International Journal of Innovative Computing, Information and Control
Volume **18**, Number **4**, August **2022**
ICIC International ©2022 ISSN 1349-4198
pp. 1265–1274

AN EXTENSION OF LABEL-ORIENTED APPROACH FOR MULTI-LABELS TEXT CLASSIFICATION

Manh Hung Nguyen

Department of Information Technology
Posts and Telecommunications Institute of Technology (PTIT)
No. 122, Hoang Quoc Viet Street, Cau Giay District, Hanoi 084, Vietnam
mhnguyen@ptit.edu.vn; nmh.nguyenmanhhung@gmail.com

Received January 2022; revised April 2022

Abstract. *Most of current text classification approaches are to classify the single-label text. These approaches are suitable and efficient to many application forms: detecting the category, the sentiment or the value of text. However, they do not support very well to the multi-labels text classification problem, such as detecting the emotions in a text, and recognizing the objects which are mentioned in a text. This paper extends the Label-Oriented (LO) approach for multi-label text classification problem, in which, a new feature of label-oriented score at the label level is added, and the label-oriented score threshold for each label is also calculated at the end of the training phase. The testing phase then calculates the label-oriented score for each considered text to compare to the threshold to detect the text is labeled to the label or not. By this way, a text may be assigned to many labels. The extended model is then evaluated in a new collected dataset of multi-label texts MEMO. The experiment results indicate that the proposed model is better than some baseline models on the considered datasets.*

Keywords: Machine learning, Supervised machine learning, Label-oriented scored, Multiple labels, Text classification

1. **Introduction.** Most of current text classification approaches are to classify the single-label text: a text belongs to only one label class. These approaches are suitable and efficient to many application forms: detecting the category, the sentiment or the value of text. However, they do not support very well to the multi-labels text classification problem, in which, a text may belong to more than one label class, such as detecting the emotions in a text, and recognizing the objects which are mentioned in a text.

Actually, we could use single-label text classifier in the context of multi-label text classification by repeating the training-testing cycle n times (n is the number of considered labels). At the time k, classify only two classes: the kth label and the rest. However, this may take longer time of re-training n time on the same training set.

The problem of text classification is already popular in the machine learning community, in which, a set of texts which are already classified with a label class, called *training set*, will be used to extract some common features of texts of the same label. If there is a new text t, the assignment of a label to the text t is based on the relationship among the text itself and the texts in the training set.

On the line from the *Label-Oriented* (LO) approach [24] which takes account of the relationship between the text and its assigned label, this paper extends this idea to classify the multi-label texts on following aspects.

DOI: 10.24507/ijicic.18.04.1265

- First, in the training phase, the model of [24] uses only one feature, that is the *label-oriented* score at the *text level*. This extension adds one more feature: the *label-oriented* score at the *label level*.
- Second, at the end of the training phase, this extension calculates the score threshold of each label.
- Third, in the classifying phase, this extension assigns any label to a text if the *label-oriented* score of the text is bigger than the threshold score. This extension enables to assign more than one label to a text. That means a label may have more than one label.

The extended model is then evaluated in a new collected dataset of MEMO – a multi-emotions dataset.

The paper is organized as follows. Section 2 presents the related works to this paper. Section 3 presents the model extended from the label-oriented approach. Section 4 presents the evaluation of the proposed model. Section 5 makes the conclusion.

2. **Related Works.** The problem of text classification has been attracting many researches. There are many approaches proposed. Let us consider these works in the last decade on three technical aspects (Table 1).

First, the technical feature of texts. Almost works use either semantic-based approaches or statistic-based approaches to extract the feature vector of texts. The semantic-based

TABLE 1. Summary of related works in the last decade

Model	**Feature + Classifier**	**Label**
Chen et al. [5]	CNN+	multi
Devlin et al. [6]	statistic + BERT	single
Erkan et al. [7]	statistic + Harmonic function	single
Glinka and Zakrzewska [8]	Labels Chain (LC)	multi
Goldberg et al. [9]	crowd-sourcing	single
Hendry et al. [10]	CNN+	multi
Joulin et al. [12]	statistic + FastText	single
Kim [13]	CNN+	single
Lau et al. [14]	re-ranking	single
Li et al. [16]	joint learning	single
Li et al. [18]	semantic + CNN+	multi
Mikolov et al. [19]	text-to-vector	single
Nam et al. [22]	CNN+	multi
Nguyen [24]	statistic + Label-Oriented (LO) score	single
Peters et al. [27]	statistic + ELMo	single
Read et al. [29]	binary relevance-based	single
Rossi et al. [30]	statistic + NB	single
Wang and Chiang [33]	SVM	multi
Wei et al. [34]	NB	multi
Yang et al. [35]	semantic	single
Yu et al. [36]	squared loss function	single
Yu et al. [37]	neighborhood rough sets	single
Zhang et al. [38]	CNN+	multi
Zhou and El-Gohary [39]	semantic	single
Our work	statistic + Label-Oriented (LO) based	multi

approaches try to extract the meaning of words, sentences in texts to measure the similarity among them (Suryapranata et al. [32], Li et al. [17], Nguyen and Tran [25], Zhou and El-Gohary [39], Yang et al. [35]). Meanwhile, the statistic-based approaches try to extract the statistical features of texts to estimate the distance among texts, mostly based on *Term Frequency-Inverse Document Frequency (TF-IDF)* [31] feature (Erkan et al. [7], Rossi et al. [30], Joulin et al. [12], Al-Anzi et al. [1]).

Second, the classifier used in these works. There are also two main tendencies: using classical classifier-based or defining their own classifier. In the first tendency, any of (or extended of any) classical classifiers could be used. For instance, Naïve Bayes (NB) is used by Rossi et al. [30], Wei et al. [34], Support Vector Machine (SVM) is used by Bekkerman et al. [2], Wang and Chiang [33].

In the second tendency, some authors could propose their own classifier, for instance, the best topic word selection and re-ranking model of Lau et al. [14]; crowd-sourcing techniques of Goldberg et al. [9]; neighborhood rough sets of Yu et al. [37]; binary relevance-based methods of Read et al. [29]; joint learning algorithm of Li et al. [16]; squared loss function of Yu et al. [36]; Labels Chain (LC) algorithm of Glinka and Zakrzewska [8]; text-to-vector of Mikolov et al. [19]; FastText of Joulin et al. [12]; convolutional and recurrent neural networks (called CNN+ genre) of Kim [13], Chen et al. [5], Zhang et al. [38]; ELMo model of Peters et al. [27]; BERT model of Devlin et al. [6]; *Label-Oriented (LO) score* approach of Nguyen [24].

Regarding the number of labels that models assign to a text, there are also two main tendencies: single-label and multi-label. Most of classical approaches are single-label. Meanwhile, multi-label classification models could be applied to the case of single-label, for instance, the works of Chen et al. [5], Glinka and Zakrzewska [8], Nam et al. [22], Wang and Chiang [33], and Wei et al. [34].

This work extents the work from Nguyen [24] to modify the Label-Oriented (LO) score model to apply in the case of multi-labels text classification: add a new feature of *label-oriented* score at the *label level*, and assign more than one label to a text as long as the score regarding those labels is bigger than the threshold score.

3. **Proposed Model.** This model extents the Label-Oriented (LO) score model [24] to apply in the case of multi-labels text classification. The algorithm is thus composed of two phases: training phase and testing phase.

3.1. **Training phase.** This phase is composed of four steps, in which, the first step is inherited from the the Label-Oriented (LO) score model [24]. The last three steps are a new extension in this work.

3.1.1. *Step 1: Calculation of label-oriented score at the text level.* Given $T=\{t_1, t_2, \ldots, t_n\}$ the set of texts in the training set; $L = \{l_1, l_2, \ldots, l_m\}$ the set of considered labels, in which, each text in T may be associated to several labels in L. The objective of this phase is to calculate the *label-oriented score* matrix where $score(x, l)$ represents the *label-oriented score* of the term $x \in V$ – V is the union of all terms in all texts in the set T – regarding the label $l \in L$. The matrix of *label-oriented score* is calculated as follows.

- For each text $t_i \in T$, split t_i into a set G_i of term (stop words could be removed).
- Take the union V of all terms in all texts in the set T.
- Calculate the Term Frequence (TF) [31] $tf(x, t)$ of the term $x \in V$ in the document $t \in T$.
- For each label $l_j \in L$, create two sets of text sample:
  - T_{l_j} is the set of all texts which are assigned to the label l_j.
  - $T_{\neg l_j}$ is the set of all texts which are not assigned to the label l_j.

- For each term $x \in V$, calculate the *Label-Oriented Term (LOT) score* of each term in the corresponding set for each label l_j as follows.
  - Calculate the *text frequency* $f_{txt}\left(x, T_{l_j}\right)$ and $f_{txt}\left(x, T_{\neg l_j}\right)$ of the term x in the two sets T_{l_j} and $T_{\neg l_j}$, respectively.

$$f_{txt}(x,T) = \begin{cases} 0 & \text{if } |T| = 0 \\ \dfrac{|\{t \in T : x \in t\}|}{|T|} & \text{otherwise} \end{cases} \tag{1}$$

  - Calculate the *Label-Oriented (LO) score* of a term x at the text level, regarding a label l, by the following formula:

$$s_{txt}^{LO}(x,l) = \begin{cases} 0 & \text{if } f_{txt}(x, T_l \cup T_{\neg l}) = 0 \\ \dfrac{2 * f_{txt}(x, T_l)}{f_{txt}(x, T_l) + f_{txt}\left(x, T_{\neg l}\right)} & \text{otherwise} \end{cases} \tag{2}$$

  - Calculate the *Label-Unoriented (LU) score* of a term x at the text level, regarding a label l, by the following formula:

$$s_{txt}^{LU}(x,l) = \begin{cases} 0 & \text{if } f_{txt}(x, T_l \cup T_{\neg l}) = 0 \\ \dfrac{2 * f_{txt}(x, T_{\neg l})}{f_{txt}(x, T_l) + f_{txt}(x, T_{\neg l})} & \text{otherwise} \end{cases} \tag{3}$$

3.1.2. *Step 2: Calculation of label-oriented score at the label level.* At the label level, we could consider the two sets T_l and $T_{\neg l}$ as two sets of terms T'_l and $T'_{\neg l}$ by splitting each text in each set into terms. From that point of view, this model counts the frequency of a term x in each set T'_l and $T'_{\neg l}$. It is obvious that if the frequency of the term x in the set T'_l is higher than that in the set $T'_{\neg l}$, then the possibility that a text containing the term x will belong to the set T_l may be higher than the possibility that the text belongs to the set $T_{\neg l}$; and vise versa, if the frequency of the term x in T'_l is lower than that in $T'_{\neg l}$, then the possibility that a text containing the term x will belong to the set T_l may be lower than the possibility that the text belongs to the set $T_{\neg l}$.

Accordingly, this fact is taken into account as follows.

- Calculate the *frequency at the label level* $f_{lbl}\left(x, T'_{l_j}\right)$ and $f_{lbl}\left(x, T'_{\neg l_j}\right)$ of the term x in the two sets T'_{l_j} and $T'_{\neg l_j}$, respectively.

$$f_{lbl}\left(x, T'\right) = \begin{cases} 0 & \text{if } x \notin T' \\ \dfrac{\text{number of appearance of } x \text{ in } T'}{|T'|} & \text{otherwise} \end{cases} \tag{4}$$

- Calculate the *Label-Oriented (LO) score* of a term x at the label level, regarding a label l, by the following formula:

$$s_{lbl}^{LO}(x,l) = \begin{cases} 0 & \text{if } f_{lbl}(x, T_l \cup T_{\neg l}) = 0 \\ \dfrac{2 * f_{lbl}(x, T_l)}{f_{lbl}(x, T_l) + f_{lbl}(x, T_{\neg l})} & \text{otherwise} \end{cases} \tag{5}$$

- Calculate the *Label-Unoriented (LU) score* of a term x at the label level, regarding a label l, by the following formula:

$$s_{lbl}^{LU}(x,l) = \begin{cases} 0 & \text{if } f_{lbl}(x, T_l \cup T_{\neg l}) = 0 \\ \dfrac{2 * f_{lbl}(x, T_{\neg l})}{f_{lbl}(x, T_l) + f_{lbl}(x, T_{\neg l})} & \text{otherwise} \end{cases} \tag{6}$$

3.1.3. *Step 3: Combination of two features.* This step combines the score at the text level (step 1) and that at the label level (step 2) into a unified score.

- Calculate the *final score* of a term x regarding a label l by the following formula:

$$score(x,l) = \left(s_{txt}^{LO}(x,l) + s_{lbl}^{LO}(x,l)\right) - \left(s_{txt}^{LU}(x,l) + s_{lbl}^{LU}(x,l)\right) \tag{7}$$

At the end of this step, the *label-oriented* score matrix S is obtained: the value of $S[i][j]$ represents the *label-oriented* score $score(x_i, l_j)$ of the term $x_i \in V$ regarding the label $l_j \in L$.

3.1.4. *Step 4: Calculation of threshold score for each label.* The objective of this step is to determine the threshold score of each label. Theoretically, it may be 0. However, in general case, it may be not. So, determining the best threshold score for a label could help better separate if a text belongs or does not belong to the label.

- Calculate the label-oriented central point of all texts in the training set which belong to the label $l_i \in L$:

$$cp^+(l_i) = \frac{1}{|T_{l_i}|} \sum_{t_j \in T_{l_i}} \sum_{x \in t_j} tf(x,t_j) * score(x,l_i) \tag{8}$$

- Calculate the label-unoriented central point of all texts in the training set which do not belong to the label $l_i \in L$:

$$cp^-(l_i) = \frac{1}{|T_{\neg l_i}|} \sum_{t_j \in T_{\neg l_i}} \sum_{x \in t_j} tf(x,t_j) * score(x,l_i) \tag{9}$$

- The threshold score of the label $l_i \in L$ is the middle point of the two central points:

$$threshold(l_i) = \frac{cp^+(l_i) + cp^-(l_i)}{2} \tag{10}$$

At the end of this step, each label $l \in L$ is associated to its threshold score $threshold(l)$.

3.2. **Classifying phase.** For a new text t, the choice of label to assign to the text is as follows.

- Split t into a set of terms $X = (x_1, x_2, \ldots, x_n)$.
- Calculate the *TF value* $tf(x_i, t)$ for each term x_i in the text t.
- For each label $l_i \in L$:
  - The *Label-Oriented Document (LOD) score* of a text t for label l is calculated as follows:

$$score_{doc}(t,l_i) = \sum_{x \in t} tf(x,t) * score(x,l_i) \tag{11}$$

  - If $score_{doc}(t,l_i) > threshold(l_i)$, then the text t will be assigned to the corresponding label l_i. This is a new extension from the model [24]: it enables to detect more than one label of a text if the text could have.

4. **Evaluation.** This section presents an experiment to evaluate the extended model in comparing to some related models on a dataset.

4.1. **Dataset.** Actually, there are many datasets for text classification which have been proposed, specially, in the field of emotion classification in text, for instance, EmoLex [21], and Semeval2017 [20]. However, most of them are single-label datasets, in which, a text is assigned to only one emotion. Consequently, they could not be used in the problem of multi-label classification. There is one dataset, Brat data [3], which is a multi-label dataset. However, this dataset is too small at the number of samples, as well as too small at the number of samples in each emotion label.

In order to evaluate the proposed model in a multi-label dataset, we build MEMO – Multiple Emotions in Texts: we collected several texts from status on social networks such as Twitter from 2018 to 2020. These texts are then labelled with emotions. The selection of emotions is mainly based on the cognitive definition of Ortony et al. [26], and the cognitive pattern of emotion of Lazarus [15].

In the datasets, there are about 4000 samples. The distribution of samples on each label is presented in Table 3. Each label has about 500 samples. The distribution of samples on the number of labels for each sample is presented in Table 2. There are more than 1000 text samples which have more than one label.

Table 2. Distribution of emotion number in the datasets

Number of labels	**Number of samples**
1 label	2683
2 labels	1031
3 labels	237
4 labels	45
5 labels	4
All	**4000**

Table 3. Distribution of emotion type in the datasets

Label	**Number of samples**
Joy	534
Sadness	574
Hope	536
Fear	513
Love	597
Disgust	551
Pride	517
Admiration	521
Anger	517
Other	796
All	**5656**

4.2. **Scenario.** The 10-folds cross validation method is used in this experiment: The dataset is randomly divided into 10 folds. Each time, the kth fold is considered as the testing set, and the 9 remain folds are regrouped into the training set. The proposed model will be compared to the following models.

- Some classical classifiers such as Support Vector Machine (SVM) [4], Naïve Bayes (NB) [11], and Decision Tree C4.5 (J48) [28]. As these classical classifiers are to classify single-label texts, we have to repeat the training-testing process in several times: At the time k, classify only two classes: the kth label and the rest. If the model indicates that the text belongs to the current considered label, then the label will be assigned to the text. By repeating this process for all considered labels, a text may be assigned to more than one label.
- Some recent model of CNN (Convolution Neuron Network) genre: model ELMo of Peters et al. [27], and model BERT of Devlin et al. [6].

4.3. **Output parameters.** Let O_i be the original label set of the text i, and E_i the extracted label set of the text i. And $C_i = O_i \cap E_i$ is the intersection set of O_i and E_i. These following parameters (Nguyen [23] based on Salton and McGill [31]) are used.

- The *precision* on a text i is

$$p_i = \frac{|C_i|}{|E_i|} * 100\% \tag{12}$$

- The *precision* on n texts in the testing set is

$$Precision = \frac{\sum_{i=1}^{n} p_i}{n} = \frac{1}{n} * \sum_{i=1}^{n} \frac{|C_i|}{|E_i|} * 100\% \tag{13}$$

- The *recall* on a text i is

$$r_i = \frac{|C_i|}{|O_i|} * 100\% \tag{14}$$

- The *recall* on n texts in the testing set is

$$Recall = \frac{\sum_{i=1}^{n} r_i}{n} = \frac{1}{n} * \sum_{i=1}^{n} \frac{|C_i|}{|O_i|} * 100\% \tag{15}$$

- The *F1-score* on all texts of the testing set is

$$F1\text{-}score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{16}$$

In this experiment, three following output parameters are observed: *Precision*, *Recall*, and *F1-score*.

4.4. **Results.** The results are presented in Table 4. Considering among compared models, the Decision Tree C4.5 brings the best performance among classical classifiers. Meanwhile, the two CNN models bring a little better result than that of classical classifiers. The extended model in this paper brings a little better result than that of CNN models.

TABLE 4. Comparison results among considered models

Model	Precision (%)	Recall (%)	F1-score (%)
Naïve Bayes [11]	45.98	61.03	52.44
Support Vector Machine [4]	59.72	56.46	58.04
C4.5 [28]	62.53	62.07	62.29
ELMo [27]	63.73	62.63	63.18
BERT [6]	64.53	64.05	64.29
Extension of LO (ours)	**65.19**	**64.70**	**64.94**

In spite of the advance of the proposed model regarding the considered works, the results generally indicate that the *F1-score* observed from considered models are not too high. There are probably three reasons for that.

First, it is because of the Twitter text characteristics: the texts are much shorter than regular text in other datasets. Therefore, the statistical approach may not be very well if the text sample is too short.

Second, the text is not in the well format and correct grammar as in the regular text. This fact may probably influence the final results.

Third, it may be the critical reason: the samples are not easy to classify. Let us consider five samples in Table 5 regarding the label *love*. The first sample is assigned to the label *love* because its verb is love, that seems to be easy. However, the second and the third

TABLE 5. Some samples in the dataset

No.	Text	Labels
1	I am in love with that color!	love
2	I want to be the person you are scared to lose.	love, hope
3	I am scared of losing you, but you are not even mine.	fear, love, sadness
4	I hope I die while I am in love, not because of love.	hope
5	Some people are meant to be in love, just not meant to be together.	other

samples have no word related to the keyword *love*, but they also imply the emotion of *love* inside. They are not easy to be classified. Moreover, the fourth and the fifth samples have the keyword *love* but they have no *love* emotion. They are not easy to classify even for human. The difficulty of the dataset may be one of reasons that the model does not get better results.

5. **Conclusion.** This paper presents an extension of *Label-Oriented* (LO) model for multi-label text classification, in which, the training phase objective is to calculate a matrix of *label-oriented score* for all terms (the first dimension) in training set for all considered labels (the second dimension) by combining two features: the *label-oriented* score at the *text level* and at the *label level*. In the testing phase, the sum of this *label-oriented score* of all terms in a text (repeating for each label) is compared to the threshold score of the label to determine whether the text should be assigned to the label or not by comparing it to the threshold score of each label. The extended model is then evaluated in a new collected dataset of MEMO – a multi-emotions dataset. The experiment results indicate that the extended model is better than some considered models.

Evaluating this model on longer texts and/or on more labels is one of our perspectives in the near future.

REFERENCES

[1] F. S. Al-Anzi, D. A. Zeina and S. Hasan, Utilizing standard deviation in text classification weighting schemes, *International Journal of Innovative Computing, Information and Control*, vol.13, no.4, pp.1385-1398, 2017.
[2] R. Bekkerman, R. El-Yaniv, N. Tishby and Y. Winter, On feature distributional clustering for text categorization, *Proc. of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'01)*, New York, NY, USA, pp.146-153, 2001.
[3] L. Buitinck, J. van Amerongen, E. Tan and M. de Rijke, Multi-emotion detection in user-generated reviews, *Proc. of European Conference on Information Retrieval (ECIR)*, 2015.
[4] C. J. C. Burges, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, vol.2, no.2, pp.121-167, 1998.
[5] G. Chen, D. Ye, Z. Xing, J. Chen and E. Cambria, Ensemble application of convolutional and recurrent neural networks for multi-label text categorization, *2017 International Joint Conference on Neural Networks (IJCNN2017)*, Anchorage, AK, USA, pp.2377-2383, 2017.
[6] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, *CoRR*, abs/1810.04805, 2018.
[7] G. Erkan, A. Hassan, Q. Diao and D. R. Radev, *Improved Nearest Neighbor Methods for Text Classification*, Technical Report, Technical Report CSE-TR-576-11, Department of Electrical Engineering and Computer Science, University of Michigan, 2011.
[8] K. Glinka and D. Zakrzewska, Effective multi-label classification method for multidimensional datasets, in *Flexible Query Answering Systems 2015. Advances in Intelligent Systems and Computing*, T. Andreasen, H. Christiansen, J. Kacprzyk et al. (eds.), Cham, Springer International Publishing, 2016.

[9] S. L. Goldberg, D. Z. Wang and T. Kraska, CASTLE: Crowd-assisted system for text labeling and extraction, *The 1st AAAI Conference on Human Computation and Crowdsourcing*, 2013.

[10] Hendry, R.-C. Chen, L.-H. Li and Q. Zhao, Using deep learning to learn user rating from user comments, *International Journal of Innovative Computing, Information and Control*, vol.14, no.3, pp.1141-1149, 2018.

[11] G. H. John and P. Langley, Estimating continuous distributions in Bayesian classifiers, *The 11th Conference on Uncertainty in Artificial Intelligence*, San Mateo, pp.338-345, 1995.

[12] A. Joulin, E. Grave, P. Bojanowski and T. Mikolov, Bag of tricks for efficient text classification, *Proc. of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Valencia, Spain, pp.427-431, 2017.

[13] Y. Kim, Convolutional neural networks for sentence classification, *Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.1746-1751, 2014.

[14] J. H. Lau, D. Newman, S. Karimi and T. Baldwin, Best topic word selection for topic labelling, *Proc. of the 23rd International Conference on Computational Linguistics: Posters (COLING'10)*, Stroudsburg, PA, USA, pp.605-613, 2010.

[15] R. S. Lazarus, *Emotion & Adaptation*, Oxford University Press, 1991.

[16] L. Li, B. Chang, S. Zhao, L. Sha, X. Sun and H. Wang, Multi-label text categorization with joint learning predictions-as-features method, *Proc. of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.835-839, 2015.

[17] X. Li, N. Liu, C. Yao and F. Fan, Text similarity measurement with semantic analysis, *International Journal of Innovative Computing, Information and Control*, vol.13, no.5, pp.1693-1708, 2017.

[18] X. Li, C. Yao, Q. Zhang and G. Zhang, Semantic similarity modeling based on multi-granularity interaction matching, *International Journal of Innovative Computing, Information and Control*, vol.15, no.5, pp.1685-1700, 2019.

[19] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, Distributed representations of words and phrases and their compositionality, *Proc. of the 26th International Conference on Neural Information Processing Systems (NIPS'13)*, USA, pp.3111-3119, 2013.

[20] S. M. Mohammad and F. Bravo-Marquez, Emotion intensities in Tweets, *Proc. of the 6th Joint Conference on Lexical and Computational Semantics (*Sem)*, Vancouver, Canada, 2017.

[21] S. M. Mohammad and S. Kiritchenko, Using hashtags to capture fine emotion categories from tweets, *Computational Intelligence*, vol.31, no.2, pp.301-326, 2015.

[22] J. Nam, J. Kim, I. Gurevych and J. Fürnkranz, Large-scale multi-label text classification – Revisiting neural networks, *CoRR*, abs/1312.5419, 2013.

[23] M. H. Nguyen, On the distinction of subjectivity and objectivity of emotions in texts, *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol.9, no.9, pp.584-589, 2018.

[24] M. H. Nguyen, A label-oriented approach for text classification, *International Journal of Innovative Computing, Information and Control*, vol.16, no.5, pp.1593-1609, 2020.

[25] M. H. Nguyen and D. Q. Tran, Estimation in semantic similarity of texts, *Journal of Information Science and Engineering*, vol.37, no.3, pp.617-633, 2021.

[26] A. Ortony, G. L. Clore and A. Collins, *The Congnitive Structure of Emotions*, The Cambridge University Press, 1988.

[27] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee and L. Zettlemoyer, Deep contextualized word representations, *Proc. of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, New Orleans, LA, pp.2227-2237, 2018.

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

[29] J. Read, B. Pfahringer, G. Holmes and E. Frank, Classifier chains for multi-label classification, *Mach. Learn.*, vol.85, no.3, pp.333-359, 2011.

[30] R. G. Rossi, R. M. Marcacini and S. O. Rezende, *Benchmarking Text Collections for Classification and Clustering Tasks*, Technical Report 395, Institute of Mathematics and Computer Sciences, University of Sao Paulo, 2013.

[31] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, Inc., New York, NY, USA, 1986.

[32] L. K. P. Suryapranata, G. P. Kusuma, Y. Heryadi and B. S. Abbas, Adaptive gamification framework with proper player type classification and effectiveness evaluation, *ICIC Express Letters*, vol.14, no.1, pp.9-14, 2020.

[33] T.-Y. Wang and H.-M. Chiang, Solving multi-label text categorization problem using support vector machine approach with membership function, *Neurocomputing*, vol.74, no.17, pp.3682-3689, 2011.
[34] Z. Wei, H. Zhang, Z. Zhang, W. Li and D. Miao, A Naïve Bayesian multi-label classification algorithm with application to visualize text search results, *International Journal of Advanced Intelligence*, vol.3, no.2, pp.173-188, 2011.
[35] L. Yang, C. Li, Q. Ding and L. Li, Combining lexical and semantic features for short text classification, *Procedia Computer Science*, vol.22, pp.78-86, 2013.
[36] H.-F. Yu, P. Jain, P. Kar and I. S. Dhillon, Large-scale multi-label learning with missing labels, *Proc. of the 31st International Conference on International Conference on Machine Learning (ICML'14)*, Bejing, China, pp.593-601, 2014.
[37] Y. Yu, W. Pedrycz and D. Miao, Multi-label classification by exploiting label correlations, *Expert Syst. Appl.*, vol.41, no.6, pp.2989-3004, 2014.
[38] X. Zhang, J. Zhao and Y. LeCun, Character-level convolutional networks for text classification, *Proc. of the 28th International Conference on Neural Information Processing Systems (NIPS'15)*, Cambridge, MA, USA, pp.649-657, 2015.
[39] P. Zhou and N. El-Gohary, Ontology-based multilabel text classification of construction regulatory documents, *Journal of Computing in Civil Engineering*, vol.30, no.4, 2016.

Author Biography

Manh Hung Nguyen received his Ph.D. in Computer Science at the University of Toulouse III Paul Sabatier, France, in 2010.

Dr. Nguyen is currently an associate professor at the Faculty of Information Technology, Posts and Telecommunications Institute of Technologies (PTIT), Hanoi, Vietnam. His research interests include artificial intelligence, multi-agent system, machine learning, distributed intelligent computing, data mining in social network.