Information Sciences 585 (2022) 41–57

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Three-way multi-granularity learning towards open topic classification

Xin Yang [a,b,*], Yujie Li [a,b], Dan Meng [a,b], Yuxuan Yang [a,b], Dun Liu [c], Tianrui Li [d]

[a] Department of Artificial Intelligence, School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China
[b] Financial Intelligence & Financial Engineering Key Laboratory of Sichuan Province, Southwestern University of Finance and Economics, Chengdu 611130, China
[c] School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China
[d] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

ARTICLE INFO

Article history:
Received 22 July 2021
Received in revised form 7 November 2021
Accepted 12 November 2021
Available online 21 November 2021

Keywords:
Three-way decision
Multi-granularity learning
Open topic
Uncertainty
Knowledge accumulation

ABSTRACT

Traditional topic classification usually adopts the closed-world assumption that all the test topics have been seen in training. However, in open dynamic environments, the potential new topics may appear in testing due to the evolution of text data over time. Considering the uncertainty and multi-granularity of dynamic text data, such open topic classification needs to detect unseen topics by mining the boundary region continually, and incrementally update the previous models by knowledge accumulation. To address these challenge issues, this paper introduces a unified framework of three-way multi-granularity learning to open topic classification based on the fusion of three-way decision and granular computing. First, we propose the multilevel granular structure of tasks from the temporal-spatial multi-granularity perspective. Then, we construct an adaptive decision boundary and use the centroids and the corresponding radius to discover unknowns by the reject option. Subsequently, we further explore the unknown topics by three-way enhanced clustering and the uncertain instances will be re-investigated in the next stage. Besides, we design a built-in knowledge base represented as the centroid of each topic to store the topic knowledge. Finally, the experiments are conducted to compare the performances of proposed models and the efficiency of knowledge accumulation with classic models.

1. Introduction

Traditional supervised learning focuses on deterministic conditions and closed world assumption [32,1], which means that the classes appeared in the test set must be known in the training set [12,3]. However, this assumption may often be violated in real-world applications. For instance, in open-world object recognition, new objects may appear constantly, and a classifier built from old objects may incorrectly classify a new object as one of the old objects [33]. This situation calls models to be more robust and adapt to the open dynamic environment, such as changes of value/attribute and even the appearance of new categories. Hence, this will present more significant challenges and broader application prospects for current machine learning researches and developments. Open dynamic situations call for *on-the-job learning*, as opposed to tra-

* Corresponding author at: Department of Artificial Intelligence, School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China.

E-mail addresses: yangxin@swufe.edu.cn (X. Yang), yujie_li@smail.swufe.edu.cn (Y. Li), mengd_t@swufe.edu.cn (D. Meng), 41810023@smail.swufe.edu.cn (Y. Yang), trli@swjtu.edu.cn (T. Li).

https://doi.org/10.1016/j.ins.2021.11.035
0020-0255/

ditional closed world environment. *On-the-job learning* proposed by Liu [26], which refers to learning after the model has been deployed in an application or during model application. Here, Liu [26] defined an open dynamic system should (1) discover unknowns and create new learning tasks from the unknowns, (2) collect training or ground-truth data through interactions with users and the environment by imitation of humans or other agents, and (3) incrementally learn the new tasks. The whole process is also needed to be carried out *on the fly* in a self-motivated and self-supervised manner.

The text classification by topics is helpful for searching, data mining, and text analysis. However, topic classification is time-consuming and error-prone, especially the open dynamic tasks such as the dialog system and real-time news reports. Most of the existing state-of-the-art methods rely on supervised algorithms with fixed training data and view tasks in isolation rather than looking at such tasks as a whole. The data is constantly changing for open topic classification tasks, so it creates uncertainty in open classes and the learned knowledge. As shown in Fig. 1, we have four available tags (known classes) for specific topics, such as exchange charge, cancel a transfer, pending top-up, and verify identity. However, there are also texts with open/unknown topics. From the perspective of multi-granularity learning, the open topic classification tasks can be divided into multiple granularity levels. In the coarser granularity level, it is necessary to distinguish these texts from the known topics as much as possible, and all these unknown topics will be considered a whole. In the finer granularity, known topics and unknown topics will be further processed according to the current feature space. The processing procedure of open topic classification tasks can be viewed as a granularity construction process that discovers knowledge from coarser granularity to finer granularity. This paper tries to connect three-way multi-granularity learning with open dynamic learning, and then deal with the uncertainty to empower the ability of learning continually in open topic classification.

For open topic classification tasks, three underlying challenges remain to be addressed. First, based on the open-world assumption, the uncertainty in an open dynamic environment needs to be further studied. Second, most of the advanced topic classification systems center on using complex structures to capture the information, which requires a long time to converge during their training stage. And last, a desirable open dynamic model can capture knowledge at different granularity levels to embody the granularity change of dynamic data. To address these issues, we propose **t**hree-**w**ay **m**ulti-**g**ranularity learning towards **open** topic classification (TWMG-Open) model. In this work, we follow the open-world assumption and the framework of three-way multi-granularity learning has been demonstrated as an effective method for open problems. Also, searching for an appropriate granularity level for decision or classification is a crucial problem [19]. This paper deliberates open topic classification with the framework of three-way multi-granularity and constructs decision-making processes according to different granularity levels of dynamic data.

In this work, the open topic classification task has been discussed in light of three-way multi-granularity learning. Besides, it tackles the entire process from open detection/discovery to open classification and considers different granularities for different kinds of problems. By constructing a built-in knowledge base, the ability of continual learning is formed and the three-way multi-granularity learning is utilized to enhance the accuracy and validity of the learned knowledge through knowledge accumulation.

The remainder of this paper is organized as follows. In Section 2, a review of three-way multi-granularity learning and open topic classification are presented. Section 3 constructs a framework of three-way multi-granularity learning towards open topic classification and introduces each part of the proposed model accordingly. Section 4 designs a series of experiments and provides the experimental analysis. Finally, the conclusion is given in Section 5.

2. Related work

2.1. Three-way decision and three-way multi-granularity learning

Three-way decision (3WD) proposed by Yao [38] is initially to describe the three regions of decision-theoretic rough sets, and further be widely investigated as a philosophy of thinking in three, a methodology of working with three, and a mech-

Texts (short)	Coarser Tags	Finer Tags
Could you help me figure out the exchange fee?	Known	Exchange charge
Please help me, there is a transaction in the incorrect account.	Known	Cancel a transfer
I have a pending top-up.	Known	Pending top-up
Please tell me the steps to verify my identity.	Known	Verify identity
…	…	
Can my PIN changed remotely without visiting a bank?	Unknown	Pseudo-label 1
I got an unexpected charge on my transfers. How can I correct this?	Unknown	Pseudo-label 2

Fig. 1. An example of open topic classification. The open topic classification can be divided into multiple granularity levels: the coarser level is to discover open topics, and the finer level is further classifying or clustering according to the current knowledge.

anism of processing through three [42,40]. In the traditional two-way classification models, an object is assigned to only two regions: the positive region for positive instances, and the negative region for negative instances. However, either region may lead to significant errors when the object involves uncertainty. 3WD can enhance binary methods to deal with uncertainty better. Recently, Yao [39] introduced a generalized trisecting-acting-outcome (TAO) model for 3WD. The framework of TAO involves three tasks: (1) construct a trisection from a whole, (2) devise strategies to process the three parts of the trisection, and (3) evaluate the effectiveness of the trisection and strategies [42].

Granular computing (GrC) is a paradigm of information processing [28]. The philosophy of GrC implies two mutually dependent tasks of structured problem solving, namely, constructing a hierarchical view and working with the associated hierarchy. Furthermore, Yao [39] proposed three Ms. of granular: multilevel, multiview and multipurpose based on granules, levels and hierarchies. Three-way granular computing is a paradigm of thinking and information processing in three granules [41]. The granular structure of three-way decision is a combination of different levels, views and purposes. From the perspective of granular computing, the static 3WD models are mono-granular models. Many extensions of three-way multi-granularity learning have been proposed. Yang et al. [36,37] presents a temporality and spatiality framework of three-way decision for hybrid data and discussed the temporality of data and the spatiality of parameters in neighborhood system. Zhang et al. [45] developed the notion of multi-granularity decision-theoretic rough sets into the hesitant fuzzy linguistic background. Qian et al. [31] proposed a generalized model of sequential three-way decisions via multi-granularity. Also, three-way multi-granularity learning have been applied to many machine learning tasks, including face recognition [22,23], sentiment classification [48,49] and attribute reduction [47].

According to the aforementioned analysis, three-way multi-granularity learning is useful to make multiple-stage of decisions when data are increasing over time. At an early stage, due to insufficient information, the majority of people prefer to delay or not to make decisions rather than make decisions immediately. With the accumulation of information, people can then make satisfactory decisions with more confidence. In this study, the three-way multi-granularity learning framework is extended to an open dynamic environment and adopted for a strategy of non-commitment with insufficient information.

2.2. Topic classification task under open dynamic environment

Topic classification is a fundamental task of natural language processing for assigning an article or a document available in text data into at least one predefined label or category according to its content. Classifying texts into topic categories is necessary and important for many applications, such as spam filtering [8], language recognition [4], Tweet topic classification [21], intents classification [6,25], etc. Although those methods have achieved good performance for specific tasks, the open dynamic environment poses higher challenges for current ML. The open dynamic environment means that the application environment may contain unseen (or open) classes that have not appeared in training before. Under this challenging situation, we need open-world learning (or short for open classification) [3]. It can detect (or reject) unseen instances that are not presented during training and update the existing model continually without retraining the entire model. Ideally, open-learning systems should be able to accomplish the following three tasks [13,26]: (1) assign each incoming example to one of the seen classes (appeared in training) and reject those examples from unseen classes (not appeared in training), (2) discover hidden unseen classes in the rejected examples, and (3) learn the new classes incrementally. The first task requires the open-learning systems to detect unseen samples and add them to the reject set R. The second task involves identifying the open classes C in the reject set R independently (or in conjunction with human experience), which is typically called open classes discovery. Lastly, open-learning systems must be able to identify open classes as well as store the acquired knowledge continuously. On the basis of the first two tasks, open-learning systems can construct a knowledge base and learn continuously, without retraining the entire model [27].

Scheirer et al. [32] proposed open risk, the concept of risk associated with open spaces. According to Fei et al. [12], open space risk was reduced by learning the closed boundary of each positive class in the similarity space. Bendale and Boult [1] reduced the open space risk through deep neural networks (DNNs) and Hendrycks et al. [16] used the Softmax probability as the confidence score. However, in both cases, the confidence threshold needs to be selected by sampling. In Shu et al. [33], the Softmax activation function was replaced with a sigmoid activation function, which is used to match classes statistically. Although statistics-based thresholds can detect characteristics of known classes, they cannot differentiate between the known classes and the open (or unknown) class. In Lin et al. [24], deep intent features were learned by evaluating the margin loss and unknown intents were detected by examining outliers locally. However, it does not have a specific decision boundary for distinguishing open classes and requires model architecture adjustments accordingly. Those existing methods require specific classifiers for identifying open classes and perform poorly in classification. Furthermore, the performance of open classification is determined mainly by the decision conditions. Furthermore, the process of manually selecting the optimal decision condition is also complex and time-consuming, which is not applicable in real-world scenarios. So for open topic classification tasks, it is necessary to use known topics as prior knowledge and learn the adaptive decision boundary. The knowledge base that stores the learned knowledge is also a pivotal part of the entire open topic classification process.

Open-learning systems require the introduction of reject options to reduce misclassification and improve classification reliability. Under the framework of the minimum risk theory, Chow [5] defined the optimal classification rules for reject options. With the development of machine learning, the applications of reject option have been widely discussed, such as hierarchical classification [18], multi-label classification [17,30], deep neural network with reject option [15,14], etc. It is worth noting that the reject option and three-way decision have much in common. One of the most significant similarities is to delay decisions to reduce the misclassification cost. Compared with the reject option, three-way decision developed multi-thresholds and integrated with the multi-granularity framework, which provides more research space for continual learning. In this study, different granularities are adopted according to different decision phrases to combine the reject option with three-way decision to improve the efficiency of the proposed model.

3. Proposed model

With the inspiration of the open-world learning paradigm, we combine multi-granularity learning with open topic classification to construct a dynamic three-way multi-granularity enhanced open topic classification model (TWMG-Open). Given the unknown topic classes arising constantly, TWMG-Open detects unknown topics from known topics in the coarser granularity and then figures out the space feature of unknowns and how new document collections with unknown topics affected the knowledge base in the finer granularity. Considered properties of dynamic data, the knowledge should be updated dynamically and continually while building a multi-granularity structure at different phrases.

To summarize, TWMG-Open constructs a complete multi-granularity learning framework towards dynamic data in dealing with open topic classification problems. Later in this section, we first describe how to build a multi-granularity structure dynamically, introduce the preliminaries and explain each component of TWMG-Open thoroughly.

3.1. The construction of multi-granularity structure

From the temporal multi-granularity perspective, we focus on how data change over time, which may include adding or deleting objects, appending or removing attributes, as well as updating or modifying attribute values. By considering the time factor (corresponding to $t_1, t_2, \ldots, t_k$), we can observe that data and objects evolve over time, and singular things frequently become dual or multiple so as to provide more information. This consideration leads to temporal multi-granularity learning. From the spatial multi-granularity perspective, the dynamic model can capture information or knowledge at multi-granularity levels (corresponding to $layer_1, layer_2, \ldots, layer_k$). It is possible for us to construct granularity spaces in multiple layers. By the inspiration of temporal-spatial multi-granularity framework [36], we apply both temporal and spatial granularity construction approaches to the open dynamic learning tasks. From both perspectives of granularity construction, it is effective to adopt appropriate decision strategies according to different multi-granularity levels.

The open dynamic environment is positioned within the framework of multi-granularity learning, which makes it possible to select better decision-making strategies with the corresponding granularity.

Definition 1. Consider the time factor $t_1 \leqslant t_i \leqslant t_n$ $(i = 1, 2, \ldots, n)$ and spatial granularity level $s_1 \leqslant s_j \leqslant s_m$ $(j = 1, 2, \ldots, m)$. Suppose in time t_i, the multilevel granularity structure is defined as $Gr^{s}_{t_i} = \left(Gr^{s_1}_{t_i}, Gr^{s_2}_{t_i}, \ldots, Gr^{s_j}_{t_i}\right)$, where $Gr^{s_j}_{t_i} = \left(U^{s_j}_{t_i}, \mathbf{x}^{s_j}_{t_i}, \mathbf{C}^{s_j}_{t_i}, \theta^{s_j}_{t_i}\right)$. $U^{s_j}_{t_i}$ denotes the entire processing objects, $\mathbf{x}^{s_j}_{t_i}$ denotes the granularity induced by x, $\mathbf{C}^{s_j}_{t_i}$ is a set of categories which consists of k different tags, and $\theta^{s_j}_{t_i}$ is the parameter of current granularity level.

Definition 1 provides two marks to indicate the temporal granularity and spatial granularity of every variable, i.e., time factor t_i and spatial level s_j, where i and j are not necessarily equal. As new samples appear constantly in the open dynamic environment, the multi-granularity structure can be built gradually through the time order in which the unknowns arrive.

In this work, the multi-granularity structure towards open topic classification tasks is constructed from temporal and spatial perspectives. According to the temporal perspective, with new samples (contain knowns and unknowns) continuing to enter into the model, the time factors $(t_1, \ldots, t_n)$ are used to indicate the temporal granularity levels to the chronological order. From the spatial perspective, an ideal open-learning system should (1) detect unknowns, (2) discover hidden open classes from unknowns, and (3) learn the new classes incrementally. Therefore, we put these tasks into the multi-granularity framework and generate multiple levels of spatial granularity (s_1, s_2, s_3) from coarser to finer.

As shown in Fig. 2, the spatial granularity will become finer and finer from s_1 to s_3 with the corresponding tasks at each step. Firstly, under the coarsest spatial granularity level s_1, the unknowns are separated from the knowns and the unknowns are treated into a single class. Secondly, under the finer spatial granularity level s_2, the single open class is further subdivided/clustered into different categories with corresponding core points. And lastly, at the most refined spatial granularity level s_3, we construct a new open space that consists of known classes and open classes simultaneously. Consequently, the multi-granularity framework towards open topic classification can be defined as follows:

Definition 2. Let $Gr^{s_j}_{t_i} = \left(U^{s_j}_{t_i}, \mathbf{x}^{s_j}_{t_i}, \mathbf{C}^{s_j}_{t_i}, \theta^{s_j}_{t_i}\right)$. $U^{s_j}_{t_i}$ be a multi-granularity structure towards open dynamic tasks, where $i = 1, 2, \ldots, n$ and $j = 1, 2, 3$. At time t_i $(t \geqslant 0)$, $Gr^{s}_{t_i}$ can be refined and expanded into three spatial granularity levels from coarser to finer:

$$Gr^{s}_{t_i} = \left\{Gr^{s_1}_{t_i}, Gr^{s_2}_{t_i}, Gr^{s_2}_{t_i}\right\}. \tag{1}$$

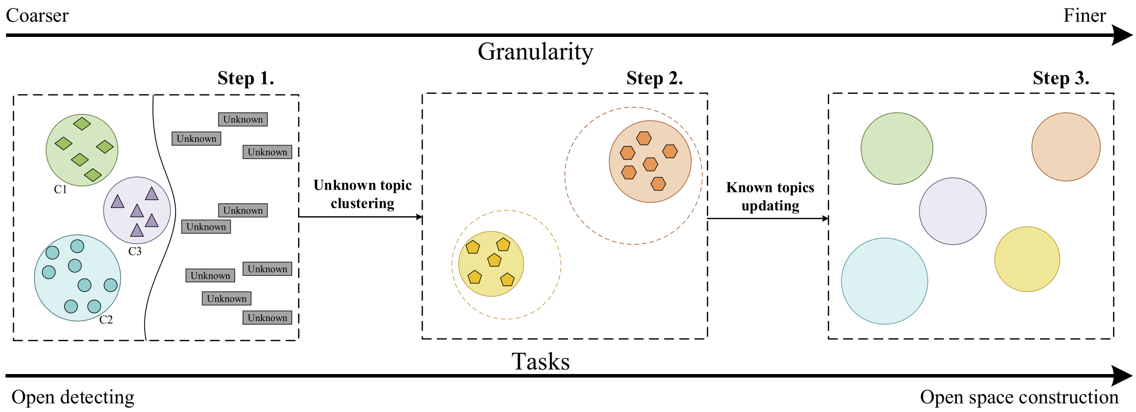

Fig. 2. Illustration of multi-granularity construction process toward open topic classification.

Therefore, we can construct the multi-granularity structure by Definition 2. As the spatial granularity getting finer, the single open class consists all the unknowns will be subdivided into different classes so as to construct the new open space. Besides, $\mathbf{C}_{t_i}$ will get enriched, where $\mathbf{C}_{t_1} \subseteq \ldots \subseteq \mathbf{C}_{t_i} \subseteq \ldots \subseteq \mathbf{C}_{t_n}$.

3.2. Preliminaries

The *lifelong and continual learning* was defined according to [26,27]. At any time point, the learner has learned a sequence of n tasks, $T_1, T_2, \ldots, T_n$ using their corresponding training data $Doc_1, Doc_2, \ldots, Doc_n$. When faced with the $(n+1)^{th}$ task T_{n+1} with its training data Doc_{n+1}, the learner can transfer the knowledge learned from the previous n tasks to help learn T_{n+1}. On-the-job learning refers to learning after the model has been deployed in an application or during model application. Here, the system should (1) discover unknowns and create new learning tasks from the unknowns, (2) collect training or ground-truth data through interactions with users and the environment and through imitation of humans or other AI agents, and (3) incrementally learn the new tasks. The whole process is carried out on the fly in a self-motivated and self-supervised manner.

Therefore, our work is divided into three corresponding parts: (1) detect unknowns in the coarser granularity level, (2) create new clustering tasks for the unknowns in the finer granularity level, and (3) keep updating KB continually to store the learned knowledge for next/new tasks.

As the bounded spherical area could significantly reduce the open space risk, the spherical shape boundary has shown superiority in many applications [12]. Hence, our work adopts an adaptive ball area to constrain all the known samples. Here, the adaptive decision boundary can be formulated as Definition 3 below.

Definition 3. *(Centroid and decision boundary)* Given the i^{th} input sentence $\mathbf{s}_i$. $\mathbf{z}_i \in \mathbb{R}^{\mathbb{D}}$ is the topic representation, where D is the dimension of $\mathbf{z}_i$. $S = \{(\mathbf{z}_1, y_1), \ldots, (\mathbf{z}_N, y_N)\}$ is the set of samples with their corresponding labels. S_k is the set of examples labeled with topic k. The centroid $c_k \in \mathbb{R}^{\mathbb{D}}$ is defined as the mean vector of embedded samples of S_k:

$$c_k = \frac{1}{|S_k|} \sum_{\mathbf{z}_i, y_i \in S_k} \mathbf{z}_i, \tag{2}$$

where $|S_k|$ is the number of samples in S_k. In addition to centroid $\mathbf{c}_k$, the radius of the decision boundary Δ_k is also used to constrain the samples into the ball area. So for each known topic class k, the samples belonged to class k should satisfy:

$$\forall \mathbf{z}_i \in S_k, \|\mathbf{z}_i - c_k\|_2 \leqslant \Delta_k, \tag{3}$$

where $\|\mathbf{z}_i - c_k\|_2$ is the Euclidean distance between $\mathbf{z}_i$ and c_k.

Hence, the centroid c_k and the radius Δ_k together formulate the spherical region of topic k. For a given sample x, we firstly find a nearest centroid from this sample to determine a topic k. Therefore, even if the samples different dimensional values lie into different topic regions, we can find the topic k by the nearest centroid from the sample x accordingly. Then, we adopt a reject option in the coarser granularity level: if this sample in the ball region of topic k, then we consider this sample as belonging to this known topic k; if this sample x does not lie in the ball region of topic k, then we consider reject option to this sample and treat the sample x as belonging to the unknown class. All the samples that do not belong to any bounded spherical region are detailed to form a separate category, i.e. the unknown class.

Next, the unknown class will be processed in a finer granularity level. Because of no pre-knowledge of topic labels, the unknowns need to be further clustered into different categories to mine the knowledge of unknown topics. As opposed to the traditional binary cluster which represents by a single set, we use two nested sets to represent the three-way cluster [43]. Suppose $\mathbf{C} = \{C_1, C_2, \dots, C_k\}$ is the family of clusters, and Z is the non-empty universe of all the topic representations. In the finer granularity, the three-way clustering was defined as below:

Definition 4. *(Three-way clustering)* A cluster is depicted by a pair of nested sets [44]:

$$C_i = \left[\underline{C_i}, \overline{C_i}\right], \tag{4}$$

where $\underline{C_i}$ is the lower bound of the cluster and $\overline{C_i}$ is the upper bound of the cluster and, $\underline{C_i} \subseteq \overline{C_i} \subseteq Z$. The set $\underline{C_i}, \overline{C_i} - \underline{C_i}$and $Z - \overline{C_i}$ can formulate three regions of C_i as positive region (POS), boundary region (BND) and negative region (NEG):

$$\begin{aligned} &POS(C_i) = \underline{C_i}, \\ &BND(C_i) = \overline{C_i} - \underline{C_i}, \\ &NEG(C_i) = Z - \overline{C_i}. \end{aligned} \tag{5}$$

Objects in $POS(C_i)$ definitely belong to the class C_i. Objects in $NEG(C_i)$ definitely do not belong to the class C_i. Objects in $BND(C_i)$ imply that their relationship with clusters is ambiguous. We need more information to make judgments and thus they may belong to one or more classes.

DBSCAN [10], a classical density-based clustering algorithm, can find several clusters by the estimated density distribution and cluster data with arbitrary shapes. The most significant advantage of DBSCAN is automatic determination of cluster size (or shape) depended on the density distributions. Firstly, DBSCAN can discover clusters in arbitrary shapes (or sizes). Secondly, DBSCAN does not need to set the number of topics in advance. Finally, it identifies clusters of different densities with different centroids, which can be stored as knowledge.

Essentially, it requires that a neighborhood of a given radius *Eps* consist of a minimum number of objects *MinPts* for each object in the cluster. Considering an arbitrary point P, DBSCAN searches all points in the ϵ neighborhood of P, where the ϵ is the maximum distance from neighborhood to the point P. Then, if there are at least *MinPts* points within ϵ-neighborhoods of point P, then point P is considered a core point. DBSCAN can find all density-reachable points for each point in the cluster and adds them into the same cluster. If there are points within the cluster and density-reachable, but their ϵ-neighborhoods is less than *MinPts*, then those points are border points. If a point cannot be reached by any other point, then it is an outlier. According to DBSCAN, three types of points are distinguishable: core points with high density, border points with low density, and noise points [43]:

Definition 5. *(Type function of naive DBSCAN)* Given any two points x and $y, sim(x,y)$ is the measure of similarity between them. $\Gamma_\epsilon(x)$ is the ϵ-neighborhood of x, where the $\Gamma_\epsilon(x) = \{y \in V \mid sim(x,y) \leqslant \epsilon\}$ and V is the set of all points. Given $\rho(x) = \mid \Gamma_\epsilon(x) \mid$ is the density value of x, the type function $Type(x)$, it is defined as:

$$Type(x) = \begin{cases} 1, & \text{core point with } \rho(x) \geqslant \text{MinPts}; \\ 0, & \text{border point with } 1 < \rho(x) < \text{MinPts}; \\ -1, & \text{noise with} \rho(x) = 1. \end{cases} \tag{6}$$

Here, our work attempts to apply the idea of three-way clustering to the DBSCAN algorithm. At the same time, the idea of three-way clustering in the finer granularity is different from the reject option in the coarser granularity, which shows the superiority of the three-way decision compared with the reject option in multi-granularity learning. Therefore, according to the above analysis, we adopt three-way clustering method combined with DBSCAN algorithm in the finer granularity level, which is also conducive to knowledge accumulation process.

Finally, consider building the knowledge base. The knowledge learned at current stage will become the prior knowledge of next stage. So as to ensure the model performance and the quality of classifications, we hope to have more reliable knowledge stored. The class C_i is defined as $POS(C_i)$ and the centroid c_i will be seen as new knowledge and added to knowledge base. The knowledge base has been used in many open tasks, and various ways of knowledge accumulation have been explored. Here, we use a simple method to accumulate knowledge and the knowledge base (KB) is defined as below:

Definition 6. *(Knowledge base)* Given the task sequence $(T_1, T_2, \dots, T_n, T_{n+1})$ and suppose model has learned the first n tasks. The knowledge of each task (i.e., centroid c_i of each topic class) has also been stored in the KB. Then those prior knowledge are used to new/target task T_{n+1} to generate new knowledge. Suppose $\mathbf{c}_n = (c_1, \dots, c_k)$ is the set of prior knowledge, and the new knowledge c_{k+1} learned from T_{n+1} will be appended into the $\mathbf{c}_n$ to generate $\mathbf{c}_{n+1}$:

$$\mathbf{c}_n^\frown \langle c_{k+1} \rangle = \mathbf{c}_{n+1}. \tag{7}$$

3.3. Three-way multi-granularity decision towards open topic classification

In TWMG-Open, the key point is to deal with the uncertainty through the framework of multi-granularity learning, which emphasizes the importance of dealing properly with uncertainty in open tasks. Uncertainty is bound to occur with ever-changing problems and data. By using multi-granularity learning as a framework, the uncertain knowledge associated with a certain level of granularity can be captured to reduce the uncertainty of open problems.

A typical TWMG-Open framework should consist of four major components (although they can be designed as a unified or pipelined framework), as described in Fig. 3.

3.3.1. Pre-training module

Among the many problems of Natural Language Processing (NLP), topic classification is one of the most important and classic task. The text representation is a basic step in the process. Various neural models have been used to learn text representation, including convolutional models, recurrent models, and attention mechanisms. More recently, studies have shown that pre-trained language models can be used to learn common language representations by relying on large amounts of unlabeled data. Among them, the BERT [9] model relies on a multi-layer bidirectional Transformer [35] and is trained with plain text in masked word prediction tasks and next sentence prediction tasks.

In this study, we use the BERT model to extract deep topic features representations. Given i^{th} input sentence s_i, all the token embeddings $[[CLS], T_1, \ldots, T_N] \in \mathbb{R}^{(\mathbb{N}+1)\times\mathbb{H}}$ are extracted from the last hidden layer. H is the size of hidden layer, $[CLS]$ is the vector for classification, N is the sequence length. Then, a mean-pooling is performed on these token embeddings to synthesize the high-level semantic features in one sentence, and get the averaged representation $\mathbf{x}_i \in \mathbb{R}^{\mathbb{H}}$[25]:

$$\mathbf{x}_i = mean - pooling[[CLS], T_1, \ldots, T_N]. \tag{8}$$

Furthermore, $\mathbf{x}_i$ is fed to a dense layer to generate strengthened feature representations $\mathbf{z}_i$:

$$\mathbf{z}_i = h(\mathbf{x}_i) = ReLU(W_h\mathbf{x}_i + b_h), \tag{9}$$

where D is the dimension of $\mathbf{z}_i, W_h \in \mathbb{R}^{\mathbb{H}\times\mathbb{D}}$ and $b_n \in \mathbb{R}^{\mathbb{D}}$ respectively denote the weights and the bias term of layer h.

Since open topic samples are scarce, we use known topics as prior knowledge to pre-train the model. In the topic classification task, the relationship matrix between tokens is relatively sparse, where each word vector only needs to be associated with a few other words. Therefore, with Softmax as the final layer in pre-trained BERT, the quality of feature representation slightly can be improved. Then, for the purpose of measuring the performance of the acquired decision boundary, the Softmax loss $\mathscr{L}_s$ is used to learn the feature representation $\mathbf{z}_i$:

$$\mathscr{L}_s = -\frac{1}{N}\sum_{1=1}^{N} log \frac{exp\left(\phi(\mathbf{z}_i)^{y_i}\right)}{\sum_{j=1}^{K} exp\left(\phi_j(\mathbf{z}_i)\right)}, \tag{10}$$

where $\phi(\cdot)$ is a linear classifier and $\phi_j(\mathbf{z}_i)$ is the output logits of j^{th} class.

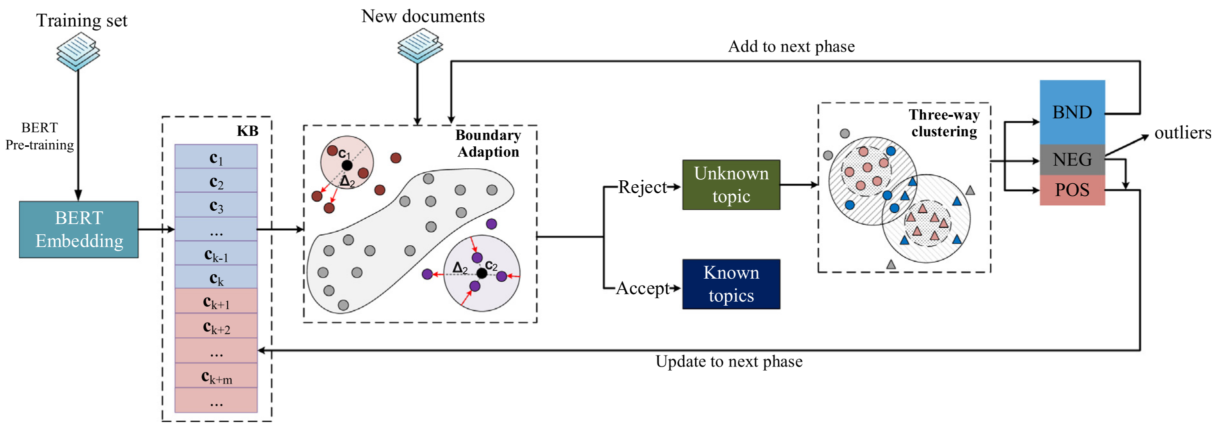

Fig. 3. Illustration of TWMG-Open architecture. Firstly, Bert is used to pre-train the model and extract feature representation. Secondly, in the coarser granularity level, the centroids $\{\mathbf{c}_i\}_{i=1}^{K}$ and the corresponding radius of decision boundaries $\{\Delta_i\}_{i=1}^{K}$ are used to detect unknowns by reject option. Thirdly, in the finer granularity level, three-way enhanced DBSCAN is used to learn the new knowledge of new classes. Finally, the whole model can learn knowledge accumulatively and continually by applying KB.

The pre-trained BERT model can be used to create contextualized word embeddings and the feature representation $\mathbf{z}_i$. After the decoder layer, a fully connected layer and a Softmax layer are added at the end of pre-training stage. These outputs can then be provided to downstream tasks.

3.3.2. Adaptive decision boundary construction

As illustrated in Fig. 3, we first pre-train the model by BERT [9] to extract topic representations (or embeddings). For each known class, we define centroid in Definition 3 and suppose that the known topic features are constrained in the closed ball areas. Then, the radius of each topic region is calculated to obtain the decision boundary. Noteworthy is the use of a new loss function for automatically learning tight decision boundaries that adaptive to feature space. The suitable decision boundaries should satisfy two conditions [46]: (1) it is necessary to cover the known topics as much as possible by a wide margin, and (2) decision boundaries must also be tight enough to prevent unknown topic samples from being identified as unknown topic samples. In order to address these problems, our work adopts a new loss function that optimizes the boundary parameters by balancing the open risks and empirical risks [32]. Determining the decision boundaries is the coarser granularity task, which allows finding unknown topics through adaptive decision boundaries.

In the coarser granularity level, the performance of open detecting is largely determined by the decision conditions. And the process of manually selecting the optimal decision condition is also complex and time-consuming, which is not applicable in real-world scenarios. So for open topic discovery, it is therefore necessary to use known topics as prior knowledge and learn the decision boundary adaptively. In the Definition 3, the decision boundary has been formulated according to the centroid c_k and the radius Δ_k of topic k. To learn the decision boundary adaptively, we want to use the centroid c_k and the radius Δ_k to confine the samples belonging to topic k to a spherical region. Here, the radius Δ_k needs to be adaptive to different feature space, so the neural network is applied to optimize the learnable boundary parameter $\hat{\Delta_k} \in \mathbb{R}$ as follows.

First, the *Softplus* activation function is used to map between Δ_k and $\hat{\Delta_k}$ [34]:

$$\Delta_k = log\left(1 + e^{\hat{\Delta_k}}\right), \tag{11}$$

where the *Softplus* activation function is totally differentiable $\hat{\Delta_k}$ and guarantees the learned radius Δ_k is above zero and, in contrast to *ReLU* activation function, it allows for bigger Δ_k if necessary.

In Definition 3, for each known topic class k, the samples belong to class k should satisfy Eq. (3). In order to balance the empirical risk and open space risk, we apply an adaptive strategy [46] to learn the decision boundaries. On the one hand, if $\|\mathbf{z}_i - c_k\|_2 > \Delta_k$, the knowns will be outside their respective decision boundaries, which may introduce empirical risk. Thus, the decision boundaries should be expanded to include more samples from known classes. On the other hand, if $\|\mathbf{z}_i - c_k\|_2 \leqslant \Delta_k$, it is likely that more knowns will be identified with the wider decision boundaries, but it may increase the open space risk. Therefore, according to the above strategies, the boundary loss $\mathscr{L}_b$ is formulated as [46]:

$$\mathscr{L}_b = \frac{1}{N}\sum_{i=1}^{N}\left[\delta_i\left(\|\mathbf{z}_i - c_i\|_2 - \Delta_{y_i}\right) + (1-\delta_i)\left(\Delta_{y_i} - \|\mathbf{z}_i - c_i\|_2\right)\right], \tag{12}$$

where y_i is the label of the i^{th} sample and δ_i is defined as:

$$\delta_i = \begin{cases} 1, & \text{if } \left\|\mathbf{z}_i - c_{y_i}\right\|_2 > \Delta_{y_i}; \\ 0, & \text{if } \left\|\mathbf{z}_i - c_{y_i}\right\|_2 \leqslant \Delta_{y_i}. \end{cases} \tag{13}$$

Then the boundary parameter $\hat{\Delta_k}$ is updated according to $\mathscr{L}_b$:

$$\hat{\Delta_k} := \hat{\Delta_k} - \eta\frac{\partial \mathscr{L}_b}{\partial \hat{\Delta_k}}, \eta \mathrm{isthelearningrate}. \tag{14}$$

The $\frac{\partial \mathscr{L}_b}{\partial \hat{\Delta_k}}$ is calculated by:

$$\frac{\partial \mathscr{L}_b}{\partial \hat{\Delta_k}} = \frac{\sum_{i=1}^{N}\delta'(y_i = k)(-1)^{\delta_i}}{\sum_{i=1}^{N}\delta'(y_i = k)} \cdot \frac{1}{1 + e^{-\hat{\Delta_k}}}, \tag{15}$$

where δ'_i equals to:

$$\delta'_i = \begin{cases} 1, & \text{if } y_i = k; \\ 0, & \text{if } y_i \neq k. \end{cases} \tag{16}$$

Here, on the radius Δ_{y_i} belonging to class k will be updated to make sure the denominator is not zero.

By incorporating the boundary loss $\mathscr{L}_b$, boundaries can adapt to the intent feature space and learn effective and proper decision boundaries. Not only does a learned decision boundary cover most of the known intent samples, but it also does not stray far from the known class centroid, which helps identify open intent samples better. Then, as the new samples need to be processed appear, the centroids c_i and the learned decision boundaries are applied to determine which samples can be accepted as knowns and which samples need to be rejected as unknown. Suppose $\mathscr{Y} = \{1, 2, \ldots, K\}$ is the set of known topics, the classification tasks performed in coarser granularity level is:

$$\hat{y}_i = \begin{cases} \text{unknown}, & \text{if } \left\|\mathbf{z}_i - c_{y_i}\right\|_2 > \Delta_{y_i}, \forall y_i \in \mathscr{Y}; \\ argmin_{k \in \mathscr{Y}}, & \text{if } \left\|\mathbf{z}_i - c_{y_i}\right\|_2 \leqslant \Delta_{y_i}, \text{otherwise}. \end{cases} \tag{17}$$

3.3.3. Three-way clustering with DBSCAN

Next, as shown in Fig. 3, we will further deal with knowns and unknowns respectively in the finer granularity. For the former, those known topic samples are simply classified according to the existing knowledge by the adaptive decision boundaries. For the latter, given the variation of granularity, this work proposed a new three-way enhanced DBSCAN clustering model, which is different from the present models [43,11] and can give the new/open topics pseudo labels according to the space feature in finer granularity level.

As mentioned in Definition 5, the DBSCAN clustering algorithm can make clusters according to density distributions automatically, which eliminates the need to know the size/shape of clusters. Meanwhile, the three-way clustering method can be applied to enhance the naive DBSCAN. Unlike two-way clusters which are modeled by a single set, the three-way clusters are described by nested sets $C_i = \left[\overline{C_i}, \underline{C_i}\right]$ in Definition 4. Based on the original type function, we have developed three different regions respectively: POS, BND and NEG. On the one hand, we adopt different strategies according to different regions to capture the uncertainty of initial core points in a finer granularity. On the other hand, through three-way multi-granularity framework, we can learn reliable knowledge from unknown/open information and store it for future classifications. The three-way enhanced DBSCAN is explained step by step as follows:

Step 0. At the beginning, to obtain initial core points, the DBSCAN algorithm is simply conducted with parameter $MinPts_{core}$. According to Definition 5, i.e., if $Type(x) = 1$, then x is core point.

Step 1. Then, to determine the positive region of class C_i, the basic idea is to compare core points density with a new threshold $MinPts_{POS}$, which is stricter than $Minpts$. Then, the core points are classified according:

$$Type\prime(x) = \begin{cases} 1, & \text{core points with } \rho(x) \geqslant MinPts_{POS}; \\ 0, & \text{core points with } MinPts_{core} \leqslant \rho(x) < MinPts_{POS}; \\ -1, & \text{border points with } \rho(x) < MinPts_{core}. \end{cases} \tag{18}$$

If an object x belongs to a high-density field of class C_i, i.e., $Type\prime(x) = 1$ in Eq. (18), then it is a typical element of class C_i and naturally should be assigned to the positive region. If an object x is in a low-density field, i.e. $Type\prime(x) = 0$ in Eq. (18), it is possible that it may belong to another or some others clusters, then x will be assigned to the boundary region. Therefore, the positive region and boundary region of C_i can be constructed as follows:

$$\begin{aligned} &POS(C_i) = \{x \mid Type'(x) = 1, x \in C_i\}, \\ &BND(C_i) = \{x \mid Type'(x) = 0, x \in C_i\}, \\ &NEG(C_i) = \{x \mid Type'(x) = -1, x \in C_i\}, \end{aligned} \tag{19}$$

where the $POS(C_i)$ is the set of all core points and $BND(C_i)$ is the set of all border points.

Step 2. Eq. (19) assigns the border objects to a corresponding boundary region (BND), but we also need to treat them further as overlapping objects because they can be a part of other clusters. Hence, the following strategy is adopted: for one point x satisfying $Tpye\prime(x) = 0$, if a neighbor of x in ϵ-neighborhood size belongs to one other class C_i, then x is considered as an *overlapping object* and then make the point x stay in the $BND(C_i)$. Therefore, the boundary region $BND(C_i)$ is updated as:

$$BND\prime(C_i) = BND(C_i) \cup \{x \mid C_i \cap \Gamma_\epsilon(x) \neq \varnothing\}. \tag{20}$$

Step 3. For the remaining points in NEG, we also want to further process to: (1) assign them to the proper clusters or (2) discard them as noises. The $AllPOS$ denotes the set of all POS. Then, given a point x in NEG, we calculate the distance $dis(x, y)$ from x to another point in the set $AllPOS$, i.e., $AllPOS = \bigcup_{i=1}^{k} POS(C_i)$. Next, the nearest core neighbor of x in $NCN(x)$ is constructed by:

$$NCN(x) = y \leftarrow argmin_{y \in AllPOS} dis(x, y), \tag{21}$$

where the $dis(x, y)$ is the distance from x to another point in the set $AllPOS$. If $NCN(x) \subseteq BND(C_j) \neq \varnothing (i \neq j)$, then $NCN(x)$ should be added to $BND(C_j)$ and wait for next stage of processing; otherwise, it will be discarded as outliers.

As a final point, it's also critical to stress that the three-way multi-granularity decision framework of this work is dynamic and continually updated. In other words, open dynamic environments require the traditional static three-way multi-granularity learning framework to be put into a more generalized open situation in light of the fact that new data are constantly emerging. In the process of knowledge accumulation, the granularity of knowledge is also changing: the transversion of knowledge granularity from coarser level to finer level is the procedure of learning. The knowledge accumulated by the proposed dynamic three-way multi-granularity learning also can be stored and become the prior knowledge for the follow-up tasks.

3.3.4. Knowledge base (KB)

Lastly, in order to make this model have ability of continual learning, a knowledge base is designed to store the topic knowledge in the finer granularity level, because only storing knowledge in finer granularity can improve the accuracy and validity of knowledge. In this work, the *knowledge* is represented by the centroid of each topic. After learning the knowledge of a new topic/class at each time stage, add the corresponding centroid to the KB. This knowledge has been stored in KB will become the prior knowledge for the next tasks. The knowledge is constantly accumulated and TWMG-Open makes use of KB to have the ability of continual learning, which can store the knowledge learned from the current task to reduce the computational cost and time cost of subsequent tasks.

Ideally, the knowledge is constantly accumulated, compiled and learned by engineers. However, it is impossible to cover all of the knowledge in any field of the world, let alone knowing that everything changes constantly. This will be a significant uncertainty and loss of opportunity during learning, leading to train the entire group of engineers over and over again. KB is used to store the knowledge acquired from the previous tasks in an appropriate representation and then for adaptive boundary decision and three-way clustering in current and future tasks.

For example, in a knowledge-grounded topic classification model, knowledge can be represented as centroids of topic classes. Hence, in vector-based open dynamic factual knowledge learning, KB can be a centroid store system, or a knowledge graph storing real-world facts, or even a dictionary representing glossary of various terms and concepts etc. Just as shown in the Fig. 3, the KB of TWMG-Open is designed to store the centroids $\{\mathbf{c}_i\}_{i=1}^{K}$ of each category and updated according to Definition 6.

In this work, in order to improve the accuracy and validity of knowledge, we use the information of finer granularity level to be stored as knowledge. Through the three-way clustering, all the categories in *POS* region serve as the prior knowledge of the subsequent tasks to guide the subsequent classification. Moreover, to reduce the difficulty of knowledge storing, we use *centroid* of each class as the representation of acquired knowledge and place it into KB, so that we can store a lot of knowledge at a relatively small cost. For all classes in $\mathbf{C}_{POS_n} = \{C_1, \ldots, C_k\}$, the centroid of each class $\mathbf{c}_{POS_n} = \{c_1, \ldots, c_k\}$, then following the Definition 6, the new knowledge c_{k+1} of T_{n+1} will be appended into the $\mathbf{c}_{POS}$ to generate $\mathbf{c}_{n+1}$:

$$\mathbf{c}_{POS_n}{}^{\frown}\langle c_{k+1}\rangle = \mathbf{c}_{POS_{n+1}}. \tag{22}$$

3.4. Algorithms

Algorithm 1: Three-way DBSCAN (3 W-DBSCAN)

Input: data, $Eps(\epsilon)$, $MinPts_{core}$ and $MinPts_{POS}$.
Initiate the universe cluster of all clusters $C_{initial} = \{C_1, C_2, ..., C_i, ..., C_n\}$, core points and border points;
Train the data with naive DBSCAN's Eps, $MinPts$;
for *each cluster* **do**
  for *each core point in cluster* **do**
    Calculate all the density-reachable samples of this core point with $Minpts_{POS}$ and add into POS;
  end
  Calculate $C_{initial} - POS$ and add the result into BND;
end
for *each core point in POS* **do**
  if *in BND* **then**
    Remove this core point from POS;
  end
end
Output POS, BND.

Algorithm 2: Three-way multi-granularity learning towards open topic classification (TWMG-Open)

Input: Training data D_{train} and data stream $D_i \in \{D_1, D_2, \ldots, D_n\}$.
1 **Initiate** Training set D_{train};
2 BERT pre-training with D_{train} and get the feature representations $\mathbf{z}_i$ by Equation (10) ;
3 **for** *each* $D_i \in \{D_1, D_2, \ldots, D_n\}$ **do**
4 Learn the decision boundary adaptively of D_{train} by Equation (12);
5 Classify D_{input} to $Unknowns_i$ and $Knowns_i$ by Equation (17);
6 Using 3W-DBSCAN (Algorithm 1) to cluster *unknown*;
7 Get POS_i and BND_i from 3W-DBSCAN;
8 **Output** POS_i;
9 $input = BND_i$;
10 $D_{train_{i+1}} = Knowns_i + POS_i$ and update KB by Equation (22);
11 $D_{input} = D_{i+1} + BND_i$;
12 **end**
13 Classify D_{input} by the nearest center in ADB;
14 **Output** all the *Knowns*.

In Algorithm 1, Step 1 is to initiate the universe region. Step 2 is to conduct DBSCAN algorithm with *Eps* and *MinPts* using input data and extract the initial clusters of input data. Steps 3–8 are to calculate the universe region and boundary region. Steps 9–13 are to remove the boundary samples from positive region and Step 14 is to output the positive region and boundary region. In the worst case, the time complexity of Step 2 is $O(n^2)$. The time complexity of Steps 3–8 is $O(n^2)$. The time complexity of Steps 9–13 is $O(n^2)$. Hence, the total time complexity is $O(n^2)$.

In Algorithm 2, Step 1 is to initiate the training data. Step 2 is to pre-train the BERT with training data and get the feature representations. Steps 3–12 are to classify input data and update knowledge base. Specifically, Step 4 is to update input data, Steps 5–7 are to classify unknowns and knowns using ADB, Steps 8–10 are to re-classify unknowns with Algorithm 1 and Steps 11–12 are to update the input data and knowledge base continually. Finally in the last time stage, Step 13 is to classify all the unclassified samples by the nearest center of ADB compulsively.

It's important to emphasize that in real-world scenarios, we don't need to implement the last step (Step 13) of Algorithm 2, since the data flows are continuous and have no end. Specifically, we conduct a dynamic experimental design. In experiments, there are five time stages, and Step 13 is adopted for final classification in the last stage. Details of the experiments are demonstrated in the Section 4.

4. Experiments

A series of experiments were conducted to demonstrate the effectiveness of proposed cost-sensitive three-way multi-granular open topic classification method. All the experiments were performed on a computer with Intel Xeon E5-2678 v3 and NVIDIA GeForce RTX 3090. The Python version is 3.7 for Windows OS x64.

4.1. Datasets

In order to verify that TWMG-Open is effective in different scenarios of topic classification, both short texts and long texts are considered. The experiments are conducted on three challenging real-world datasets to evaluate our approach. We have pre-processed all datasets into balanced datasets, and the number of samples under each label is equal. In pre-training, we use the pre-trained model BERT. The stopwords have been removed from the BERT dictionary, so stopwords nor symbols have been removed from these datasets. More details of each dataset are described below:

Banking77. The Banking77 dataset comprises 13,083 customer service queries labeled with 77 intents [2] form a banking domain. There are some overlaps among some categories and each sample contains 12 words on average.

StackOverflow. StackOverflow conducts an annual survey of people on the website to gather information such as programming languages, salaries, code style, and various other information. The StackOverflow dataset was published in Kaggle.com. This dataset consists of 3,370,528 samples (contain question titles and corresponding contents) from Stackoverflow.com. In our experiments, we randomly select 20,000 question titles with 20 different tags.

TC20. The text classification 20 (TC20) was published in Kaggle.com. This dataset consists of 19306 long texts with 20 different domains/topics. In our experiments, we deleted texts that were too long (more than 500 words), then finally got 17909 samples with 20 classes and 169 words per text on average.

4.2. Experimental settings

According to the same settings as in [33,24], we maintain some classes as unknown (open) for testing purposes and insert them back into the system during runtime. The datasets are divided into three types: training, validation, and test. Initially, we vary the proportion of known classes by 25%, 50%, and 75% in the training set. We consider the remaining classes as unknown classes and do not include them in the training set. Finally, for testing, we use both known and unknown classes. Moreover, given the streaming data, we design five-round experiments for each known class ratio, in which a certain proportion of unknown classes would randomly appear in each round.

Naive DBSCAN, Agglomerative Clustering [29], and Mean Shift [7] were compared with different large datasets to evaluate TWMG-Open's performance. The naive DBSCAN is a classical clustering method that differs from the three-way enhanced DBSCAN and only has single *MinPts*. Agglomerative Clustering is a classical and well-established technique in unsupervised machine learning among many in-depth and practical unsupervised machine learning techniques. Mean Shift is a density-based unsupervised clustering algorithm. The idea of Mean Shift is to assume that the datasets of different classes are under different probability density distributions and find the fastest direction in which the density of any points increases, and the points converging to the same local maximum value are considered to be members of the same class. Also, a notable recent unsupervised text classification approach, Semantic Clustering by Adopting Nearest neighbors (SCAN) [50], is used for comparison at each round, too. The intuition behind SCAN is that neighbors in representation space often share the same label. This regularity can be leveraged as a weakly supervised signal for training models.

To achieve optimality, we implement the BERT model [9] (consists of a 12-layer transformer) and make use of most of its suggested settings in experiments. The training procedure is sped up by freezing all BERT parameters except for the last couple of transformer layers, and the learning rate is 2e-5. And the Adam algorithm [20] is used to optimize the boundary parameters of the boundary loss $\mathscr{L}_b$ with a learning rate of 0.05. For baseline models, we hope that each model will achieve the best algorithm performance. Each model has its own characteristics. In comparison, some models are density-based and other models are distance-based. Consequently, the parameter settings are adjusted during the experiments and identified to give satisfactory results. And all models use the same data set, epoch, and batch size. Under different known classes ratio (25%, 50%, and 75%) and different datasets, the parameter settings of TWMG-Open and benchmark models are outlined in Table 1.

4.3. Results analysis

The score of accuracy, the macro F1 score, and the normalized mutual information (NMI) are used to evaluate the overall performances of TWMG-Open. All the metrics are calculated across all classes (consist of known classes and unknown classes) at the finer granularity level.

In order to compare the performance with the overall TWMG-Open, we use BERT, CNN and RNN as benchmark models. These three deep neural network models are applied to topic classification in experiments. Table 2 shows the comparisons of different models. Among them, NaiveDBSCAN, AC, MS and SCAN are mainly comparing the unknowns clustering quality; BERT, CNN and RNN are mainly comparing the overall performances; and TWMG-Open without KB is mainly comparing the effect of KB in the model. Table 3 to Table 5 show the results of each round on different datasets. Fig. 4 shows the improvement of the calculation efficiency of the model by calculating the computation time. To verify the ability of continual learning of the model, Fig. 5 shows the comparison of model performance before and after KB ablation.

Table 2 shows the overall experimental results of TWMG-Open compared with the TWMG-Open without KB, three clustering models and three neural network models. The Hungarian algorithm assigns optimal labels for calculating ACC and all metrics are calculated in the last round. Because the benchmark models and the TWMG-Open without KB do not have the ability of continual learning, labels are assigned in each round. In this case, even if the same category is detected in different rounds, it will still be given the same label. That is, they are considered to be the same category. However, TWMG-Open has a

Table 1
Parameter setting.

		Banking77			StackOverflow			TC20		
	Methods	eps	minPts	spreadPts	eps	minPts	spreadPts	eps	minPts	spreadPts
25%	TWMG-Open	9.5	10	2	4.3	15	2	3.7	28	1
	TWMG-Open (NonKB)	9.5	10	2	4.6	15	2	4.6	25	2
	NaiveDBSCAN	9.5	8	\	5.5	22	\	4.5	25	\
	AC	1.0	\	\	2.1	\	\	1.5	\	\
	MS	0.332	\	\	0.295	\	\	0.22	\	\
50%	TWMG-Open	12	12	2	8	25	2	6	20	1
	TWMG-Open (NonKB)	12	12	2	7.5	20	2	7.5	20	2
	NaiveDBSCAN	12.5	8	\	8.2	23	\	7.5	23	\
	AC	1.1	\	\	2.6	\	\	2.6	\	\
	MS	0.385	\	\	0.345	\	\	0.31	\	\
75%	TWMG-Open	15	15	2	10	30	2	9.2	30	2
	TWMG-Open (NonKB)	15	15	2	9.5	30	2	9.5	30	2
	NaiveDBSCAN	15.5	8	\	10	20	\	10	20	\
	AC	1.6	\	\	3.1	\	\	3.1	\	\
	MS	0.435	\	\	0.405	\	\	0.38	\	\

Table 2
Overall results (F1, ACC, NMI of all test samples).

	Methods	Banking77			StackOverflow			TC20		
		F1	ACC	NMI	F1	ACC	NMI	F1	ACC	NMI
25%	TWMG-Open	38.44	45.92	**67.68**	31.02	**59.14**	**51.42**	22.72	**33.84**	**43.82**
	TWMG-Open (NonKB)	33.01	36.81	57.09	28.84	31.37	31.77	23.45	30.23	37.67
	NaiveDBSCAN	37.62	41.30	59.29	**52.48**	53.78	47.11	**28.26**	33.76	40.04
	AC	**46.89**	**48.17**	63.49	36.06	36.52	29.30	30.03	34.72	40.21
	MS	36.79	40.69	60.52	25.55	28.37	25.47	28.11	34.55	40.95
	Bert	10.30	23.80	\	9.93	23.04	\	10.29	23.47	\
	CNN	11.39	21.75	\	9.70	22.36	\	10.09	22.16	\
	RNN	9.91	19.83	\	9.99	22.79	\	6.55	15.92	\
50%	TWMG-Open	55.31	60.81	**74.72**	**60.82**	**61.49**	**54.25**	41.49	**51.03**	**54.47**
	TWMG-Open (NonKB)	52.33	55.66	68.90	53.76	55.48	48.95	44.57	48.40	49.85
	NaiveDBSCAN	54.63	57.64	69.19	54.12	56.11	47.70	**45.16**	48.60	48.30
	AC	**59.90**	**61.07**	71.45	58.22	57.84	49.96	51.15	53.68	54.10
	MS	50.90	54.07	67.97	46.39	46.26	44.85	41.80	45.95	49.76
	Bert	31.89	45.20	\	30.68	44.15	\	28.91	42.02	\
	CNN	24.02	34.01	\	29.43	42.08	\	29.19	41.02	\
	RNN	18.94	28.00	\	29.41	42.89	\	20.91	30.60	\
75%	TWMG-Open	74.59	**77.32**	**83.67**	**79.55**	**78.55**	**68.90**	**64.45**	**66.04**	**64.08**
	TWMG-Open (NonKB)	73.59	75.91	81.49	69.43	70.52	61.88	61.72	64.25	62.36
	NaiveDBSCAN	**75.08**	77.07	81.78	70.20	71.42	60.18	62.88	64.80	61.95
	AC	73.29	73.28	79.25	73.52	70.41	64.37	61.34	60.93	60.59
	MS	69.09	70.35	77.46	67.00	63.47	62.79	58.02	58.26	58.61
	Bert	59.81	68.42	\	57.93	66.67	\	54.87	62.36	\
	CNN	45.36	53.40	\	55.24	63.49	\	43.90	50.31	\
	RNN	39.29	45.90	\	55.09	62.30	\	32.04	37.96	\

knowledge base and thus, we only assign labels in the last round. Therefore, if a category is detected in different rounds, the dynamic model will mistakenly identifies this category as two different classes, which will cause misclassification when calculating ACC and F1, and it is a bit unfair to TWMG-Open. It can be seen from the Table 2 that when the ratio of known categories is set to 25%, the NMI of TWMG-Open on the Banking77 dataset and TC20 dataset are highest; when the ratio of known categories increase to 50% and 75%, TWMG -Open's NMI on all datasets are best, and F1 and ACC are also improved significantly. As shown in Table 2, BERT, CNN and RNN cannot well adapt to the open dynamic environment, which further demonstrates that the open dynamic environment poses a greater challenge to the existing models. As the number of known classes increases, the effectiveness of TWMG-Open also increases. The NMI of TWMG-Open is optimal in multiple datasets, reflecting the ability to continually extract correct knowledge and learning knowledge. see Table 4.

Table 3 to Table 5 show the results of each round on different datasets. Three classical clustering model are compared with TWMF-Open. Also, a notable recent unsupervised text classification approach, SCAN, is used for comparison, too. SCAN has been shown to work well for image classification. Here, we adapt SCAN for use with text data to tackle the task of unsu-

Table 3
Results of each round on Banking77.

	Banking77	Round 1		Round 2		Round 3		Round 4		Round 5	
	Methods	Num	NMI	Num	NMI	Num	NMI	Num	NMI	Num	NMI
25%	TWMG-Open	830	**80.40**	1394	**76.75**	1710	**73.64**	1737	**72.07**	871	**57.36**
	TWMG-Open (NonKB)	830	80.40	1139	76.52	1296	69.31	1408	70.61	1869	40.45
	NaiveDBSCAN	1600	65.79	1600	67.54	1600	67.00	1742	64.82	\	\
	AC	1600	71.26	1600	72.17	1600	72.12	1742	71.47	\	\
	MS	1600	68.08	1600	67.57	1600	68.82	1742	68.08	\	\
50%	TWMG-Open	1040	**84.72**	1439	82.59	1534	**81.27**	1655	**81.35**	874	**61.52**
	TWMG-Open (NonKB)	1040	84.72	1276	**83.40**	1379	80.46	1543	79.02	1304	49.59
	NaiveDBSCAN	1600	73.82	1600	76.06	1600	76.08	1742	74.24	\	\
	AC	1600	77.28	1600	77.64	1600	77.86	1742	76.68	\	\
	MS	1600	73.33	1600	74.48	1600	74.99	1742	73.11	\	\
75%	TWMG-Open	1224	**90.89**	1449	88.83	1480	90.65	1592	89.77	797	**67.41**
	TWMG-Open (NonKB)	1224	90.89	1365	**89.48**	1415	**90.77**	1526	**90.38**	1012	61.79
	NaiveDBSCAN	1600	84.42	1600	85.30	1600	85.70	1742	84.07	\	\
	AC	1600	82.80	1600	83.35	1600	84.03	1742	82.29	\	\
	MS	1600	81.30	1600	81.47	1600	82.52	1742	80.96	\	\
	SCAN	1600	63.77	1600	63.73	1600	62.93	1742	62.07	\	\

Table 4
Results of each round on StackOverflow.

	StackOverflow	Round 1		Round 2		Round 3		Round 4		Round 5	
	Methods	Num	NMI	Num	NMI	Num	NMI	Num	NMI	Num	NMI
25%	TWMG-Open	695	73.22	1049	67.4	2791	53.06	2971	43.6	2494	**12.62**
	TWMG-Open (NonKB)	678	**73.61**	692	**75.44**	839	**71.01**	869	**73.35**	6922	7.08
	NaiveDBSCAN	2500	49.69	2500	49.14	2500	50.22	2500	50.67	\	\
	AC	2500	33.72	2500	35.33	2500	33.53	2500	33.27	\	\
	MS	2500	31.48	2500	29.84	2500	31.31	2500	31.31	\	\
50%	TWMG-Open	1215	86.37	1886	75.71	2461	63.01	2513	58.26	1925	**17.72**
	TWMG-Open (NonKB)	1181	**86.60**	1466	**83.10**	1648	**79.34**	1686	**77.95**	4019	12.26
	NaiveDBSCAN	2500	51.79	2500	51.21	2500	51.03	2500	51.54	\	\
	AC	2500	52.47	2500	52.40	2500	55.01	2500	54.02	\	\
	MS	2500	51.16	2500	52.05	2500	51.69	2500	52.00	\	\
75%	TWMG-Open	1521	**94.91**	1733	95.10	2301	85.71	2243	81.13	2202	**22.42**
	TWMG-Open (NonKB)	1491	94.91	1620	**95.91**	1704	**94.15**	1668	**94.00**	3517	18.38
	NaiveDBSCAN	2500	63.23	2500	64.60	2500	64.36	2500	64.20	\	\
	AC	2500	65.69	2500	68.04	2500	67.12	2500	66.55	\	\
	MS	2500	67.72	2500	68.46	2500	67.80	2500	68.27	\	\
	SCAN	2500	27.99	2500	27.13	2500	29.49	2500	29.03	\	\

pervised text classification. The tables show the number of samples classified in each round and the NMI for that round. We enforced classification on all delayed samples in the last round, so the result of the last round was decreased. However, the data is accumulated continually in the open dynamic environment, and the time stages are endless in the real world. Therefore, we mainly focus on the results of the first four rounds instead of the mandatory classification at the last round. From Table 3 to Table 5, we can see that our proposed model (both with and without KB) can achieve better NMI than all the traditional benchmark models in each dataset. In addition, the number of samples classified by TWMG-Open in each round is less than that of the benchmark models, so it can be concluded that TWMG-Open can make delayed decisions on samples that are difficult to be classified through three-way multi-granularity learning, and thus greatly improving the quality of decisions.

Then the performance of TWMG-Open with or without KB was compared. From Table 3 to Table 5, the intact TWMG-Open has been able to achieve the optimal results of NMI in Banking77 dataset and TC20 dataset, and the number of samples classified in each round is close to that of the TWMG-Open without KB. However, when using the StackOverflow dataset, TWMG-Open without KB works is better than intact TWMG-Open. With no KB, TWMG-Open delays a large number of samples in each round, which results in improved performance. Therefore, it can be explained that knowledge accumulation can reduce the samples of delay through knowledge accumulation to further reduce the cost of delay in the finer granularity level.

Next, the ablation experiments of KB are conducted to reflect the significance of KB. In reducing time consumption, we compare the decision-making time of TWMG-Open with KB and without KB. As shown in Fig. 4, the red line represents

Table 5
Results of each round on TC20.

	TC20	Round 1		Round 2		Round 3		Round 4		Round 5	
	Methods	Num	NMI	Num	NMI	Num	NMI	Num	NMI	Num	NMI
25%	TWMG-Open	1327	**54.11**	1530	**52.80**	1885	**52.08**	2078	**52.09**	2135	**20.71**
	TWMG-Open (NonKB)	1486	51.16	1789	47.98	1941	46.74	2145	45.85	1594	17.77
	NaiveDBSCAN	2200	42.95	2200	43.47	2200	42.76	2355	44.15	\	\
	AC	2200	44.96	2200	43.99	2200	44.04	2355	44.74	\	\
	MS	2200	44.50	2200	44.08	2200	45.15	2355	45.40	\	\
50%	TWMG-Open	1347	**71.05**	1479	**69.14**	1591	**68.62**	1929	67.64	2609	**30.92**
	TWMG-Open (NonKB)	1493	69.56	1493	68.64	1753	64.80	1810	**67.89**	2406	24.66
	NaiveDBSCAN	2200	52.20	2200	52.78	2200	52.10	2355	53.21	\	\
	AC	2200	57.72	2200	58.69	2200	57.47	2355	57.57	\	\
	MS	2200	52.70	2200	52.58	2200	51.87	2355	53.01	\	\
75%	TWMG-Open	1647	**75.39**	1781	**74.23**	1850	**73.56**	2262	70.74	1415	**38.47**
	TWMG-Open (NonKB)	1647	75.39	1739	73.97	1880	72.02	2056	**72.97**	1633	37.28
	NaiveDBSCAN	2200	64.37	2200	63.76	2200	63.20	2355	65.57	\	\
	AC	2200	61.64	2200	62.79	2200	62.72	2355	64.13	\	\
	MS	2200	61.02	2200	61.05	2200	61.43	2355	62.12	\	\
	SCAN	2200	39.25	2200	38.70	2200	38.72	2355	38.70	\	\

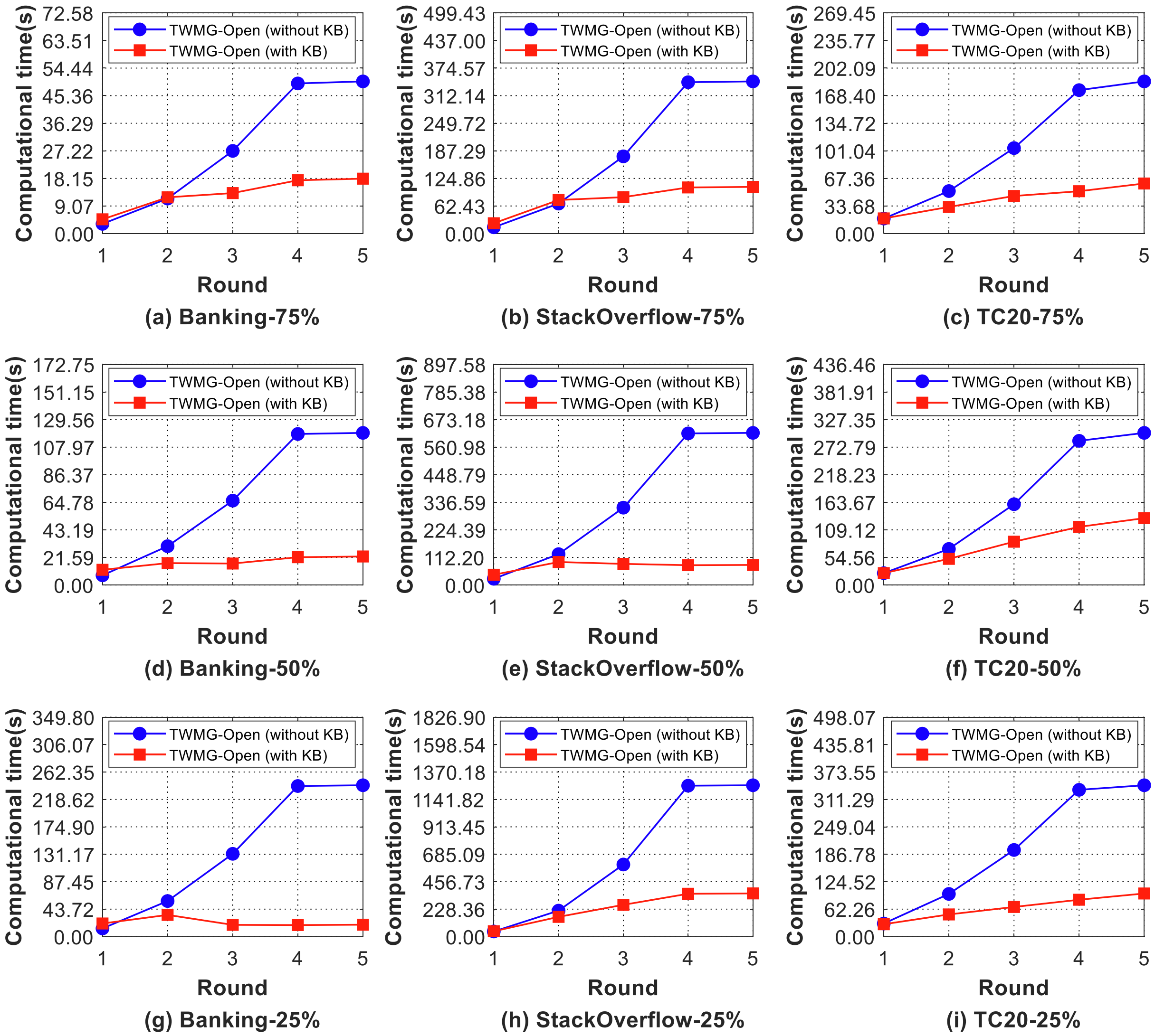

Fig. 4. Computational time of decision-making.

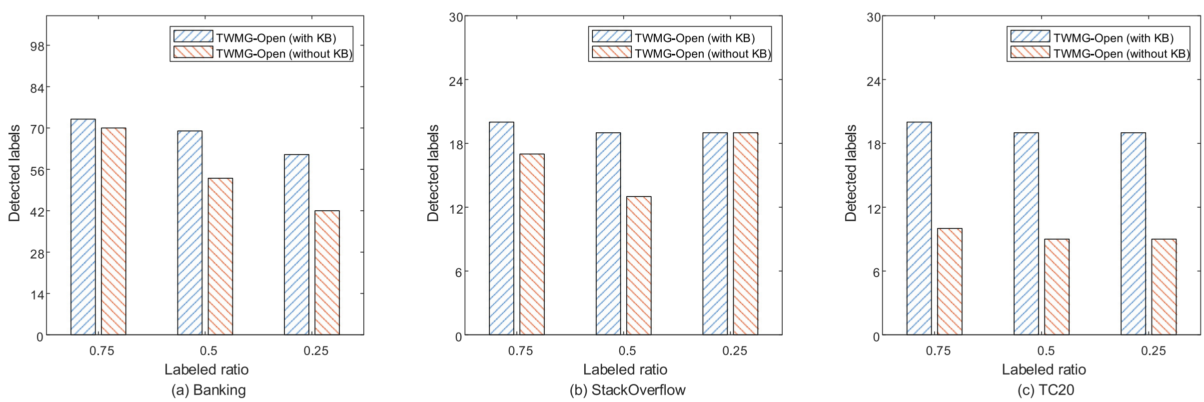

Fig. 5. Number of detected labels.

the time costs of the original TWMG-Open, while the blue line represents the time costs of the TWMG-Open without KB. It shows that KB greatly decreases the time costs in decision-making, which improves the efficiency so that the model has the ability of continual learning through knowledge accumulation.

Lastly, we compare the number of detected labels of each dataset. As shown in Fig. 5, for all datasets, the TWMG-Open with KB can find more valid labels than the model without KB, and the number of labels found is closer to the actual number of labels. Moreover, as the ratio of known categories increases from 25% to 75%, the performance of the entire TWMG-Open is constantly improved.

Thus, it can be found that the KB plays a significant role in knowledge accumulation and can improve computational efficiency and give the model the ability of continual learning. Besides, protecting data privacy is a key challenge in real application and it requires that the data used in previous tasks cannot be used by subsequent tasks. It is suitable to apply KB in such cases, so that (1) the learned knowledge instead of the data can be used in subsequent tasks, and (2) there is no need to training with previous data repeatedly. When using KB, there is no need to store a large amount of data, so it can greatly reduce the pressure of data storage.

5. Conclusions

For an open dynamic task, we are interested in detecting and managing the uncertainty, and implementing the three-way multi-granularity structure in knowledge accumulation. On the basis of three-way multi-granularity learning, the open topic classification was investigated in different levels of granularity to conduct a more efficient approach of learning knowledge continually. Compared with traditional static open systems, we used three datasets to conduct a series of experiments, which showed that TWMG-Open could improve the efficiency of knowledge accumulation by putting multiple tasks of the open dynamic environment into a framework of three-way multi-granularity learning. However, there are two challenges we need to address in future works: first, during human knowledge acquisition, some previous knowledge might be inevitably incorrect, which needs to be updated or revised; second is to interact with humans in the real world and proactively acquire the knowledge from models.

CRediT authorship contribution statement

Xin Yang: Conceptualization, Methodology, Writing – original draft. **Yujie Li:** Software, Writing – original draft. **Dan Meng:** Writing – review & editing. **Yuxuan Yang:** Software, Writing – review & editing. **Dun Liu:** Writing – review & editing. **Tianrui Li:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61773324, 61876157), the Humanity and Social Science Youth Foundation of Ministry of Education of China (No. 20YJC630191), the Fintech Innovation Center of Southwestern University of Finance and Economics, and the Financial Intelligence & Financial Engineering Key Laboratory of Sichuan Province.

References

[1] A. Bendale, T. Boult, Towards open world recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1893–1902.
[2] I. Casanueva, T. Temčinas, D. Gerz, M. Henderson, I. Vulić, Efficient intent detection with dual sentence encoders, in: Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI, 2020.
[3] Z.Y. Chen, B. Liu, Lifelong machine learning, Synthesis Lectures Artif. Intell. Mach. Learn. 12 (3) (2018) 1–207.
[4] M.J. Cheok, Z. Omar, M.H. Jaward, A review of hand gesture and sign language recognition techniques, Int. J. Mach. Learn. Cybern. 10 (1) (2019) 131–153.
[5] C. Chow, On optimum recognition error and reject tradeoff, IEEE Trans. Inf. Theory 16 (1) (1970) 41–46.
[6] A. Cohan, W. Ammar, M. van Zuylen, F. Cady, Structural scaffolds for citation intent classification in scientific publications, in, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, pp. 3586–3596.
[7] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 603–619.
[8] E.G. Dada, J.S. Bassi, H. Chiroma, A.O. Adetunmbi, O.E. Ajibuwa, et al, Machine learning for email spam filtering: review, approaches and open research problems, Heliyon 5 (6) (2019) e01802.
[9] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, pp. 4171–4186.
[10] M. Ester, H.P. Kriegel, J. Sander, X.W. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the 2nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 96, 1996, pp. 226–231.

[11] F. Min, S.M. Zhang, D. Ciucci, M. Wang, Three-way active learning through clustering selection, Int. J. Mach. Learn. Cybern. 11 (5) (2020) 1033–1046.
[12] G.L. Fei, B. Liu, Breaking the closed world assumption in text classification, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 506–514..
[13] G.L. Fei, S. Wang, B. Liu, Learning cumulatively to become more knowledgeable, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1565–1574.
[14] Y. Geifman, R. El-Yaniv, Selective classification for deep neural networks, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 4885–4894.
[15] Y. Geifman, R. El-Yaniv, Selectivenet: A deep neural network with an integrated reject option, in: Proceedings of the 36th International Conference on Machine Learning, 2019, pp. 2151–2159.
[16] D. Hendrycks, K. Gimpel, A baseline for detecting misclassified and out-of-distribution examples in neural networks, in: Proceedings of the 5th International Conference for Learning Representations (ICLR), 2017.
[17] R. Herbei, M.H. Wegkamp, Classification with reject option, The Canadian Journal of Statistics/La Revue Canadienne de Statistique (2006) 709–721.
[18] P.X. Huang, B.J. Boom, R.B. Fisher, Hierarchical classification with reject option for live fish recognition, Mach. Vis. Appl. 26 (1) (2015) 89–102.
[19] H.R. Ju, W. Pedrycz, H.X. Li, W.P. Ding, X.B. Yang, X.Z. Zhou, Sequential three-way classifier with justifiable granularity, Knowl.-Based Syst. 163 (2019) 103–119.
[20] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Proceedings of the third International Conference for Learning Representations (ICLR), 2015.
[21] K. Lee, D. Palsetia, R. Narayanan, M.M.A. Patwary, A. Agrawal, A. Choudhary, Twitter trending topic classification, in: 2011 IEEE 11th International Conference on Data Mining Workshops, 2011, pp. 251–258.
[22] H.X. Li, L.B. Zhang, B. Huang, X.Z. Zhou, Sequential three-way decision and granulation for cost-sensitive face recognition, Knowl.-Based Syst. 91 (2016) 241–251.
[23] H.X. Li, L.B. Zhang, X.Z. Zhou, B. Huang, Cost-sensitive sequential three-way decision modeling using a deep neural network, Int. J. Approximate Reasoning 85 (2017) 68–78.
[24] T.E. Lin, H. Xu, A post-processing method for detecting unknown intent of dialogue system via pre-trained deep neural network classifier, Knowl.-Based Syst. 186 (2019) 104979.
[25] T.E. Lin, H. Xu, H. Zhang, Discovering new intents via constrained deep adaptive clustering with cluster refinement, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 8360–8367..
[26] B. Liu, Learning on the job: Online lifelong and continual learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 13544–13549..
[27] B. Liu, S. Mazumder, Lifelong and continual learning dialogue systems: Learning during conversation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021..
[28] D. Liu, X. Yang, T.R. Li, Three-way decisions: beyond rough sets and granular computing, Int. J. Mach. Learn. Cybern. 11 (5) (2020) 989–1002.
[29] F. Murtagh, P. Legendre, Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion, J. Classif. 31 (3) (2014) 274–295.
[30] I. Pillai, G. Fumera, F. Roli, Multi-label classification with a reject option, Pattern Recogn. 46 (8) (2013) 2256–2266.
[31] J. Qian, C.H. Liu, D.Q. Miao, X.D. Yue, Sequential three-way decisions via multi-granularity, Inf. Sci. 507 (2020) 606–629.
[32] W.J. Scheirer, A. de Rezende Rocha, A. Sapkota, T.E. Boult, Toward open set recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (7) (2012) 1757–1772.
[33] L. Shu, H. Xu, B. Liu, Doc: Deep open classification of text documents, in, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2911–2916.
[34] M. Tapaswi, M.T. Law, S. Fidler, Video face clustering with unknown number of clusters, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 5027–5036.
[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017.
[36] X. Yang, T.R. Li, D. Liu, H. Fujita, A temporal-spatial composite sequential approach of three-way granular computing, Inf. Sci. 486 (2019) 171–189.
[37] X. Yang, Y.Y. Zhang, H. Fujita, D. Liu, T.R. Li, Local temporal-spatial multi-granularity learning for sequential three-way granular computing, Inf. Sci. 541 (2020) 75–97.
[38] Y.Y. Yao, Three-way decisions with probabilistic rough sets, Inf. Sci. 180 (3) (2010) 341–353.
[39] Y.Y. Yao, Three-way decision and granular computing, Int. J. Approximate Reasoning 103 (2018) 107–123.
[40] Y.Y. Yao, Tri-level thinking: models of three-way decision, Int. J. Mach. Learn. Cybern. (2019) 1–13.
[41] Y.Y. Yao, Three-way granular computing, rough sets, and formal concept analysis, Int. J. Approximate Reasoning 116 (2020) 106–125.
[42] Y.Y. Yao, Set-theoretic models of three-way decision, Granular Comput. 6 (1) (2021) 133–148.
[43] H. Yu, L.Y. Chen, J.T. Yao, X. Wang, A three-way clustering method based on an improved dbscan algorithm, Phys. A 535 (2019) 122289.
[44] H. Yu, C. Zhang, G.Y. Wang, A tree-based incremental overlapping clustering method using the three-way decision theory, Knowl.-Based Syst. 91 (2016) 189–203.
[45] C. Zhang, D.Y. Li, J.Y. Liang, Multi-granularity three-way decisions with adjustable hesitant fuzzy linguistic multigranulation decision-theoretic rough sets over two universes, Inf. Sci. 507 (2020) 665–683.
[46] H.L. Zhang, H. Xu, T.E. Lin, Deep open intent classification with adaptive decision boundary, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 14374–14382..
[47] L.B. Zhang, H.Z. Li, X.Z. Zhou, B. Huang, Sequential three-way decision based on multi-granular autoencoder features, Inf. Sci. 507 (2020) 630–643.
[48] Y.B. Zhang, D.Q. Miao, J.Q. Wang, Z.F. Zhang, A cost-sensitive three-way combination technique for ensemble learning in sentiment classification, Int. J. Approximate Reasoning 105 (2019) 85–97.
[49] Y.B. Zhang, Z.F. Zhang, D.Q. Miao, J.Q. Wang, Three-way enhanced convolutional neural networks for sentence-level sentiment classification, Inf. Sci. 477 (2019) 55–64.
[50] W. Van Gansbeke, S. Vandenhende, S. Georgoulis, M. Proesmans, L. Van Gool, Scan: Learning to classify images without labels, European Conference on Computer Vision (2020) 268–285.