Analyzing The Strength Between Mission And Vision Statements And Industry Via Machine Learning

Faleh Alshameri, Marymount University, USA
Nathan Green, Marymount University, USA

ABSTRACT

Mission and vision statements are critical to a company's success both from a company's long-term goals and appearance to potential customers. We analyze a collection of 772 mission and vision statements from companies via natural language processing. This data is hand annotated into 15 industry types. We show the distinctiveness and connectiveness of each industry via text processing and machine learning techniques. The extracted features of each industry are a telling and guiding indicator of what that industry embraces. We show high predictive power via machine learning to determine an industry by looking only at the mission and vision statements.

Keywords: CLUTO; Information Extraction; Machine Learning; Mission and Vision Statements; Organizational Performance; Strategic Management Process; Text Mining

1. INTRODUCTION

A critical part of the strategic management process requires the development of a vision and mission statement. According to Papulova (2014) there are four basic elements in a comprehensive model: environmental scanning, strategy formulation, strategy implementation and evaluation and control. The vision and mission statement remain part of the strategy formulation, along with clarification of the strategic objectives and development process. Effective managers should use these set strategic objectives to implement, monitor and evaluate the operational strategy. Customers, suppliers, and shareholders often base their decision to choose a business on the performance of the organization and the strength of their mission and vision. In fact, several studies show that high performance organizations do a lot better than lower performance organizations, including writing better mission and vision statements (Khalifa, 2011).

In this paper we derive useful information from 772 text documents from the dataset. We explore how to meaningfully categorize mission and vision statements into clusters. To achieve this goal, we examine the dataset in various ways by extracting key features from documents in the dataset to promote comprehension of the documents' contents. In order to predict an industry based on mission and vision statements, and to discover the relationship between industries, we empirically test 6 different machine learning classifiers.

2. TEXTUAL DATA MINING

Researchers have begun to develop new ways of identifying and extracting data clutter from the large volumes of data now available due to the rise in web usage. This makes navigating, summarizing, and organizing the information easier for users so that they can discover interesting relationships in the data. Large and semi-formatted datasets can be efficiently and effectively extracted through a variety of computational techniques. Clustering algorithms provide a relational picture, improving analysis by showing characteristics that provide meaning and determine the most important features (Al-Hassan, Alshameri & Sibley, 2013).

Extracting relevant and useful information from big data can be attained more efficiently by developing fast and high-quality document clustering algorithms (Alshameri, Greene & Srivastava, 2012). Small meaningful clusters are

organized to provide meaningful navigation and browsing mechanisms. These clusters greatly improve the process of retrieval by cluster-driven dimensionality reduction, term-weighting, or query expansion. The clusters can both explain characteristics of data distribution and at the same time, show how the datasets inter-relate. Not only does this improve the analysis and extraction of data, but also the added meaning increases a deeper context of the work.

3. LITERATURE REVIEW

By categorizing the profile of mission and vision statements of the top fortune companies in the context of globalization, innovation, and strategy, Alshameri et al. (2012) highlight the meaning that text mining can illuminate. Each cluster was categorized by their industry description, illustrating analysis of the existing trends within the industry.

Textual data mining was also used by Al-Hassan et al. (2013), clustering the datasets of legal statements of the 500 fortune companies' websites, with expectations of clusters aligned to their industrial classification code.

Adhikaria, Betancourtb & Alshameri (2014) used Clustering Toolkit (CLUTO), a text clustering software, to analyze the responses to the SEC's Proposed IFRS Roadmap. While traditional content analysis was used to look into the amount of support for the proposal, CLUTO allowed a more comprehensive look at the underlying motivations for the responses and isolated specific characteristics. CLUTO's computational power helped to uncover inherent underlying structures in big data sets.

The importance of the vision and mission statements development for businesses is also prevalent in Eastern European countries. Papulova (2014) researched how vision and mission development is considered in the Slovak Republic. Only 4.5% of companies do not have a mission or vision statement. Over 90% of businesses have a vision with almost 70% confirming the importance and significance.

Employee behavior and attitudes are also impacted by the mission and vision statements. Darbi (2012) provides research on how employees take ownership of the vision and mission, impacting the leadership and strategic knowledge and perception of their role within the organization.

Law and Breznik (2018) explore the key values embedded in the mission statements of 200 of the best ranked airlines worldwide companies, and how values are correlated to their strategies. They used content and network analysis methods to extract six content dimensions from the mission statements.

Lin, Huang, Zhu & Zhang (2019) conducted a quantitative study on the mission statements of 100 Chinese and American companies by establishing a corpus of mission statements based on the appraisal system. They analyzed the similarities and differences in the mission statements of Chinese and American enterprises, and provided suggestions for enterprises for effective mission statement development. They concluded that the mission statement of Chinese companies are society oriented and emphasize the social roles of an organization. On the other hand, the American companies' mission statements pay more attention to customer and partner relationships.

4. WHAT IS THE MISSION STATEMENT?

The mission statement can be defined in several ways. According to Bart (2001), the mission statement explains the purpose of the organization's existence. Some say the mission statement differentiates one organization from another (David & David, 2003). The mission of a company must be a contribution that characterizes the identity of the company (Cardona & Rey, 2006). The mission statement can also include the social contribution and underlying motivation that the business undertakes (Collis & Rukstad, 2008).

Many scholars and experts agree that there are numerous ways to develop mission and vision statements in an organization. However, the key issue of creating an effective mission and vision lies in the value, meaning and significance of the message as well as the implementation of the concepts (Papulova, 2014).

All types of businesses from the public sector, not-for-profit, private, for profit, a multinational or small and medium scale enterprise consider the mission and vision statement part of the strategic management process, impacting the organizational performance (Darbi, 2012).

Typically, the mission statement and vision statement are defined before developing the strategy. The vision and mission statements often include an explanation of why a specific strategy has been chosen. The mission statement defines the area in which a strategy is created. On the other hand, the vision determines direction and indicates the future strategic focus and what the future would look like if all the strategic goals were attained. Thus, when a vision is developed, consideration should include the environment that will be created in the future from the organization's functional outcomes.

The mission typically will answer why an organization exists, clarifying the identity, purpose, and direction (Henry, 2008). The combination of the vision and mission can drive the organization's culture, impacting the attitudes, work ethic and values that are reflected in the actions and inactions of the employees. Karami (2001) suggests that a mission statement can increase customer satisfaction by including customer value within the mission statement itself.

Using the mission statement as a communication tool for the stakeholders, the mission statement's main goal serves as a tool to guide strategic planning (Bartkus, Glassman & McAfee, 2004). Financial performance has been found to directly correlate to employee behavior and the employees' positive response to the internal policies and programs tied to the mission statement (Bart, 2001).

4.1 Examples of Mission and Vision Statements

The following examples of mission statements for four of top fortune companies, 3M, IBM, Walmart and Home Depot extracted from their websites.

3M: "*3M is committed to actively contributing to sustainable development through environmental protection, social responsibility and economic progress.*"

IBM's corporate mission is "*to lead in the creation, development and manufacture of the industry's most advanced information technologies, including computer systems, software, networking systems, storage devices and microelectronics.*"

Walmart's corporate mission is "*to save people money so they can live better.*"

Home Depot: "*The Home Depot is in the home improvement business and our goal is to provide the highest level of service, the broadest selection of products and the most competitive prices.*"

5. METHODOLOGY

5.1 Data Collection

The research dataset consisted of 772 Fortune companies structured vision and mission statements gathered from two websites:

- http://www.missionstatements.com/company_mission_statements.html
- http://www.company-statements-slogans.info/fortune-500-companies.htm

The first website contains statements of mission and vision segmented into different categories, such as constructions, legal, manufacturing, real estate, technology, etc.

The second website contains Fortune 500 companies mission statements listed in alphabetical order and includes more detailed information relating to each of the firms.

123

Each vision statement was preprocessed. This includes tokenization which separated out punctuation, and conjunctions, and compound words. After that stop words were removed. For stop words we used traditional English stop words provided by the Natural Language Toolkit (NLTK) (Loper & Bird, 2002). The removal of stop words allows us to give more importance to terms that are less common in regular text.

5.2 Comparing Visions

First, in order to compare industries, we hand annotated each vision statement to include one of 15 industries. These vision statements were then grouped together and treated as one document per industry.

To create feature vectors of each industry document we use the common information retrieval technique of term frequency–inverse document frequency (tf-idf) (Ramos, 2003). Tf-idf ranks the importance of a term to a document relative to other documents.

Having a value for each term, we can turn each document into a vector of values representing the terms in the document. To find out if two industries are similar, we take the cosine similarity measure between the two vectors. If the cosine value is equal to 1, they are the same industry, if the cosine value is 0, the industries share no terms in common.

5.3 Predictive Models

In order to predict an industry based on mission and vision statements we set up a test to empirically test 6 different machine learning classifiers. For each one we used standard default training flags and did not tune any parameters. We tested the following models Support Vector Machines, Naïve Bayes, Decision Trees, Random Forests, K Nearest Neighbors, and Neural Networks. The purpose was not to hyper parameterize a model but to show that our vector set created from tf-idf is a distinct enough feature set for these off-the-shelf models to be able to distinguish mission and vision statements by industry type.

If our hypothesis was incorrect and industries did not have identifiable and common elements to their mission and vision statements, the machine learning models on random data should achieve an accuracy of around 6%, by arbitrarily guessing 1 out of 15 industry types.

In order to evaluate we trained our models using an 80/20 split (80% for training, 20% for testing). We repeated this procedure 100 times, each time using random sampling without replacement. The scored report in the results section is the average of the 100 runs. We report the final averages per industry as well as the total across industries.

5.4 Clustering

To see relations between industries, we applied Ward's hierarchical clustering (El-Hamdouchi & Willett, 1986). This clustering is intended to show us how industries are related and to explain misclassification errors. These errors occur when two industries are inherently close to each other. This would be indicated as two siblings in a hierarchical cluster. Additionally, hierarchical clustering provides a check to make sure our underlying features are correct. Industries that are close (Transportation and Airlines) should be in either sibling or parent/child relationships.

For Ward's algorithm, we used the data for all 15 industries and used the tf-idf scores as the feature vector or determining geometric distance. Additionally, we included a dendrogram to visually see the relations between industries.

5.5 Key Terms

To analyze the important subjects in each industry we examined each vector set to find the terms with the highest tf-idf scores, which are the most important and distinctive terms in each industry. These were then ranked and analyze

both as a sanity check of our methodology and as an analysis of what each industry views and important. For evaluation we examine the top 6 terms ranked by importance for each industry.

6. RESULTS AND DISCUSSION

If the mission and vision statements were not identifiable by industry type, our models would achieve around 6% accuracy. Our best machine learning model will assign the correct industry 56% of the time, an 83.3% percent increase over random selection.

Table 1. Predictive Models

15 Industries	SVM	Naïve Bayes	Decision Tree	Random Forest	KNeighbors	Neural Network
Airlines	65.19%	0.00%	59.54%	24.86%	30.86%	39.35%
Automotive	44.54%	0.00%	35.08%	28.04%	22.55%	43.62%
Entertainment	30.19%	0.00%	10.33%	7.41%	8.90%	33.33%
Finance and Investment	63.12%	27.39%	45.95%	58.82%	62.63%	60.55%
Food	49.96%	4.54%	45.19%	51.26%	37.31%	51.92%
Gas and Electric Utilities	69.37%	50.15%	58.09%	63.13%	74.09%	67.96%
Healthcare	76.88%	52.03%	62.60%	67.02%	71.25%	72.49%
Insurance	47.76%	0.00%	36.72%	41.06%	2.97%	53.25%
Manufacturing	53.91%	28.96%	37.40%	41.45%	39.49%	49.24%
Marketing and Business Products	36.52%	0.00%	25.00%	12.83%	33.28%	38.83%
Real Estate	47.65%	1.45%	35.64%	29.61%	43.97%	48.09%
Retailing	41.79%	0.25%	24.34%	23.02%	20.25%	41.66%
Services	29.45%	0.00%	17.10%	18.36%	9.43%	32.33%
Technology	63.71%	94.76%	41.69%	50.72%	63.66%	58.06%
Transportation	62.59%	0.00%	53.73%	47.04%	0.00%	57.19%
Total	**56.83%**	**34.42%**	**42.25%**	**45.90%**	**47.92%**	**54.77%**

Looking across the different predictive models we can see some interesting results. If a row of data has consistently high results, this indicates that industry used distinctive concepts across their mission statements. If a row has consistently low scores, it indicates that the industry does not contain distinctive terminology and those mission statements are being confused with a different industry. For instance, technology is particularly high. This means the tf-idf feature vector was a clean decision boundary for that industry. The same can be seen in other industries such as Gas and Electric, Healthcare, Insurance, and Finance and Investment. Each one of those industries has carved out a unique identity in their mission statements.

On the other hand, there are a few industries that do not seem to have a unique identity. Top among those would be Services. From looking at their mission statements, this result is logical. Services are a provider for the other industries listed so their mission' statements are often tied to outside industries. This dilution of key terms makes them less distinctive. This same phenomenon can be seen with Marketing and Business Products. Their core service is to provide value to other industries. This makes it harder to correctly predict their industry from their mission statements.

6.1 Keywords

In the table below, the first important indicator is that many industries mention their industry by name in the mission statements. For instance, Transportation, Entertainment, Insurance and Food are all self-referencing. Other industries such as Manufacturing and Healthcare use a version of the industry, Manufacture and Health, as key words in their mission.

If a company wants to target a specific industry, this keyword extraction is a good starting point. One can determine the values of an industry through some of the keywords. For instance, Entertainment is the only industry focusing on creativity. Gas and Electric is the only industry talking about reliability. There are certain words that are used heavily across industry such as clients. While possibly important to mention, this will not make a mission stand out.

Table 2. Keywords based on industries.

Transportation	Technology	Entertainment
Transportation	Communication	Entertainment
Customers	Clients	Associates
Logistical	Information	Properties
Southern	Meet	Creativity
Worldwide	Use	Project
Public	Advance	Brands
Manufacturing	**Airlines**	**Services**
Manufacture	Priority	Clients
Brands	Different	Professional
Consumer	Right	Range
Use	Inclusive	Advice
Performing	Result	Opportunities
Leader	Respect	Talent
Gas and Electric Utilities	**Finance and Investment**	**Healthcare**
Energy	Clients	Health
Environmental	Members	Carefully
Reliability	Financial	Hospitality
Fuel	Credited	Educate
Nature	Organization	Living
Performing	Life	High
Insurance	**Retailing**	**Real Estate**
Insurance	Stores	Real
Protecting	Place	Constructive
Retirement	Price	Home
Risk	Merchandise	Clients
Mutual	Home	Professional
Management	Sell	Highest
Food	**Automotive**	**Marketing and Business**
Food	Leader	Products
Every	Advance	Clients
Brands	Growing	Firm
Consumer	Provide	Public
Ensure	Brands	Laws
Ethical	Years	Creativity
		Professional

6.2 Clusters

For the most part, the results of Ward clustering seem to be as expected. Services matches with Business Products, Transportation matches with Airlines, Food and Retail are related. We can see a possible reason why Services and Marketing and Business Products had low predictive power. They are clustered together on the lowest level, indicating that these two could be merged. If merged into one industry, it would be more distinctive and more likely to be predicted correctly by traditional machine learning algorithms.

Figure 1. Hierarchal clusters for the industries.
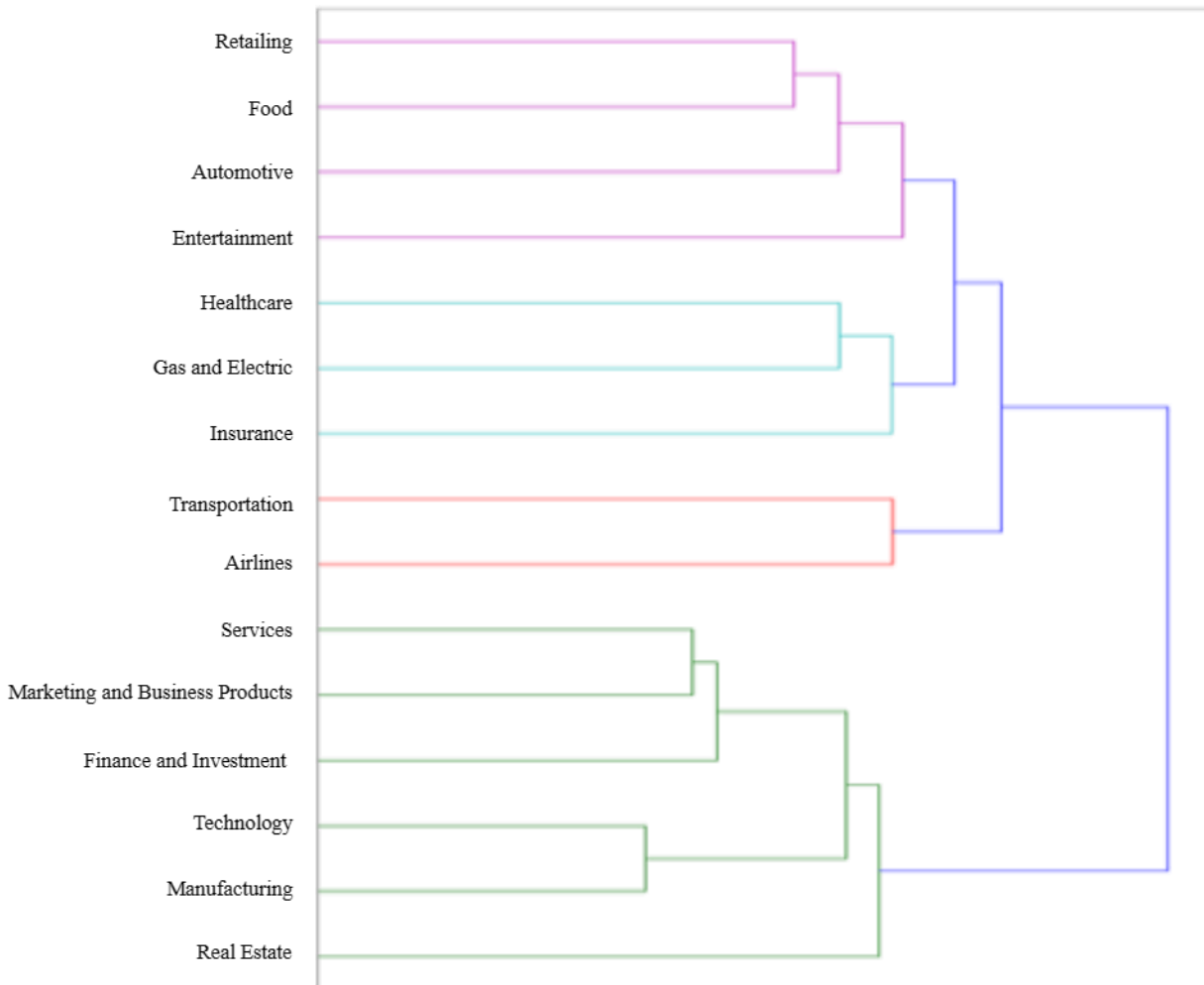
7. CONCLUSION

We have shown the distinctive terminology for each of the 15 industry types as well as the connectiveness of different industries with each other. We accomplished this by using hierarchical cluster along with 6 different machine learning models. Our models were able to predict industry type a given industry from mission and vision statements with accuracy well above random chance. This indicates a strong relation between mission and vision statements to their given industry. This work can be used for automatic classification of industry types in the future and also help guide businesses on key aspects of their chosen industry if they want to be in sync with other companies.

8. FUTURE WORK

This work focused on analysis based off our industry code annotation for 15 industries. Future work will include additional annotations for company size, for-profit status, and geographic region. We anticipate that these additional annotations will give further insight on how companies choose to represent themselves both externally and internally through missions and vision statements. While this work focused on terminology via word choice, future work can include further language analysis using underlying syntax and structure.

AUTHOR BIOGRAPHIES

Dr. Faleh Alshameri is an Associate Professor of Computer Information Systems. He received his PhD from the Volgenau School of Engineering and Information Technology, George Mason University. His research interests include Text Mining, Image Mining, Data Science, and Big Data Analytics.

Dr. Nathan Green Received his PhD from Charles University in Computational Linguistics. His research interests are in Natural Language Processing, Machine Learning, and corpora construction. He is currently an Assistant Professor at Marymount University in Arlington Virginia.

REFERENCES

Adhikaria, A., Betancourtb, L., & Alshameri, F. (2014). The SEC's proposed IFRS roadmap: An analysis of comment letters using content analysis and textual software. *Journal of International Accounting, Auditing and Taxation*, *23*, 98-108.

Al-Hassan, A., Alshameri, F., & Sibley, E. (2013). A research case study: Difficulties and recommendations when using a textual data mining tool, *Information & Management*, *50*, 540-552.

Alshameri, F., Greene, G., & Srivastava, M. (2012). Categorizing top fortune company mission and vision statements via text mining. *International Journal of Management & Information Systems, 16*(3), 227-238.

Bart, C. (2001). Measuring the mission effect in human intellectual capital. *Journal of intellectual capital*, *2*(3), 320-30.

Bartkus, B., Glassman, M., & McAfee, R. (2004). A comparison of the quality of European, Japanese and US mission statements: A content analysis. *European Management Journal*, *22*(4), 393-401.

Cardona, P., & Rey, C. (2006). Management by missions: how to make the mission a part of management", *Problems and Perspectives in Management*. *4*(1), 164-74.

Collis, D., & Rukstad, M. (2008). Can you say what your strategy is?. *Harvard Business Review*, *89*(4), 82-90.

Darbi, W. (2012). Of mission and vision statements and their potential impact on employee behavior and attitudes: The case of a public but profit-oriented tertiary institution. *International Journal of Business and Social Science, 3*(14), 95-109.

David, F., & David F. (2003). It's time to redraft your mission statement. *Journal of business strategy*, *24*(1), 11-14.

El-Hamdouchi, A., & Willett, P. (1986). Hierarchic document classification using Ward's clustering method. *In proceedings of the Ninth International ACM SIGIR Conference on Research and Development in Information Retrieval*, 149-156.

Henry, A. (2008). Understanding Strategic Management. *Oxford, UK: Oxford University Press*, Ch. 1, 11-12.

Khalifa, A. (2011). Three Fs for the mission statement: what's next?. *Journal of Strategy and Management*, *4*(1), 25-43.

Karami, A. (2001). Business strategy, environment analysis, and company performance: the evidence from the UK electronic industry. *International Business and Economic Research Conference*, CO, October 8-12. Colorado, USA.

Law, K. & Breznik, K. (2018). What do airlines mission statements reveal about value and strategy?. *Journal of Air Transport Management*, *70*, 36-44.

Lin, Q., Huang, Y., Zhu, R., & Zhang, Y. (2019). Comparative analysis of mission statements of chinese and american fortune 500 companies: A study from the perspective of linguistics. *Sustainability*, *11*(18), 4905-4112.

Loper, E., &Bird, S. (2002). NLTK: the natural language toolkit. *In Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics*, 63-70.

Papulova, Z. (2014). The significance of vision and mission development for enterprises in slovak republic. *Journal of Economics, Business and Management, 2*(1), 12-16.

Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. *In First International Conference on Machine Learning*, New Brunswick:NJ, USA.