Visualizing Paragraph Closeness for Academic Writing Support

Stephen T. O’Rourke, Rafael A. Calvo
School of Electrical and Information Engineering, University of Sydney
{stepheno, rafa}@ee.usyd.edu.au

Abstract

In this paper, we describe a novel visualization to support formative assessment in academic writing. The visualization makes use of text mining techniques to provide insight on the flow of topics in an essay. We propose that visualization can be used to mitigate many of the problems associated with the subjectivity of essay assessment by bringing greater insight to an essay’s latent features. The proposed visualization method involves a process of non-negative matrix factorization (NMF), to uncover topics in an essay, followed by multidimensional scaling (MDS), to map the topic closeness of the essay’s paragraphs. We evaluate the visualization method with a corpus of 44 short essays written by university students.

1. Introduction

Formative assessment refers to the process of helping a learner to improve their performance by providing them with feedback on the quality of their work [1]. In academic settings, an essay is considered by many educators to be one of the most useful tools for assessing a learner’s writing ability and attainment of learning outcomes. Essay assessment, however, is a highly subjective process that is prone to several types of errors [2], leading to many inconsistencies in the grades given by different assessors. While some aspects of essay writing can be assessed based on objective features, such as the word count or number of spelling errors, others require a more subjective interpretation, such as the flow of topics and ideas. These aspects of essay writing are somewhat inexact and cannot be easily abstracted without a more in-depth analysis.

North [3] discusses how visualization can be used to bring greater insight to qualitative features of complex datasets. A qualitative view can help to aid users in completing subjective tasks, such as the identification of latent structures and relationships [4], by increasing user speed and accuracy. Visualization has previously been applied in many educational scenarios to bring greater insight to learning tasks, such as the awareness of participation in collaborative groups [5].

We propose that visualization can be used to mitigate many of the problems associated with the subjectivity of essay assessment by bringing greater insight to an essay’s latent features. Building on the well established rubric of the MASUS procedure [6] for assessing academic writing skills, we hope to have identified features relevant to essay assessment. Importantly, it should be noted that we are not attempting to visually score an essay based on an optimal essay formula, but rather use visualization to aid in its subjective interpretation.

In the next section, we provide a comparison with previous work to highlight the novelty of our contribution. Section 3 presents a short description of the algorithms used to analyze the essay features. In Section 4, the visualization is introduced and explained using examples. In Section 5, we provide an evaluation of the visualization and assess to what extent it allows one to capture features relevant to the MASUS procedure. Finally, in Section 6, we conclude the paper and give directions for future work.

2. Background

Since the emergence of computer-assisted essay assessment in the 1960’s, many lines of research have emerged to support the various stages of the writing process. This research has resulted in the development of a number of software tools which are focused on providing automated essay scoring (AES), to grade essays [7-9], as well as automated feedback, to identify common problem [10, 11] and highlight features for reflection [12].

The adoption of AES is one approach that has been proposed to bring greater consistency and objectivity to the essay assessment process. While AES has been
shown to have inter-assessor correlations comparable to those of humans [13], many scholars are still highly critical of the validity and robustness of the approach. AES systems typically use essays manually scored by humans to train an essay specific assessment model, which can then be used to approximate the grades for unscored essays. Some scholars suspect that the reportedly high inter-assessor correlations between AES and human assessors may be indirectly due to features which commonly occur in well written essays, such as the use a rich vocabulary [14], rather than a direct assessment of the essay answer. Critics of AES are also concerned about the affects of students ‘unintelligently’ writing to the audience of a machine, attempting to match its ideal ‘formulaic’ model at the expense of logical argument [15]. Overall many scholars are still unconvinced that AES tools can simulate the complexities needed to accurately score an essay.

Due to the associated problems with AES, along with solid pedagogical reasoning, many researchers have focused on going beyond the automated scoring of an essay to instead provide a learner with automated feedback. Automated feedback tools typically use the same techniques as AES to extract essay features, but instead, attempt to meaningfully translate this information to help improve a learner’s performance. The Writer’s Workbench tool provides feedback on spelling, style and diction by analyzing English prose and suggesting possible improvements [11]. The Sourcer’s Apprentice Intelligent Feedback tool [10] provides automatic feedback on sourcing by detecting citations and plagiarized sentences and suggesting ways to resolve them. More recently, the Glosser tool [12] sought to provide automated feedback by highlighting important essay features and using thought provoking question to promote reflection.

Our approach differs from previous tools in that it doesn’t attempt to score or highlight certain features for reflection, but rather, use these features to visualize the essay in a new way. The essay visualization contains exactly the same information as its equivalent essay text, but uses visual techniques to emphasize holistic features of the essay content. Thus, the visualization provides insight into an essay’s latent features while still maintaining the subjectivity of the assessment process.

3. Analyzing Topic Flow

Text mining is the process of applying linguistic and statistical techniques to extract information from unstructured texts. We use text mining techniques to model the topic mixture of a document’s paragraphs and map it to a two-dimensional space for visual consumption. In short, the automated mapping approach involves performing the following steps on a document: 1) preparation of a term-by-paragraph matrix, 2) creation of a topic model using non-negative matrix factorization (NMF), 3) projection of the topic model to a two-dimensional space using multidimensional scaling (MDS), and finally, 4) visualization of the paragraphs.

NMF is a matrix factorization technique which can be used to model a document’s paragraphs as a mixture of the topics they span [16]. The NMF model is generated from a term-by-paragraph matrix of weighted elements. The matrix elements are weighted using the TF-IDF scheme [17], which weighs a term by its local frequency in a paragraph relative to its global frequency in the whole document. TF-IDF gives greater emphasis to the more ‘important’ terms in a document while also eliminating the ‘noise’ of common terms. The NMF topic model is generated by decomposing the term-by-paragraph matrix into the product of two k rank non-negative matrices, W and H. Given that k represents the number of latent topics in a document, W becomes a term-by-topical matrix, representing the mixture of terms in each topic, and H becomes a topic-by-paragraph matrix, representing the mixture of topics in each paragraph. A review of algorithms for performing NMF is available in the paper [18].

The topic closeness of two paragraphs can be calculated by measuring the ‘distance’ between their topic mixtures in the NMF model. For the purposes of the essay visualization, the measure of cosine similarity is used to calculate the distance between paragraphs. By calculating the distances between all the paragraphs in a document, the relative flow of topics can be deduced.

In order to meaningfully visualize topic flow, MDS is used to transform the paragraph distances to a two-dimensional representation. MDS is a technique for representing distances between high dimensional objects as distances among points in a low dimensional geometric space [19]. For example, given the distances between all the cities in a country, MDS could be used to plot the relative location of each city on a map. In a similar way, we use MDS to reduce the observed complexity of the paragraphs’ high dimensional topic mixture to a two-dimensional map-like representation. The MDS algorithm works by relating the distances between paragraphs to their estimated Euclidean distances in two-dimensional space. This transformation is performed using a procedure called iterative majorization [20]. The iterative majorization algorithm undertakes a least-squares approach to MDS by attempting to minimize a loss function called stress.
4. Overview of the Visualization

A well written essay should have a clear and logical flow of ideas, inherently linked through paragraphs. To assess this aspect of essay writing, we propose the use of visualization to map the flow of topics among paragraphs. The proposed visualization is intended to help assess essay features related to the structure and development of the answer, and the control of academic writing, as described in the MASUS procedure [6].

Figure 1 illustrates a ‘map’ of the topic flow generated from the introduction paragraphs of this paper. A node-link diagram is used to represent the document. The paragraph nodes are plotted on a circular grid with the diameter of the grid equal to the maximum possible distance between paragraphs (i.e. no topic overlap). Arrows and text labels are used to convey the sequence of paragraphs. The similarity between the paragraphs’ topic mixtures is solely represented by their relative proximity with the actual orientation of axes having no meaning.

Figure 1: A map of the topic flow generated from the introduction paragraphs of this paper.

The five-paragraph essay is a writing paradigm that has been proposed to structure argument in academic writing [21]. The format of a five-paragraph essay consists of an introduction paragraph, stating the topic thesis and introducing the main supporting subtopics, three body paragraphs, each of which present different supporting subtopics, and a concluding paragraph, which restates the thesis topic and summarizes the supporting subtopics. When mapping the topic flow of an ideally structured five-paragraph essay, one might expect a circular layout of sequential paragraphs, indicating a smooth change in topic over the essay, with the introduction and conclusion paragraphs positioned on similar points. In contrast, a poorly structured essay might present many rough shifts in topic, with paragraphs positioned somewhat randomly around the map.

Figure 2 illustrates the maps of two short essays which have been assessed according to the MASUS procedure. The essay on the left was given a high grade while the essay on the right was given a low grade. Examining these maps in depth, it can be observed that the topic flow of the high grade essay appears to resemble that of an ideal five paragraph essay, while the topic flow of the low grade essay appears somewhat disorganized. In a formative assessment scenario, the rough topic shifts in the low grade essay might be identified as a problem. This could, for example, indicate a lack of information linking two subtopics, which may require the insertion of an additional paragraph or transitional sentence.

Whilst the scenario described above makes sense for the five-paragraph essay, it is important to keep in mind that there are many genres and paradigms for writing that do not follow this same strict structure. Furthermore, essay assessment is subjective and it is unlikely that a strictly optimal set of visual heuristics could be identified. Such work is more in the domain of AES and is not considered here. Instead the map is intended to capture the topic flow an essay, leaving its interpretation to the users themselves or those who require familiarity with the essay, rather than as a basis for external judgment.

5. Evaluation

The visualization method was evaluated using a text corpus of 44 short essays handwritten in a timed assessment by university students. The corpus includes its associated essay marks, which were assessed according to the MASUS procedure.
The aim of the experiment was to determine whether there are in fact features identifiable in the maps which relate to topic flow. We performed a qualitative analysis of the features pertaining to good and poor topic flow by comparing how accurately rough topic shifts could be identified using the maps. In order to have an objective measure on which to evaluate our results, a subset of contaminated documents was generated. The contaminated documents were generated by randomly mixing paragraphs from essays of different topics.

In the authors’ opinion, the introduced rough topic shifts were clearly evident in the maps of the contaminated documents. A sample map of a document from the contaminated essay subset is shown in Figure 3. Looking at the contaminated map, the foreign paragraph (circled) can easily be visually differentiated from the original essay. Although this experiment presents an extreme case, it demonstrates a realistic scenario where the map could be useful for reviewing the topic flow of an essay.

Presumably though, the usage of the visualization method on essays with a larger number of paragraphs might become problematic. This is partly due to the nature of the MDS algorithm, which in any optimal configuration is still ultimately restricted by the complexity of the dataset. Similarly, in other more lengthy writing genres, such as a thesis or book, it would also be difficult to abstract meaning from a map. However, using the same concept and techniques, larger documents could be broken up and mapped as separate chapters, or alternatively, larger text passages could be used to create a chapter map of a whole document. Alternatively, at the other end of the scale, a shorter document such as an abstract could be broken up and mapped at a sentence level.

Figure 2: A map of an essay with a high grade (left) and an essay with a low grade (right). Note that the low grade essay has many rough shifts in topics while the high grade essay has a gradual topic flow that starts and finishes on a similar topic mixture.

Figure 3: An essay contaminated with a foreign paragraph (circled) of a different topic.
6. Conclusions and Future Work

In this paper, we presented how visualization can be used to capture latent essay features. We described a novel visualization, motivated by the MASUS procedure, for assessing aspects relating to the topic flow of an essay. The visualization method involves a process of NMF, to uncover topics in an essay, followed by MDS, to map the topic closeness of the essay’s paragraphs.

The maps were evaluated using a corpus of short essays written by university students. The evaluation presented a realistic scenario, which demonstrated the usefulness of the visualization for detecting rough topic shifts. Critical to this analysis was the ability of the visualization to illustrate the topic flow of an essay as a whole, as opposed to just consecutive paragraphs.

In future work, we will evaluate the visualization in such a way that directly measures the insight gained by users in a formative assessment scenario. This evaluation will ultimately determine whether the maps are in fact achieving their purpose. We predict that the topic flow of an essay can be subjectively assessed faster and more accurately with the aid of visualization.

7. References


