

Visualising Student Satisfaction

Samuel Cunningham-Nelson, Mahsa Baktashmotlagh and Wageeh Boles.

Queensland University of Technology

Corresponding Author Email: samuel.cunninghamnelson@qut.edu.au

CONTEXT

Student satisfaction is an important metric used in teaching and education, and is used in most educational institutions. There are many ways to measure student satisfaction (Elliott, 2002), however student satisfaction is often given as both a numerical score on a Likert scale as well as a text comment which contains further information. Although the numerical scores are often used, the free text comment is an invaluable source of information, often providing further pointers towards possible teaching enhancements.

PURPOSE

In this study, we explore the use of machine learning techniques to visualise student satisfaction. This visualisation will be exploited in the context of the following 2 research questions,

1. How can we use visual representations of comments to examine student satisfaction?
2. What impact can this have for educators?

APPROACH

Using a dataset of over 20 subjects, many student comments and reviews were available for analysis. Primarily, using an analysis method called Latent Dirichlet Allocation (LDA) (Blei, et al., 2003), topics can be extracted from these comments about subjects. Sentiment analysis is then used to find the positivity and negativity of certain comments.

RESULTS

Using the approach mentioned, comments from two subjects were analysed to demonstrate the capability of the process. For each subject, two figures were generated, one about the subject, and one focussing on a keyword or topic of interest. Both figures contain 9 automatically determined keywords as the focus of the plot. For each keyword, the length of the bar represents the frequency of keywords. The positive and negative sentiment is also represented on the same figure. From these figures, the reader can identify keywords related to a particular positive or negative sentiment.

CONCLUSIONS

In this work, we have shown a process which automatically analyses and produces a visual representation for student satisfaction. The produced visuals can be used by lecturers, subject coordinators or managers to compare and review several subjects at once. As a lecturer, you would not need to know how this system works, but have access to student feedback. With the student feedback, these plots can be automatically generated. The visualisations can lead to quick detection of both positive and negative aspects within a subject, thus prompting appropriate action.

KEYWORDS

Sentiment Analysis, Student Satisfaction, Machine Learning, Student Surveys

Introduction

Student feedback is very important for improving teaching practice and allowing students to express their views and ideas. Part of a teacher's role is to manage and address the expectations of students, and feedback and ratings from evaluation allows that to be done (Cheong Cheng, 1997). From feedback, pointers and information can be determined and hopefully influence the teaching style and method of delivery.

One form of feedback is determining whether the student had a positive, negative or neutral experience in a subject. Sentiment analysis is a method which aims to determine this positive, negative or neutral sentiment of a text statement. Many of these surveys ask for a rating on a Likert scale, which will show sentiment, but this number does not provide information about the sentiment of important components mentioned within a response. For example, a response may say "The lectures were informative. The tutorials were difficult". With a single rating or number, we can't determine which teaching activity students were satisfied or not satisfied with.

In this work, we explore some text analysis techniques to automatically analyse student responses. The main technique used here, known as Latent Dirichlet Allocation (LDA), allows a document to be classified into several topics, and each topic represented by some keywords (Blei, et al., 2003). This is an important initial step for producing a visualisation of student satisfaction.

Research Questions

In this study, we explore the use of machine learning techniques to visualise student satisfaction. We aim to develop a method which will improve on existing techniques used for visualisation, allowing the produced figures to be generated with an education use as the primary objective. This visualisation will be exploited in the context of the following 2 research questions,

1. How can we use visual representations of comments to examine student satisfaction?
2. What impact can this have for educators?

Motivation

Being able to visualise a student's perception of a topic or perception of a subject of interest is very powerful. Quite often lecturers and universities have access to a large number of responses from students and being able to easily summarise these would be useful in a number of applications. As an example, in the Reframe framework (QUT, 2015) students are asked to rate their overall satisfaction with the subject, and provide a free text comment. These satisfaction scores are often used to compare subjects and assess student satisfaction, but student satisfaction ratings can be flawed and may not be accurate. For the student perception of subjects to be examined properly, all the text comments need to be read. With large subjects, this can be quite time consuming for a lecturer to read through the comments and summarise the key points. If a Dean or manager is trying to assess multiple subjects, reading through all the comments is an almost impossible task, and there is no easy way to compare two subjects together. In this work, we develop a visualisation method for students' text comments, and compare our results to the outputs of two existing text analysis software.

Method

The Data Set

Introduced in a previous work (Cunningham-Nelson, Baktashmotlagh, & Boles, 2016) we use a dataset consisting of over 20 subjects across four years. QUT runs two student surveys titled 'Pulse' and 'Insight' (QUT, 2015) obtaining students' feedback in the early weeks of the semester and at the end of semester. In each of these surveys, students are asked to rate their views on three statements. They respond to each of these questions on a 5 point Likert scale. For this work, we are focusing on the final statement "I am satisfied with this unit so far" in both surveys. An open-ended response is then presented to the students, allowing them to give feedback or further information on the scores that they give. This feedback often contains suggestions for teaching improvements which can be used by teaching staff for future iterations of the subject. An example of one of these responses is,

Satisfaction rating: 5. "Great structured unit. Very well organised and great learning environments"

This shows a satisfaction score, and a comment which seems to reflect the student's satisfaction in the subject. This however is not always the case. The comment below shows an example where this is not necessarily the case,

Satisfaction rating: 3. "Way too much new information all at once, which is then built on the very next week, whether we have a solid foundation or not..."

This satisfaction score given shows that the student is ok with the subject, but this is not really reflected in the comment. It is important that we consider both the satisfaction scores as well as the comments given by the students.

Choosing a Subset of Data

The total data set consists of many subjects, across many semesters, and the two surveys. Depending on the desired purpose, the data can be used wholly, limited to a faculty, subject, semester, etc. The data chosen depends on the type of level of information the user wants. For analysis in this paper, the data used has been limited to single subjects within a single semester. Both the "Pulse" and "Insight" survey data was used.

Pre-Processing the Data

When text data is being analysed, it is important to pre-process it, to ensure that small inconsistencies don't have a large effect on the results. For this analysis, several actions were performed on the data, including,

- Changing the text to be all in lowercase, so that "Hello" is treated in the same manner as "hello".
- Removing the stop words (i.e. the, and, a) which are present in these responses, as these are not important for the overall meaning.
- Lemmatizing the words (using the base of each word). For example "running" would become "run" for the purposes of analysis.

Determining Keywords

After cleaning and preparing the data, methods which could then determine keywords were investigated. One obvious method to determine keywords would be to use a frequency count to see the words which are mentioned most frequently. This however does not group words of similar topics together, and would perhaps include words that are not necessarily keywords. Using the LDA method previously mentioned (Blei, et al., 2003), we can split the

document into a set number of topics, and this will group similar terms under one keyword. The implementation and testing of this was performed in Python. In this study, we have selected 9 keywords, and run the LDA methods 10 times as the keywords returned can be slightly different each time. The most common 9 keywords are selected as the keywords to represent the subject.

Finding Sentiment of a Response

After the keywords have been selected, the sentiment of each keyword is determined. Sentiment refers to the positivity or negativity of a word. Common examples where sentiment analysis can be used are movie reviews (Jong, 2011) or product ratings. Twitter data is often used a sample dataset to train and test sentiment analysis models (Pak & Paroubek, 2010).

When finding the sentiment in this work, the responses are initially broken into individual sentences, to isolate each idea. Using a dataset known as “AFINN” (Nielsen, 2011) we can see the sentiment of individual words. AFINN has a list of words which have been rated with an integer between positive and negative 5, depending on satisfaction. For example, the word “amazing” is labelled as a +4, and the word “conflict” is labelled as a -2. The sentiment of each word is added up across the sentence, to give an indication of the students satisfaction. If the overall score is negative, the sentiment is deemed to be negative, and if the score is positive, then the sentiment is also positive. The sentiment calculated is then linked with the keywords that occur within a sentence. The sentiment and keywords are then combined into a figure, together with frequency of occurrence. This provides an initial view of the subject overall.

Focussed Keyword Feedback

Taking this further, we can narrow down the data used to include only responses with one keyword within the selected unit. We can then search for additional keywords, which are mentioned often in conjunction with the keyword of interest. This is used as an additional visual, to give more information into reasons why a keyword may have a positive or negative sentiment.

Results

The analysis techniques described above can be used to identify subject overall feedback or feedback specific to a keyword. As a use case, two subjects in single semesters were selected to visualise student perception by using keyword specific data and identifying positive and negative sentiment.

Example 1

For the first example, two plots show an overall view of the subject, automatically identifying keywords within all the students’ responses, as well as the associated sentiment with that keyword. The first subject selected was a first-year engineering subject that had free text comments given by 72 students. These comments can range from positive to negative, providing suggestions and allowing students to express how they felt about the subject. Figure 1 below shows a whole subject level analysis of these responses.

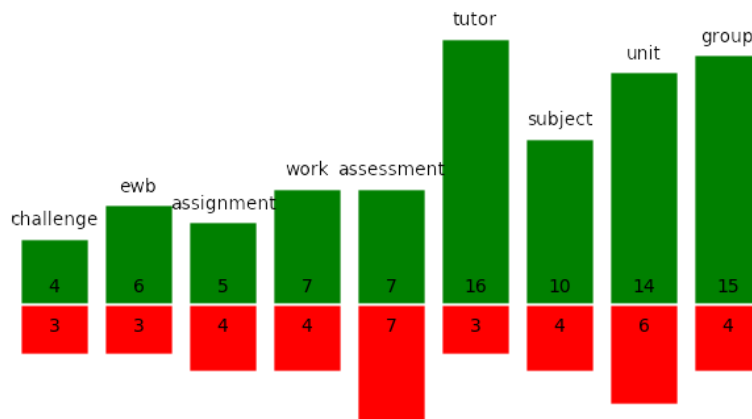


Figure 1 – Example 1: Single Subject Analysis

In this graph, nine keywords have been extracted, which are nine “topics”. These topics summarise the text corpus. Let’s focus in on the keyword “tutor” as an example. From this visual representation, we can see that 16 people mentioned tutor in a positive sense, and 3 negatively. This makes a total of 19 responses using the keyword tutor in either a positive or negative sentiment. We can conclude from this representation, that most students are content with tutorials. If desired, we could examine the negative responses for further detail.

Looking again at the graph above, we can see that the keyword “assessment” has equal positive and negative sentiments. This makes it an important keyword that we might want to focus more on. Figure 2 below shows a focussed analysis of responses which contained “assessment”.



Figure 2 - Example 1: Focused Keyword Analysis

From this graph we can see keywords (topics) that students are mentioning in the same responses that mention assessment. It is important to note that in this situation, most students use the words unit and subject interchangeable. Looking at the keyword “group” first as an example. This suggests that most people are happy with group components of the subject, and possibly the group components of “assessment”. A part of one of the student’s comments which also reflects this is shown below,

“... manageable load of work but would not have wanted anymore assessment. Also happy with the way the load is split so that if there is trouble with groups you can still have the chance to pass if you put the work in...”

Figure 2 also shows that some students negative feeling associated with the word “time”. One response which reflects this is below,

“Mixed response from assessments. Good insight into today's problems but waste of time”

Example 2

Let's consider another example of a different subject. This is a second-year engineering subject. Figure 3 below again shows another 9 automatically obtained keywords for this subject. We invite the reader here to stop and think about the possible conclusions that could be drawn from this information. Figure 4 is included also, giving a focussed keyword analysis for the word “lecture”, for further consideration.

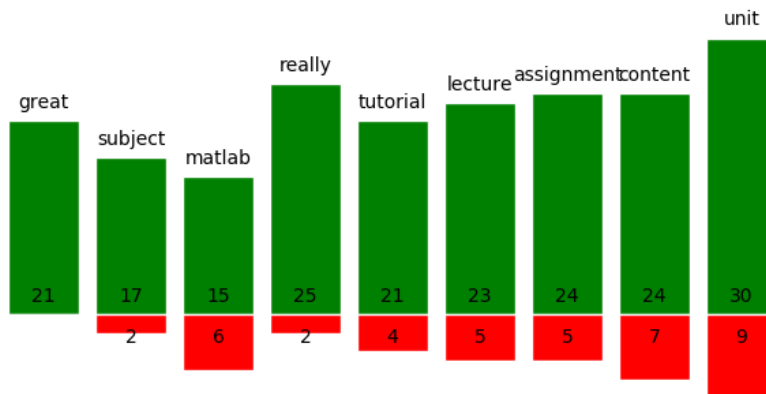


Figure 3 - Example 2: Single Subject Analysis

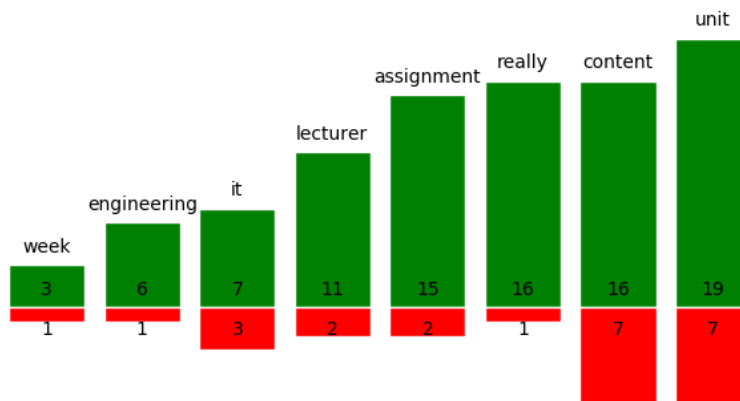


Figure 4 - Example 2: Focused Keyword Analysis

Examining Figure 3 above, we can see that students have a positive outlook on most aspects of the subject. Keywords such as “matlab” were found to be key terms in responses for the subject, and the programming language MATLAB was taught in the subject. Analysing Figure 4 further shows some keywords commonly used in responses with lecture. For example, “lecturer” is mentioned mostly in a positive light. This is reflected in the comment below,

“It's great to have such a passionate and inspiring lecturer...”

Content might be an area which we could examine and investigate more closely. Again, selecting a comment, we can see this reflected below,

“Main Lecture: Sometimes hard to understand what he is saying, would prefer if all the content was contained in the slides...”

These visualisation techniques are able to inform the decisions which we make about teaching strategies or which comments we may want to read more closely, without reading the entire set of comments. This is quite powerful, especially when dealing with a large number of comments.

Comparison with other Visualisation Techniques

Other text visualisation software is available; and we have selected two pieces of software to demonstrate some capabilities and limitations with current visualisation tools. Leximancer and a Word Cloud generator are discussed and shown below. The capabilities of each piece of software are presented, and compared against the solution presented in this study. The same dataset used in Example 1 is used for a consistent comparison.

Leximancer

Leximancer (Smith & Humphreys, 2006) can analyse a given text input. Leximancer has many options. It allows the user to pick a number of topics, and also allows the user to combine terms and topics. Figure 5 shows an example of an output which Leximancer can produce from the “example 1” data. The figure shows relationships between words, for example “group” and “work” are linked closely, as they were often mentioned together in the context of group work.

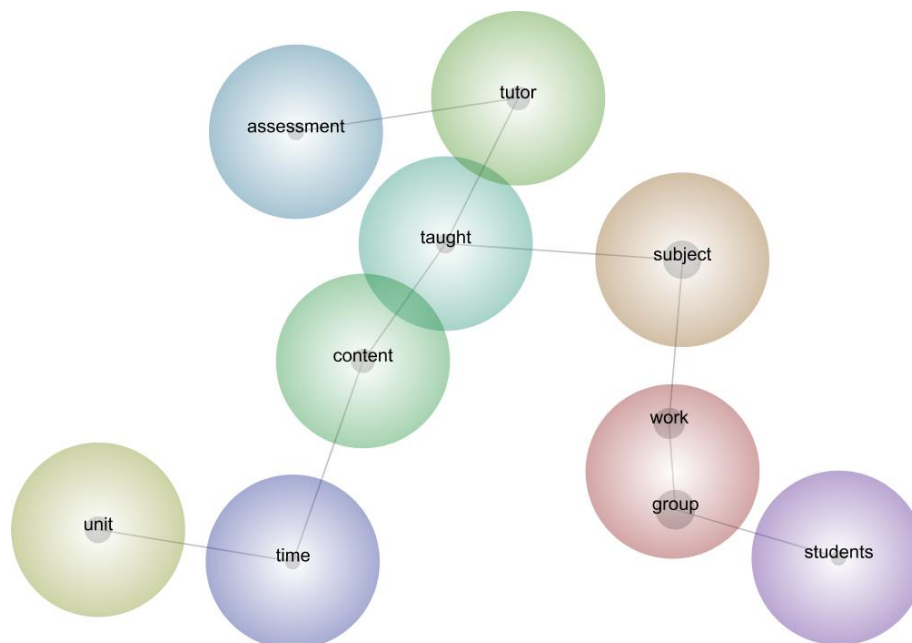


Figure 5 - Leximancer Diagram for Example 1

Leximancer however has several limitations when it comes to visualising student satisfaction. Firstly, a knowledge of the software is required to generate this type of plot, as well as a software license. Some manual labelling and grouping is required to generate these plots. The plot is also lacking in information about the frequency of occurrences, as well as sentiment information for student satisfaction. You can find frequency information, but it is not represented on the same plot.

Word Cloud

Word Clouds are another commonly used tool for visualising text. They are available to use for free and require very little knowledge of the platform to use. The cloud conveys the key words and frequencies, the more commonly occurring words being represented larger. This allows important words to be brought to the reader’s attention easily.

The word cloud is lacking in several of the same aspects of the Leximancer representation for this application. Sentiment is not conveyed, and some important aspects can sometimes become hidden amongst the other words.



Figure 6 - Word Cloud for Example 1

Conclusions and Recommendations

In this paper, we present a consistent way to automatically analyse and produce a visual representation of student satisfaction for a subject. This can be useful when a reader is dealing with a large number of comments, or would like to compare two subjects together visually. In these visualisations, keywords are extracted automatically and grouped into topics which are most relevant to students, as conveyed in their responses. The sentiment is then found for each of these selected keywords, meaning that the reader can examine the proportion of comments which mention the selected keyword in a positive or negative light. The responses can then be examined even further, selecting a keyword to investigate further, and finding which other keywords are mentioned often in conjunction with the selected keyword. This allows for patterns and enable possible improvements in selecting and implementing teaching strategies to be extracted, helping both educators and students.

It is important to emphasise that the entire process for creating the figures is automated, apart from the user selecting a keyword which they want for the second stage of further analysis. This is quite powerful, as it allows scalability for any number of comments, and larger comments sets may provide more insights. Future work includes looking at additional ways in which sentiment can be found, using a scale as opposed to a singular positive or negative result. Being able to improve how the keywords or entities are detected, as well as investigating possible links between these words would help to strengthen this work.

Being able to visualise student satisfaction from comments and feedback is a powerful tool. The produced visuals can be used by lecturers, subject coordinators or managers to compare and review several subjects at once. As a lecturer, you would not need to know how this system works, but have access to student feedback. The lecturer also has the option of changing more detailed parameters based on their needs. With the student feedback, these plots can be automatically generated. The visualisations can lead to quick detection of both positive and negative aspects within a subject, thus prompting appropriate action.

References

- Blei, D. M., Edu, B., Ng, A. Y., Edu, A., Jordan, M. I., & Edu, J. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 993-1022.
- Cunningham-Nelson, S., Baktashmotlagh, M., & Boles, W. (2016). Linking numerical scores with sentiment analysis of students' teaching and subject evaluation surveys: Pointers to teaching enhancements. 27th Annual Conference of the Australasian Association for Engineering Education: AAEE 2016, (p. 187). Coffs Harbour.

- Elliott, K. M. (2002). Student satisfaction: An alternative approach to assessing this important concept. *Journal of Higher Education Policy and Management*, 197-209.
- Nielsen, F. Å. (2011, March). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903.
- Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *LREc. QUT*. (2015). *Protocols: QUT's Evaluation Framework*. Brisbane.
- Shin, K. M. (2002). Student Satisfaction: An alternative approach to assessing this important concept. *Journal of Higher Education Policy and Management*, 97-209.
- Smith, A. E., & Humphreys, M. S. (2006). Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behavior research methods*, 38(2), 262-279.

Acknowledgements

Data for this study was obtained from the QUT Pulse and Insight evaluation surveys. The authors would like to thank QUT students and staff for this information.