Thread Structure Prediction for MOOC Discussion Forum

Chengjie Sun[1(✉)], Shang-wen Li[2], and Lei Lin[1]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
{cjsun, linl}@insun.hit.edu.cn

[2] MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139, USA
swli@mit.edu

Abstract. Discussion forums are an indispensable interactive component for Massive Open Online Courses (MOOC). However, the organization of current discussion forums is not well-designed. Trouble-shooting threads are valuable for both learners and instructors, but they are drowned out in the forums with huge amounts of threads. This work first built a labeled data set for trouble-shooting thread structure prediction by crowdsourcing and then proposed methods for trouble-shooting thread detection and thread structure prediction on the data set. The output of this work can be used to spot trouble-shooting threads and show them along with structure tags in MOOC discussion forums.

Keywords: Thread structure prediction · Crowdsourcing · Lightly supervised learning · MOOC

1 Introduction

Discussion forum is critical to MOOC because it provides interactive features for MOOC. Students are supposed to use discussion forum to shape their learning communities and peer learning environment. However, some researchers have indicated that students of MOOC are not well engaged in discussion forums [1]. It is believed that the interface and organization of current discussion forums are not well designed. For example, threads for different purposes (such as content-related questions, social activities and general questions) are juxtaposed and lack informative tags. As a consequence, it's difficult for users to find target information.

Some MOOC sites, like Udacity, have provided a way for a questioner to tag the role of the posts that replied to his questions, such as whether a post is an answer. But most of the posts are still lacking in role tags because either the tagging function isn't available or users just don't provide tags. Besides, a post may need different tags to support different functions. It would be helpful if automatic methods could be used to assign informative tags to posts in MOOC forums according to different purposes.

This work proposes solutions to automatically predict the structure within a thread for MOOC forum. The heterogeneous and diverse background of the learners in MOOC makes the contents in the forum more challenging to analyze compared to other

W. Che et al. (Eds.): ICYCSEE 2016, Part II, CCIS 624, pp. 92–101, 2016.
DOI: 10.1007/978-981-10-2098-8_13

online forums. With the structures in place, current contents in MOOC forum can be reorganized. For example, we can assign each post a semantic tag to show its role in a thread. The structure of a thread is related to its type. However, there are many thread types. We focused on trouble-shooting threads in this work to illustrate the process of automatic thread structure prediction.

Trouble-shooting threads refer to threads whose first post is asking for help. They contain the problems encountered by learners during the study process in MOOC, and posts in a trouble-shooting thread form a "learning conversation." They are valuable for both learners and instructors. For course instructors, these threads can be used to uncover learners' confusions and provide better explanations to these confusions in later instructions. Learners can find out whether a question they want to pose already exists in an established thread, along with the answer. Although trouble-shooting threads are very important, they are drowned out in the MOOC forums. No explicit tags are given to make them easy to discover.

Machine learning methods are good choices for automatic thread structure prediction. But in MOOC forum domain, there is a lack of labeled data for this task. In our work, crowdsourcing was used for labeling data instead of experts, which makes the annotation process easy to replicate and extend. Once we have the annotated data, supervised learning methods can be applied. We also propose a lightly supervised method for thread structure prediction, which can be used when there is no labeled data, a small amount of labeled data, or labeled data in another domain. The performances of the two different solutions were also compared in this work.

The remainder of this paper is organized as follows. In Sect. 2, related work is discussed. Section 3 defines the problem. Data set building process is described in Sect. 4. The proposed methods are given in Sect. 5. Section 6 shows the experiments settings and results. Section 7 concludes our work.

2 Related Work

Online forum is a rich knowledge resource that has drawn lots of interest from researchers. Forums in online education have been researched extensively even before MOOC came into being.

The users of online education forum before MOOC usually came from traditional classrooms or remote education and the number of users was about one hundred. [2] proposed a rule-based recommendation framework for a class forum with 110 registers in "Comtella Discussions platform", which can save students' time by pointing the student to relevant posts. In order to help learners to improve collaboration learning management, [3] inferred learner collaboration levels by the Expectation-Maximization clustering method with the activities of learners in forum. [4] analyzed the patterns of annual, sessional, daily and hourly user behaviors in online forums with a large-scale multi-year sample of Charles Sturt University online supported forum. [4] showed how to manage students' activities by using data mining methods to discover behavior patterns in education forums. [5] proposed a genre classification system to classify a posting as an announcement, a question, clarification, interpretation, conflict, assertion, etc. The data set came from a discussion forum of Moodle CMS used by a public senior high school in Taiwan during 2004 and 2005.

With the recent popularity of MOOC, MOOC forum has drawn a lot of researchers' attention. Forums record explicit students' activities. It is valuable for student behavior analysis and enhancement of teaching effectiveness. Currently, research on forum in MOOC mainly analyzes the forum from a macro perspective. The behaviors of learners in MOOC forums were used to evaluate the learners' engagement [6, 7] and predicate their drop off probabilities [8]. A few research efforts focus on the content analysis of posts in MOOC forum. For example, [9] defined a post classification standard for MOOC forums and annotated a data set according to the standard.

Although there is little direct research on thread structure analysis for MOOC forum, some research on thread structure analysis in other online forums are closely related to this work. [10] learned online discussion structures by a conditional random fields (CRF) method. Because only the replying structure was learned, thread types weren't considered in their work. [11, 12] learned a more complicated thread structure specially for trouble-shooting threads over a technical web forum. They assumed the trouble-shooting threads were pre-selected. [13, 14] extracted question-answer pairs from online forum threads, which could be taken as an application of thread structure prediction. Our research distinguishes itself from previous work, because we predict the thread structure after thread classification.

3 Problem Definition

The target of this work is to predict the thread structure for MOOC forum. Thread structure is related to thread types. It's necessary to know the type of a thread in order to predict its structure correctly. This work focuses on the trouble-shooting thread. This section defines the trouble-shooting type thread and thread structure prediction problem.

Formally, let $T = \{X_0, X_1, \ldots, X_n\}$ be a set of thread discussions from online forum; each thread X_n consists of individual posts $\{p_0, p_1, \ldots, p_{(m-1)}\}$ arranged in chronological order.

3.1 Trouble-Shooting Thread Definition

If the initiator post p_0 of a thread X is asking for help, then the thread X is considered as a trouble-shooting thread. This definition is very similar to "Question thread" defined in [13].

3.2 Thread Structure Prediction Definition

The target of thread structure prediction is to assign each p_i a structure tag t_i which consists of two parts: Dialogue Act (DA) class (listed in Table 1) and Link Parent (LP, Post p_i is said to be the link parent post of p_j if and only if p_j is posted later than p_i and contains an immediate follow-up discussion of p_i). LP tag is denoted by the value of the relative position between the current post and its LP. The DA classes are shown in Table 1; their detailed descriptions can be found in [12].

4 Data Set Construction

In this section, we describe how a data set is built by crowdsourcing for thread structure prediction using the threads in a MOOC course forum in edX (2013 spring course MITx 7.00x, henceforth "7.00x"). There were two stages in the whole annotation: trouble-shooting thread selection and thread structure annotation.

4.1 Trouble-Shooting Thread Selection

1000 threads (with number of replies larger than 1 and less than 10) were randomly selected from 29619 threads in the discussion forum of course 7.00x.

We designed a human intelligence task (HIT) to recruit online workers (turkers) on Amazon mechanical turk (AMT) and asked turkers to decide the intention of a forum thread. Turkers need to label whether a thread is intending to ask for help.

Table 1. Dialogue act classes

Category	Sub-category
Question	Question-question
	Question-add
	Question-correction
	Question-confirmation
Answer	Answer-answer
	Answer-add
	Answer-confirmation
	Answer-correction
	Answer-objection
Resolution	Resolution
Reproduction	Reproduction
Other	Other

Each thread was assigned to 3 turkers, and the final results were obtained by majority voting in order to minimize the effect of spammers and improve the reliability of labeling. We paid \$0.01 for each thread and a total of \$30 was paid for this task. 78 turkers attended this task and they completed it in 5 days.

To evaluate the quality of crowdsourcing annotation result, an expert was asked to annotate the same data. The two annotation results are shown in Table 2. The Cohen's kappa value between the two annotations is 0.812.

Table 2. Trouble-shooting thread annotation result

		Turkers	
		Yes	No
Expert	Yes	561	64
	No	26	349

4.2 Thread Structure Annotation

With the threads labeled as trouble-shooting in the previous stage, we further implemented a HIT where turkers were requested to assign the structure tag for each post in a thread (except the initial post). The structure tag of a post consists of two parts as defined in Sect. 3.2: Link Parent label and Dialogue Act label (one of the 12 sub-category labels in Table 1).

561 trouble-shooting threads (agreed by Expert and Turker in Table 2) were chosen for this annotation stage. There are 1977 posts and the average number of posts per thread is 3.5. The average number of words per post is 42.

In this task, we paid $0.05 for each thread and each thread was assigned to 5 turkers. $140.25 was paid. A total of 166 turkers were involved in this task, and 125 of them had some familiarity with courses material.

A majority voting method was used to obtain the final annotation results. The results were compared with an expert's sample annotation results (15 threads/55 posts) to calculate modified Cohen's kappa values for Link Parent label and Dialogue Act label. They were 0.76 and 0.51 respectively. This data set is called "MOOC data set" in the remainder of this paper.

5 Method

Our solution for thread structure prediction includes 2 steps: thread classification and thread structure prediction.

5.1 ME Model for Thread Classification

This step is actually a binary classification problem. The aim is to detect where a thread is a trouble-shooting thread. Maximum entropy (ME) model was used to address this problem.

5.2 Methods for Thread Structure Prediction

Because a post's role in a thread is influenced by its context or history, thread structure prediction task was formulated as a sequence labeling problem in this work. Considering the supervised learning method, the CRF model is a good choice according to previous work [10, 11].

Supervised learning works fine if we have a large number of labeled data. But the reality is that labeled data are hard to find when one is faced with a new problem or new domain. To deal with this situation, we proposed a lightly supervised machine learning method to predict the structure of a trouble-shooting thread.

Lightly supervised learning is a kind of compromise between unsupervised learning and semi-supervised learning. It can estimate the model parameters with *a priori* knowledge and unlabeled data. There are several frameworks that can utilize *a priori* knowledge to do model parameter estimation. We used the Generalized Expectation (GE) criteria framework, which was proposed by McCallum [15] and is suitable for combination with discriminative model.

In practice, GE criteria were used as a term in the object function to involve the feature constraints (*a priori* knowledge) into model parameter estimation. Different score functions could be defined to express the model preferences on some features. For example, Formula 1 defined a KL divergence function to calculate the differences between the prior distribution $\tilde{\Phi}$ and model distribution $\mathrm{E}_{(\mathrm{p}(\mathrm{y_U}|\mathrm{x};\theta))}[\Phi(\mathrm{x},\mathrm{y_U})]$ of feature $\Phi(\mathrm{x},\mathrm{y_U})$.

$$S(E_{(p(y_U|x;\theta))}\,[\Phi(x,y_U)]) = -D_{KL}\,(\tilde{\Phi}||E_{(p(y_U|x;\theta))}\,[\Phi(x,y_U)]) \tag{1}$$

The feature constraints in the GE criterion could be obtained in the following manner: assigned by domain experts; calculated from feature annotation data; calculated from sample annotation data.

The GE criterion needs to be combined with the concrete machine learning model to estimate the model parameters. So a method combining CRF and GE criteria (GE-CRF) for thread structure prediction was proposed in this work. The object function of the proposed GE-CRF is defined as formula 2.

$$O(\theta) = \log p(y_L|x;\theta) + S(E_{(p(y_U|x;\theta))}\,[\Phi(x,y_U)]) + \log p(\theta) \tag{2}$$

In formula 2, θ represents the parameters of the CRF model; $\log p(\theta)$ is the regularization term to constrain the size of θ; $\log p(y_L|x;\theta)$ is used to calculate the likelihood of labeled samples. It can be removed if there is no labeled sample. The Mallet toolkit was used to implement the proposed method.

5.3 Feature Description

Features used for trouble-shooting thread classification were borrowed from [13], including: number of question marks; number of question words (5W1H); N-gram features (1-g to 5-g); authorship and number of posts in the current thread.

Features used for thread structure prediction are drawn largely from the work of [11, 12]. Three categories of features were involved: structural features, semantic features and author features. The detailed feature descriptions are shown in Table 3.

Table 3. Features for thread structure prediction

Feature category	Feature name	Feature description
Structure features	Initiator	Whether the author of current post is the initiator of the thread
	Position	The position of current thread
Semantic features	qmark	# of question marks in a post
	emark	# of exclamation marks in a post
	url	# of URLs in a post
	PostSim	The relative position of the most similarity post
Author feature	PageRank	PageRank value of the author of current post

6 Experiment

This section reports our experimental results for the two steps mentioned in Sect. 5.

6.1 Thread Classification Results

With the ME model, the data set annotated in Sect. 4.1 and the features mentioned in Sect. 5.3, 10-fold cross validation average results for trouble-shooting threads are shown in Table 4. The results are comparable with results in [13].

Table 4. Trouble-shooting thread classification results

Classes	Precision	Recall	F-measure
Trouble-shooting	0.789	0.955	0.864
Non trouble-shooting	0.842	0.485	0.615
All	0.799	0.799	0.799

6.2 Thread Structure Prediction Experiment

In this part, we compared the prediction performances between supervised CRF and GE-CRF with different setting as shown in Table 5. Accuracy was used as our evaluation metric. In Table 5, all results were average over 10-fold cross validation on the MOOC data set except the results in row 2 and row 4. CNET was a data set built by [12] in another domain which had the same tag set as the MOOC data set.

The position-based baseline method proposed by [11] achieved an accuracy of 0.47. It classified all the first posts of a thread as “0 + Question-question” and all the second posts of a thread as “1 + Answer-answer”.

We can see that supervised CRF trained on the MOOC data set obtained the best performance. The lower accuracy of CRF trained on the CNET data set indicates that the label distribution differs between the two data sets. The results in row 5 confirm this observation. Figure 1 shows the dialogue act category distribution differences between the MOOC data set and the CNET data set.

For GE-CRF, the method used to obtain the feature constraints is vital to the final performance. Here we present 3 ways to obtain the feature constraints and compare their performances: (1) Obtain feature constraints from an existing data set in another domain (the CNET data set was used); (2) Obtain feature constraints by expert assignments; (3) Obtain feature constrains from labeled MOOC data. The third way (row 7) realized the highest accuracy in all GE-CRF setting. So the feature constraints calculated from the data with identical distribution as the test set are most effective.

The score function of GE also affected the results. In Table 5, KL denoted the KL divergence score function and L2 denoted the squared difference function. KL's performance was better than L2's. But for KL, every label needs to be assigned a constraint value, which is not convenient when the feature constraints are assigned by experts.

Table 5. Thread structure prediction results

Method	Training data	Accuracy
Position-based		0.47
CRF	MOOC	0.576
CRF	CNET	0.521
GE-CRF	Feature constraints calculated from labeled CNET data + unlabeled MOOC data	0.423 (L2) 0.461 (KL)
GE-CRF	Feature constraints assigned by an expert + unlabeled MOOC data	0.495 (L2)
GE-CRF	Feature constraints calculated from labeled MOOC data + unlabeled MOOC data	0.501 (L2) 0.517 (KL)

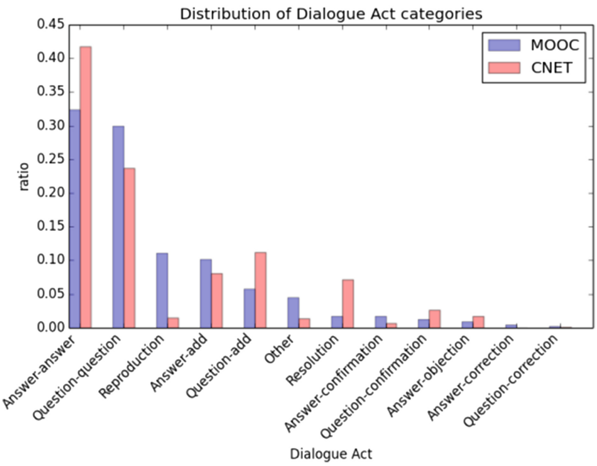

Fig. 1. Dialogue act category distributions of MOOC data set and CNET data set (Color figure online)

The thread structure prediction performance of this work is lower than what was achieved in CNET data set in [11]. The reason for that may be 2-fold: (1) Interactions in MOOC forum threads are more diverse than in CNET forum. The performance of a position-based baseline could be a kind of evidence: in the CENT data set it was 0.515, while it was 0.47 in the MOOC data set. (2) The annotation consistency of the CNET data was higher than that of the MOOC data because the kappa values of the CNET data (LP: 0.78, DA: 0.59) are higher than those of the MOOC data.

7 Conclusion

This work defined the trouble-shooting thread selection problem and thread structure prediction problem for MOOC forums. ME model was used to address the trouble-shooting thread selection problem and CRF and GE-CRF were adapted for thread structure prediction problem. The contributions of this paper include: First, We built an annotated data set by crowdsourcing for understanding the interaction of trouble-shooting threads in MOOC forums. Our practice showed that crowdsourcing is a cost effective way to annotate forum data. Second, we proposed a framework for thread structure analysis from scratch, which includes two steps: thread classification and structure prediction. Third, we provided supervised and lightly supervised methods for thread structure prediction in different situations and compared their performances.

Acknowledgment. This work is sponsored by Quanta Computers, Inc. under the Qmulus Project and National Natural Science Foundation of China (61572151 and 71573065).

References

1. Lori, B., David, P., Jennifer, D., Glenda, S., Ho, A., Seaton, D.T.: Studying learning in the worldwide classroom: Research into edx's first mooc. Res. Pract. Assess. **8**, 13–25 (2013)
2. Abel, F., Bittencourt, I.I., Henze, N., Krause, D., Vassileva, J.: A rule-based recommender system for online discussion forums. In: Nejdl, W., Kay, J., Pu, P., Herder, E. (eds.) AH 2008. LNCS, vol. 5149, pp. 12–21. Springer, Heidelberg (2008)
3. Anaya, A.R., Boticario, J.G.: A data mining approach to reveal representative collaboration indicators in open collaboration frameworks. In: International Working Group on Educational Data Mining, pp. 210–219 (2009)
4. Dringus, L.P., Ellis, T.: Using data mining as a strategy for assessing asynchronous discussion forums. Comput. Educ. **45**, 141–160 (2005)
5. Lin, F.-R., Hsieh, L.-S., Chuang, F.-T.: Discovering genres of online discussion threads via text mining. Comput. Educ. **52**, 481–495 (2009)
6. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J.: Engaging with massive online courses. In: Proceedings of the 23rd International World Wide Web Conference, pp. 687–698 (2014)
7. Wen, M., Yang, D., Rosé, C.: Linguistic reflections of student engagement in massive open online courses. In: Proceedings of the International Conference on Weblogs and Social Media (2014)

8. Ramesh, A., Goldwasser, D.: Modeling learner engagement in MOOCs using probabilistic soft logic. In: NIPS Workshop on Data Driven Education, pp. 1–7 (2013)
9. Stump, G.S., Deboer, J., Whittinghill, J., Breslow, L.: Development of a framework to classify MOOC discussion forum posts : methodology and challenges. In: NIPS Workshop on Data Driven Education, pp. 1–20 (2013)
10. Wang, H., Wang, C., Zhai, C., Han, J.: Learning online discussion structures by conditional random fields. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 435–444 (2011)
11. Wang, L., Lui, M., Kim, S.N., Nivre, J., Baldwin, T.: Predicting thread discourse structure over technical web forums. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 13–25 (2011)
12. Kim, S., Wang, L., Baldwin, T.: Tagging and linking web forum posts. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning, pp. 192–202 (2010)
13. Hong, L., Davison, B.D.: A classification-based approach to question answering in discussion boards. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, pp. 171–178. ACM Press, New York (2009)
14. Cong, G., Wang, L., Lin, C.-Y., Song, Y.-I., Sun, Y.: Finding question-answer pairs from online forums. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 467–474. ACM Press, New York (2008)
15. McCallum, A., Mann, G., Druck, G.: Generalized expectation criteria (2007)