





# LTWNN: A Novel Approach Using Sentence Embeddings for Extracting Diverse Concepts in MOOCs

Zhijie Wu<sup>1</sup>, Jia Zhu<sup>2</sup>, Shi Xu<sup>1</sup>, Zhiwen Yan<sup>1</sup>, and Wanying Liang<sup>1</sup>

<sup>1</sup> School of Computer Science, South China Normal University, Guangzhou, China  
{2019022683,xushi,2019022686,sylvialaung}@scnu.edu.cn

<sup>2</sup> Zhejiang Normal University, Zhejiang, China  
jiazhu@zjnu.edu.cn

**Abstract.** As a global online education platform, Massive Open Online Courses (MOOCs) provide high-quality learning content. It is a challenging issue to design a key course concept for students with different backgrounds. Even though much work concerned with course concept extraction in MOOC has been done, those related works simply utilize external knowledge to get the relatedness of two different candidate concepts. Furthermore, they require the input to belong to multi-document and severely rely on seed sets, in which their model shows poor performance when input is a single document. Addressing these drawbacks, we tackle concept extraction from a single document using **LTWNN**, a novel method **L**earning to **W**eight with **N**eural **N**etwork for **C**ourse **C**oncept **E**xtraction in **M**OOCs. With LTWNN, we make full use of external knowledge via making relatedness between each candidate concept and document by introducing an embedding-based maximal marginal relevance (MMR), which explicitly increases diversity among selected concepts. Moreover, we combine the inner statistical information and external knowledge, in which the neural network automatically learns to allocate weight for them. Experiments on different course corpus show that our method outperforms alternative methods.

**Keywords:** MOOC · Course concept extraction · Learning to weight · Neural network · Diversity

## 1 Introduction

The rapid development of modern networks and the Internet 2.0 have spawned many online open platforms, of which Coursera and Xuetang Online are online curriculum education platforms, providing great convenience for learners within doors. Following this trend, a large amount of knowledge data are created, including course videos and their subtitles. However, it is difficult for learners to understand and analyze the knowledge from a global perspective, while course concepts

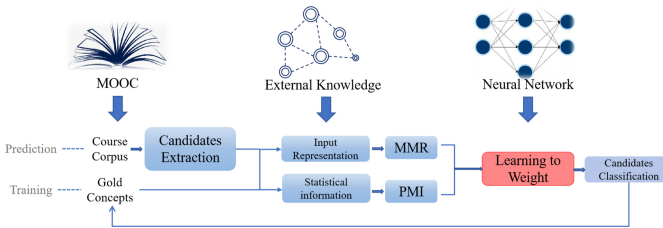
---

Supported by organization x.

can describe the knowledge points contained in these classrooms or textbooks. Understanding the overall concept makes it easier to learn the subject and assist in understanding the text for learners.

Although quite a few researches [5, 14, 19, 25] on course concepts extraction from teaching materials and course subtitles have been done, the problem of concept extraction from course subtitles in MOOCs is far from solved. Course concept extraction is non-trivial and challenging due to three reasons, including the single short context problem, the low-frequency problem, and the poor diversity of concepts.

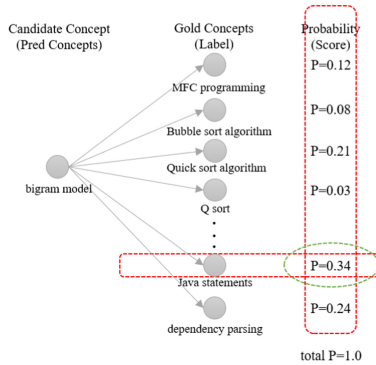
Related research topics, including keyphrase extraction [6, 12, 15, 21, 22] and term extraction [8, 11] are popular and valid in the information retrieval domain. Pan et al. [19] introduce external knowledge to explore the relationships between different concepts that have the same meanings. However, simply making use of external knowledge by using relatedness of candidates via word embedding resulting in being unable to utilize the global embedding feature. Furthermore, their work is based on multi documents, while ours is both simple and only requires the current document (single document), rather than an entire corpus. Also, their work relies heavily on seed sets, yet these seed sets are limited to acquire in some cases.



**Fig. 1.** Our proposed framework LTWNN. Note that we do not need extract candidates in training phase

To address the above problems, we propose learning to weight using sentence embedding with neural networks for course concept extraction in this paper, as shown in Fig. 1. The critical aspect of our idea is that it cannot only improve the diversity of extracted course concepts by introducing external knowledge but also automatically learn to weight to leverage inner statistical information and external expertise. First, we extract some keyphrases as candidates by the Part-of-speech (POS) rule template, and we introduce external knowledge to represent each document by sentence embedding model. Then, to improve the diversity of extracted concepts, we introduce the MMR algorithm and change the formula to fit our task. Next, we combine with the score of MMR and statistical information (i.e., PMI), and then our model learns to weight by neural network classifier (e.g., MLP). Finally, in the prediction phase, the MMR score and PMI score of each candidate concept will be the input of the trained model. Note

that we do not care label that the model predicts, we just select the value of the maximal probability of result, as shown in Fig. 2. After the prediction of each candidate concept, we rank them by the selected maximal probability.



**Fig. 2.** For the candidate concept ‘bigram model’, its maximal probability corresponding to the result of classification ‘Java statements’. Here we just select the value  $P = 0.34$  as score of concept.

The main contributions of our model are summarized as follows:

- We propose to introduce the MMR algorithm and utilize external knowledge to calculate relatedness between candidate concepts and documents, which validly solves the diversity of extracted concepts.
- We propose to combine inner statistical information and external knowledge properly, in which we apply neural networks to learn to weight for each feature information automatically.
- We propose LTWNN, which incorporates neural networks into the course concept extraction model without relying on multi-document corpus and seed sets.

## 2 Related Work

### 2.1 Course Concept Extraction (CCE)

Based on the keyphrases extraction, Pan et al. [19] compared the task with keyword extraction and designed a novel graph-based propagation process. Chen et al. [5] extended Pan’s approach to upgrading the quality of candidate concepts via a novel automated phrase mining method called AutoPhrase [24]. Moreover, based on Pan’s approach, Yu et al. [25] achieved course concept expansion with an interactive game.

Different from these architectures are listed above that regard CCE as a ranking problem, Lu et al. [14] applied deep learning in CCE by setting three

types of tag for educational textbooks. Their proposed model mainly adopts a gated recurrent unit (GRU) network. Simultaneously, their application scenario is national curriculum standards of mathematics, which is different from ours, for that the colloquial of course data brings more difficulties for our task. All the above approaches bring valuable references for our work course concept extraction.

## 2.2 Word and Sentence Embeddings

We introduce external knowledge via embeddings in this paper. Next, we review the development of embeddings. Word embedding (word2vec) [16] is proposed to improve the semantic via representing words as vectors in continuous vector space. To make up for the weakness of word2vec, GloVe [20] is proposed to train the embedding model based on global vocabulary. GloVe integrated Global Matrix Factorization into word2vec, which enriches the semantic and syntax information between words.

The represent of entire sentences and documents is needed to get relatedness between two sentences. Similar to word2vec, Skip-Thought [9] provides sentence embeddings trained to predict neighbor sentences. Based on the Skip-Thought, Logeswaran et al. [13] proposed Quick-thoughts via classifying neighbor sentence, but not generating a new sentence. The Quick-thoughts features a much faster training than Skip-Thought. Different from general word vectors, Sent2Vec [18] produces words and N-gram vectors that can be integrated to form sentence vectors after special training. Additionally, experiments conducted by [1] suggest that sentence representation based on averaged word vectors is effective. This property is used in our embedding method, for the reason that it is accessible and valid.

## 3 LTWNN: Learning to Weight with Neural Networks

Next, we will clearly describe every procedure of the proposed method. Note that the extraction of candidates has been described in Sect. 2 and so that it will not be described in detail.

### 3.1 Statistical Information

Statistical information is usually regarded as an important quantization indicator for extracting keyphrases, including TFIDF [21], Log-likelihood (LL) [7], and Pointwise Mutual Information (PMI) [6]. Due to the existence of a single short document, in our paper, we adopt PMI to get enough statistical features. The basis of these methods is that if the constituents of a multi-word candidate phrase form a collocation rather than co-occurring by chance, it is more likely to be considered a phrase [10]. Specifically, for the  $N$ -gram candidate concept  $P = \{c_1, c_2, \dots, c_n\}$ , where  $N > 1$ , the PMI will be calculated by

$$PMI(c_1, c_2) = \frac{2 \times freq(c_1, c_2)}{freq(c_1) + freq(c_2)} \quad (1)$$

where  $freq(P)$  indicates the frequency of the candidate concept  $P$  on one document  $d \in Cor$ . For the candidates that belong to  $N$ -gram ( $N > 2$ ), the PMI is defined as

$$PMI_t = max(\{PMI(P, B)\}) \quad (2)$$

where  $P = \{c_1, c_2, \dots, c_i\}$  and  $B = \{c_{i+1}, c_{i+2}, \dots, c_N\}$ .

### 3.2 From Embedding to Candidate Concepts with MMR

The problem of low-frequency and single short context leads to some apparent weakness. For example, most candidates appear only once (i.e., the  $freq(c_1 + c_2) = 1$ ), which shows that the semantic relatedness between each candidate concept provided by internal statistics is limited. Therefore, we propose to represent candidates via introducing information from external knowledge.

Typical embedding methods (e.g., word, sentence and document embedding) show great performance on capturing semantic relatedness between different words within the shared vector space. Word embeddings represent each phrase and word via low-dimensional space vector, the relatedness between two phrases can be reflected by their cosine distance of their vectors. Here, we use trained word embedding  $vec = \{v_{w_1}, v_{w_2}, \dots, v_{w_i}\}$ , where  $v_{w_i}$  is real-valued vector of each word  $w_i$ . Then, for the each candidate consist of  $L$  length,  $P = \{char_1, char_2, \dots, char_i\}$ , we get its vectors  $vp = \{v_1, v_2, \dots, v_L\}$  is the corresponding word vector of  $char_i$  from  $vec$ .

Getting word vector from external knowledge is helpful to improve semantic relatedness for low-frequency words, and it improves the probability of extracting the informal expression “Q sort” of “quick sort”. However, it brings new problems for us. For example, we can extract the concept “bubble sort algorithm” and “heap algorithm”, while another concept “algorithm methods” is also extracted just because it contains a key-word “algorithm”. Pan et al.[19] called the issue “overlapping problem”, they simply introduced a penalty factor to overcome the problem. In fact, the method may incorrectly filter those gold concepts containing “algorithm”, for that it is hard to control the proper value of the penalty factor.

To address the problem described above, inspired by [4], we introduce Maximal Marginal Relevance (MMR), which is one of the simplest and most effective solutions to balance query-document relevance and document diversity. Next, we show how to adapt the MMR algorithm to our task course concept extraction.

The original MMR is used to improve diversity in the information retrieval and recommendation domain. Specifically, based on the all retrieved documents  $R$ , for a given input query  $Q$ , and initial set  $S$  that receives the good answer for  $Q$  in each iteration via computing MMR as described in formula (3), where  $Sim$  represent cosine similarity between two documents or query,  $\lambda$  is a balance factor that controls relevance and diversity of result,  $D_i$  and  $D_j$  are retrieved documents.

$$MMR := \arg \max_{D_i \in R \setminus S} [\lambda \cdot Sim_1(D_i, Q) - (1 - \lambda) \max_{D_j \in S} Sim_2(D_i, D_j)] \quad (3)$$

In order to use MMR here, we change the formula to fit our task [2], as follows:

$$\text{MMR} := \arg \max_{C_i \in C \setminus K} [\gamma \cdot \widehat{\text{cos}}_{sim}(D_i, doc) - (1 - \gamma) \max_{C_j \in K} \widehat{\text{cos}}_{sim}(D_i, D_j)] \quad (4)$$

where  $C$  is the set of candidate concepts,  $K$  is the set of extracted concepts,  $doc$  represents full embedding of each course corpus preprocessed (it will be described as follow),  $D_i$  and  $D_j$  are embeddings of candidate concepts  $i$  and  $j$ , respectively. The  $\gamma$  will be set as 0.5 to ensure that the relatedness and diversity parts of the equation have equal importance. Note that  $\widehat{\text{cos}}$  is a normalized cosine similarity [17], described by the following equations.

$$\widehat{\text{cos}}(C_i, doc) = 0.5 + \frac{n \text{cos}_{sim}(C_i, doc) - n \text{cos}_{sim}(C, doc)}{\sigma(n \text{cos}_{sim}(C, doc))} \quad (5)$$

$$n \text{cos}_{sim}(C, doc) = \frac{\text{cos}_{sim}(C_i, doc) - \min_{C_j \in C} \text{cos}_{sim}(C_j, doc)}{\max_{C_j \in C} \text{cos}_{sim}(C_j, doc)} \quad (6)$$

To compute the cosine similarity between each candidate concept and the corresponding entire course corpus, we need to calculate the full embedding of each document (i.e., single video corpus). Compared with word embeddings, it has been proved that sentence embedding is able to retain key sentence information, which will improve semantic relatedness between one concept and corresponding corpus document.

### 3.3 Learn to Weight and Concepts Ranking

**Learn to Weight.** To properly allocate weight for each feature information, we apply a Multi-Layer Perceptron to predict concept label  $y_c$  (i.e., 0, 1, 2, ...,  $n$ ), where  $n$  is the total label number of gold concepts, as followed:

$$p(y_c | c) = \text{MLP}(c), \quad (7)$$

where  $c$  is a candidate concept to be classified.

**Concepts Ranking.** At the prediction phase,  $X = [PMI; MMR]$  is used as input, we get the classification probability via MLP, as followed:

$$pro = \text{softmax}(\text{ReLU}(XW_h + b_h)) \quad (8)$$

$$score = \max(pro) \quad (9)$$

Again, we do not care classification label for each candidate but focus on maximal classification probability. Here, we get the maximal value of the function, and then select the Top-K as concepts by ranking score.

**Table 1.** The four datasets we use. Columns are: the domain of each dataset; the number of documents (i.e., course subtitles); the number of gold concepts; the average number of candidates per document; the average number of tokens per document; the average number of gold concepts per document.

Dataset	Domain	Documents	GoldCon	AvgCan	AvgTok	AvgCon
CSZH	Computer Science	2849	5309	27.73	804.23	1.86
CSEN	Computer Science	690	4096	85.58	1560.79	5.93
EcoZH	Economic	455	3663	133.11	2394.55	8.05
EcoEN	Economic	381	3652	72.36	896.46	9.58

## 4 Experiments

### 4.1 Dataset and Experiments Setup

We evaluate the proposed model in online MOOC datasets. The datasets<sup>1</sup> include two course corpus with Computer Science and Economic domains in two different languages. The statistics of the MOOC datasets are reported in Table 1. For our method, in the training phase, given gold concept as a classification result, for the extraction of statistical feature PMI, we calculate the frequency of words based on the entire document (i.e., all documents). For the concept  $N = 1$  (i.e., the length of a concept is 1), we directly set PMI to 0.001. In the evaluation phase, we extract feature information on a single document separately, but not the entire corpus.

### 4.2 Evaluation Measure

In this paper, we select Mean Average Precision (MAP) as an evaluation metric. Considering the precision of the ranking item, we select the R-precision [26], which is also a standard information retrieval metric that is different from Recall and Precision. Specifically, given a ranking list with  $K$  candidate concepts, it computes the number of gold concepts (i.e., precision) over  $K$  highest-ranked candidates, and the real value of  $K$  will be considered in the experiment.

### 4.3 Comparison Method with Baseline Models

We compare the proposed method LTWNN against the following baselines:

**PMI** [6]: In the Point-Mutual-Information (PMI) method, we directly rank each candidate concept based on the score calculated by the method described in Sect. 3.1.

**TextRank**<sup>2</sup> [15]: TextRank is a well-known graph-based algorithm inspired by PageRank [3]. It regards each candidate as a vertex and word relatedness as an

<sup>1</sup> The source dataset is released on <http://moocdata.cn/data/concept-extraction>.

<sup>2</sup> <https://github.com/boudinfl/pke>.

**Table 2.** Comparison of proposed method with CGP on the four datasets. MAP and R-precision at  $K (= 5, 10, 15)$  are reported. Two ablation experiments about diversity are reported.

K	Method	CSZH		CSEN		EcoZH		EcoEN	
		MAP	$R_p$	MAP	$R_p$	MAP	$R_p$	MAP	$R_p$
5	CGP	0.516	0.238	0.509	0.249	<b>0.561</b>	<b>0.297</b>	0.566	<b>0.335</b>
	LTWNN	<b>0.705</b>	<b>0.409</b>	<b>0.547</b>	<b>0.256</b>	0.433	0.235	<b>0.632</b>	0.295
	W/O PMI	0.601	<b>0.413</b>	0.533	0.256	0.395	0.232	0.554	0.298
	W/O MMR	0.473	0.358	0.347	0.135	0.462	0.251	0.470	0.243
10	CGP	0.369	0.135	0.393	<b>0.276</b>	<b>0.398</b>	<b>0.181</b>	0.483	<b>0.254</b>
	LTWNN	<b>0.603</b>	<b>0.337</b>	<b>0.403</b>	0.181	0.274	0.162	<b>0.504</b>	0.216
	W/O PMI	0.506	0.336	0.388	0.181	0.276	0.158	0.450	0.223
	W/O MMR	0.399	0.311	0.182	0.140	0.326	0.159	0.314	0.191
15	CGP	0.308	0.102	0.213	0.135	<b>0.306</b>	<b>0.140</b>	<b>0.455</b>	<b>0.224</b>
	LTWNN	<b>0.581</b>	<b>0.317</b>	<b>0.341</b>	<b>0.159</b>	0.230	0.134	0.445	0.202
	W/O PMI	0.572	0.316	0.334	0.158	0.234	0.136	0.413	0.202
	W/O MMR	0.376	0.299	0.076	0.109	0.287	0.133	0.264	0.180

edge. As an undirected weight graph, TextRank iteratively computes the rank value of each vertex.

**CGP**<sup>3</sup> [19]: Concept Graph Propagation is the state-of-the-art method in the course concept extraction of the MOOC dataset. They construct a concept graph for each course corpus, which is similar to TextRank. Different from TextRank, they calculate concept scores with PMI and external knowledge via generalized voting scores.

#### 4.4 Result Analysis

As shown in Table 2, at the overall level, our method LTWNN outperforms available methods on three of the four datasets in MAP and R-precision.

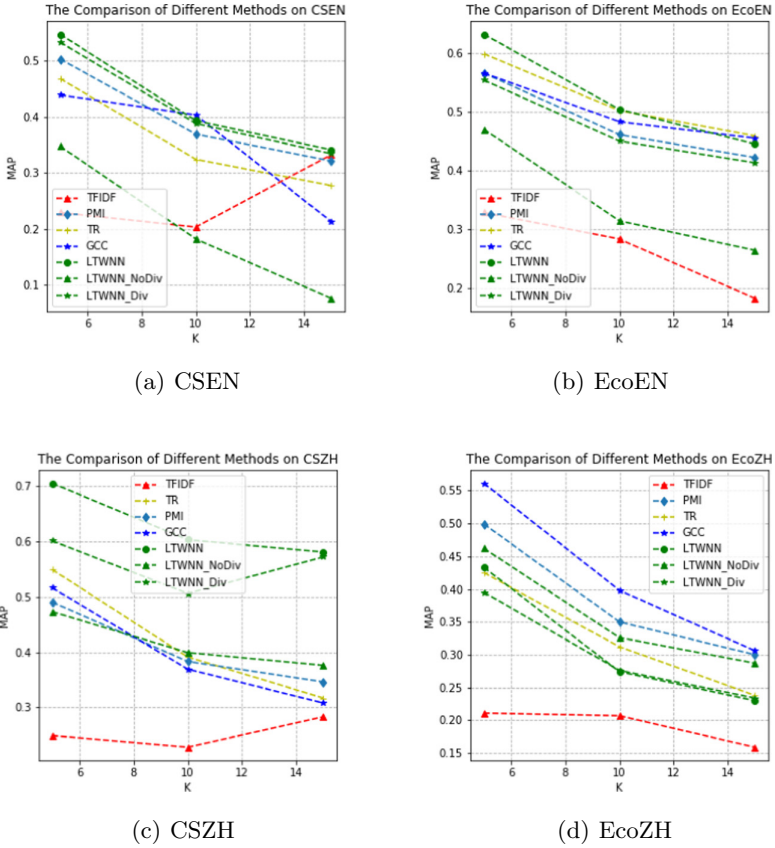
For the performance on English data, LTWNN outperforms other methods at the  $K = 5$ . Moreover, when the  $K = \{10, 15\}$ , LTWNN shows similar performance with the state-of-the-art model.

For the performance on dataset CSZH, LTWNN shows apparent robustness and effectiveness over other methods. From the information described in Table 1, we know the average number of concepts per document is only 1.86, which indicates that the phenomenon of low-frequency and poverty of diversity on the dataset is more obvious than others. Thus, the experiment suggests that LTWNN is effective in solving the problem of low-frequency and poverty of diversity on a single document.

The performance of LTWNN on dataset EcoZH shows worse than available model CGP and PMI. We conduct an experiment of diversity factor influence

<sup>3</sup> <https://github.com/thukg/concept-expansion-snippet>.





**Fig. 3.** Comparison of our method with baselines on the four dataset at MAP metric, these green lines are the performance of our method.

on dataset EcoEN, as shown in Fig. 4 and formulate (4), with the raise of  $\lambda$  (i.e., the diversity decrease), the performance of LTWNN increase continuously. As shown in experiments, the robustness and the effectiveness of Textrank in dataset CSZH is more evident than that in dataset CSEN and EcoEN, for that the average number of tokens per document in the former dataset is smaller than that in the latter.

**Ablation Study.** In our approach, the diversity of concepts plays a critical role in improving course concept extraction. As can be seen in Fig. 5, we show a concrete example, utilizing one 300-dimensional vector representing a single document and a 300-dimensional vector for each candidate concept. Then, We select the top-10 gold concepts out of 23 candidates, and the closer candidate is to the document vector, the higher the probability score it is a gold concept. Furthermore, as shown in Table 2 and Fig. 3, the comparison (except dataset

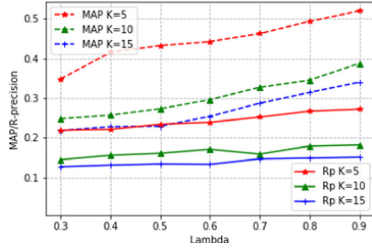


Fig. 4. The study of diversity factor influence on dataset EcoZH.

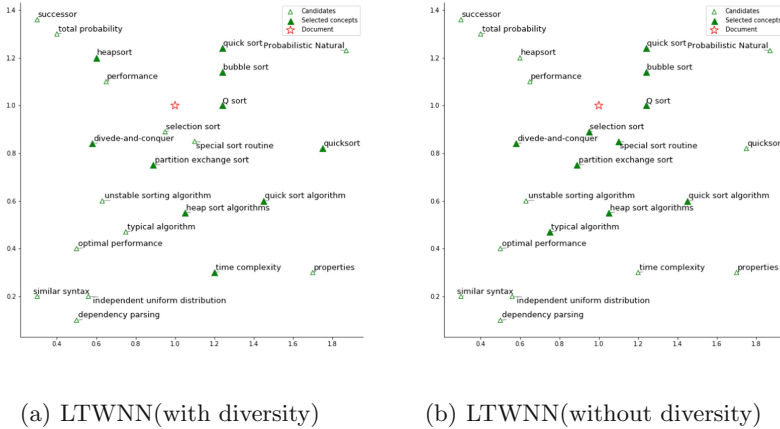


Fig. 5. The effect of diversity on the distribution of the extracted concepts. Embedding space (Visualization based on multidimensional scaling with cosine distance on the original  $Z = 300$  dimensional embeddings) of one documents, which includes sentence embedding and word embedding of candidates “Q sort”, “unstable sorting algorithm”, and so on.

EcoZH) of LTWNN-Without-PMI and LTWNN-Without-MMR suggests that poverty of diversity hampers their performance.

### 5 Conclusion and Future Work

This study demonstrates how the course concept is extracted from the MOOC corpus, in which each online course may attract more than 100,000 learners [23]. Due to the attribution of open-course, the learners have diverse knowledge backgrounds. The study is aimed at extracting core knowledge for different background students. The content from MOOC courses is usually rich and complex, which is difficult for students to understand and analyze the knowledge from a global perspective. Course-related concepts represent the core knowledge, which will help students grasp the core knowledge.

Moreover, constructing educational knowledge graph based on the course concept entity is helpful for students and teachers, including makes personal education and deep knowledge tracking. And with the course concept extraction, we can build an interaction machine to help students better grasp core knowledge.

In future work, incorporating other external knowledge such as topic knowledge that classifies course knowledge into several groups is an available method to further improve the performance of course concept extraction.

**Acknowledgment.** This work was supported by the National Natural Science Foundation of China (No. 62077015).

## References

1. Adi, Y., Kermany, E., Belinkov, Y., Lavi, O., Goldberg, Y.: Fine-grained analysis of sentence embeddings using auxiliary prediction tasks. arXiv preprint [arXiv:1608.04207](https://arxiv.org/abs/1608.04207) (2016)
2. Bennani-Smires, K., Musat, C., Hossmann, A., Baeriswyl, M., Jaggi, M.: Simple unsupervised keyphrase extraction using sentence embeddings. arXiv preprint [arXiv:1801.04470](https://arxiv.org/abs/1801.04470) (2018)
3. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine (1998)
4. Carbonell, J., Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 335–336 (1998)
5. Chen, P., Lu, Y., Zheng, V.W., Chen, X., Yang, B.: KnowEdu: a system to construct knowledge graph for education. IEEE Access **6**, 31553–31563 (2018)
6. Church, K., Hanks, P.: Word association norms, mutual information, and lexicography. *Comput. Linguist.* **16**(1), 22–29 (1990)
7. Dunning, T.E.: Accurate methods for the statistics of surprise and coincidence. *Comput. Linguist.* **19**(1), 61–74 (1993)
8. Hisamitsu, T., Niwa, Y., Tsujii, J.: A method of measuring term representativeness-baseline method using co-occurrence distribution. In: COLING 2000: The 18th International Conference on Computational Linguistics, vol. 1 (2000)
9. Kiros, R., et al.: Skip-thought vectors. In: Advances in Neural Information Processing Systems, pp. 3294–3302 (2015)
10. Korkontzelos, I., Klapaftis, I.P., Manandhar, S.: Reviewing and evaluating automatic term recognition techniques. In: Nordström, B., Ranta, A. (eds.) GoTAL 2008. LNCS (LNAI), vol. 5221, pp. 248–259. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-85287-2\\_24](https://doi.org/10.1007/978-3-540-85287-2_24)
11. Li, S., Li, J., Song, T., Li, W., Chang, B.: A novel topic model for automatic term extraction. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 885–888 (2013)
12. Liu, Z., Huang, W., Zheng, Y., Sun, M.: Automatic keyphrase extraction via topic decomposition. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 366–376 (2010)
13. Logeswaran, L., Lee, H.: An efficient framework for learning sentence representations. arXiv preprint [arXiv:1803.02893](https://arxiv.org/abs/1803.02893) (2018)

14. Lu, W., Zhou, Y., Yu, J., Jia, C.: Concept extraction and prerequisite relation learning from educational data. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 9678–9685 (2019)
15. Mihalcea, R., Tarau, P.: TextRank: bringing order into text. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 404–411 (2004)
16. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint [arXiv:1301.3781](https://arxiv.org/abs/1301.3781) (2013)
17. Mori, T., Sasaki, T.: Information gain ratio meets maximal marginal relevance. In: les actes de National Institute of Informatics Test Collections for Information Retrieval (NTCIR) (2002)
18. Pagliardini, M., Gupta, P., Jaggi, M.: Unsupervised learning of sentence embeddings using compositional n-Gram features. arXiv preprint [arXiv:1703.02507](https://arxiv.org/abs/1703.02507) (2017)
19. Pan, L., Wang, X., Li, C., Li, J., Tang, J.: Course concept extraction in MOOCs via embedding-based graph propagation. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 875–884 (2017)
20. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
21. Ramos, J., et al.: Using TF-IDF to determine word relevance in document queries. In: Proceedings of the First Instructional Conference on Machine Learning, vol. 242, pp. 133–142, New Jersey, USA (2003)
22. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage.* **24**(5), 513–523 (1988)
23. Seaton, D.T., Bergner, Y., Chuang, I., Mitros, P., Pritchard, D.E.: Who does what in a massive open online course? *Commun. ACM* **57**(4), 58–65 (2014)
24. Shang, J., Liu, J., Jiang, M., Ren, X., Voss, C.R., Han, J.: Automated phrase mining from massive text corpora. *IEEE Trans. Knowl. Data Eng.* **30**(10), 1825–1837 (2018)
25. Yu, J., et al.: Course concept expansion in MOOCs with external knowledge and interactive game. arXiv preprint [arXiv:1909.07739](https://arxiv.org/abs/1909.07739) (2019)
26. Zesch, T., Gurevych, I.: Approximate matching for evaluating keyphrase extraction. In: Proceedings of the International Conference RANLP-2009, pp. 484–489 (2009)