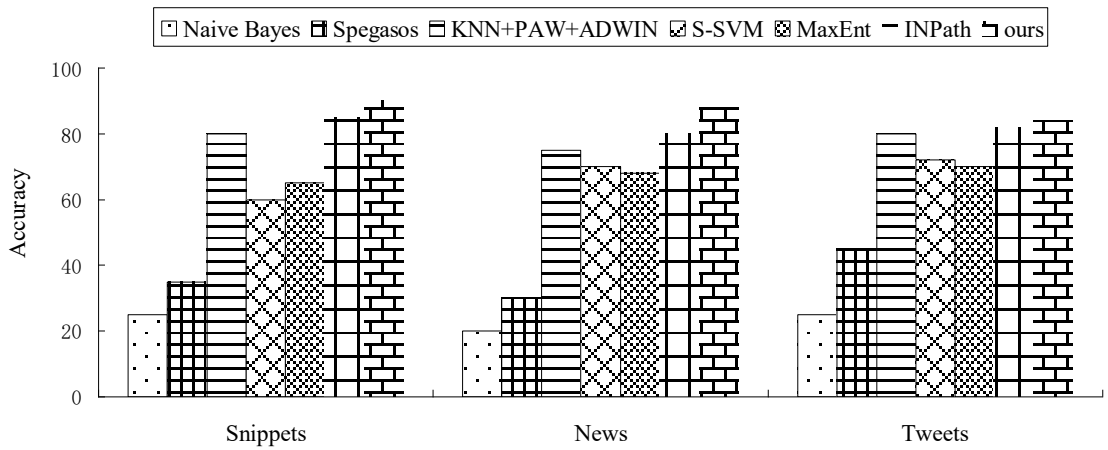

Fig. 3. Classification accuracy on data sets.

E. Classification Performance

Figure 3 shows the classification accuracy of the algorithm proposed in this paper and the benchmark algorithm on the three data sets. The algorithm proposed in this paper has achieved a certain degree of improvement on the data sets *Snippets*, *News* and *Tweets*. It shows that the algorithm proposed in this paper is effective for short text data stream classification.

When concept drift occurs in the short text data stream, the algorithm uses multiple different data blocks to construct multiple base classifiers, containing more concept drifts that have occurred, which can more effectively adapt to the concept drifts. The algorithm proposed in this paper alleviates the high-dimensional and sparse problems of short text data streams with feature extension, adopts the semantic distance between data blocks to judge whether concept drift occurs, and uses the latest data blocks to update the

ensemble model. Therefore, the method proposed in this paper is more suitable for the classification of short text data streams and has strong robustness and better classification performance.

V. Conclusions

This paper presents a new short text data stream classification algorithm. For the high-dimension and sparseness of short text, the feature space selected in this paper takes into account the difference in data distribution on the data set and reduces the dimension of the feature space, thereby alleviating the classification impact caused by sparse features and weak concept description. To effectively detect and adapt to the concept drifts in short text data streams, a concept drift detection method is also used based on topic distribution. By accumulating the semantic distance between each short text in the new and current data block, it is judged whether concept drift occurs in the new data block. The ensemble classification model is updated according to whether the concept drift has occurred. Experimental results show that the classification performance of the short text data stream classification algorithm proposed in this paper is more effective than other algorithms on the same data sets.

References

[1] A.M. Khedr, "EDCP: Effective Decomposable Closest Pair Algorithm for Distributed Databases," Engineering Letters, vol. 28, no.3, pp. 930-938, 2020

[2] Y. Yan, L. Zhang, T. Feng, P. Xie, and G. Xin, "Location Big Data Partition and Publishing Method Based on Sampling and Adjustment," Engineering Letters, vol. 28, no.2, pp. 280-289, 2020

[3] B. Amel, "Managing Early Aspects Interaction. Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2019", 22-24 October, San Francisco, USA, pp. 100-106, 2019.

[4] C. D. Molina-Machado, J. D. Martinez-Vargas, E. Giraldo, "Dynamic State Estimation for Large Scale Systems Based on a Parallel Proximal Algorithm," Engineering Letters, vol. 28, no.2, pp. 347-351, 2020

[5] Q. Yuan, G. Cong, and N. M. Thalmann, "Enhancing Naive Bayes with Various Smoothing Methods for Short Text Classification," In: Proceedings of the Web Conference, pp. 645-646, 2012.

[6] P. Wang, J. Xu, B. Xu, C. Liu, H. Zhang, F. Wang, and H. Hao, "Semantic Clustering and Convolutional Neural Network for Short Text Categorization," In: Proceedings of International Joint Conference on Natural Language Processing, pp. 352-357, 2015.

[7] L.W. Gao, S.G. Zhou, and J.H. Guan, "Effectively Classifying Short Texts by Structured Sparse Representation with Dictionary Filtering," Information Sciences, pp. 130-142, 2015.

[8] M. Haddoud, A. Mokhtari, and T. Lecroq, "Combining Supervised Term-weighting Metrics for SVM Text Classification with Extended Term Representation," Knowledge and Information Systems, vol. 49, no.3, pp. 909-931, 2016.

[9] P.V. Bicalho, M. Pita, and G. Pedrosa., "A General Framework to Expand Short Text for Topic Modeling," Information Sciences, pp. 66-81, 2017.

[10] N.D. Doulamis, A.D. Doulamis, and P.C. Kokkinos, "Event Detection in Twitter Microblogging," IEEE Transactions on Systems, Man, and Cybernetics, vol. 46, no.12, pp. 2810-2824, 2016.

[11] J. Flisar, V. Podgorelec, "Improving Short Text Classification using Information from DBpedia Ontology," Fundamenta Informaticae, vol. 172, no.3, pp. 261-297, 2020.

[12] Bouaziz A, Costa Pereira C D, Pallez C D, et al. "Interactive Generic Learning Method (IGLM): a new Approach to Interactive Short Text Classification," ACM Symposium on Applied Computing, 2016: 847-852.

[13] Z. Ren, M.H. Peetz, S. Liang, "Hierarchical Multi-Label Classification of Social Text Streams," International ACM Sigir Conference on Research and Development in Information Retrieval, pp. 213-222, 2014.

[14] P.P. Li, L. He, and H.Y. Wang, "Learning from Short Text Streams with Topic Drifts," IEEE Transactions on Systems, vol. 48, no.9, pp. 2697-2711, 2018.

[15] J. Gama, P. Medas, and G. Castillo, "Learning with Drift Detection," Brazilian Symposium on Artificial Intelligence, pp. 286-295, 2004.

[16] B. Kurlej and M. Wozniak, "Active Learning Approach to Concept Drift Problem," Logic Journal of IGPL, vol. 20, no.3, pp. 550-559, 2012.

[17] J. Sun and H. Li, "Dynamic Financial Distress Prediction using Instance Selection for the Disposal of Concept Drift.," Expert Systems with Applications, vol. 38, no.3, pp. 2566-2576, 2011.

[18] I. Hegedűs, R. Ormandi, and M. Jelasity, "Massively Distributed Concept Drift Handling in Large Networks," Advances in Complex Systems, vol. 16, no.4, pp.1-28, 2013.

[19] P. Li, X. Wu, and X. Hu, "Learning Concept-drifting Data Streams with Random Ensemble Decision Trees," Neurocomputing, pp. 68-83, 2015.

[20] P. Loeffel, C. Marsala, and M. Detyniecki, "Classification with a Reject Option under Concept Drift: The Droplets Algorithm," In: Proceedings of IEEE International Conference on Data Science and Advanced Analytics, pp. 1-9, 2015.

[21] C. Chen, Y. Li, and T. Hong, "A GA-based Approach for Mining Membership Functions and Concept-drift Patterns," In: Proceedings of Congress on Evolutionary Computation, pp. 2961-2965, 2015.

[22] B. Mirza, Z. Lin, and N. Liu, "Ensemble of Subset Online Sequential Extreme Learning Machine for Class Imbalance and Concept Drift," Neurocomputing, pp. 316-329, 2015.

[23] T.S. Sethi and M. Kantardzic, "On the Reliable Detection of Concept Drift from Streaming Unlabeled Data," Expert Systems with Applications, pp. 77-99, 2017.

[24] M. Tennant, F.T. Stahl, and O.F. Rana, "Scalable Real-time Classification of Data Streams with Concept Drift," Future Generation Computer Systems, pp. 187-199, 2017.

[25] J.D. Silva, E.R. Hruschka, and J. Gama, "An Evolutionary Algorithm for Clustering Data Streams with a Variable Number of Clusters," Expert Systems with Applications, pp. 228-238, 2017.

[26] D.Y. Deng, K.W. Lu, and H.K. Huang, "Attribute Reduction for Concepts and Concept Drifting Detection in Heterogeneous Data," Chinese Journal of Electronics, vol. 46, no.5, pp.1234-1239, 2018.

[27] A. Bifet and R. Gavalda, "Learning from Time-Changing Data with Adaptive Windowing," In: Proceedings of SIAM International Conference on Data Mining, pp 443-448, 2007.

[28] C.Q. Yang, "Research on Short Text Classification based on its Own Features," Hefei: Hefei University of Technology, 2016.

[29] H.Y. Wang, "Research on the Short Text Stream Classification based on the Corpuse Extension from Wikipedia," Hefei: Hefei University of Technology, 2019.

[30] S. Shalevshwartz, Y. Singer, and N. Srebro, "Pegasos: Primal Estimated Sub-Gradient Solver for SVM," Mathematical Programming, vol. 127, no.1, pp. 3-30, 2011.

[31] I. Frias-Blanco and J.D. Campo-Avila, "Ramos-Jimenez G, et al. Online and Non-Parametric Drift Detection Methods based on Hoeffding's Bounds," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no.3, pp. 810-823, 2015.