

Visually Exploring Sentiment and Keywords for Analysing Student Satisfaction Data

Samuel Cunningham-Nelson, Mahsa Baktashmotlagh, Wageeh Boles.
Queensland University of Technology
Corresponding Author Email: samuel.cunninghamnelsonn@qut.edu.au

STRUCTURED ABSTRACT

CONTEXT

Measuring student satisfaction is an important task, which can lead to improvements in teaching strategies and techniques. There are several ways that satisfaction can be measured. One of the most common ways, is for students to rate or score a subject on a Likert scale. This can give educators an initial grasp on subject satisfaction, however this doesn't provide insight into reasons why the score was given. Free text comments are often given by students along with the score but summarising or finding meaningful information from these comments can be quite time consuming, especially for large classes. These comments can be used for educators to improve their teaching practice.

PURPOSE

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

1. How can we visualise keywords, sentiment, and relationships between these keywords?
2. What benefit would this have for students and educators?

APPROACH

Following on from previous work completed, we used several text analysis techniques to process initially clean and process text data. The dataset used for this analysis is student satisfaction survey data obtained from systematic university evaluations of units. These consist of both a satisfaction score, and a free text response, allowing students to give detailed feedback. We identified keywords, and the sentiment of these keywords, as well as position of these keywords to relate certain ideas together. This process is completed automatically, so a user or lecturer can use and benefit from the visualisations without understanding the technical aspects of this text analysis process.

RESULTS

A high-level graph is generated, for each unit, which includes keywords, sentiment of the keywords, and demonstrates connections between the keywords. A large amount of comments were used, and analysed using the various methods presented. Results analysing units revealed that keywords such as "assessment" and "tutorials" were prominently featured in the generated visualisations.

CONCLUSIONS

Using these generated visualisations, large amounts of comments are compared with various subjects. This comparison is valuable in identifying and linking key factors related to teaching and learning approaches used, as well as those relating to the teaching environment. These visualisations are shown in the context of engineering specific learning activities such as programming or electronic circuit analysis, and also compare the teaching activity terms between various engineering and non-engineering units. Pointers to certain approaches to teaching can have an impact for learning and allow the educator to ensure there is constructive alignment between learning activities, assessment and unit outcomes. This will improve students' overall learning experience and outcomes.

KEYWORDS

Student Satisfaction, Machine Learning, Conceptual Understanding

Introduction

Overview of Student Feedback

As part of the systematic enhancement of the teaching environment and providing students with a positive learning experience, universities run regular student surveys. These surveys measure student satisfaction of teaching and units of study.

There are several ways that satisfaction can be measured. One of the most common ways, is for students to rate or score a unit on a Likhert scale. However, relying on the score alone doesn't provide sufficient insight into the reasons why the score was given. Therefore, the free text comments given by students along with the score are very valuable in providing such an insight.

Summarising or finding meaningful information from these comments can be quite time consuming, especially for large classes. Also, factors affecting student satisfaction in different units or different teaching periods are difficult to compare, beyond the scores. Further, people also tend to put a much higher weight to negative comments, resulting in a negativity bias (Rozin & Royzman, 2001). This can lead to educators taking actions in response to the negative comments, without due consideration to students' suggestions hidden within the positive comments.

Why visualize student comments

The importance and value of finding a systematic and easy method for summarising and visualising students' comments can be demonstrated by referring to the known "start-start-continue" strategy (Hoon, Oliver, Szpakowska, & Newton, 2015). One implementation of this involves distributing sticky notes of three colours to students in the class. They are directed to put aspects of teaching or the unit that they want the teacher to: keep doing (continue), stop doing (stop) and suggestions to begin doing (start). Therefore, if student comments are summarised in a few broad clusters, actions can be taken, without the load of additional surveys and polling. We present several different methods which educators can use to visualise student feedback data, as the starting point for taking enhancement actions.

Research Questions

Visualising student satisfaction data in various ways can provide invaluable information. In this paper the focus has been narrowed to answer two research questions,

1. How can we visualise keywords, sentiment, and relationships between these keywords?
2. What benefit would this have for students and educators?

APPROACH

Following on from our previous work, we used several text analysis techniques to initially clean and then process text data. The dataset used for this analysis is student satisfaction survey data obtained from systematic university evaluations of units. We identify keywords, and the sentiment of these keywords, as well as position of these keywords within sentences to relate certain ideas together. This process is completed automatically, so a user or lecturer does not need to know about the technical aspects of this text analysis process, yet they would be able to use it and benefit from it.

Method

Student opinion surveys are conducted at most universities – often given to students towards the end of the teaching period. For this study, the data we previously introduced is used (Cunningham-Nelson, Baktashmotlagh, & Boles, 2016). In this data, a satisfaction rating on a

Likert scale from one to five, as well as a free text comment is given by a student. An example of this is, Satisfaction rating: 5. *“Great structured unit. Very well organised and great learning environments”*

The satisfaction score is often the commonly used metric, as it is easy to perform statistical analysis on, and identify trends or problems. However, the majority of the useful information in these student satisfaction surveys comes from the free-text responses that students give. These often contain constructive criticism, and advice for educators to take on board, and improve their practice.

Visual Tables

Although the student satisfaction score provides some overall information about the unit, it does not provide specific actionable suggestions. Information about particular activities would prove more useful. For example, we could also ask students: “How satisfied are you with lectures?” This would give more actionable information for an educator, but it may discourage students from participating in the survey, especially if many of these questions were asked. Closed type responses also limit students’ ability to express comments or opinions which may vary from normal or expected ones.

A basic table presentation which an educator may be given was extended upon, by initially selecting several keywords that an educator may be interested in. We then took advantage of an Entity-Sentiment detection service provided by Google. This can determine the keyword or entities in one review and associate a sentiment score with each of these. We can then display that graphically, with some colour information to reinforce its use as a “stop-start-continue” example – creating a visual table.

Visual Connections

In addition to identifying the sentiment of selected keywords, being able to visualise connections between keywords can provide further insight. Students may leave a comment such as *“the lecture content is disconnected from the tutorial”*. Being able to represent this relationship between the lecture and the tutorial, especially for a commonly occurring theme, would provide further information for educators to act on, as well as see which keywords are most closely linked.

To generate this visualisation, the keywords to be represented need to be decided first. These could be selected manually, depending on the educators’ interest. However, to automate the process, we have systematically determined the keywords in this work. Using a statistical analysis method known as Latent Dirichlet Allocation (LDA) these keywords can be selected. LDA allows a set of text documents to be summarised into several topics, and each topic represented by a set number of words (Blei, et al., 2003).

After the keywords representing the data are found by LDA, a co-occurrence matrix is generated. The co-occurrence matrix contains the keywords along the rows, and columns. When two of the keywords are mentioned within a single comment, one is added to the value in the corresponding cell. This matrix will have large numbers when keywords are often mentioned together, and smaller numbers for infrequent relationships.

The co-occurrence matrix is then visualised using Gephi, which is an open source software for graph and network analysis (Bastian, Heymann, & Jacomy, 2009). Examples of where Gephi is used include capturing the relationships within recipes (Teng, Lin, & Adamic, 2012; Jacomy, Venturini, Heymann, & Bastian, 2014) and analysing conversations on twitter (Bruns, 2012).

Visual Histograms

With a large quantity of text responses, or text responses across multiple units, deciding on teaching actions or directions moving forward can be a difficult task. Being able to easily

visualise all the key elements of a unit, and whether students rate them with a positive or negative sentiment overall can provide valuable information for an educator. We developed this visualisation method previously but will discuss here briefly. Further details on this method can be found in (Cunningham-Nelson, Baktashmotlagh, & Boles, Visualising Student Satisfaction, 2017).

Similarly, to the previous visual connections method, the first step is to automatically generate the keywords to be displayed in the visualisation, using LDA. These keywords represent several main ideas that the students mention throughout their feedback.

In each comment, the sentiment of each keyword needs to be determined. One way to do this is to use the dataset, "AFINN" (Nielsen, 2011). AFINN contains a large table of words, and an integer rating between +5 and -5. For each sentence, the sentiment can then be calculated by adding the rating for each individual word across the sentence. The sentence is then decided as positive or negative depending on the final overall value. The sentiment of each keyword, positive or negative, can then be grouped together and visualised.

Results

We examined and compared the results for the three different visualisation methods introduced previously in this paper.

Visual Tables

Student evaluation feedback is often presented in table format, with comments present next to scores that students give. For the first visualisation method, we present additional information to the data that is just given by students. Sentiment scores are used for keywords of interest, and ranked on a scale of -1, to +1. This scale is shown in Table 1 below, with colour to highlight this information.

Table 1 – Sentiment Key

	Completely Negative	Neutral	Completely Positive
Sentiment	-1.00	0.00	+1.00

Table 2 shows the visual tables visualisation method, with subsections of several chosen comments. The first column in the table shows these comments selected. The second comment represents the overall satisfaction score students gave the unit. This score ranges from 1 to 5, 5 being very satisfied. The overall satisfaction can be useful as an initial metric, but it fails to provide further information about why the student was satisfied or dissatisfied. For this unit, we have chosen five keywords based on key teaching activities or materials which we wanted to investigate further: lecture, tutorial, MATLAB, video and assignment. The following columns represent the calculated sentiment of each entity, using the sentiment scale mentioned previously.

This visualisation method displayed a much richer form of information, allowing a reader to see information at a quick glance, that could only be obtained by reading some comments carefully. For example, from the first comment we can see the keyword *tutorial* is mentioned, in a neutral sentiment, and the keyword *video* in a positive light. This is verified by reading the comment, as we can see a strong positive association with videos, and the word tutorial is used neutrally, with a suggestion.

By quantifying the sentiment for each comment and each keyword, we performed comparison operations such as finding the average for each of the keywords (note: the averages are calculated from the entire table, not only the presented rows). This allows comparisons to be done between different semesters of the same unit, or different units. It also means that educators can make informed decisions based on presented evidence. For example, in Table 2 we can clearly see the overall satisfaction of *assignments* is negative, prompting us to further investigate why this could be the case.

Table 2 - Selected Survey Entries

Comments	Overall Sat	Lecture Sat	Tutorial Sat	MATLAB Sat	Video Sat	Assignment Sat
AVERAGE	3.67	0.19	0.14	-0.06	0.49	-0.15
<i>...The pre-tutorial videos are amazing but it would be beneficial to have the tutorials recorded as well.</i>	5		0.00		0.90	
<i>Great teaching team, but sometimes lectures and tutorials were a bit slow-paced.</i>	5	-0.20	-0.80			
<i>...my only negative is that the assignment's wording can be a bit complicated and hard to comprehend.</i>	4					-0.20
<i>... I do like the themed assignments (like the BASA space mission).</i>	4					0.20
<i>Good matching of tutorials and both lectures however the tutorial is often a large step up from the lecture content...</i>	4	0.70	0.40			

Visual Connections

Connections in student satisfaction comments can provide valuable information for educators to use, beyond comments on their own. These connections may represent links between keywords which are having a positive, or negative impact on the satisfaction of students. More prominent connections are represented by darked lines in the diagrams.

Figure 1 and Figure 2 show connection diagrams, with comments separated based on the overall satisfaction score given by students. Figure 1 was created using only comments from students who gave the unit an overall negative satisfaction score (1 or 2 out of 5). Looking at the diagram presented we can see the two strongest links are between; assessment ↔ lecture and week ↔ lecture. One conclusion we may be able to draw from this, is that some students who dislike the lectures also dislike the assessment.

Figure 2 was created using only comments from students who gave the unit an overall neutral or positive satisfaction score (3, 4 or 5 out of 5). Several strong links can be seen between; unit ↔ tutorial; assignment ↔ unit; assignment ↔ lecture; assignment ↔ tutorial; lecture ↔ subject and tutorial ↔ subject.

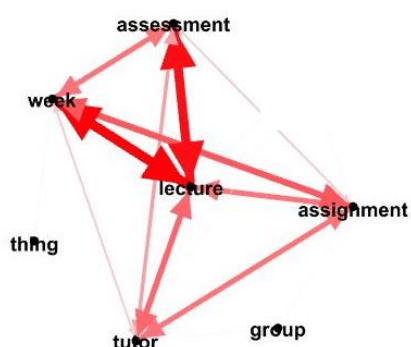


Figure 1 – Negative Connections

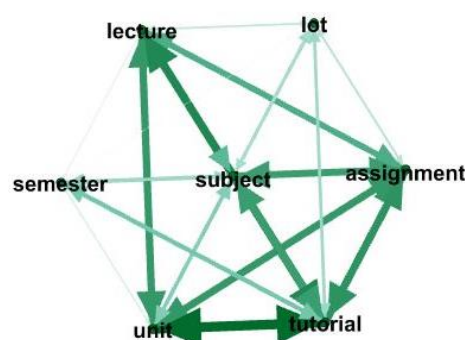


Figure 2 - Positive Connections

The connection for the positive diagram are much more abundant. This is most likely due to the larger number of positive comments present. One conclusion we could draw from the strong connection, is that the tutorials have a large overall impact on the unit, when students rate it positively. We can also see a connection between the assignments, and the unit,

lectures and tutorials as well. This points positively for constructive alignment between learning activities and assessment.

Visual Histograms

Being able to have a high-level view about the number of students can provide the ability to quickly compare sets of student comments. Nine keywords are presented in Figure 3, including *lecture*, *tutorial* and *assignment*. These represent the main ideas students mention in their comments.

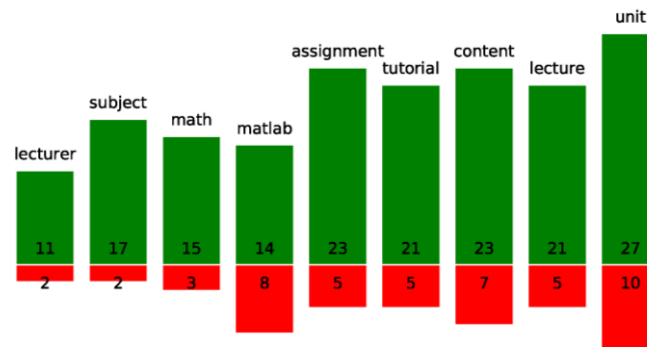


Figure 3 - Visual Histogram

For example, from Figure 3 the keyword *assignment* is mentioned 28 times in total. Out of these 28, 23 mentions are in a positive context, and 5 mentions in a negative context. This separation allows a reader to easily identify aspects that students mention more frequently, as well as the sentiment of each of these aspects. This can allow for an easy comparison between semesters, or between units.

Comparison and Discussion

The visual tables method allows a user to extend beyond satisfaction scores and comments, to focus on keywords of interest. These keywords are quantified on a continuous scale for each comment and allows a reader to identify comments of interest easily amongst others, based on keyword sentiment.

The visual connection diagrams allow readers to identify possible links between keywords mentioned by students. By displaying a diagram of connections where students were satisfied overall, and another diagram where students were dissatisfied overall, positive and negative connections can be made between keywords. This allows an educator to identify possible causes for problems, or reasons for satisfaction.

The visual histogram allows for a higher level of comparison for units. This visualisation is presented, counting the sentiment for particular keywords in a set of comments. It is easy to visualise which keywords are mentioned more frequently by students, and which have the largest number of positive or negative comments. The histogram also provides an easy comparison between units, or between teaching periods.

All three visualisation methods presented use colour to represent positive and negative sentiment. This directly links with the stop-start-continue method discussed previously, and allows educators to visually identify actionable aspects of the unit being considered.

Conclusions and Recommendations

Student unit reviews are common practice at most universities, however they often lack detailed methods for analysis and comparison. In this paper, we presented three methods for extending and visualising student comments obtained from evaluation surveys. The methods

presented can all be automated, and generated for many units, providing a large benefit, even for users who are not familiar with data analytics methods.

Summarising and visualising student satisfaction survey comments is a valuable exercise for educators, allowing them to compare and easily identify actionable information from a potentially large number of comments. Each of the three visualisations presented show the information in a different form and can lead to different (insights) from the comments analysed.

Each visualisation method presented is colour-coded, and links back to the over-arching premise, of applying the stop-start-continue principles in the classroom. This allows educators to quickly decide on an action to take.

Improving on pre-processing and spelling correction techniques would help to increase the meaningful output of the results. In addition to this, more investigation into ways to cluster and group similar comments together would provide some additional insights, and possible additional visualisation or comparison metrics to be explored.

Three visualisation methods were presented, which can automatically be generated from student satisfaction scores and comments. The methods extend beyond the usual feedback provided to educators, allowing them to make informed and logical decisions based on metrics and graphical representations. These could be automated and presented to educators, as part of a report when they receive the student evaluation survey data.

References

- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an Open Source Software for Exploring and Manipulating Networks. *ICWSM*, 8, 361-362.
- 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.
- Bruns, A. (2012). How long is a tweet? Mapping dynamic conversation networks on Twitter using Gawk and Gephi. *Information, Communication & Society*, 15(9), 1323-1351.
- 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.
- Cunningham-Nelson, S., Baktashmotlagh, M., & Boles, W. (2017). Visualising Student Satisfaction. *28th Annual Conference of the Australasian Association for Engineering Education (AAEE 2017)*. Manly, Sydney, Australia: Australasian Association for Engineering Education.
- Hoon, A., Oliver, E., Szpakowska, K., & Newton, P. (2015). Use of the 'Stop, Start, Continue' method is associated with the production of constructive qualitative feedback by students in higher education. *Assessment and Evaluation in Higher Education*, 40(5), 755-767.
- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PloS one*, 9(6).
- Nielsen, F. Å. (2011, March). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.
- Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review*, 5(4), 296-320.
- Teng, C.-Y., Lin, Y.-R., & Adamic, L. A. (2012). Recipe Recommendation Using Ingredient Networks. *Proceedings of the 4th Annual ACM Web Science Conference*, (pp. 298-307).

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.