



Text GCN-SW-KNN: a novel collaborative training multi-label classification method for WMS application themes by considering geographic semantics

Zhengyang Wei, Zhipeng Gui, Min Zhang, Zelong Yang, Yuao Mei, Huayi Wu, Hongbo Liu & Jing Yu

To cite this article: Zhengyang Wei, Zhipeng Gui, Min Zhang, Zelong Yang, Yuao Mei, Huayi Wu, Hongbo Liu & Jing Yu (2021) Text GCN-SW-KNN: a novel collaborative training multi-label classification method for WMS application themes by considering geographic semantics, Big Earth Data, 5:1, 66-89, DOI: [10.1080/20964471.2021.1877434](https://doi.org/10.1080/20964471.2021.1877434)

To link to this article: <https://doi.org/10.1080/20964471.2021.1877434>



© 2021 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the CASEarth Strategic Priority Research Programme.



Published online: 24 Feb 2021.



Submit your article to this journal [↗](#)



Article views: 2325



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 4 View citing articles [↗](#)

Text GCN-SW-KNN: a novel collaborative training multi-label classification method for WMS application themes by considering geographic semantics

Zhengyang Wei^a, Zhipeng Gui^{a,b,c}, Min Zhang^{b,c}, Zelong Yang^{b,c}, Yuao Mei^b, Huayi Wu^{b,c}, Hongbo Liu^d and Jing Yu^d

^aSchool of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China; ^bState Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; ^cCollaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan, China; ^dInformation Technology Department, Chongqing Geomatics and Remote Sensing Center, Chongqing, China

ABSTRACT

Without explicit description of map application themes, it is difficult for users to discover desired map resources from massive online Web Map Services (WMS). However, metadata-based map application theme extraction is a challenging multi-label text classification task due to limited training samples, mixed vocabularies, variable length and content arbitrariness of text fields. In this paper, we propose a novel multi-label text classification method, Text GCN-SW-KNN, based on geographic semantics and collaborative training to improve classification accuracy. The semi-supervised collaborative training adopts two base models, i.e. a modified Text Graph Convolutional Network (Text GCN) by utilizing Semantic Web, named Text GCN-SW, and widely-used Multi-Label K-Nearest Neighbor (ML-KNN). Text GCN-SW is improved from Text GCN by adjusting the adjacency matrix of the heterogeneous word document graph with the shortest semantic distances between themes and words in metadata text. The distances are calculated with the Semantic Web of Earth and Environmental Terminology (SWEET) and WordNet dictionaries. Experiments on both the WMS and layer metadata show that the proposed methods can achieve higher F1-score and accuracy than state-of-the-art baselines, and demonstrate better stability in repeating experiments and robustness to less training data. Text GCN-SW-KNN can be extended to other multi-label text classification scenario for better supporting metadata enhancement and geospatial resource discovery in Earth Science domain.

ARTICLE HISTORY

Received 3 August 2020
Accepted 10 January 2021

KEYWORDS

Web map service; multi-label text classification; semantic distance; text graph convolutional network; collaborative training; ML-KNN; application theme extraction

1. Introduction

Accurate map application theme classifications, e.g. disasters and ecosystems, can facilitate online geospatial resources discovery and assist end users from different disciplines to find target map resources effectively.

Volunteered Geographic Information (VGI) and Open Government Data Initiatives promote the advancement of public accessible online geospatial resources, including massive map

services in the form of Open Geospatial Consortium (OGC) WMS. However, due to the limit of the existing metadata description mechanisms, including International Standards Organization (ISO) 19119¹ and Content Standard for Digital Geospatial Metadata (CSDGM) from Federal Geographic Data Committee (FGCD),² there is no explicit description of map application themes for WMS (Zhang, Gui, Cheng, Cao, & Wu, 2019). As the result, end users from cross-disciplinary fields of Earth and Social Sciences cannot retrieve maps that match their desired topics efficiently and accurately using existing full-text indexing and query techniques. Although existing metadata fields may depict application themes implicitly, such as title, keywords and abstract, there are usual plain texts with variable lengths, arbitrary content description but no strict content regulation. Thus, extra text processing is needed to extract theme labels for metadata text for further facilitating query and discovery. The extraction process can be treated as a multi-label text classification problem since a map layer may link with multiple application themes, such as *climate* and *disasters*. Compared to existing text classification problems, achieving accurate multi-label classification of map application themes is faced with following challenges:

- (a) **Lack of labeled training samples.** Unlike many widely used datasets, such as IMDB and OpinRank, there is no existing large-scale labeled dataset publicly available for WMS metadata classification. Deep learning techniques are usually data hungry methods (Berger, 2014; Li, Gui, Cheng, Wu, & Qin, 2019). Limited size of sample sets makes it unsuitable for models that require intensive training to achieve a high accuracy rate.
- (b) **Metadata texts with variable lengths and formats.** There is no explicit content constrain for the metadata fields that may contain theme information, including Title, URL, Abstract and Keywords etc. As a result, the fields may be filled with insufficient or irrelative information, e.g. very short description or even missing fields, which has a negative impact on the classification accuracy. Moreover, the text may consist of limited sentences or just a phrase or word simultaneously, so the classification methods based on sequence orders or grammar might not be applicable.
- (c) **Geoscience terminologies mixed with generic vocabularies.** Besides the daily expression and vocabularies, there are also many geoscience domain-specific terminologies embedded in WMS metadata, such as *NDVI* and *neve*. These terms are often closely related to the map application themes. Underutilization of geographic semantics may impede the understanding and extraction of the useful information effectively.

These features cause low accuracy of application theme classification of the existing methods (Zhang et al., 2019). Therefore, this paper proposed a novel multi-label classification method, Text GCN-SW-KNN, which utilizes semi-supervised collaborative training and geographic semantics corporately to address aforementioned issues. To be more specific, the developed collaborative training model consists of two base models, i.e. a basic ML-KNN model and a modified Text GCN model, Text GCN-SW. Text GCN-SW uses the shortest semantic distance between themes and vocabularies in metadata to adjust the adjacency matrix of the heterogeneous word document graph. The experiments of

classification performance comparison, stability test and training data size analysis verify the effectiveness of the proposed method.

The remainder of this paper is organized as follows. [Section 2](#) reviews the related works. [Section 3](#) describes the proposed method. Experiments in [section 4](#) demonstrate the strength of proposed methods. Then, [Section 5](#) concludes this article and points out future research.

2. Literature review

2.1. Geospatial resource discovery

A large number of Spatial Data Infrastructures (SDIs), have been built by international organizations or government sections to facilitate Earth Science data sharing and discovery, including Data.gov,³ Global Change Master Directory (GCMD) from National Aeronautics and Space Administration (NASA)², National Centers for Environmental Information (NCEI) from National Oceanic and Atmospheric Administration (NOAA) and Global Earth Observing System of Systems (GEOSS) Clearinghouse (Liu et al., 2011). These SDIs provide standardized APIs and web portals to support efficient keyword-based search and spatiotemporal query, e.g. temporal and spatial coverage, data providers, spatial projections, even any text in the metadata, with the help of the state-of-the-art full-text and spatiotemporal indexing mechanisms. However, Text-matching-based solutions are limited by the metadata, and may be incapable of distinguishing the retrieved geospatial resources with similar metadata descriptions. So, different strategies have been adopted to refine query by introducing extra information for decision-making. Continuous performance monitorings of WMSs are used to provide quality evaluation for selecting WMSs which provide same or similar maps (Gui, Cao, Liu, Cheng, & Wu, 2016; Wu, Li, Zhang, Yang, & Shen, 2011). A cloud-based search broker, GeoSearch, integrates data visualization, interactive filtering technologies, and service quality information to help end users narrow down the retrieved candidates (Gui et al., 2013a). Image contents of WMS layers (Yang, Gui, Wu, & Li, 2019) and user relevance feedback were also used in retrieval to deal with semantic gaps in human-computer interactions (Hu, Gui, Cheng, Qi, & Wu, 2016; Li et al., 2019). However, most of these methods were limited to similarity matching, and is unable to perceive geographic semantics in map services (Yang et al., 2019).

In the view of these problems, semantics and context analysis of metadata became a research hotspot. Combining semantic ontology is an effective approach of modeling the latent semantic relationships among data and refining the classification of OGC services (Gui, Yang, Xia, Liu, & Lostritto, 2013b; Li et al., 2011). The Semantic Web for Earth and Environmental Terminology (SWEET) (Raskin & Pan, 2005) helps to build domain-specific ontology graphs for accurate and systematic geospatial resource description, e.g. hydrology ontology. ESIP Semantic testbed has been also incorporated to develop semantic support system for data discovery. These methods improved the accuracy of geospatial resources retrieval for domain experts to some extends. However, there is still a large gap between domain-specific geographic semantics expression and search criteria used by mass or cross disciplinary users. An effective application theme labeling mechanism is still needed to close this gap by making the data description easily to be understood by common users, and in turn, achieve efficient data discovery.

A Labeled Latent Dirichlet Allocation (LLDA) model was proposed to assign themes to metadata records so as to address the problem of metadata topic heterogeneity caused by multiple standards (Hu, Janowicz, Prasad, & Gao, 2015). An unsupervised application themes classification and metadata extension mechanism was proposed to better support WMS retrieval (Zhang et al., 2019) by using geographic semantics. These methods can assist end users in obtaining service resources quickly in the desired fields and have great reference value for our work. However, the proposed unsupervised classification algorithms are limited due to the complexity and diversity of metadata expression. It may lead to a low adaptability owing to the empirical parameter settings, the short text and even the absence of metadata fields.

2.2. Text multi-label classification

Multi-label learning is the key to achieving WMS application theme classification, and it has received significant attention in the past few years. According to whether the method transforms the problem into single-label classification/regression problems or extends it into specific learning algorithms, multi-label learning methods can be categorized into two major types, i.e. problem transformation methods and algorithm adaptation methods (Tsoumakas & Katakis, 2007).

The simplest problem transformation strategy is to convert the multi-label classification into several binary classification problems, which is known as the binary relevance (BR) method (Tsoumakas & Katakis, 2007). Specifically, SVM (Godbole & Sarawagi, 2004), Naïve Bayes (John & Langley, 1995) and many other algorithms can be used to tackle binary classification problems. Deep learning methods can also be applied to BR method, including CNN and RNN. Classifier Chain (CC) method is closely related to the BR method and involves Q binary classifiers linked along a chain (Read, Pfahringer, Holmes, & Frank, 2011). These methods are simple and easy to understand, but ignore correlations between labels. Another type of problem transformation method is the label power-set method (LP), whose basis is to combine entire label sets into an atomic label to form a single-label problem (Read, Pfahringer, & Holmes, 2008; Tsoumakas & Katakis, 2007). CNN-RNN model converted the multi-label classification problem into label sequence predictions, where the label sequence is the assignment of ordered labels to a text and LSTMs are used for label sequence predictions (Chen, Ye, Xing, Chen, & Cambria, 2017). Problem transformation methods are easy to implement and any traditional efficient classification algorithm can be used as the basic classifier. However, as the size of the label sets increases, the algorithmic complexity also increases rapidly, so many researchers turn their attention to algorithm adaptation methods.

Algorithm adaptation methods focus on adapting, extending, and customizing the existing machine learning algorithm for the task of multi-label learning (Madjarov, Kocev, Gjorgjevikj, & Deroski, 2012). The boosting algorithm BoosTexter is proposed for multi-label classification problems, which can be divided into two algorithms, i.e. Boost.MH and AdaBoost.MR (Schapire & Singer, 2000). Boost.MH is designed to minimize Hamming loss, while AdaBoost.MR is to find a hypothesis which ranks the correct labels at the top of the ranking. Rank-SVMs was used to handle multi-label problems with a large margin ranking system that shares a lot of common properties with SVMs (Elisseeff & Weston, 2002). ML-KNN, the multi-label version of KNN, utilizes the maximum a posteriori principle to

determine the label set for those unseen instances (Zhang & Zhou, 2007). Binary Relevance K-Nearest Neighbor (BR-KNN) extends the KNN algorithm so that independent predictions are made for each label, following a single search of the K nearest neighbors (Spyromitros, Tsoumakas, & Vlahavas, 2008). Sigmoid function was also used on the output layer of neural network models instead of using a rectified linear unit as activation function, which has shown significant improvements on Convolutional Neural Network (CNN) and Gate Recurrent Unit (GRU) models (Berger, 2014). XML-CNN went beyond other deep learning methods for multi-class classification by using a dynamic max pooling scheme, a binary cross-entropy loss that is more suitable for multi-label problems (Liu, Chang, Wu, & Yang, 2017). A neural network initialization method was proposed to embed the label co-occurrence information between the hidden and output layers with the initial weights set to upper bound (Kurata, Xiang, & Zhou, 2016). In general, compared with problem transformation methods, algorithm adaptation methods can achieve higher classification accuracy with lower computational cost.

2.3. Deep learning methods

Problem transformation methods and algorithm adaptation methods can allow many popular text classification methods to be used for the text multi-label classification problem, especially some deep learning method (Berger, 2014; Chen et al., 2017; Kurata et al., 2016).

CNNs with one-dimensional convolutions have been directly used for sentence classification (Kim, 2014). The model of character-level convolutional networks has been used as an effective method for text classification (Zhang, Zhao, & Lecun, 2015). Long Short-Term Memory recurrent neural network (LSTM), a specific type of Recurrent Neural Network (RNN) with a more complex computational unit, has been applied to learn the text representation and have obtained stronger results on a variety of sequence modeling tasks (Liu, Qiu, & Huang, 2016; Tai, Socher, & Manning, 2015). A unified model called C-LSTM utilizes CNN to extract a sequence of higher-level phrase representations, and utilizes a LSTM to obtain the sentence representation (Zhou, Sun, Liu, & Lau, 2015). Attention-based LSTM for aspect-level sentiment classification has been proposed, which can concentrate on different parts of a sentence when different aspects are taken as input (Wang, Huang, Zhu, & Zhao, 2016). Attention mechanisms have also been applied at the word or sentence-level to enable it to attend differentially to more and less important content when constructing the document representation and achieving text classification (Yang, Yang, Dyer, He, & Hovy, 2016; Zhou, Shi, Tian, Qi, & Xu, 2016). These methods mainly focus on local word sequences, but ignore the global word co-occurrence information in a corpus. However, many WMS metadata texts are composed of phrases and words, rather than complete sentences. So, the local word sequences may not help with the classification, and the global word co-occurrence information is useful for classification.

Graph Neural Networks (GNNs) have been explored for text classification (Henaff, Bruna, & Lecun, 2015; Kipf & Welling, 2017), which is capable of encoding both graph structure and node features. GNNs are computationally efficient in general, but they either viewed a document or a sentence as a graph of word node. So, the not-routinely-available inter-document relations are still desired to be modeled (Yao, Mao, & Luo, 2018). Text Graph Convolutional Networks (Text GCN) is proposed furtherly which uses heterogeneous word document graph and then turn document classification into a node classification problem

(Yao et al., 2018). Text GCN achieves promising results because it is capable to capture global word co-occurrence information and utilize limited labeled documents.

Therefore, we choose Text GCN to perform the classification. However, it doesn't take geographic semantics into consideration, leading to poor classification results when there are some words with high term frequency-inverse document frequency (TF-IDF), which may not be related to a certain theme. Meanwhile, the labeled training samples are limited, so a semi-supervised training mechanism is required. Due to the reasons above, the basic idea of this paper is to build a semi-supervised collaborative training model, Text GCN-SW-KNN, to improve multi-label classification accuracy of map application themes with limited labeled training samples. Text GCN-SW-KNN consists of two base models, i.e. Text GCN-SW and ML-KNN. The proposed Text GCN-SW in this paper is upon Text GCN by considering geographic semantics. In the view of our application scenario, the shortest distance between the words to the themes is an important measurement for the geography semantics similarity. So, we constructed a geographic semantic network based on the SWEET and WordNet for the shortest distance calculation. In Text GCN-SW, the shortest distance is regarded as a part of weights to adjust the adjacency matrix of Text GCN to improve classification accuracy. Meanwhile, the activation function of output layer and the loss function are modified to make it suitable to tackle with the multi-label classification problem.

3. Methodology

To facilitate cross-domain geospatial resources discovery, appropriate application theme setting is critical. According to the definition of Group on Earth Observations (GEO),⁴ geospatial resources can be classified into nine Societal Benefit Areas (SBAs), including *Agriculture, Biodiversity, Climate, Disasters, Ecosystems, Energy, Health, Water, and Weather*, which researchers in different fields are interested in. In addition, there are also a large amount of geological data provided by United States Geological Survey (USGS) and other contributors available online in the form of OGC WMS (Gui et al., 2016), such as USGS Mineral Resources Online Spatial Data.⁵ Considering the search demands in the geological field, we combine *Geology* with other nine themes from SBAs and define ten application themes in this paper. Based on the 10 themes, we propose our multi-label classification model for WMS metadata text for supporting geospatial resource retrieval furtherly.

3.1. Model architecture

The overall architecture of our collaborative training model for multi-label text classification, Text GCN-SW-KNN, is illustrated in Figure 1, which consists of two base models, ML-KNN and Text GCN-SW. More specifically, besides the widely-used ML-KNN model, we proposed a Text GCN-based multi-label text classification method, which combines the geographic semantics to improve the structure of GCN. The two base models work collaboratively to achieve semi-supervised classification.

As illustrated in Figure 1, to build the collaborative training model, the construction of the base model Text GCN-SW is the key. Firstly, the shortest geographic semantic distances from the words to the themes are calculated using the semantic network constructed with the SWEET and WordNet, which is the basis of our proposed method.

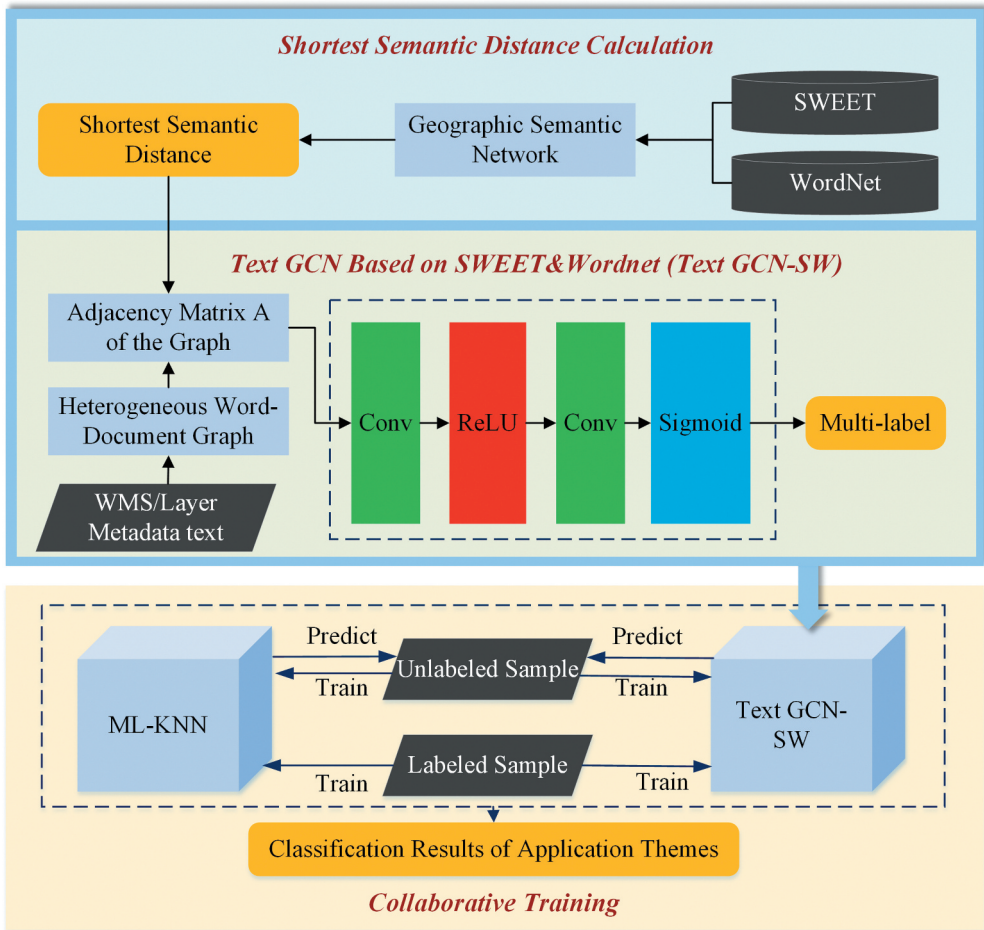


Figure 1. Architecture of the proposed collaborative training model Text GCN-SW-KNN by cooperating ML-KNN and text GCN-SW, where the first two tiers show the working mechanism of Text GCN-SW.

Then, the shortest semantic distances are used as a part of the weights to modify the adjacency matrix of the original Text GCN model to reduce the impact of topic-irrelevant words on classification results. Meanwhile, to build the new base model, Text GCN-SW, we also adjusted the activation function of output layer and the loss function to achieve multi-label classification. Finally, semi-supervised collaborative training model is built upon the two base models, i.e. Text GCN-SW, and widely-used ML-KNN, to achieve multi-label classification with higher accuracy and limited samples.

3.2. Shortest semantic distance calculation

The calculation of the semantic distances between feature words and themes is the basis for measuring the potential theme associations. The WMS metadata obtained through *GetCapabilities* operation contains a large amount of geoscience domain-specific terminologies as well as daily vocabulary and generic feature words. Both of them are

important for application theme classification. Therefore, we use SWEET (Raskin & Pan, 2005) and WordNet (Fellbaum & Miller, 1998) corporately to calculate the semantic distances. The ontological model SWEET is for measuring geoscience-related concepts and terminologies, while the widely-used semantic network WordNet is for daily expression. The calculation process includes the following two steps.

(1) Find an alternative word B of the feature word A in both SWEET and WordNet

To achieve theme matching, the first step is to find an alternative word of the feature word in two semantic networks. The search starts from SWEET first, which consists of a collection of domain-specific ontologies for the Earth and environmental sciences and supporting areas, modeled in the web ontology language (OWL).⁶ Since all theme words are included in SWEET, it can help to find the matched themes efficiently and accurately. The distance of each edge of Wordnet is defined according to the principles that the distance between two words with the same meaning should be 0, and the distance between a pair of hypernyms, hyponyms, entailments, or antonyms should be 1. If the feature word A is included in SWEET as shown in Figure 2(a), the alternative word B is A itself, and the shortest distance D_1 between A and B should be 0; while if the feature word A is not included in SWEET as shown in Figure 2(b), WordNet is used to search the hypernym or hyponym of the feature word A iteratively until finding a word that is defined in SWEET. This word is treated as the alternative word B of the feature word A , and the shortest distance D_1 between A and B is calculated by the distance in WordNet. In Figure 2(b), the hypernym of feature word *Neve* is *Ice* which is in SWEET, so we can regard *Ice* as the alternative word B and $D_1 = 1$; in Figure 2(c), the hypernym of feature word *Virga* is *Precipitation*, and *Sleet* in SWEET is one of the hyponyms of *Precipitation*. So, *Sleet* is the alternative word B and $D_1 = 2$.

(2) Calculate the shortest distance between the feature word A and the theme T

The second step is to calculate the shortest distance between the feature word A and the theme D_3 . To achieve that, we use SWEET to calculate the shortest distance D_2 between the alternative word B and the theme T , then D_3 can be calculated as the sum of two distances, i.e. $D_3 = D_1 + D_2$.

The method of finding the shortest path in SWEET is to generate an undirected graph based on the network structure defined by SWEET and WordNet. Dijkstra algorithm is used to find the shortest path between alternative word B and the theme T . In SWEET, there are many kinds of relationships between two ontologies. Among all the relationships, we define the distance between two ontologies as 0 if their relationship is *equivalentClass* or *sameAs*; while the distance is set as 1 if belongs to one of the following relationships: *disjointWith*, *approximates*, *differentFrom*, *equivalentProperty*, *inverseOf*, *subclassOf*, *hasSource*, *range*, *subPropertyOf*, *domain*, *hasBaseUnit*, *hasRole*, *hasAstronomicalBody*, *hasOperand*, *hasUnit*, *hasPeriod*, *strongerThan*, *largerThan*, *greaterVerticalExtentThan*, *fartherThan*, *largerScaleThan*, *moreActiveThan*, *warmerThan*, and *moreFrequentThan*. As shown in Figure 2(a), *Glacier* is included in SWEET, so $D_1 = 0$, $D_2 = 3$, and then the shortest distance between *Glacier* and theme *Water* D_3 is 3. While in Figure 2(b), *Neve* is not included in SWEET, its alternative word B is *Ice*, so $D_1 = 1$, $D_2 = 2$, and then the shortest distance between *Neve* and theme *Water*

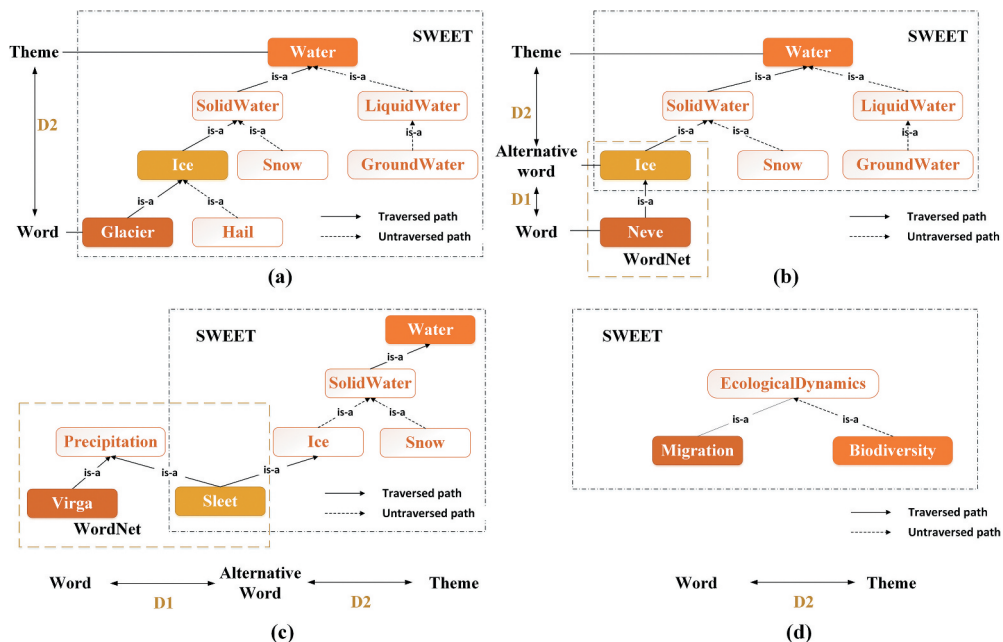


Figure 2. Exemplary shortest path between feature words and theme within SWEET and WordNet. (a) Feature word A is included in SWEET; (b) Feature word A is not included in SWEET but in WordNet; (c) Feature word A and alternative word B are two subclasses away; (d) Feature word A and Theme T are two subclasses away.

D_3 is 3. Figure 2(c) shows the situation that feature word *Virga* and alternative word *Sleet* are two subclasses away. Since $D_1 = 2$ and $D_2 = 3$, the shortest distance between *Virga* and theme *Water* D_3 is 5. As shown in Figure 2(d), theme *Biodiversity* is a subclass of *EcologicalDynamics*, and for the feature word *Migration*, $D_3 = D_2 = 2$.

3.3. Text GCN based on SWEET & WordNet (Text GCN-SW)

We proposed a modified Text GCN as a base model for collaborative training by utilizing the calculated shortest semantic distance. A GCN is a multi-layer neural network that operates directly on a graph and induces embedding vectors of nodes based on properties of their neighborhoods (Kipf & Welling, 2017). In Text GCN (Yao et al., 2018), a graph $G = (V, E)$ consists of a set of nodes $V(|V| = n)$, including word nodes and document nodes, and a set of edges E , which explicitly model the global word co-occurrence. The feature matrix $X = I$ is an identity matrix where every word or document is represented as a one-hot vector as the input to Text GCN. The adjacency matrix A of the graph G and its degree matrix D are introduced to construct the GCN. A two-layer GCN is computed as formula 1:

$$Z = f(X, A) = \rho_2(\tilde{A}\rho_1(\tilde{A}XW_0)W_1) \quad (1)$$

where $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is the normalized symmetric adjacency matrix, W_i is weight matrixes that can be trained by gradient descent, and ρ_i is activation functions. Particularly, for

a two-layer Text GCN, ρ_1 is defined as RELU function, ρ_2 is defined as *Softmax* function, and the loss function is defined as the cross-entropy error over all labeled.

In order to address our problem better, we propose an improved Text GCN based on SWEET & WordNet, named Text GCN-SW, which changes the adjacency matrix A of the graph, the activation function of the output layer and the loss function of original Text GCN to achieve high accuracy in multi-label text classification.

(1) Adjacency matrix A

In Text GCN, the weight of the edge between a document node and a word node is the term frequency-inverse document frequency (TF-IDF) of the word in the document, where term frequency is the number of times the word appears in the document, inverse document frequency is the logarithmically scaled inverse fraction of the number of documents that contain the word. In order to integrate geographic semantics into the network, in our Text GCN-SW, we change the weight of the edge from TF-IDF to TF-IDF weighted by geographic semantic distance. It helps the model to reduce the influence of topic-irrelevant words that could lead to uncertainty and ensure the efficiency of gradient descent (Ruder, 2016). While, the weight of edge between two words remains unchanged, i.e. calculated by point-wise mutual information (PMI), a popular measure for word associations. The adjacency matrix A is defined in Equation 2:

$$A_{ij} = \begin{cases} \text{PMI}(i, j) & i, j \text{ are words, } \text{PMI}(i, j) > 0 \\ s(j) \cdot \text{TF} - \text{IDF}_{ij} & i \text{ is document, } j \text{ is word} \\ 1 & i = j \\ 0 & \text{other} \end{cases} \quad (2)$$

$$d(j) = \min_{y_t \in Y} d(j, y_t), \quad s(j) = \frac{1}{d(j) + 1} \quad (3)$$

where y_t is the t th element in the set of themes Y , and $d(j, y_t)$ denotes the shortest distance between word j and a theme y_t , and $d(j)$ is the shortest distance among all the themes for word j because our model is multi-label classification method rather than a binary classification model for each theme. By considering the label correlation, it can avoid the misclassification problem in binary relevance method to a certain extent (Berger, 2014). $s(j)$ calculated as Equation 3 is to adjust the weight of the edge between a document node and the word node for j . In turn, Text GCN can capture the co-occurrence of global word along with the geographic semantics.

(2) Activation function of the output layer and Loss function

To achieve multi-label classification, the activation function of the output layer is changed to a sigmoid activation function as shown in Equation 4, which can produce a probability for each of the potential labels. Meanwhile, binary cross-entropy loss is used as the loss function as Equation 5.

$$Z = f(X, A) = \text{sigmoid}(\tilde{A}\text{RELU}(\tilde{A}XW_0)W_1) \quad (4)$$

$$L = -\frac{1}{n} \sum_i^n \sum_j^L [Y_{ij} \ln Z_{ij} + (1 - Y_{ij}) \ln (1 - Z_{ij})] \quad (5)$$

where $\text{sigmoid}(x) = \frac{1}{1+e^{-x}}$, Y is the label indicator matrix whose index can be 0 to n , and L is the dimension of the output features, which is equal to the number of themes. The weight matrix W_0 and W_1 can be trained through gradient descent to minimize the loss.

3.4. Collaborative training

The basic idea of the collaborative training algorithm is that two classifiers can provide useful information mutually to improve the performance of the classifiers by re-training them iteratively using the increasing annotated data obtained from the classification results of unlabeled data (Blum & Mitchell, 1998). When one classifier cannot classify a sample confidently, while another classifier may have enough information to correctly classify it. Based on the proposed Text GCN-SW and the widely-used ML-KNN, our co-training algorithm is illustrated in Table 1.

Each sample in the labeled training sample list or the unlabeled test sample list is associated with the attributes including id, label, document. For each iteration, current labeled training sample set is used to train both ML-KNN and Text GCN-SW models. The two models can generate two classification results for each test sample, then if the two results are same, the sample should be removed from the test sample list and added to the training sample list for next iteration. The iteration will terminate when the classification results of the two base models are same, or when the results of the two iterations do not change, or when the number of iterations exceeds the maximum iteration limit. For those samples that cannot be classified into the same labels by two models, we choose the classification results of ML-KNN as their final labels to achieve the best overall accuracy.

Table 1. Co-training algorithm based on Text GCN-SW and ML-KNN.

Input: labeled training sample list (<i>train_list</i>) and unlabeled test sample list (<i>test_list</i>)	
Output: newly labeled test sample list <i>test_list</i>	
1:	Initialize iteration counter $r = 0$ for the co-training model
2:	Copy a new list <i>incon_list</i> = <i>test_list</i> for test samples with inconsistent labels by base classifiers
3:	Initialize changing detection flag <i>changed</i> = true for <i>incon_list</i>
4:	while (<i>incon_list</i> .length() > 0 and <i>changed</i> != false and $r < \text{max_iteration}$) do
5:	Reset <i>changed</i> = false
6:	Train ML-KNN using <i>train_list</i> as base classifier C_1
7:	Train Text GCN-SW using <i>train_list</i> as base classifier C_2
8:	for each i in range [0, <i>incon_list</i> .length()-1] do
9:	Assign <i>incon_list</i> [i].label = C_1 .classify(<i>incon_list</i> [i].document)
10:	Create a new label <i>GCN_label</i> = C_2 .classify(<i>incon_list</i> [i].document)
11:	if <i>incon_list</i> [i].label == <i>GCN_label</i> then
12:	<i>train_list</i> .add(<i>incon_list</i> [i])
13:	<i>incon_list</i> .remove(<i>incon_list</i> [i])
14:	<i>changed</i> = true
15:	$r++$
16:	for each j in range [0, <i>test_list</i> .length()-1] do
17:	if <i>test_list</i> [j] corresponded test sample in <i>train_list</i> then
18:	Assign <i>test_list</i> [j].label = <i>train_list</i> .getItemById(<i>test_list</i> [j].id).label
19:	else if <i>test_list</i> [j] corresponded test sample in <i>inconsistent_list</i> then
20:	Assign <i>test_list</i> [j].label = <i>incon_list</i> .getItemById(<i>test_list</i> [j].id).label

4. Experiment

In order to verify the feasibility of our method, we defined three experiments. The overall performance evaluation experiment verifies the classification performance of our method by comparing it with eight baselines. The stability assessment experiment demonstrates the stability of the model incorporated with geographic semantics. The training data setting experiment analyzes the applicability of the model to different proportions of training data. The last two experiments also are the basis to verify the feasibility of the base model selection in collaborative training model design.

4.1. Data Collection and preprocessing

The experiment data is 46,298 OGC WMSs acquired via Topic-focused Web Crawler (Gui et al., 2016, 2013b). Excluding the WMSs without metadata text content, the number of WMSs available is 40,722, including 210,732 layers. These WMSs come from 989 service providers including NASA, NOAA, USGS, and other research institutes, governmental sectors as well as universities. Since a WMS may provide multiple map layers that belong to different themes, we generate both the service-level and layer-level themes via our collaborative training model by using different text fields in the metadata. *Title*, *URL*, *Abstract* and *Keywords* of each WMS are extracted from the XML-based metadata document as shown in Figure 3 for service theme classification, obtained through *GetCapabilities* operation. *Name*, *Title*, *Abstract*, *Keywords* and *Attribution* of each WMS layer are extracted for layer theme classification. All these fields are treated equally and combined into a document for constructing the heterogeneous word document graph. We selected 501 WMS metadata records and 1460 Layer metadata records in random, respectively, for our experiments. The frequency histogram of the number of labels for each WMS and each layer on the selected 501 samples are shown in Figure 4. Among 501 WMS and 1460 layers, more than 70% of the samples just have one label, while there are also many samples with several labels, up to seven labels. The number of labels for each service and layer are 1.4031 and 1.37 in average, respectively.

4.2. Experimental design

4.2.1. Baseline

We select seven text classification methods as well as the proposed Text GCN-SW as the baselines to verify the performance of our collaborative training method. These state-of-the-art methods adopt different architectures, and have different working mechanisms and unique features.

- **Text Graph Convolutional Network (Text GCN)** can capture global word co-occurrence information with a heterogeneous word document graph and utilize limited labeled documents, even a simple two-layer Text GCN demonstrates promising results (Yao et al., 2018).
- **Multi-Label K-Nearest Neighbor (ML-KNN)** is the multi-label version of KNN. According to the number of identified neighboring instances belonging to each

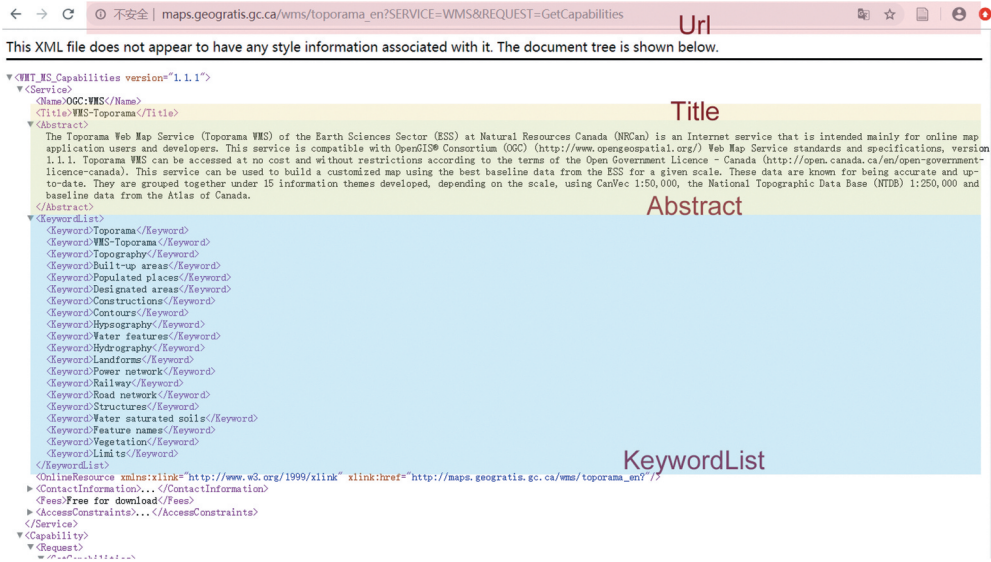


Figure 3. An exemplary WMS metadata document in XML format, including URL, title, abstract and keyword, etc.

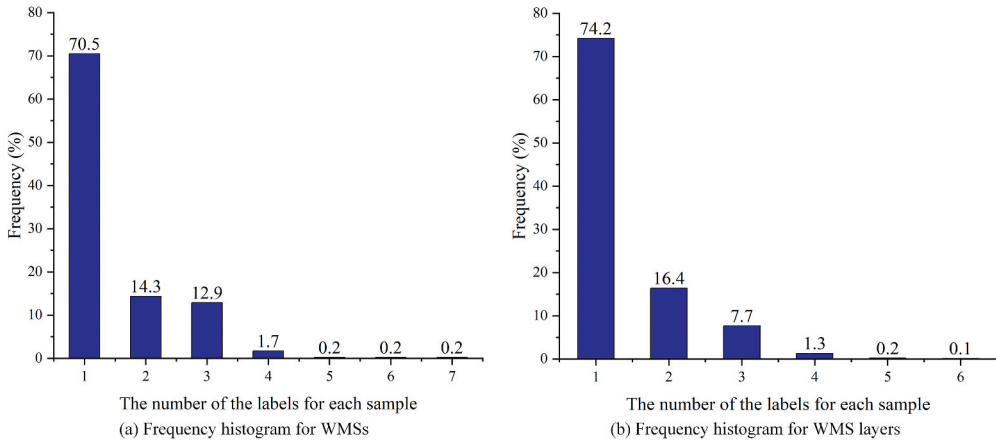


Figure 4. Frequency histogram of the number of labels for each WMS and each layer on the selected 501 samples.

possible class, ML-KNN utilizes the maximum a posteriori (MAP) principle to determine the label set for the test instance (Zhang & Zhou, 2007).

- **Binary Relevance K-Nearest Neighbor (BRKNN)** extends the KNN algorithm so that independent predictions are made for each label, following a single search of the K nearest neighbors. It can avoid redundant time-intensive computations (Spyromitros et al., 2008).
- **Long Short-Term Memory (LSTM)** is a popular recurrent neural network. It uses the last hidden state as the representation of the whole text in text classification and improve the performances by exploring common features (Liu et al., 2016).

- **Convolutional Neural Network (CNN)** can achieve sentence-level classification tasks by using a slight variant of the CNN architecture (Kim, 2014).
- **Bi-directional Long Short-Term Memory (Bi-LSTM)** is a bi-directional LSTM, which focuses on capturing the most important semantic information in a sentence (Zhou et al., 2016).
- **C-LSTM** combines the strengths of CNN and LSTM. In this method, the sequence of higher-level phrase representations is extracted by CNN, and the sentence representation is obtained via LSTM (Zhou et al., 2015). The learned semantic sentence representations are effective for classification.

4.2.2. Evaluation metrics

Six metrics are used to evaluate our model, including Hamming loss, average accuracy, Jaccard similarity, precision, recall, and F1-score. Hamming Loss measures the inconsistency between the predicted label and the actual label of the sample. Since Hamming loss counts the number of misclassified labels, the smaller the Hamming loss, the better the model performance. Average accuracy measures the average prediction performance of the classifier on the entire labels of a sample. If the predicted result of a sample is exactly the same as the actual label combination, the accuracy is 1, otherwise it is 0. Jaccard similarity measures the proportion of sample which is correctly assigned. Recall and precision measure the proportion of samples that are correctly classified. F1-score is calculated by combining the recall and the precision. The calculations of these criteria are shown in formula 6.

$$\begin{aligned}
 \text{Hamming Loss}(x_i, y_i) &= \frac{1}{N} \sum_{i=1}^N \frac{\text{xor}(x_i, y_i)}{M} \\
 \text{AverageAccuracy}(x_i, y_i) &= \frac{1}{|D|} \sum_{i=0}^{|D|-1} 1(x_i = y_i) \\
 \text{Jaccard Similarity Score}(x_i, y_i) &= \frac{|x_i \cap y_i|}{|x_i \cup y_i|} \\
 \text{Recall} &= b/a, \text{ Precision} = b/c \\
 \text{F1} &= 2 \times \text{Precision} \times \text{Recall} / (\text{Recall} + \text{Precision})
 \end{aligned} \tag{6}$$

where M is the total number of labels, N is the total number of samples, y_i and x_i represent the true label and the predicted label of the i th sample, respectively, and xor represents the XOR relationship between the predicted label and the actual label. $|D|$ is the total number of test samples. a is the total number of samples whose actual labels include the i th theme category, and b is the total number of samples in the i th category that is correctly predicted, and c is the total number of samples predicted as the i th category.

4.2.3. Parameter settings

In experiments, we set the parameters of Text GCN-SW-KNN, Text GCN-SW, Text GCN model as follows. Initial learning rate is set as 0.02 and dropout rate is set as 0.5. Number of epochs to train is 200 and tolerance for early stopping is 10. Number of units in first hidden layer is 200. The maximum Chebyshev polynomial degree is 3.

For ML-KNN, the parameter k is set as 4 for both WMS metadata and layer metadata. While, k is set as 3 for BRKNN.

For LSTM, Bi-LSTM, CNN, and C-LSTM, the parameters are set as follows. Dropout keep probability is 0.5 and learning rate is 0.001. The batch size is 25 and the number of epochs is 50. L2 regularization lambda is 0.001. For the CNN-related parameters of CNN and C-LSTM, the embedding size is 256; the filter sizes are 3, 4, and 5, respectively, and the number of filters is always 128 for per filter size. As for the parameters of the LSTM part, the number of the LSTM cells is 2 for LSTM, Bi-LSTM and C-LSTM. For both LSTM and Bi-LSTM, the number of hidden units in the LSTM cell is 128.

4.3. Overall performance evaluation

We use the aforementioned six evaluation metrics to evaluate the performance of the proposed method Text GCN-SW-KNN and eight baselines introduced in Section 4.2. 80% data are used for training and 20% data are used for testing. The results of Text GCN-SW-KNN, Text GCN-SW, Text GCN, LSTM, Bi-LSTM, CNN, and CLSTM are the average of 10 repeating experiments. The conditions for the termination of co-training iteration are as described in Section 3.4. The maximum iteration limit is set as 20. The iteration times for WMSs and layers usually 2 or 3 times and 3 or 4 time before the termination. The comparison results for service-level and layer-level classifications are visualized in Figure 5, and also shown in Tables A1 and A2 in Appendix, respectively.

As illustrated in Figure 5, our model outperforms all baselines on all six metrics except recall and Hamming loss. More specifically, for service-level classification, Text GCN-SW-KNN achieves higher accuracy, Jaccard similarity, precision and F1-score than any other methods; while for layer-level classification, Text GCN-SW-KNN achieves higher performance than other methods, but the recall is lower than ML-KNN, Text GCN and Text GCN-SW. In our collaborative training model, since the requirement for precision is high, some correct labels may be also excluded, and the recall rate decreases accordingly. When comparing the result of Text GCN-SW and Text GCN, Text GCN-SW can outperform Text GCN on all six metrics for the classification results of the two-levels, which illustrates that the integration of geographic semantics improves the classification effect by differentiating the contribution of different words. Especially, the results show that Text GCN-SW without co-training can achieve higher F1 than ML-KNN while Text GCN cannot. The performance of Text GCN-SW for layer-level classification is better than that on service-level, due to the fact that the edges in WMS text graph are fewer than layer text graph, which limits the message passing among the nodes. LSTM, Bi-LSTM and C-LSTM gain inferior results because they model consecutive word sequences explicitly, but many WMS metadata texts are composed of phrases and words, rather than complete sentences, and the effect of sequence information on classification results is subtle. In general, the experiments prove that Text GCN-SW outperformed the original Text GCN methods with the improvement of combining the geographic semantics, and Text GCN-SW-KNN can yield promising results with highest accuracy and F1-score.

4.4. Stability assessment

Because of the uncertainty of the input order, the classification results and accuracies of the model obtained in each training may be different, so the stability of the models needs to be verified. In general, collaborative training can ensure the stability of the algorithm, but the stability of the base model also influence the final stability. So, this experiment analyzes the

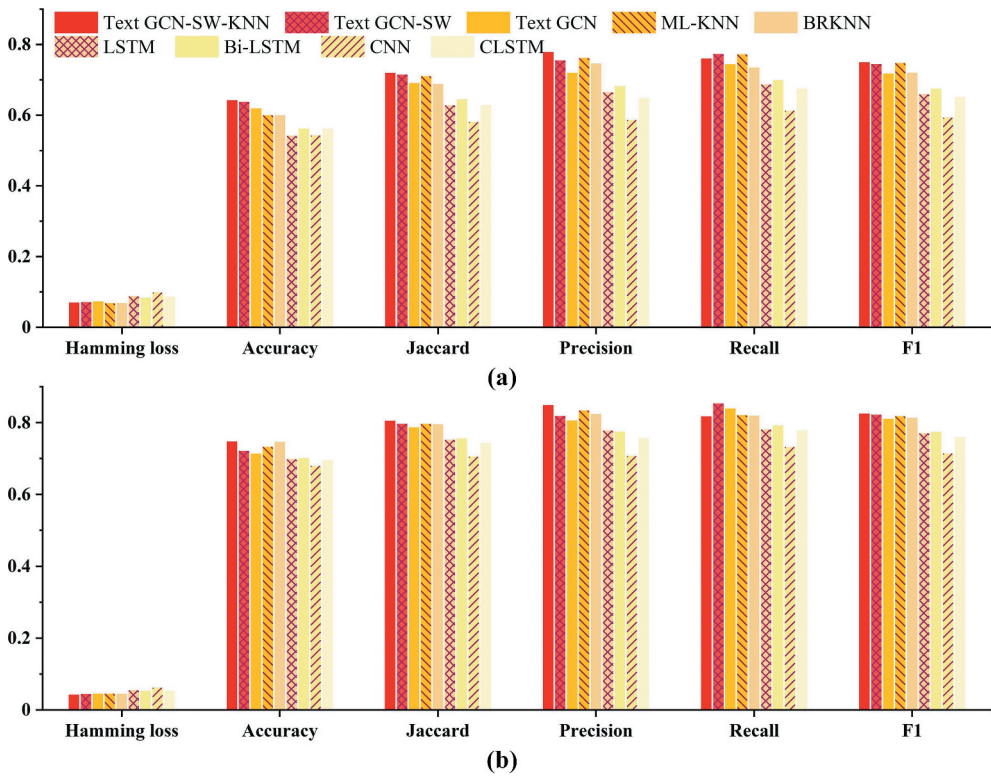


Figure 5. Performance comparison of text GCN-SW-KNN and eight baselines for both service-level and layer-level classification, respectively. (a) is the result for WMS and (b) is for WMS layer.

stability of the improved base model Text GCN-SW by comparing it with the original Text GCN. Fifty repeating experiments for Text GCN and Text GCN-SW are conducted for both service-level and layer-level classifications to evaluate the stability of the proposed method and standard deviation of F1-score is used as the criteria. 80% data are used for training and the left is used for testing. In Figure 6, the line charts show the F1-score details of Text GCN and Text GCN-SW in 50 repeating experiments, and the box-plots and frequency distribution histograms depict the corresponding stability and performance from statistical views.

According to Figure 6, we can find that Text GCN-SW has better F1-score and better F1-score stability compared with Text GCN. For service-level classification, the standard deviation of F1-score for Text GCN is about 0.04205, while that for Text GCN-SW is 0.007899. For layer-level classification, the standard deviation for Text GCN is about 0.00835, while that for Text GCN-SW is 0.0065. For both service and layer classification, the values of Q3-1.5IQR of Text GCN-SW in the box-plot are higher than the medians and mean values of Text GCN, which demonstrates the improvement of the adjacency matrix enhanced both the stability and performance of the algorithm. That is because some topic-irrelevant words, e.g. using, copyright and scientific, with high TF-IDF may not be related to a certain theme category, and their high weights can lead to increasing uncertainty in classification. Our base model Text GCN-SW can use semantic distance to adjust the weight accordingly, and in turn avoids misleading and enhances the overall

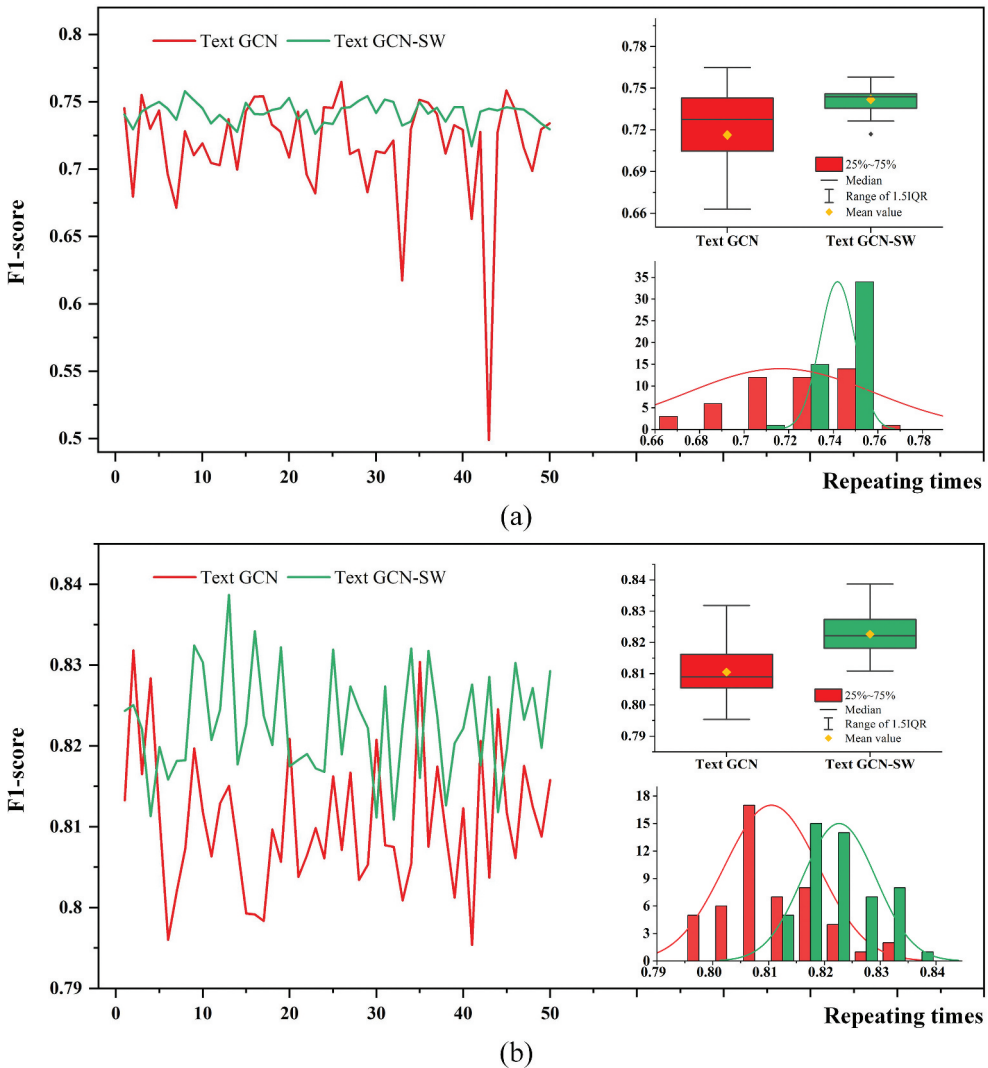


Figure 6. Stability of F1-score in repeating experiments for Text GCN and Text GCN-SW. (a) is the result for WMS and (b) is for WMS layer.

stability effectively. Furthermore, the stability difference between the two methods is more significant in service-level classification. It might be caused by the fact that the text of the WMS metadata is shorter in general and it is particularly important to avoid uncertain information in the situation with relatively few features. Therefore, the stability of Text GCN-SW is better than that of ordinary Text GCN, which make Text GCN-SW more suitable for our application scenario as a base model for collaborative training.

4.5. Performances with different proportions of the training data

To evaluate the influences of the size of the labeled data to the model performances, we tested four selected methods, i.e. Text GCN, Text GCN-SW, ML-KNN and LSTM, which

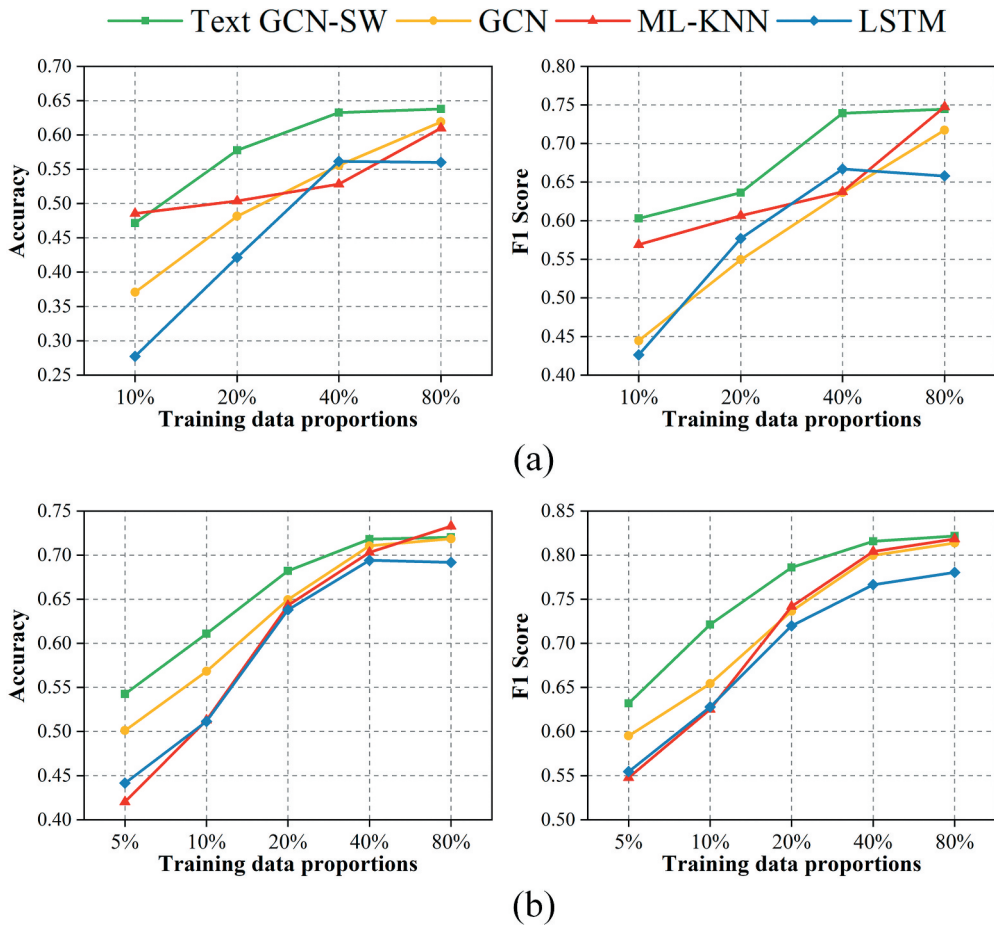


Figure 7. Accuracy and F1 for Text GCN, Text GCN-SW, ML-KNN and LSTM with different proportions of the training data: (a) is for WMS and (b) is for WMS layer.

performed relatively well in previous experiments, with different proportions of the training data. Figure 7 shows test accuracies and F1-scores with 10%, 20%, 40%, 80% of the WMS metadata as the training data for service-level classification, and with 5%, 10%, 20%, 40%, 80% of the layer metadata as the training data for layer-level classification. Since collaborative training can allow inexpensive unlabeled data to augment a much smaller set of labeled examples, the results of the collaborative training method are better than the non-cooperative training method when the training data are lack in general. Therefore, we did not compare our collaborative training method, Text GCN-SW-KNN, with other non-cooperative training methods. The performance comparison of the base models under different training data proportions explains the reason for our base model selection.

In general, Text GCN-SW can gain higher accuracies and F1-scores than other three methods for both the service and layer classifications. Meanwhile, when the proportion of the training data decreases, the performance advantages over other methods will significantly manifest, except that of ML-KNN with 10% training data. The experiment results demonstrate that compared with other methods, Text GCN-SW can propagate

document label information to the entire graph well for classification with very limited training data (Yao et al., 2018). So that, it is suitable for cases where there are only few labeled samples, and ensure the performance of our semi-supervised collaborative training model Text GCN-SW-KNN under small proportions of the training data. When the proportion of the training data is relatively high, the F1-score and accuracy obtained by another base model ML-KNN increase significantly and even outperform other methods, which guarantees the effectiveness of collaborative training to some extent as well. Therefore, the cooperation of the two base models contributes to the performance of our Text GCN-SW-KNN method.

5. Conclusion

In this study, we proposed a novel multi-label text classification method, Text GCN-SW-KNN, for extracting WMS application theme from WMS metadata. It adopts collaborative training mechanism by combining Text GCN with semantic information from SWEET and WordNet through an improved base model Text GCN-SW. The results show that integrating geographic semantics has three advantages: 1) achieving higher accuracy and F1 than state-of-the-art comparison methods, 2) improving the stability compared with Text GCN, and 3) having better performance with limited labeled metadata. Meanwhile, the semi-supervised collaborative training of ML-KNN and Text GCN-SW can further improve the classification results. Thus, the proposed classification method can be applied to the geportal or the service catalog to assist end users to acquire desired WMSs more efficiently. Moreover, it can be extended to the classification problems of other online geospatial resources that with metadata text descriptions.

Further study can be conducted from the following three aspects. Different treatments for texts from different metadata fields, such as *Title*, *URL*, *Abstract*, *Attribution* and *Keywords*, will be considered. For example, we might consider adjusting the weight of the corresponding parts of different fields or construct several graphs for each field and merge them to operate the classification. Besides, syntactic and sequential contextual information can be included to better understanding the meaning of the metadata text. In addition, currently, we only define 10 first-level application themes, which cannot support refined retrieval and distinguish the results of different sub-themes. We will extract the fine-grained sub-themes of the first-level themes, and build a two-level theme directory with LDA topic analysis accordingly.

Notes

1. <https://www.iso.org/standard/59221.html>.
2. <https://gcmd.gsfc.nasa.gov/>.
3. <http://catalog.data.gov/dataset>.
4. http://www.aprsaf.org/data/feature/f_086_3.pdf.
5. <https://mrdata.usgs.gov/>.
6. <http://www.w3.org/TR/owl-ref/>.

Data availability statement

The code and data that support the findings of this study are openly available in GitHub at <https://github.com/ZPGuiGroupWuhu/Text-based-WMS-Application-Theme-Classification/tree/master>.

Funding

This paper is supported by National Natural Science Foundation of China [No. 41971349, No. 41930107, No. 42090010 and No. 41501434] and National Key Research and Development Program of China [No. 2017YFB0503704 and No. 2018YFC0809806].

Notes on contributors



Zhengyang Wei is an undergraduate student in the School of Remote Sensing and Information Engineering, Wuhan University. Her research interests include text classification and computer vision.



Zhipeng Gui is an Associate Professor of Geographic Information Science in the School of Remote Sensing and Information Engineering, Wuhan University. His research interests are high-performance spatiotemporal data mining, geovisual analytics and Distributed Geographic Information Processing (DGIP), especially on 1) Spatiotemporal point pattern analysis and GeoAI; 2) High-performance geocomputation and spatial cloud computing; 3) Geospatial service chain modeling and optimization; 4) QoGIS-aware monitoring and evaluation of geospatial web services. He now serves as the Co-chair of International Society for Photogrammetry and Remote Sensing (ISPRS) Working Group V/4: Web-based Resource Sharing for Education and Research.



Min Zhang has received her master degree from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University in 2019. Now, she is working in Tencent. Her research interests include geographic theme extraction and geovisual analytics.



Zelong Yang has received his Ph.D. degree in cartography and geographic information system from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His current research interest is on the geospatial information analysis, using natural language process (NLP), deep learning (DL), multimodality fusion and analysis, and cloud computing.



Yuao Mei is a master student in the School of Remote Sensing and Information Engineering, Wuhan University. His research interest focuses on GeoAI especially for population spatialization.



Huayi Wu is currently a full Professor in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His scientific interests include high-performance geospatial computing and intelligent geospatial web services.



Hongbo Liu is an Engineer in Chongqing Geomatics and Remote Sensing Center. He received his master degree from the School of Resource and Environmental Sciences, Wuhan University in 2011. His research interests include big geospatial data analysis and online geospatial resource sharing.



Jing Yu is a Senior Engineer in Chongqing Geomatics and Remote Sensing Center. She received her master degree from the School of Remote Sensing and Information Engineering, Wuhan University in 2006. Her research interests include big geospatial data analysis and geoportals design.

ORCID

Zhipeng Gui  <http://orcid.org/0000-0001-9467-9680>

References

- Berger, M. J. (2014). Large scale multi-label text classification with semantic word vectors. *Technical Report*.
- Blum, A., & Mitchell, T. (1998, July 24–26). Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory* (pp. 92–100). Madison, WI. doi:10.1145/279943.279962.
- Chen, G., Ye, D., Xing, Z., Chen, J., & Cambria, E. (2017, May 14–19). Ensemble application of convolutional and recurrent neural networks for multi-label text categorization. In *2017 International joint conference on neural networks (IJCNN)* (pp. 2377–2383). Anchorage, AK. doi:10.1109/IJCNN.2017.7966144.

- Elisseeff, A., & Weston, J. (2002). A kernel method for multi-labelled classification. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems 14 (NIPS 2001)* (pp. 681–687). Vancouver, Canada: The MIT Press. doi:[10.7551/mitpress/1120.003.0092](https://doi.org/10.7551/mitpress/1120.003.0092)
- Fellbaum, C., & Miller, G. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
- Godbole, S., & Sarawagi, S. (2004). Discriminative methods for multi-labeled classification. *Advances in Knowledge Discovery and Data Mining*, (vol), 3056. doi:[10.1007/978-3-540-24775-3_5](https://doi.org/10.1007/978-3-540-24775-3_5)
- Gui, Z., Cao, J., Liu, X., Cheng, X., & Wu, H. (2016). Global-scale resource survey and performance monitoring of public OGC web map services. *ISPRS International Journal of Geo-Information*, 5(6), 88. doi:[10.3390/ijgi5060088](https://doi.org/10.3390/ijgi5060088)
- Gui, Z., Yang, C., Xia, J., Li, J., Rezgui, A., Sun, M., ... Fay, D.. (2013a). A visualization-enhanced graphical user interface for geospatial resource discovery. *Annals of GIS*, 19(2), 109–121. doi:[10.1080/19475683.2013.782467](https://doi.org/10.1080/19475683.2013.782467)
- Gui, Z., Yang, C., Xia, J., Liu, K., & Lostritto, P. (2013b). A performance, semantic and service quality-enhanced distributed search engine for improving geospatial resource discovery. *International Journal of Geographical Information Science*, 27(6), 1109–1132. doi:[10.1080/13658816.2012.739692](https://doi.org/10.1080/13658816.2012.739692)
- Henaff, M., Bruna, J., & Lecun, Y. (2015). Deep convolutional networks on graph-structured data. *Computer Science*, 10 pp. *arXiv:1506.05163*.
- Hu, K., Gui, Z., Cheng, X., Qi, K., & Wu, H. (2016). Content-based discovery for web map service using support vector machine and user relevance feedback. *PLoS ONE*, 11(11), e0166098. doi:[10.1371/journal.pone.0166098](https://doi.org/10.1371/journal.pone.0166098)
- Hu, Y., Janowicz, K., Prasad, S., & Gao, S. (2015). Metadata topic harmonization and semantic search for linked-data-driven geoportals: A case study using ArcGIS online. *Transactions in GIS*, 19(3), 398–416. doi:[10.1111/tgis.12151](https://doi.org/10.1111/tgis.12151)
- John, G., & Langley, P. (1995, August 18–20). Estimating continuous distributions in Bayesian classifiers. In *Proceedings of the 11th conference on uncertainty in artificial intelligence* (pp. 338–345). Montreal, Quebec, Canada.
- Kim, Y. (2014, October 25–29). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1746–1751). Doha, Qatar. doi:[10.3115/v1/D14-1181](https://doi.org/10.3115/v1/D14-1181).
- Kipf, T. N., & Welling, M. (2017, April 24–26). Semi-supervised classification with graph convolutional networks. In *5th international conference on learning representations* (p. 14). Toulon, France.
- Kurata, G., Xiang, B., & Zhou, B. (2016, June 12–17). Improved neural network-based multi-label classification with better initialization leveraging label co-occurrence. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 521–526). San Diego, CA. doi:[10.18653/v1/N16-1063](https://doi.org/10.18653/v1/N16-1063).
- Li, M., Gui, Z., Cheng, X., Wu, H., & Qin, K. (2019). A content-based WMS layer retrieval method combining multiple kernel learning and user feedback. *Acta Geodaetica et Cartographica Sinica*, 48(10), 1320–1330.
- Li, W., Yang, C., Nebert, D., Raskin, R., Houser, P., & Wu, H. (2011). Semantic-based web service discovery and chaining for building an arctic spatial data infrastructure. *Computers & Geosciences*, 37(11), 1752–1762. doi:[10.1016/j.cageo.2011.06.024](https://doi.org/10.1016/j.cageo.2011.06.024)
- Liu, J., Chang, W., Wu, Y., & Yang, Y. (2017, August 7–11). Deep learning for extreme multi-label text classification. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval* (pp. 115–124). Shinjuku, Tokyo, Japan. doi:[10.1145/3077136.3080834](https://doi.org/10.1145/3077136.3080834).
- Liu, K., Yang, C., Li, W., Li, Z., Wu, H., Rezgui, A., & Xia, J. (2011, June 24–26). The GEOSS clearinghouse high performance search engine. In *Proceedings - 19th international conference on geoinformatics, geoinformatics 2011 (Geoinformatics 2011)*. Shanghai, China. doi:[10.1109/Geoinformatics.2011.5981077](https://doi.org/10.1109/Geoinformatics.2011.5981077).
- Liu, P., Qiu, X., & Huang, X. (2016, July 9–15). Recurrent neural network for text classification with multi-task learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (pp. 2873–2879). New York, NY.

- Madjarov, G., Kocev, D., Gjorgjevikj, D., & Deroski, S. (2012). An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition*, 45(9), 3084–3104. doi:10.1016/j.patcog.2012.03.004
- Raskin, R. G., & Pan, M. J. (2005). Knowledge representation in the semantic web for earth and environmental terminology (SWEET). *Computers & Geosciences*, 31(9), 1119–1125. doi:10.1016/j.cageo.2004.12.004
- Read, J., Pfahringer, B., & Holmes, G. (2008, December 15–19). Multi-label classification using ensembles of pruned sets. In *Proceedings of the 8th IEEE international conference on data mining (ICDM 2008)* (pp. 995–1000). Pisa, Italy: ICDM. doi:10.1109/ICDM.2008.74.
- Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2011). Classifier chains for multi-label classification. *Machine Learning*, 85(3), 333–359. doi:10.1007/s10994-011-5256-5
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv:1609.04747*.
- Schapiro, R. E., & Singer, Y. (2000). Boostexter: A boosting-based system for text categorization. *Machine Learning*, 39(2/3), 135–168. doi:10.1023/A:1007649029923
- Spyromitros, E., Tsoumakas, G., & Vlahavas, I. (2008). An empirical study of lazy multilabel classification algorithms. *Artificial Intelligence: Theories, Models and Applications*, 5138, 401–406. doi:10.1007/978-3-540-87881-0_40
- Tai, K. S., Socher, R., & Manning, C. D. (2015, July 26–31). Improved semantic representations from tree-structured long short-term, memory networks. In *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing* (pp. 1556–1566). Beijing, China. doi:10.3115/v1/P15-1150
- Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(3), 1–13. doi:10.4018/jdwm.2007070101
- Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016, November 1–5). Attention-based LSTM for Aspect-level Sentiment Classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing (EMNLP 2016)* (pp. 606–615). Austin, TX. doi:10.18653/v1/D16-1058.
- Wu, H., Li, Z., Zhang, H., Yang, C., & Shen, S. (2011). Monitoring and evaluating the quality of Web Map Service resources for optimizing map composition over the internet to support decision making. *Computers & Geosciences*, 37(4), 485–494. doi:10.1016/j.cageo.2010.05.026
- Yang, Z., Gui, Z., Wu, H., & Li, W. (2019). A latent feature-based multimodality fusion method for theme classification on web map service. *IEEE Access*, 8, 25299–25309.
- Yang, Z., Yang, D., Dyer, C., He, X., & Hovy, E. (2016, June 12–17). Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. (pp. 1480–1489). San Diego, CA. doi:10.18653/v1/N16-1174.
- Yao, L., Mao, C., & Luo, Y. (2018). Graph convolutional networks for text classification. *arXiv:1809.05679*.
- Zhang, M., Gui, Z., Cheng, X., Cao, J., & Wu, H. (2019). A text-based WMS domain themes extraction and metadata extension method. *Geomatics and Information Science of Wuhan University*, 44(11), 1730–1738.
- Zhang, M. L., & Zhou, Z. H. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7), 2038–2048. doi:10.1016/j.patcog.2006.12.019
- Zhang, X., Zhao, J., & Lecun, Y. (2015, December 7–12). Character-level convolutional networks for text classification. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems 28 (NIPS 2015)* (pp. 649–657). Montreal, Quebec: The MIT Press.
- Zhou, C., Sun, C., Liu, Z., & Lau, F. C. M. (2015). A C-LSTM neural network for text classification. *Computer Science*, 1(4), 39–44.
- Zhou, P., Shi, W., Tian, J., Qi, Z., & Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2, 207–212. doi:10.18653/v1/P16-2034

Appendix

Table A1. Performance metrics of text GCN-SW-KNN and eight baselines for WMS service-level classification.

Method	Hamming loss	Accuracy	Jaccard	Precision	Recall	F1
Text GCN-SW-KNN	0.07009	0.64200	0.71968	0.77842	0.75975	0.74959
Text GCN-SW	0.07118	0.63800	0.71483	0.75475	0.77300	0.74460
Text GCN	0.07345	0.61900	0.69053	0.71975	0.74383	0.71744
ML-KNN	0.06818	0.60000	0.71050	0.76183	0.77200	0.74750
BRKNN	0.06909	0.60000	0.68850	0.74617	0.73450	0.72024
LSTM	0.08800	0.54200	0.62822	0.66520	0.68705	0.65940
Bi-LSTM	0.08436	0.56200	0.64490	0.68223	0.69900	0.67459
CNN	0.09872	0.54300	0.58083	0.58650	0.61300	0.59346
CLSTM	0.08709	0.56200	0.62833	0.64783	0.67475	0.65027

Table A2. Performance metrics of text GCN-SW-KNN and eight baselines for WMS layer-level classification.

Method	Hamming loss	Accuracy	Jaccard	Precision	Recall	F1
Text GCN-SW-KNN	0.04256	0.74760	0.80543	0.84909	0.81787	0.82515
Text GCN-SW	0.04433	0.72164	0.79705	0.81880	0.85357	0.82263
Text GCN	0.04542	0.71411	0.78679	0.80596	0.83969	0.81050
ML-KNN	0.04545	0.73288	0.79663	0.83436	0.82095	0.81826
BRKNN	0.04483	0.74658	0.79561	0.82420	0.81935	0.81393
LSTM	0.05490	0.69863	0.75269	0.77853	0.78140	0.77082
Bi-LSTM	0.05327	0.70171	0.75628	0.77518	0.79255	0.77448
CNN	0.06235	0.67995	0.70570	0.70772	0.73241	0.71444
CLSTM	0.05355	0.69552	0.74349	0.75654	0.77834	0.75937