Student Performance Evaluation of Multimodal Learning via a Vector Space Model

Subhasree Basu, Yi Yu, Roger Zimmermann
School of Computing, National University of Singapore
Singapore 117417
{sbasu,yuy,rogerz}@comp.nus.edu.sg

ABSTRACT

Multimodal learning, as an effective method to helping students understand complex concepts, has attracted much research interest recently. Our motivation of this work is very intuitive: we want to evaluate student performance of multimodal learning over the Internet. We are developing a system for student performance evaluation which can automatically collect student-generated multimedia data during online multimodal learning and analyze student performance. As our initial step, we propose to make use of a vector space model to process student-generated multimodal data, aiming at evaluating student performance by exploring all annotation information. In particular, the area of a study material is represented as a 2-dimensional grid and pre-defined attributes form an attribute space. Then, annotations generated by students are mapped to a 3-dimensional indicator matrix, 2-dimensions corresponding to object positions in the grid of the study material and a third dimension recording attributes of objects. Then, recall, precision and Jaccard index are used as metrics to evaluate student performance, given the teacher's analysis as the ground truth. We applied our scheme to real datasets generated by students and teachers in two schools. The results are encouraging and confirm the effectiveness of the proposed approach to student performance evaluation in multimodal learning.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; K.3.1 [**Computers and Education**]: Computer Uses in Education—*Distance learning*

Keywords

E-learning; multimodal learning software; vector space representation

1. INTRODUCTION

Advancements in computer technology have promoted information representation in different media modalities such as text, audio, image and video. Such diverse representations help to vividly explain complex concepts better than merely using textbooks. Along this trend, a variety of multimedia software has been developed with the aim of facilitating student learning [8]. Various websites like Coursera[1], Khan Academy[2], Black Board[3], *etc.*, also aid in the process of computer based learning. Such software as well as the websites have the following advantages over conventional media: i) Various graphical images or video clips are able to hold students' attention. ii) Rich multimedia-based materials are able to interest students to actively learn and perform better in class [6]. Adding to these, Internet based learning also facilitates distant learning across the globe. Courses from well known universities like Stanford and Duke are now available to students all over the world via Coursera.

On the other hand, computer-aided learning necessitates literacy skills which go beyond traditional media [5]. Students should obtain the ability to read, view, understand, comment on and analyze a broad range of information and knowledge in multimodal contextual surroundings. The widespread use of smartphones and other mobile devices brings e-learning opportunities closer to the students. This necessitates the efficient evaluation of students' performance and a timely feedback through online interactions with teachers.

We are building a system that is able to i) upload student-generated data related to study activities to our server, and ii) evaluate student performances on the server side and give some feedback to students. In this paper, we present a case study of student performance evaluation, where the learning activities are conducted via multimodal annotation, using a multimodal analysis software—MMA [5] (a commercial software for student education). We aim to evaluate whether students create meaningful annotations associated with visual and verbal elements together. Two schools (School A – a primary school and School B – a secondary school) have been invited to participate in our preliminary study. Students in both the schools were organized into 10 study groups of 3 to 4 students each. They collaborate over the Internet to annotate the learning material and express their understanding of the contents. They use the Google Hangouts plug-in that comes with the MMA software.

The vector space model [7] is often used in information filtering, retrieval and indexing. In this research, based on analyzing unique datasets exported by the MMA software, we apply a vector space model to represent each multimodal

[1]https://www.coursera.org/
[2]https://www.khanacademy.org/
[3]http://www.blackboard.com/

WISMM '14, November 7, 2014, Orlando, Florida, USA.

http://dx.doi.org/10.1145/2661714.2661723.

analysis generated by the collaboration of each study group. The teacher's analysis is used as the ground truth. On this basis, each student analysis is evaluated to reveal the overall performance of each student group. Our promising experimental results demonstrate that the proposed method is able to effectively evaluate students' performance of multimodal learning.

The remainder of the paper is structured as follows. Section 2 discusses some related work. Section 3 describes the system model, explaining how the overall performance evaluation system works. Section 4 presents multimodal data analysis and addresses how to explore a vector space model to represent each analysis produced by student study groups. Section 5 shows some results of students' performance based on several benchmarks. Finally, Section 6 concludes the paper and points out future work.

2. RELATED WORK

Several multimodal interfaces [4] and techniques have been suggested to provide various abilities to monitor and record students' learning activities. A multimodal learning environment makes it easier to improve learning performance by combining different media contents. Accordingly, analyzing and designing the performance of students who are learning in such environment is gaining importance. Learning analytics [2] is one such emerging research topic interrelated with data mining, knowledge modeling and information retrieval, covering various education settings. An educational system—LEMMA [1] has been developed to present tutorials for topics on rotational dynamics in multimedia 3D learning environments. A decision tree method is utilized to evaluate student's performance in courses. Attendance, class test, seminar and assignment marks were obtained from the students' management system to predict their performance. Online learning websites also sometimes uses various multimodal games and interactive videos as an aid for learning [3]. In all of these Learning Management Systems (LMS), the main drawback is the assessment of the student performance. We propose to evaluate the student performance by applying a vector space model in this work. Different annotation information is represented in a unified way, which facilitates the use of standard metrics in performance evaluation. Our representation of the system is applicable to any exercise which involves the students to annotate the learning material. Hence it can be effectively adapted to Internet-based multimodal applications as well.

3. SYSTEM MODEL

In this section, we explain the proposed system model of multimodal learning and student performance evaluation. The overall process is shown in Fig. 1. A teacher prepares in advance both learning materials and attribute space, which are stored on the server side. At the client side, learning material and attribute space are fetched and presented to students via the MMA software [5], which is employed to produce a multimodal analysis using annotations in the interactive learning situation. The MMA software [5] can also be implemented as a web application, with the learning material on the browser. Students perform an annotation-based study, annotating the learning material with the provided attribute space (catalog). The study results are exported to CSV files and sent back to the server for analysis. In the analysis, both the learning material and student analysis

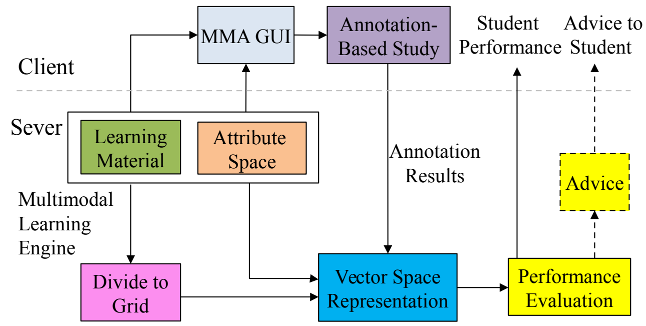

Figure 1: System model: multimodal learning and student performance evaluation.

are divided into grids, and processed to get a vector space representation as described in Algorithm 1. The teacher's analysis is stored a priori on the server as a ground truth. By comparing the analysis of a student against that of a teacher, a score will be given for the student, together with suitable feedback. Different learning is supported by allowing teachers to change the catalog for the study.

A snapshot of the software for the first case is presented in Fig. 2. A teacher pre-defines categories (text analysis and image analysis) and types within a category for the learning material. For example, the available types for the category "text analysis" is 5Ws1H (When, Where, Who, Why, What, How). All available attributes are shown on the right side of Fig. 2, under the heading Catalog. Meanwhile, an image and its text description are presented to students on the left side. Students are required to annotate parts of the text as well as the image based on their understanding.

Each annotation is divided into two steps:

- A salient object is marked by a shape (triangles, rectangles, *etc.*). There are various shapes available (not shown in Fig. 2) for the annotations.
- The category of an annotation is automatically determined based on whether it is superimposed on text or an image. A type is selected by a student to further annotate the marked object with more details.

The system stores both the object position (by using a shape covering the object) and its attribute (category and type). In other words, each annotation is composed of a tuple ⟨shape (object position), attribute (object category/type)⟩. A complete set of annotations in a study forms an analysis.

The screen display for the second case (not shown) is different because it is used for another learning, where both learning materials and attributes for annotations are different. However, with the attributes being defined, the MMA software can record student performance in the same way.

4. PROPOSED METHOD

The teacher provides his/her analysis as a ground truth, based on which the performance of a student group is evaluated. Each analysis is composed of multiple annotation tuples. But the number of annotation tuples varies in each group analysis, and so does the order of annotations. In addition, an object might be assigned different attributes. A simple method to comparing two analyses requires to iteratively compare each annotation in one analysis to every annotation in another analysis for each attribute. The shapes

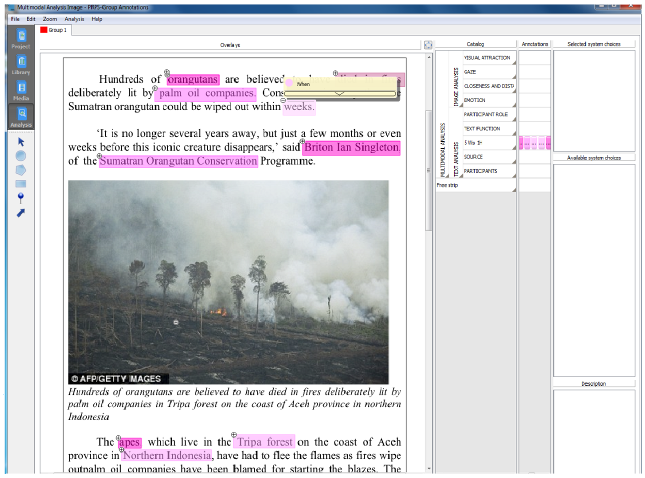

Figure 2: An example of annotation-based multimodal learning with the MMA software.

are manually placed by the students and teachers. Hence the shapes corresponding to the same object do not completely overlap, which increases the difficulty of performance evaluation.

In our initial study, we choose a vector space model to represent all analyses in a unified form, as follows.

- Representation of the study material by grid points. As described before, each annotation contains a shape defining the area of an object and an attribute describing its property. By dividing the whole analysis area into grid points, each annotated object can be represented by grid points in its coverage, with each grid point assigned the same attribute as the object.

- Representation of attributes of each grid point by a vector. Each grid point might be assigned multiple attribute values. To uniquely represent all information, each attribute is represented as a bit in a vector corresponding to the attribute space.

In this way, each analysis is represented as a 2-dimensional attribute matrix, each element of which is a vector. When the number of attributes is few, attributes at a grid point can be represented as bits in an integer. An attribute matrix defined in this way is actually a 3-dimensional matrix. The algorithm converting an analysis Γ with annotation tuples to an attribute matrix is described in Algorithm 1. Let i, j be the indices of grid points in the x-axis (transverse direction) and y-axis (portrait direction) of the study material, respectively, and k be the dimension of attributes. Each element $X_{i,j,k}$ of an attribute matrix X is a bit, '1' meaning the k^{th} attribute is set at the grid point (i, j) and '0' otherwise.

Figs. 3 and 4 show two attribute matrices, whose annotation results correspond to the study material in Fig. 2. The former is annotated by a student group and the latter is the ground truth by the teacher. An attribute matrix X is displayed as follows: all attributes at the same grid point (i, j) are represented as bits of an integer, whose logarithmic value is shown in the z-axis direction. Based on the two figures, it is easy to learn where a student analysis differs from a teacher's template, in either annotated objects (shape's place) or assigned attributes (height in the figure).

Algorithm 1 Generating an attribute matrix

```
procedure GenAttributeMat(Γ: an analysis)
    Initiate X as all zero 3-dimensional matrix.
    for t = 1, 2, 3, ··· do                ▷ Iterate each annotation
        (s_t, a_t) ← t^th annotation tuple in Γ.
        s_t is the shape, a_t is the attribute.
        k ← attribute index of a_t.
        for (i, j) in shape s_t do         ▷ Iterate each grid point
            X_{i,j,k} ← 1.
        end for
    end for
    Return X as an attribute matrix.
end procedure
```

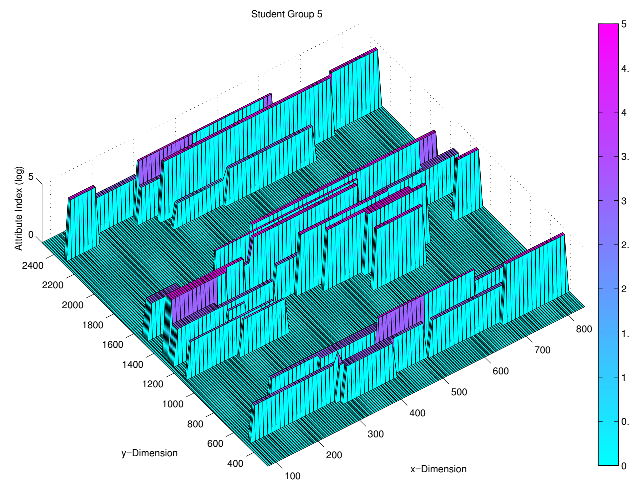

Figure 3: Multimodal analysis result by a student group (case 1).

5. EXPERIMENTAL RESULTS

School A, a primary school, and School B, a secondary school, were invited to take part in the experiments. Students in each school were divided into 10 groups. Students conducted their group study, annotating the learning material by collaborating with group members. The collaboration took place over the Internet using Google Hangouts to facilitate communication.

We computed the similarity between two analyses through their attribute matrices X and Y, where X is a student analysis and Y is the ground truth by a teacher. By representing each analysis as a 3-dimensional indicator matrix, their similarity can be estimated using standard metrics like recall, precision and the Jaccard index. In our evaluation, we computed the three metrics using the following equations:

$$Recall(X, Y) = \sum_{i,j,k} X_{i,j,k} \& Y_{i,j,k} / \sum_{i,j,k} Y_{i,j,k}, \quad (1)$$

$$Precision(X, Y) = \sum_{i,j,k} X_{i,j,k} \& Y_{i,j,k} / \sum_{i,j,k} X_{i,j,k}, \quad (2)$$

$$Jaccard(X, Y) = \sum_{i,j,k} X_{i,j,k} \& Y_{i,j,k} / \sum_{i,j,k} X_{i,j,k} | Y_{i,j,k}. \quad (3)$$

The three metrics are similar in that their numerators are the same, indicating a correctly annotated attribute at each

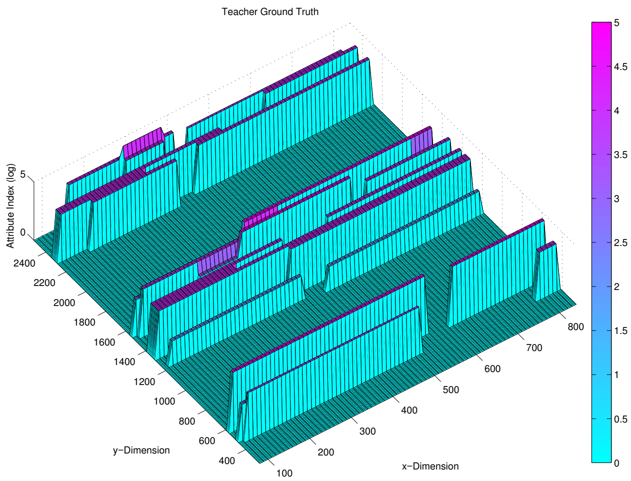

Figure 4: Multimodal analysis ground truth by a teacher (case 1).

Table 1: Student performance evaluation I (recall and precision).

	Recall		Precision	
School	School A	School B	School A	School B
Group 1	0.1786	0.1636	0.4175	0.8356
Group 2	0.2160	0.1413	0.4411	0.7513
Group 3	0.2274	0.0575	0.5072	1.0000
Group 4	0.4166	0.0241	0.5510	1.0000
Group 5	0.3955	0.4232	0.4558	0.8503
Group 6	0.2667	0.0429	0.5601	1.0000
Group 7	0.0000	0.3810	0.0000	0.8844
Group 8	0.2493	0.8876	0.6278	0.8046
Group 9	0.2083	0.3294	0.6628	0.9492
Group 10	0.2627	0.1978	0.5193	0.7898

grid point. But their denominators are different, counting the teacher's ground truth in recall, counting the students' annotations in precision, and counting both in the Jaccard index. Therefore, a high recall is associated with a low false negative (objects required by teachers are correctly annotated), a low precision represents a high false positive (non-relevant objects are annotated by students), and Jaccard index can be regarded as an overall performance.

Recall and precision by 10 groups of students from two primary schools are summarized in Table 1. Recall tends to be low while precision is relatively high. In other words, few non-relevant objects are annotated, but the correctly annotated objects are also few. This indicates that students are conservative in their annotation decisions. Results of the Jaccard index are shown in Table 2, which is lower than both recall and precision. The relative order of the student groups' performance in Tables 1 and 2 is consistent with the subjective evaluation of the groups by their teachers. Therefore, the proposed method represents a useful tool for them.

6. CONCLUSIONS AND FUTURE WORK

In multimodal learning, student-generated data involve multiple media and are annotated with shapes covering objects and attributes describing properties. How to efficiently grade students' performances and give timely suggestions is getting very important for online multimodal learning. In this work, we proposed to leverage a vector space model to represent each analysis as a 3-dimensional indicator matrix, which facilitates the evaluation of student performance in the multimodal learning using standard metrics. Experiments in two different case studies conducted in two schools confirm that the proposed method is applicable to different educational settings, if only the attribute space is predefined. In the future, we will implement the whole system to provide students' performance evaluation online, consider the sparsity of student annotations and refine the proposed algorithm to improve its running efficiency.

Table 2: Student performance evaluation II (Jaccard Index).

School	School A	School B
Group 1	0.1430	0.1585
Group 2	0.1696	0.1350
Group 3	0.1862	0.0575
Group 4	0.3110	0.0241
Group 5	0.2686	0.3938
Group 6	0.2205	0.0429
Group 7	0.0000	0.3630
Group 8	0.2172	0.7302
Group 9	0.1883	0.3237
Group 10	0.2113	0.1879

7. ACKNOWLEDGMENTS

This research has been supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office through the Centre of Social Media Innovations for Communities (COSMIC).

8. REFERENCES

[1] M. Breisinger, T. Hollerer, J. K. Ford, and D. Folsom. Implementation and Evaluation of a 3D Multi Modal Learning Environment. In *Educational Multimedia, Hypemedia and Telecommunications*, pages 2282–2289, 2006.

[2] T. Elias. Learning Analytics: Definitions, Processes and Potential. In *Learning 23*, pages 134–148, 2011.

[3] R. W. Lau, N. Y. Yen, F. Li, and B. Wah. Recent Development in Multimedia e-Learning Technologies. *World Wide Web*, 17(2):189–198, March 2014.

[4] T. Martin, A. Boucher, and J.-M. Ogier. Multimodal Analysis of Recorded Video for e-Learning. In 13^{th} *ACM International Conference on Multimedia*, pages 1043–1044, 2005.

[5] K. L. O'Halloran, A. Podlasov, A. Chua, and M. K. L. E. Interactive Software for Multimodal Analysis. *Visual Communication*, 11:363–381, 2012.

[6] Y. C. J. Phua and L. C. Chew. What Do Secondary School Students Think About Multimedia Science Computer Assisted Assessment (CAA)? In *Computer Assisted Assessment*, July 2012.

[7] G. Salton, A. Wong, and C. Yang. A Vector Space Model for Automatic Indexing. *Information Retrieval and Language Processing*, 18(11):613–620, 1975.

[8] R. Zimmermann and Y. Yu. Social Interactions over Geographic-aware Multimedia Systems. In 21^{st} *ACM International Conference on Multimedia*, pages 1115–1116, 2013.