

A Word-Space Visualization Approach to Study College of Engineering Mission Statements

Sreyoshi Bhaduri

*Department of Engineering Education
Virginia Tech
Blacksburg, USA
sreyoshi@vt.edu*

Tamoghna Roy

*Bradley Department of Electrical and Computer Engineering
Virginia Tech
Blacksburg, USA
tamoghna@vt.edu*

Abstract— Most higher education institutions have a mission statement that is developed strategically by the institutions and often reflect the college's unique mission which sets it apart from peer institutions. Through this study, we describe the use of a Machine Learning and Natural Language Processing based textual data analytics to understand the word choices in the mission statements of U.S. based colleges of engineering. Our purpose is to understand the key similarities and differences between the choice of words used in the mission statements of the two groups: public colleges and private colleges of engineering. We were specifically interested in studying the terms related to diversity and inclusion and see the trends in the use of specific terms relating to diverse communities, intersections, minority populations, and the like. In this research study, we used a Word2Vec approach to visualize the words from mission statements for 59 colleges of engineering in the United States. The contribution of this research is in the form of a visualization mapping the vector space model for word usage and complete vocabulary of pertinent words from the statements analyzed. The preliminary results of this study will help inform current state of vocabulary used in mission statements in the colleges of engineering across the United States. Ultimately, such analyses can help administrators in the development of strategies on the formation of mission and vision statements for universities by allowing insight into vocabulary currently used, to understand what words/terms may not be adequately addressed. Additionally, we are excited to present a contemporary textual data analytical technique of Natural Language Processing using a word vector representation tool such as word2vec for analyzing textual data in the field of engineering education.

Keywords— word2vec; text analytics; mission statements

I. INTRODUCTION

The field of engineering education is rapidly growing, and with this increase, there is a surge in the amount of data being generated in the field. Analyzing such data, such as year-end reports, responses to open-ended survey questions, student blog posts or journal entries, and similar education data artefacts can yield interesting insight into student behavior, trends useful for learning analytics, and the like. However, analyzing these texts are time and resource intensive to inductively analyze large bodies data for individual instructors or researchers in educational contexts [1]. Natural Language Processing (NLP) techniques coupled with Machine Learning which employ algorithmic data analytical capabilities of machines to sift information from large text based data sets may be useful for

individual researchers in the field who are interested in gaining analytical overviews of such textual datasets in time and resource efficient ways.

In this paper we present a contemporary textual data analytical technique which incorporates Natural Language Processing and Machine Learning to a data set comprising mission statements from colleges of engineering in the United States. Prior work on mission statements [e.g., 2, 3] have been found to be of interest to the community, and this specific research expands upon part of an ongoing larger research [4] which approaches analysis of mission statements from colleges of engineering through innovative quantitative and qualitative approaches.

II. BACKGROUND

Analyzing mission statements may provide insight into what university administrations value, and may also help provide insight into the learning environment created at the colleges. In this research paper, we will describe preliminary results of our analysis using Natural Language Processing and Machine Learning techniques to visualize the words from a dataset comprising mission statements from 59 colleges of engineering in the United States: 29 of these colleges are Public and 30 of them are Private colleges.

A. What can Mission Statements tell us about a college?

Mission statements can be described as public documents [2]. These statements are usually readily accessible through the internet [3], and are often regarded as unique since they are intended to convey university-specific information reflecting the university's overall vision and purpose [5]. Mission statements are often developed through strategic planning in institutions. There has been considerable interest in the mission statements of universities and colleges. Multiple research studies have acknowledged mission statements to be important in describing institutions intent and goals [6, 7], and have argued that institutions need to be more strategic in developing statements which truly reflect their characteristics [8]. This call for mission statements to be developed more strategically addresses certain criticism of these statements by authors such as Kreber and Mhina [2] who cite Detomasi [9] to describe how the latter suggest that mission statements are “embarrassingly vague, and largely comprised of academic peties, dull platitudes, and odes of self-congratulation” (p.31). Newsom and Hayes [10] echo this statement in their work analyzing

mission statements for 114 US-based schools and colleges, after which they concluded the statements to be vague and lacking clear purpose.

Despite the contradicting views in existing literature on whether or not mission statements are truly reflective of the institution's intent and goals, there have been ongoing research in understanding existing themes important to various institutions through analyzing mission statements both nationally and internationally [2, 3, 4]. This renewed focus on mission statements and their importance may perhaps be attributed in part to the accreditation systems in place which demand that institutions have unique mission statements which highlight their specific purpose and institutional goals. Morphew and Hartley describe mission statements as being ubiquitous to institutions and reason "accreditation agencies demand them, strategic planning is predicted on their formulation, and virtually every college and university has one available for review" [11]. Creamer and Ghoston [3] acknowledge views by Delucchi [12] and Morphew and Hartley [11] both of who maintain that there is little research on mission statements in higher education, and add insight based on their search for such research by stating "a more accurate statement is that research literature is dispersed across a variety of disciplines...in fields as diverse as consumer studies, business, Christian education, and educational policy and management". However, Creamer and Ghoston state that their search did not reveal any research on mission statements specific to engineering or STEM contexts [3]. Thus, despite the wide interest and prior research into mission statements of universities in general, mission statements for STEM colleges particularly engineering remain limited. This paper presents preliminary results from our study which extends a larger research project (introduced in [4]) aimed at addressing this gap by trying to analyze and compare across mission statements for public versus private colleges of engineering in the United States.

B. Text Analytics for Engineering Education Data

As previously stated, our objective is to analyze the mission statements from colleges of engineering across the United States, to understand the word usage and specific vocabulary used by administrators to describe the learning experiences or mission of such institutions. This analysis may be particularly of use to prospective faculty or students since it allows them a perspective into the college's values and mission, and helps them understand if it aligns with their own. It may also help in understanding the differences in the current state of public and private colleges of engineering in the United States, based on the vocabulary differences in terms of the words included in the statements. However, it is often difficult for individuals to analyze this information due to the large volume of the text-based data which makes it time and resource-intensive for individual instructors to manually analyze, and compare across.

Text analytics through use of Natural Language Processing may be useful in such situations for automated textual analysis which can be successful in leveraging meaningful data from large volumes of text, and thus help inform research, practice and pedagogy. The term Natural Language Processing can be described as any kind of computer-based manipulation of a

natural language [13]. Natural languages such as English, German or Hindi can be distinguished from can be connoted "artificial languages" such as those used in programming. Bird, Klein & Loper [13] observe that the former have evolved over time, passing from generation to generation with rules that may be hard to pin down explicitly (pg.1). Natural Language Processing can thus be understood as an interdisciplinary field in which computers are used to perform useful tasks involving human language [14].

Natural Language Processing techniques have been used in conjunction with machine learning tools for varied content analysis related tasks. One of the earlier definitions of machine learning was provided in the IBM Journal of Research and Development by Samuel [15] who in describing machine learning using a game of checkers, stated that these types of studies are concerned with the "programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning." (pg.71).

Specific to text-based content analysis, supervised machine learning techniques of classification have been used alongside Natural Language Processing techniques. For example, researchers have employed Natural Language Processing and Machine Learning tools for a wide range of tasks such as classification of suicide notes into genuine or illicit [16], and for analyzing over 300 State of the Union addresses by Presidents of the United States of America to determine a timeline of trends in topics of national interest [17]. Natural Language Processing and Machine Learning have been employed in the classrooms for a wide range of tasks. From automating the grading process to analyzing student learning in the classroom, researchers have successfully demonstrated how Natural Language Processing techniques coupled with machine learning can be useful to gain a deeper understanding from data sets. Examples of such studies include use of machine learning and Natural Language Processing bases systems for the purpose of assessment [18] or to understand student vocabulary [1, 19]. Using Natural Language Processing tools with supervised machine learning for a predictive analysis, for example, Robinson, Yeomans, Reich, Hulleman, and Gehlbach [20] presented a model which could predict course completion in MOOCs based on word choices of students enrolled in the course. In a study conducted for an Engineering Education classroom, Variawa et al. [1] demonstrated how an automated method could be used to develop course-specific vocabulary. In this study we employ the use of word2vec, to visually represent the words from mission statements of public and private engineering colleges onto a two-dimensional space.

III. DATA AND METHODS

A. Description of Data

The researchers identified a total of 60 engineering colleges through random sampling from data publicly available by the American Society for Engineering Education in a book called the *Engineering College Profiles and Statistics Book*. Thirty of

the colleges were public while the rest were private colleges of engineering. Following identification of the colleges, the researchers created an Excel database for the colleges wherein they assigned a column for mission statements. The mission statements were extracted specifically from the engineering school or engineering college websites. In one of the 30 schools identified for Public engineering college, no mission statement was found after a thorough search on the school's website. That college was excluded from the analysis, and a total of 59 engineering college/schools were analyzed [4].

TABLE 1 DESCRIBING THE PROPORTIONS OF THE PUBLIC AND PRIVATE UNIVERSITIES ANALYZED

Control	Number (Proportion)
Public	29 (49.126%)
Private	30 (50.814%)

As can be seen from Figure 1 below, the colleges of engineering that were chosen randomly, seem to be spread geographically across the United States, but have a higher density in the East Coast. Thus, because the colleges of engineering were chosen randomly, a limitation of this preliminary research is that there was inadequate representation from all states/geographical regions. Another limitation of this research is that besides the control (i.e., Public versus Private) other characteristics of the colleges were not considered.

FIGURE 1: DEPICTING THE COLLEGES OF ENGINEERING CHOSEN FOR THIS STUDY (GREEN: PRIVATE; RED: PUBLIC)



B. Data Pre-Processing

Prior to analysis some pre-processing of the data was performed. We loaded all the mission statements onto an Excel File. We then used Python 2.7 with NLTK tool-kit to pre-process the dataset. First all the punctuation marks were removed from the document and then all the words are converted to the lower case.

C. Visualization using Word2Vec

We used word2vec embedding to visualize the words used by the colleges of engineering on a 2-dimensional plane. The word2vec was developed by data scientists at Google [21] in order to provide a state of the art embedding for words, by

training words on Google News corpus. The underlying hypothesis of this representation is that words in similar contexts have similar meanings. Mikolov, et. al., [21] describe how the similarity of word representations go beyond simple syntactic regularities. They provide as an example: “vector (“ King”) – vector (“ Man”) + vector (“ Woman”) results in a vector that is closest to the vector representation of the word Queen”

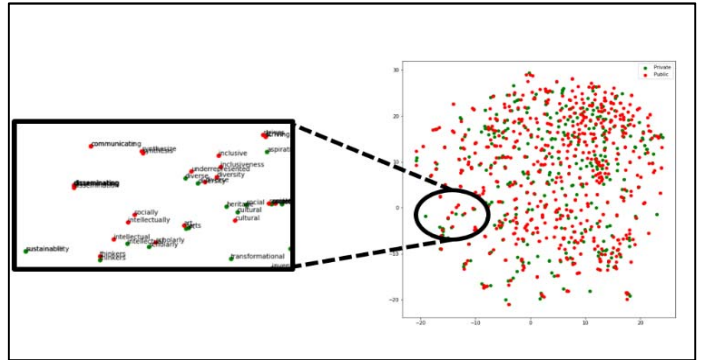
Since the introduction of this seminal embedding scheme in 2013, it has been used in various tasks for visualization purposes [22]. In this research we used Word2Vec embedding to visualize the vocabulary generated from the mission statements in an order to compare across the public and private universities.

IV. PRELIMINARY RESULTS

The purpose of this study was two-fold: to visually map the word choices among private and public colleges of engineering in the United States; and to understand the extent of focus on diversity for both the institution types.

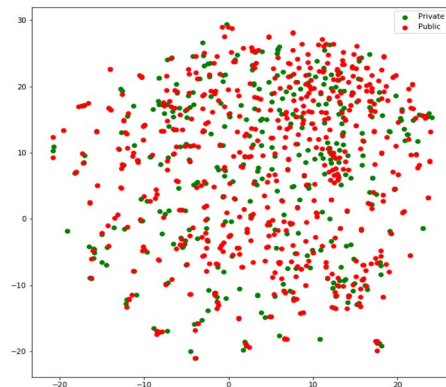
A. Mapping Word Usage for Public and Private Colleges

FIGURE 2: MAGNIFIED PORTION OF THE REPRESENTATION (GREEN: PRIVATE; RED: PUBLIC)



Using the word2vec embedding we found that there was an overlap in the word choices used by public and private colleges. In Figure 2 we present a closer examination of a section of the dots to magnify the words that the dots represent in order to provide a clearer idea of the visualization. The Figure 3 below provides an overview of the representation distribution.

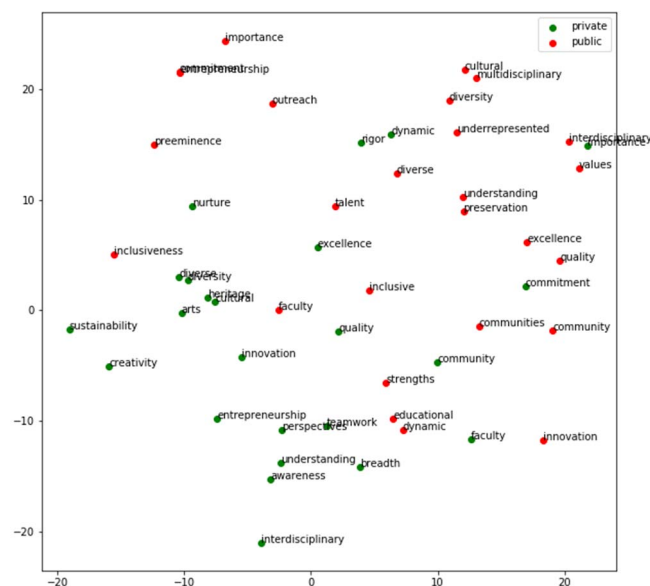
FIGURE 3: WORD2VEC VISUALIZATIONS OF THE VOCABULARY FOR MISSION STATEMENTS FROM COLLEGE OF ENGINEERING IN THE UNITED STATES. (GREEN: PRIVATE; RED: PUBLIC)



B. Extent of focus on Diversity and Inclusion

The second part of our analysis was focused on the words represented in the vocabulary extracted from mission statements of colleges of engineering which highlighted diversity and inclusion related terminology. Figure 4 plots the word vectors which are similar to the word 'diversity'. Cosine distance is taken as the similarity metric. Cosine distance is taken as the metric which is typically used to measure the similarity between word vectors [23, 24]. Words are considered "similar" to the reference word "diversity" if the cosine distance

FIGURE 4: WORD VECTORS WHICH ARE SIMILAR TO THE WORD 'DIVERSITY' BASED ON THE COSINE DISTANCE THRESHOLD OF 0.3 (GREEN: PRIVATE; RED: PUBLIC)



between them is greater than a certain threshold (0.3 in this case).

As can be seen from the visualization of the word-space, the public colleges of engineering address issues such as “inclusive”, “under-represented”, “diverse”, “cultural”. Similarly, in the private college of engineering, these constructs are emphasized through words such as “heritage”, “diverse”, “cultural”. Words related to inclusivity are however missing from the private college of engineering lexicon.

V. IMPLICATIONS AND FUTURE WORK

This research presents an exploratory textual analytical approach to analyzing mission statements of engineering colleges, using word embedding from word2vec. This research has strong implications for the future of engineering education research by successfully demonstrating how Natural Language Processing and Machine Learning can be used for summarizing data from large textual corpora. This strategy of employing such techniques for quick descriptive overviews of large textual datasets can be of interest to administration for gaining faster overviews related to their policies based on available textual data. This research can also be useful for engineering education

researchers who are interested in analyzing large corpora with constraints of limited time and resources.

Specific to our context of mission statements, as can be seen from preliminary results of our ongoing research, there exists minor (if any) differences in the corpora comprising words from mission statements of private colleges versus those from public colleges of engineering in the United States, for the colleges represented in our sample. We have described in an earlier paper [4] in detail a comparative analysis of the lexicon used by colleges in drafting mission statements. Through this work on visually representing the words in a two dimensional space, we find very little variation in the spread of word choices, and minimal outliers. This seems to indicate that administrators leading efforts to draft mission statements may be driven by similar goals. The focus on diversity and inclusion related constructs as depicted through word choice usage in the mission statements is also minimal for both the sets of institutions.

Future work in this area can help to gain a deeper understanding of comparisons across public and private engineering colleges by including more mission statements from colleges as part of the analysis. We hope to additionally consider individual institutional characteristics to understand differences which may arise due to differences in institutional goals. We also hope to use more graphic representation methods to delve further in our analysis and reveal clusters in the dataset.

REFERENCES

- [1] Variawa, C., McCahan, S., & Chignell, M. (2013). "An Automated Approach for Finding Course-specific Vocabulary." Paper presented at the 2013 ASEE Annual Conference & Exposition, Atlanta, Georgia.
- [2] Kreber, C., & Mhina, C. (2007). "The values we prize: A comparative analysis of mission statements." *Higher Education Perspectives*, 3(1).
- [3] Creamer, E. G., & Ghoston, M. (2013). "Using a mixed methods content analysis to analyze mission statements from colleges of engineering." *Journal of Mixed Methods Research*, 7(2), 110-120.
- [4] Bhaduri, S., & Roy, T. (2017). "Demonstrating Use of Natural Language Processing to Compare College of Engineering Mission Statements." Paper presented at the 2017 ASEE Annual Conference & Exposition, Columbus, Ohio.
- [5] Kibuuka, H. E. (2001). "Vision and mission statement in Christian higher educational management in Eastern Africa." *Journal of Research on Christian Education*, 10(1), 87-114.
- [6] Tierney, W. G. (1999). *Building the responsive campus: Creating high performance colleges and universities*: Sage.
- [7] Young, R. B. (2001). Colleges on the cross roads: A study of the mission statements of Catholic colleges and universities. *Current Issues in Catholic Higher Education*, 21(2), 65-81.
- [8] Barnett, R. (2003). *Beyond All Reason: Living with Ideology in the University*: ERIC.
- [9] Detomasi, D. (1995). Mission Statements: One More Time. *Planning for Higher Education*, 24(1), 31-35.
- [10] Newsom, W., & Hayes, C. (1991). "Are Mission Statements Worthwhile?" *Planning for Higher Education*, 19(2), 28-30.
- [11] Morpew, C. C., & Hartley, M. (2006). "Mission statements: A thematic analysis of rhetoric across institutional type." *The Journal of Higher Education*, 77(3), 456-471.
- [12] Delucchi, M. (1997). "Liberal arts" colleges and the myth of uniqueness. *The Journal of Higher Education*, 68(4), 414-426.
- [13] Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*: "O'Reilly Media, Inc."

- [14] Jurafsky, D., & Martin, J. H. (2007). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (Vol. 26).
- [15] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.
- [16] Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., & Leenaars, A. (2010). Suicide Note Classification Using Natural Language Processing: A Content Analysis. *Biomedical Informatics Insights*, 2010(3), 19-28. doi:10.4137/BII.S4706
- [17] Savoy, J. (2015). Text clustering: An application with the State of the Union addresses. *Journal of the Association for Information Science and Technology*, 66(8), 1645-1654.
- [18] Magliano, J. P., & Graesser, A. C. (2012). Computer-based assessment of student-constructed responses. *Behavior Research Methods*, 44(3), 608-621.
- [19] Variawa, C. (2014). *Investigating the Language of Engineering Education*. University of Toronto.
- [20] Robinson, C., Yeomans, M., Reich, J., Hulleman, C., & Gehlbach, H. (2016). *Forecasting student achievement in MOOCs with natural language processing*. Paper presented at the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge.
- [21] Mikolov, T., Yih, W. T., & Zweig, G. (2013, June). *Linguistic Regularities in Continuous Space Word Representations*. In *Hlt-naacl* (Vol. 13, pp. 746-751).
- [22] Luo, Jingyi, et al. "Predicting Student Grade Based on Free-Style Comments Using Word2Vec and ANN by Considering Prediction Results Obtained in Consecutive Lessons." *International Educational Data Mining Society* (2015).
- [23] Levy, Omer, Yoav Goldberg, and Israel Ramat-Gan. "Linguistic Regularities in Sparse and Explicit Word Representations." *CoNLL*. 2014.
- [24] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). "Distributed representations of words and phrases and their compositionality." In *Advances in neural information processing systems* (pp. 3111-3119).