# Automatic Grading of Short Answers for MOOC via Semi-supervised Document Clustering

Shumin Jing
Warner School of Education
University of Rochester, Rochester, NY, USA
jingshumin@gmail.com

## ABSTRACT

Developing an effective and impartial grading system for short answers is a challenging problem in educational measurement and assessment, due to the diversity of answers and the subjectivity of graders. In this paper, we design an automatic grading approach for short answers, based on the non-negative semi-supervised document clustering method. After assigning several answer keys, our approach is able to group the large amount of short answers into multiple sets, and output the score for each answer automatically. In this manner, the effort of teachers can be greatly reduced. Moreover, our approach allows the interaction with teachers, and therefore the system performance could be further enhanced. Experimental results on two datasets demonstrate the effectiveness of our approach.

## Keywords

Clustering, semi-supervised learning, short-answer grading

## 1. INTRODUCTION

Grading short answers is a challenging problem in the conventional educational measurement and assessment [6, 4], due to the diversity of answers and the subjectivity of graders. Especially, in the era of the massive open online course (MOOC), this problem becomes critical. MOOC provides plenty of courses, and has attracted over 10 million users during the past few years. However, traditional assessments are not suitable for MOOC. For example, in most MOOC platforms, short answers appear frequently in various quizzes and exams. Obviously, hiring lots of graders is not a feasible solution. Thus, it is very necessary to develop an automatic grading system for short answers. The automatic grading system for short-answers has been widely studied during the past decade [2]. Most recently, a system named "Powergrading" was presented by Microsoft Research, which achieved quite impressive performance [1].

We would argue that clustering is a straightforward solution to automatic grading. For short answer grading, the motivation of using clustering is that, the similar short answers should have high similarity values, while the dissimilar ones should have low similarity values. Therefore, those similar short answers could be assigned into the same group. We can then infer the final scores of those answers according which groups they belong to.

In this paper, we aim to design an automatic grading approach for short answers. Our approach is expected to solve the assessment challenge in MOOC. Moreover, it can also be applied to traditional educational assessment scenario, to reduce the efforts of teachers. We will present the methodology of our approach, discuss its influence in online education, and report the quantitative results and analysis.

## 2. METHODOLOGY

### 2.1 Feature Representation

In our problem, each short answer can be treated as a short document. Let $W = \{f_1, f_2, \cdots, f_m\}$ denote a complete vocabulary set of the short answers after the stopwords removal and words stemming operations. We can get the term-frequency vector $X_i$ of short answer $d_i$ as follows

$$X_i = [x_{1i}, x_{2i}, \cdots, x_{mi}]^\top \tag{1}$$

$$x_{ji} = t_{ji} \times \log(\frac{n}{\text{idf}_i}) \tag{2}$$

where $t_{ji}$, $\text{idf}_i$, $n$ denote the term frequency of word $f_j$ in short answer $d_i$ , the number of short answers containing word $f_j$, and the total number of documents in the corpus, respectively.

By using $X_i$ as a column, we can construct the term-short-answer matrix $X$.

### 2.2 Semi-Supervised Clustering for Short-answer Grading

We observe that, the label information of short answers is neglected in the basic document clustering approach. However, by leveraging the expertise of teachers, we can usually get some useful information. For example, teachers will tell us which two answers are essentially similar to each other, although they look quite different on the first sight.

To make use of such useful information, we propose a semi-supervised document clustering approach. The basic idea

is to add some constraints, including positive ones and negative ones. The former one shows us which short answers are similar, and we can always put them into the same cluster. On the other hand, the latter one tells us which short answers cannot be grouped together.

Inspired by the semi-supervised clustering algorithm [3, 5], we present the non-negative semi-supervised document clustering (SSDC) algorithm for short-answer grading as follows.

Let $A = X^\top X$ denote the document (e.g., short-answer) similarity matrix. In our approach, we first employ the symmetric non-negative tri-factorization as follows

$$A = QSQ^\top \tag{3}$$

where $Q$ is the cluster indicator matrix. Each element in $Q$ represents the degree of association of the short-answer $d_i$ with cluster $j$. The cluster membership information is determined by seeking an optimization matrix $S$.

In the semi-supervised setting, we are given two sets of pairwise constraints on the short-answers, including the must-link constraints $C_{ML}$ and cannot-link constraints $C_{CL}$. Every pair in $C_{ML}$ means this pair of short-answers should belong to the same cluster; every pair in $C_{CL}$ means this pair of short-answers should belong to different clusters.

Then, the objective function of SSDC algorithm is

$$\begin{aligned} &J = \arg\min \|\bar{A} - QSQ^\top\|^2 \\ &s.t., S \geq 0, Q \geq 0, \end{aligned} \tag{4}$$

where $\bar{A} = A - R_+ + R_-$. $R_+$ and $R_-$ are two penalty matrices, considering the two constraint sets $C_{CL}$ and $C_{ML}$.

The problem (4) can be solved efficiently using the standard gradient descent algorithm. The update rules of $S$ and $Q$ are given below

$$S_{ij} = S_{ij} \frac{(Q^\top \bar{A} Q)_{ij}}{(Q^\top QSQ^\top Q)_{ij}} \tag{5}$$

$$Q_{ij} = Q_{ij} \frac{(\bar{A} QS)_{ij}}{(QSQ^\top QS)_{ij}}. \tag{6}$$

After obtaining the optimized $S$ and $Q$, we can use them to infer the cluster labels for each short answer.

Finally, we can assign the score for each short-answer. For example, we know that the score of one template answer is 8.0. If another short-answer and this template answer belong to the same cluster, then the score of this short-answer should be close to 8.0. We also design a weighting strategy to adjust this score, based on the distance to the template answer.

# 3. EXPERIMENTS

We utilize the data set provided by Microsoft Research, which is also analyzed in the paper (Basu, Jacobs & Vanderwende, 2013). It contains the responses from 100 and 698 crowdsourced workers to each of 20 short-answer questions. These questions are taken from the 100 questions published by the United States Citizenship and Immigration Services as preparation for the citizenship test. It also contains labels of response correctness (grades) from three judges for a subset of 10 questions for the set of 698 responses (3 x 6980 labels).

**Table 1: The Results on MSR Dataset and MOOC Dataset.**

| Method | MSR Dataset | MOOC Dataset |
|---|---|---|
| DC | 85.2% | 74.1% |
| Semi-supervised DC | 87.5% | 78.9% |

Besides, we also collect some short answers from MOOC websites. We will evaluate the performance of our approach on both datasets.

We evaluate the performance of our approach on the MSR dataset and MOOC dataset. As we have the ground truth information, we can report the accuracy of clustering algorithms. Table 1 shows the accuracies of our approach and the baseline method DC under different settings. It shows that our semi-supervised document clustering method always achieves better performance than DC on two datasets.

# 4. CONCLUSIONS AND FUTURE WORK

We studied the educational assessment problem in MOOC. In this paper, we proposed an automatic grading approach for short answers. By leveraging the benefits of document clustering, our approach was able to assign a large amount of short answers into different groups, and infer their scores accordingly. Moreover, we designed a semi-supervised approach, which is able to incorporate the expertise of teachers. The proposed approach fits the requirements of MOOC. Results on two datasets showed the effectiveness of our approach. Our paper provides an effective solution to the educational assessment problem. In the future, we will design more computer-aided systems to address the educational assessment problem.

# 5. REFERENCES


[1] S. Basu, C. Jacobs, and L. Vanderwende. Powergrading: a clustering approach to amplify human effort for short answer grading. *TACL*, 1:391–402, 2013.

[2] C. Brew and C. Leacock. Automated short answer scoring. *Handbook of automated essay evaluation: Current applications and new directions*, (136), 2013.

[3] Y. Chen, M. Rege, M. Dong, and J. Hua. Non-negative matrix factorization for semi-supervised data clustering. *Knowl. Inf. Syst.*, 17(3):355–379, 2008.

[4] P. Ihantola, T. Ahoniemi, V. Karavirta, and O. Seppala. Review of recent systems for automatic assessment of programming assignments. *Proceedings of the 10th Koli Calling International Conference on Computing Education Research*, 2010.

[5] L. Lelis and J. Sander. Semi-supervised density-based clustering. In *ICDM 2009, The Ninth IEEE International Conference on Data Mining, Miami, Florida, USA, 6-9 December 2009*, pages 842–847, 2009.

[6] C. R. Reynolds, R. B. Livingston, V. L. Willson, and V. Willson. Measurement and assessment in education. *Pearson Education International*, 2010.