Dec 10th, 12:00 AM

A Study on Job Satisfaction Factors in Retention and Turnover Groups using Dominance Analysis and LDA Topic Modeling with Employee Reviews on Glassdoor.com

Jongseo Lee
School of Integrated Technology, Yonsei University, Semanticweb7@yonsei.ac.kr

Juyoung Kang
Ajou University College of Business Administration, jykang@ajou.ac.kr

A Study on Job Satisfaction Factors in Retention and Turnover Groups using Dominance Analysis and LDA Topic Modeling with Employee Reviews on Glassdoor.com

Short Paper

Jongseo Lee
School of Integrated Technology,
Yonsei University
Incheon, Korea
semanticweb7@yonsei.ac.kr

Juyoung Kang
Department of e-Business,
Ajou University
Suwon, Korea
jykang@ajou.ac.kr

Abstract

HR analytics is an important area for the application of big data analysis techniques, and the organizational insight that it provides enables effective management of employees. In this paper, we analyze employee review data posted on a representative third-party employee review website. We identify the relative importance of factors affecting job satisfaction and then extract topic differences after classifying employees according to retention and turnover. First, LDA Topic Modeling by adopting n-grams is performed on unstructured text data to analyze employee review data. Second, a dominance analysis is conducted to examine the relative importance of job factors. We found that the "Culture and Values" and "Senior Management" factors have the highest influence on both retention and turnover. Our model follows a novel approach in applying the analysis of reviews and text mining to the HR domain and will be of practical relevance for enhancing employee retention.

Keywords: Employee Review Data Analysis, Job Satisfaction, Text mining, Dominance Analysis, LDA Topic Modeling, Inductive Reasoning

Introduction

Applying "big data" analysis techniques in the critical application area of HR Analytics, provides insight into organizations, helping to manage workforces swiftly and effectively, and to reach business goals efficiently. According to Deloitte's business and HR trends, HR Analytics is one of the top 10 key topics (Bersin 2016). A survey conducted by MIT and IBM showed an 8 percent higher annual growth rate for companies that performed high-level HR Analytics (Sen 2015). HR Analytics has become an important factor for companies seeking to gain a competitive advantage, and "super bosses," or Chief Executive Officers, are already integrating it into their corporate strategies.

Companies rely on talented people with diverse skills and knowledge; these people are difficult to replace in complex business environments (Beugelsdijk 2008). One of the most important challenges facing corporations is to build and maintain a strong talent pipeline that will keep them competitive (Stahl et al. 2012). Talent management has six main branches: development and training; recruitment and selection; talent review; compensation and rewards; performance management; and retention (Stahl et al. 2012). Retention is an important component of talent management. A successful company knows that talent management processes must have broad ownership at all levels of management, including the CEO level, in addition to HR. Senior leaders must participate actively in the talent management process and must

prioritize the hiring, succession planning, leadership development and retention of key employees. For effective retention management, it is necessary to understand the factors that affect employees' job satisfaction and to develop strategies to improve them.

Inductive reasoning is frequently conducted within organizational science, for more effective detection of phenomena. Although the field is increasingly acknowledging inductive approaches (Spector et al. 2014), the bulk of organizational research remains strongly oriented toward hypothetico-deductive confirmatory studies. The era of big data has further strengthened inductive reasoning research (McAbee et al. 2017). The strategic goal when obtaining and analyzing big data is to facilitate informed organizational decision-making, commonly referred to as data-driven decision-making (Provost et al. 2013). Indeed, big data analytics allows organizations to scrap a one-size-fits-all model and to tailor customer experiences to the individual.

Job satisfaction analysis is a key part of HR Analytics. Job satisfaction is recognized as an important concept in the field of organizational behavior. In particular, there have been many studies on job satisfaction factors. Whanger (2002) analyzed the relative importance of the five predictive variables that influence organizational commitment, job satisfaction, and intrinsic motivation. Most of the previous research related to job satisfaction is based on the deductive approach, and similar approaches have led to limitations in the deriving of new facts. Therefore, a new directional approach is needed that involves inductive reasoning. In addition, the existing job satisfaction research does not guarantee the representativeness of the sample, which is limited to a few companies. Previous research has shown that data collection methods mainly center on surveys; as such, researchers only collect information on pre-set questions, leading to a possibility of social desirability bias. Social desirability bias is a social science research term that describes a particular type of response bias, that is, the tendency of survey respondents to answer questions in a manner that will be viewed favorably by others. Big data can be employed in a new, multi-dimensional analysis in this field.

Corporate review data has merits. First, it can improve the representativeness of the sample because it can collect information about numerous companies. Second, it can obtain candid opinions of employees about their companies. According to Lakin (2015), people seeking information about a company trust the reviews on sites, such as Glassdoor, more than traditional information such as corporate or national guidelines. Ji et al. (2017) considered that Glassdoor's 1,112,476 employee ratings, of 14,282 public firms from 2008-2015, are a more direct way to examine the corporate culture that employees experience, rather than upper echelon proxies. We extracted the reviews for all S & P 500 firms (2016) and analyzed a total of 217,379 reviews. We consider that this large amount of data represents normality. Performing exploratory data analysis (EDA) based on big data from Glassdoor will help establish new directional indicators in organizational research to aid the understanding of job satisfaction.

In this paper, we classify employees according to retention and turnover, identify the relative importance of factors affecting the job satisfaction of the employee group, and extract the topic differences to present a new direction for talent management. To do this, we analyze the data on employees' reviews of companies, is performed on unstructured text data for the analysis group in order to analyze enterprise review data in depth. In addition, dominance analysis is performed on the relative importance of job satisfaction factors by sector. Our study provides two contributions. First, we apply inductive reasoning to organizational research by using big data, and perform various analyses for diverse groups using text mining and dominance analysis. Second, through these analyses, we identify key points to help improve employee retention.

The composition of this study is as follows. In Chapter 2, we describe the research on job satisfaction and turnover, deductive and inductive reasoning based organizational research, LDA topic modeling and dominance analysis. Chapter 3 describes specific research models and procedures, and discusses the analysis and results of the research, while Chapter 4 draws conclusions and indicates future directions for the research.

Literature Review

Inductive Reasoning in Organizational Research

Deductive reasoning is the dominant research model in the organizational sciences (Locke 2007). Moqbel et al. (2013) investigates the effects of organizational members' use of social networking sites, job satisfaction, organizational commitment, and job performance using structural equation modeling. Okpara (2004) investigated the extent to which personal characteristics such as gender, age, education, income, and experience predict IT managers' job satisfaction in Nigeria. A standard multiple regression analysis was performed with specific aspects of jobs as the dependent variables and with personal characteristics as independent variables.

Various studies using inductive reasoning have been conducted in organizational research. Obalola et al. (2012) used a narrative-inductive approach to make important contributions to the literature on corporate social responsibility and organizational effectiveness. According to McAbee et al. (2017), big data analytics can affect organizational theory (McAbee et al. 2017). Luo et al. (2016) extracted anonymous employee reviews for textual analysis from Glassdoor and revealed the relationship between employee satisfaction and company performance. Indeed, organizations have already begun to employ big data analytics to identify and observe unsafe work behavior, in order to inform future training needs. In a recent study, Guo et al. (2016) implemented a behavior-based safety (BBS) approach to observe and analyze patterns of unsafe behavior at a construction site using video data (a form of unstructured data). As such, the value of inductive reasoning research using exploratory data analysis (EDA) is recognized for its value in a variety of organizational studies.

Job Satisfaction and Turnover

Job satisfaction is a measurement of an employee's attitude towards his or her experience of the job, and of the satisfaction with that experience, or a positive emotional state (Locke 1976). Job satisfaction can be operationally defined as follows. We can examine the individual's overall satisfaction with the job, examine various factors of the job, and then define the job satisfaction factors as a combination of the measured factors (Wanous et al. 1972). A general reference to job satisfaction indicates the overall satisfaction with the job. In the case of job satisfaction factors, it consists of tasks themselves or content that affects specific tasks. For example, there are factors such as Work/Life Balance, Culture and Values, and Career Opportunities. In particular, an analysis of job satisfaction that considers the employee's satisfaction with each of the job's various factors is called job satisfaction factors theory. There have been many studies on the relationship between job satisfaction and job satisfaction factors in various perspectives. Wanous et al. (1972) analyzed the method of finding job satisfaction effectively and reviewed nine operational definitions of job satisfaction. In addition, Rice et al. (1991) examined factor importance, job satisfaction, and job satisfaction factors. However, despite the various studies on job satisfaction, there is a lack of discussion on the relative importance of job satisfaction by factor. This study is expected to provide a new perspective on the theory of job satisfaction.

Since employee turnover significantly affects the performance of an organization, researchers have for a long time been working to develop a prediction model for turnover. In these studies, job satisfaction and organizational commitment were suggested as the leading factors in the turnover intention rate. The "satisfaction of job performance" hypothesis is based on the expectation theory (Vroom 1982). In other words, according to this theory, high performance will bring greater rewards, and greater rewards will definitely be attractive to everyone. This is direct evidence of job satisfaction (Lawler III et al. 1976;Locke 1976; Locke et al. 2004). In addition, explanations based on expectation theory imply that job satisfaction mediates between performance and turnover intention or turnover, rather than being directly related to it. In other words, if performance decreases, job satisfaction decreases, and therefore, the attractiveness of the job disappears, turnover intention increases, and moreover, turnover occurs. The study of Tett et al. (1993) provides interesting empirical results. According to their results from calculating True-Score Correlations, although job satisfaction is more related to turnover intention than organizational commitment, organizational commitment has a greater effect on job satisfaction than on actual turnover. Cho (2012) investigated the influence of organizational commitment and professional commitment on professionals' intention to leave their organizations for reasons of professional advancement.

Dominance Analysis

Dominance analysis is an analytical method to judge independent variables of high relative importance that affect dependent variables in multiple regression analysis (Budescu 1993). Relative importance means that the influence of the variables on the explanatory power (R^2) of the regression equation among the predicted variables is expressed as a proportionate contribution. Although many studies have proposed a method of calculating the relative importance of predictive variables, dual dominance analysis has been proposed as the most efficient method with the fewest logical defects (LeBreton et al. 2004). The specific analysis procedure is as follows. First, we make all the possible combinations of $2n$-1 for n explanatory variables and find all the increments of R^2 that change every time the variable changes. Then, the average of the increment of R^2 for each variable is obtained and the marginal contribution of each explanatory variable is evaluated. The variables with relatively high importance are identified according to the magnitude of the marginal contribution of explanatory variables. An example of using precedent analysis in organizational research was demonstrated when Whanger (2002) conducted a study analyzing the relative importance of the five predictors of organizational commitment, job satisfaction, and intrinsic motivation. LeBreton (2004) used dominance analysis to determine the relative influence of job satisfaction factors on job satisfaction, withdrawal cognitions, withdrawal behavior and the big five personality traits. Kath (2013) evaluated the relative importance of five potential predictors of nurse manager stress using role stress theory and job demands-resources theory.

Although many researchers use regression coefficients from a multiple regression analysis, those coefficients actually represent the amount of unique variance predicted by any one predictor. If the predictors are correlated, then regression coefficients do not adequately represent how well one predictor performs when compared to all combinations of the other predictors. This is, however, precisely what dominance analysis calculates: how much variance is accounted for as compared to all combinations of other predictors in the model. Hence, we use dominance analysis to calculate the relative importance of several job satisfaction factors. Since we apply dominance analysis to organizational research by using big data, our research is different from other studies.

Latent Dirichlet Allocation (LDA) Topic Modeling

LDA topic modeling is a probabilistic modeling method for classifying documents based on information related to the topic of the text in a large number of unstructured documents (Blei et al. 2003). The LDA estimates two probability distributions through the document, assuming that the documents comprise the probability distributions of the topics and the probability distributions of the words in the topic. LDA topic modeling is a text mining technique that extracts topics in a document using a Dirichlet distribution estimation in a large amount of document data. LDA has been used in various fields based on efficiency. Topic modeling has been used for various studies. Grimmer (2010) used it effectively to identify and measure the priorities of US Senate policies. In addition, Zhao et al. (2011) compared Twitter's ability to provide quick information on all news with traditional media. Hattori et al. (2013) extracted tip information that had been classified by topic from social networking services. Dickinger et al. (2017) analyzed the discrimination between positive and negative reviews based on single word items and the sector-specific relevance of hidden topics. In this study, LDA topic modeling is used to identify the weights of topics related to the job content included in the “pros” review of the retention group and the “cons” review of the turnover group.

Research Methodology

Our research methodology includes techniques from dominance analysis and text mining. As shown in Figure 1, it consists of two main analyses. It contains several major components: data collection from Glassdoor.com, text pre-processing, LDA topic modeling, and dominance analysis using job satisfaction factors. LDA topic modeling with n-grams is performed. Wallach (2006) develops a bigram topic model on the basis of the hierarchical Dirichlet language model (MacKay et al. 1995), by incorporating the concept of topic into bigram models. Gildea et al. (1999) tested their novel statistical language model using unigrams, bigrams, and trigrams. Similarly, we adopt uni, bi, and trigrams for LDA topic modeling. Dominance analysis is performed on the relative importance of job satisfaction factors by economic sector. Smart (2016) proposed statistical analysis regarding company culture matters for employee turnover, and

used a simple linear probability model to estimate which factors predict that workers will leave their employer. Job satisfaction factors are a set of Glassdoor employer ratings, including Overall Rating, Culture and Values, Work/Life Balance, Career Opportunities, Compensation and Benefits, and Senior Management. Incorporating these factors, Schmiedel et al. (2016) validated job satisfaction factors using review text. We adopt these same job satisfaction factors for dominance analysis.

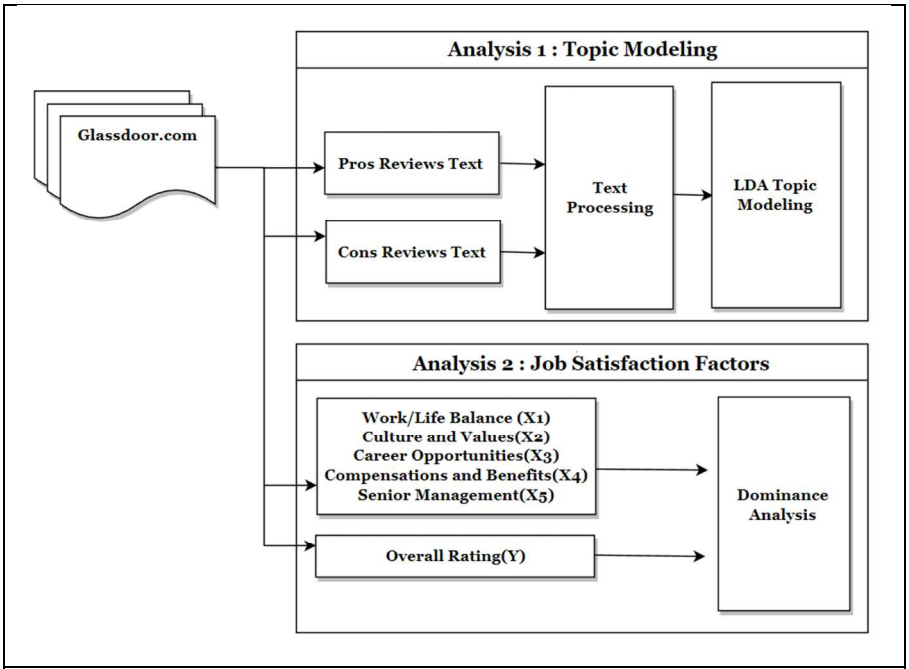

Figure 1. Research Model

Data Collection from Glassdoor.com

Glassdoor currently lists approximately 500,000 companies with more than eight million company reviews, company ratings, CEO approval ratings, and salary reports. The employment positions held by Glassdoor reviewers vary. For each individual review, Glassdoor provides several metrics to gauge job satisfaction, the most prominent of which is the overall rating, ranging on a scale from 1 to 5. Employees can optionally evaluate their employer on five additional criteria: (i) Work/Life Balance, (ii) Culture and Values, (iii) Career Opportunities, (iv) Compensation and Benefits, and (v) Senior Management. There is also an indicator for retention/turnover. Using this information, we can know whether each review is written by a former employee or current employee. Finally, there are positive and negative text reviews where the reviewer can give opinions about the employer or the job he or she wants to include. Therefore, we can analyze the job satisfaction across five specific ordinal metrics, as well as positive and negative text reviews.

Table 1. Glassdoor Dataset Summary Statistics

Sector	Number of firms	Number of reviews	Mean	Median	Maximum	Minimum
Consumer Discretionary	79	27,654	305.1	193	1,339	3
Consumer Staples	36	12,955	359.9	268	872	5
Energy	40	13,724	343.1	198.5	1233	3
Financials	65	28,912	444.8	356	3,579	10
Health Care	57	24,813	435.3	418	909	5

Industrials	70	39,955	570.8	703.5	909	3
Information Technology	67	21,620	322.7	180	952	5
Materials	27	9,555	353.9	267	853	16
Real Estate	30	22,710	757.0	569.5	8,467	7
Telecommunications Services	5	2,705	541.0	506	822	132
Utilities	23	12,776	555.5	711	890	42
Total	499	217,379				

Programming in R, we developed a Web crawling algorithm to crawl the review data from Glassdoor.com. We extracted reviews for all S & P 500 firms (2016) and crawled a total of 277,977 reviews. We also excluded reviews lacking data in any of the five job satisfaction factors because we use dominance analysis on these factors. Our final dataset is comprised of 217,379 reviews covering 499 companies, representing 11 economic sectors based on the Global Industry Category Standard (GICS). In Table 1, we provide descriptive statistics on the number of reviews, by sector. Industrial, Financials, Consumer Discretionary, Health Care, Real Estate, and Information Technology are the six sectors with the most reviews, respectively.

Multivariate Analysis through EDA

From the 217,379 employee review data sets, we performed a multivariate analysis to ascertain differences by satisfaction factors and by sector. In Figure 2, we present correlation coefficients between satisfaction factors. The Culture and Values factor, Career Opportunities, and Senior Management factors show higher correlation coefficients (dark blue) than the Work/Life Balance and the Compensation and Benefits factors (light blue). Especially, the correlation coefficients with overall ratings of (0.59, 0.76, 0.72, 0.57, and 0.75) indicate that a difference in influence between job satisfaction factors exists.

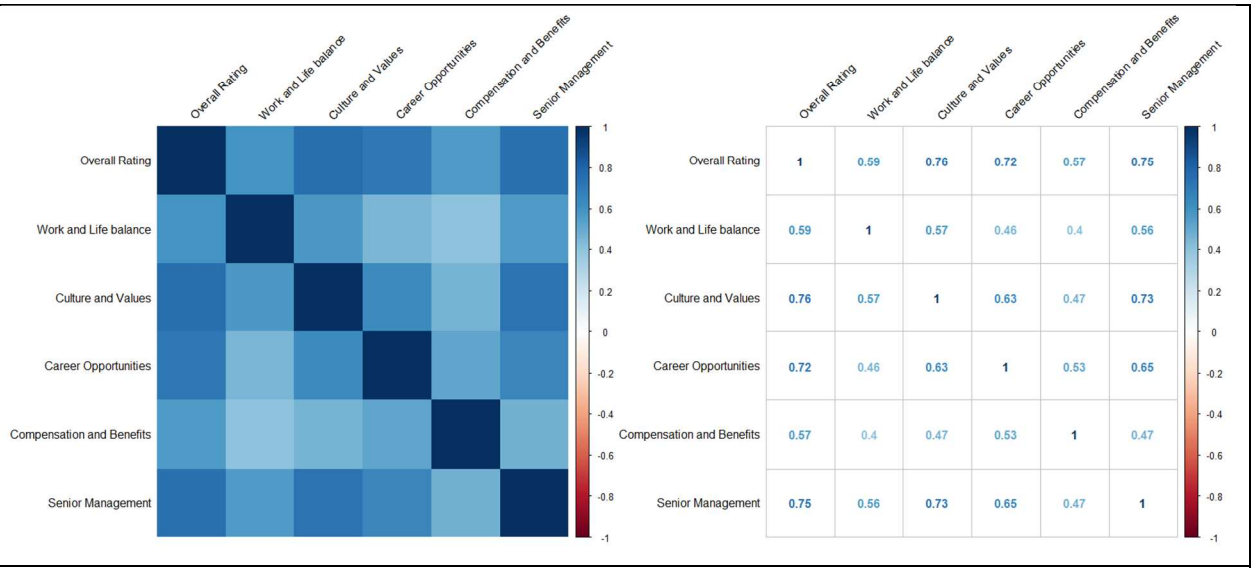

Figure 2. Correlation Coefficients between Satisfaction Factors

In Figure 3, we present the differences between the rating ratios of job satisfaction factors by sector. Among job satisfaction factors, the positive evaluation ratio of the overall rating for the Culture and Values factor is the highest in all the sectors. In the Materials sector, the positive evaluation ratio of the

overall rating for the Career Opportunities factor is relatively higher than in the others. In the Information Technology sector, the positive evaluation ratio of the overall rating for the Compensation and Benefits factor is relatively higher than the same factor in the other two sectors. Lastly, the positive evaluation ratio of the overall rating for the Work and Life Balance factor, in the Consumer Staples sector, is also relatively higher. Through these EDA, we find differences in job satisfaction factors by sector.

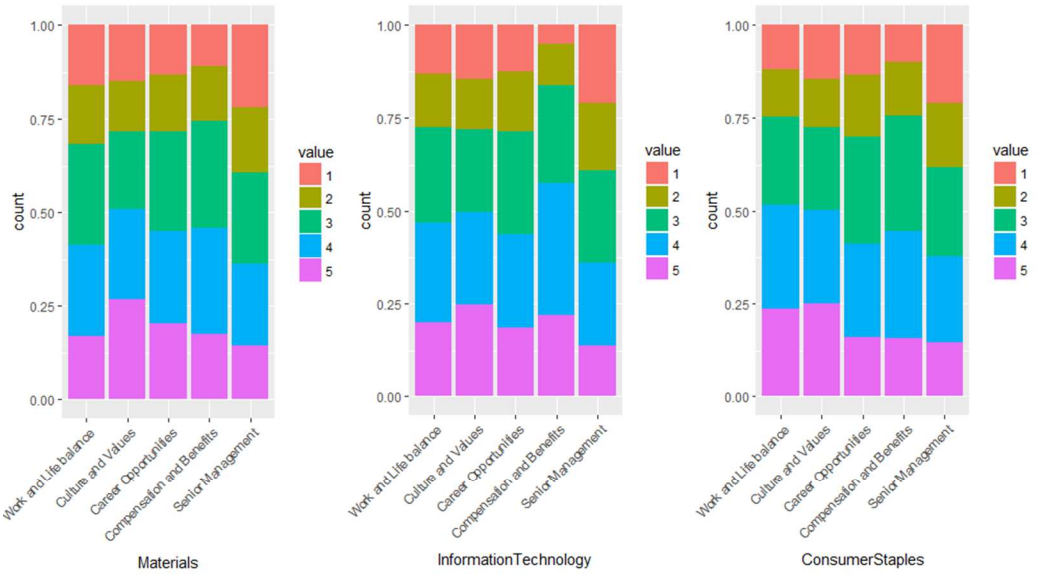

Figure 3. Difference of Rating Ratio of Job Satisfaction Factors by Sector

Topic Modeling Results

We perform LDA topic modeling to find which topics are relevant for the retention employee group with positive ratings (4, 5 points) and the turnover employee group with negative ratings (1, 2 points). We apply *n*-grams to LDA topic modeling for convincing experiments. First, we analyze the "pros" reviews in the retention group. In Figure 4, we present, by *n*-gram, topic composition ratios for retention employees. The Culture and Values topic has the largest share for each *n*-gram (33.2%, 46.4%, and 43.8%). "work environment," "flexible work schedule," "great company culture," "friendly work environment," and "fun" are extracted as highly relevant keywords for the Culture and Values topic. In Table 2, we present, by *n*-gram, the most relevant keywords by retention employees. "balance of work," "big profits," "tuition reimbursement," "insurance," and "opportunities" are extracted from relevant topics to the overall topic. Since "work life balance" and "flexible work schedule" can positively influence employee retention, organizations should reflect results for new strategies for employee retention. Second, we analyze the "cons" reviews in the turnover group. In Figure 5, we present, by *n*-gram, topic composition ratios for turnover employees. The Senior Management topic has the largest share for each *n*-gram (45%, 51.9%, and 47.8%). "senior management," "leadership," "poor management," "management doesn't listen," and "management doesn't care" are extracted as highly relevant keywords for the Senior Management topic. In Table 2, we present, by N-gram, most relevant keywords by turnover employees. "poor work life," "stressful work environment," "low pay," "minimum wage," and "limited growth opportunities" are extracted from relevant topics in the overall topic. Since "management doesn't care," "poor work life," and "limited growth opportunities" can influence employee turnover negatively, senior management should consider communicating more with their employees in order to find a compromise for preventing turnover.

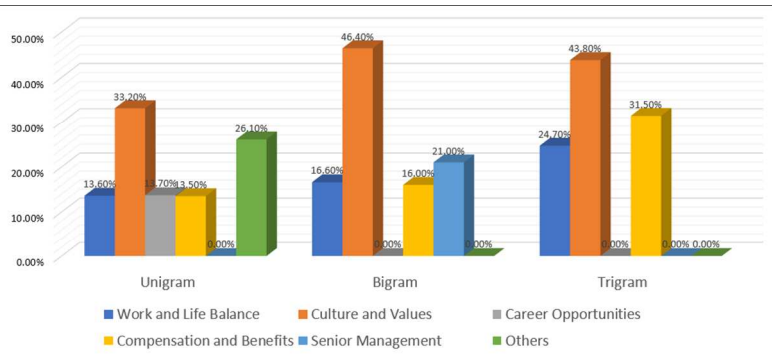

Figure 4. Topic Modeling Composition of Pros for Retention with Positive Ratings

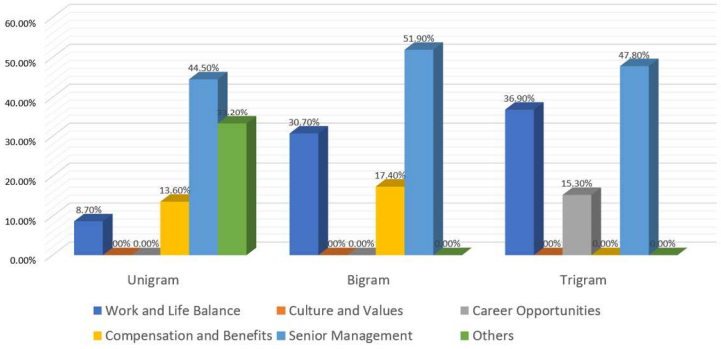

Figure 5. Topic Modeling Composition of Cons for Turnover with Negative Rating

Table 2. Extracted Relevant Keywords for Employee Retention and Turnover

Group	Unigram	Bigram	Trigram
Retention Employees With Positive Ratings (Pros)	work, benefits, career, growth, flexible, culture, opportunities, free, fun, pay, friendly, good, great, time, health, insurance, vacation, advancement, ...	work life, life balance, great benefits, work environment, flexible work, tuition reimbursement, great people, friendly people, balance good, great pay, part time, ...	work life balance, great work environment, life balance good, good benefits salaries, great company culture, friendly work environment, flexible work schedule, ...
Turnover Employees With Negative Ratings (Cons)	leadership, management, culture, people, time, hours, poor, life, sales, low, lack, benefits, balance, pay, layoffs, terrible, managers, growth, training, ...	senior management, upper management, long term, work life, life balance, low pay, long hours, poor management, minimum wage, senior leadership, ...	work life balance, poor work life, high turnover rate, limited growth opportunities, stressful work environment, long term employees, management doesn't care, ...

Relative Importance Ranking of Job Satisfaction Factors

We perform dominance analysis to check the relative importance of job satisfaction factors by sector and by retention and turnover employee groups, as shown in Figure 6. Among the job satisfaction factors, the relative importance of the Culture and Values factor is the highest overall in both the retention group (25.87%) and the turnover group (27.26%). In particular, the relative importance of Culture and Values is the highest in the Financials, Consumer Discretionary, Materials, Information Technology, and Industrials sectors. Focusing on the Culture and Values factor may be an efficient way to prevent employee turnover. According to the topic modeling results in the Culture and Values factor, "forward thinking company" and "open door policy" are extracted as relevant keywords for a positive influence. On the contrary, "highly political environment" and "culture sucks communication" are extracted as relevant

keywords for a negative influence on employee turnover. Combining the dominant analysis with the topic modeling results, a progressive culture has a positive impact, and if the culture is conservative or lacks communication within the organization, it has a negative impact. The relative importance of the Compensation and Benefits and the Work and Life Balance factors are generally lower in both the retention group (13.03% and 13.64%) and the turnover group (13.60% and 12.35%) compared to that of other factors. Through group comparison, we find that the relative importance of the Culture and Values factor is 1.39% higher in the turnover employee group than the retention employee group; and the relative importance of the Career Opportunities factor is 0.88% smaller in the turnover employee group than the retention employee group. The relative importance of the Career Opportunities factor is high in the Industrials, Information Technology, Consumer Staples, and Real Estate sectors. The Senior Management factor has high relative importance overall in both the retention employee group (24.44%) and the turnover employee group (24.65%), and is the highest factor for the turnover employee group (26.04%) in the Utilities sector. The relative importance of the Senior Management factor is high in the Materials, Utilities, Energy, and Consumer Staples sectors.

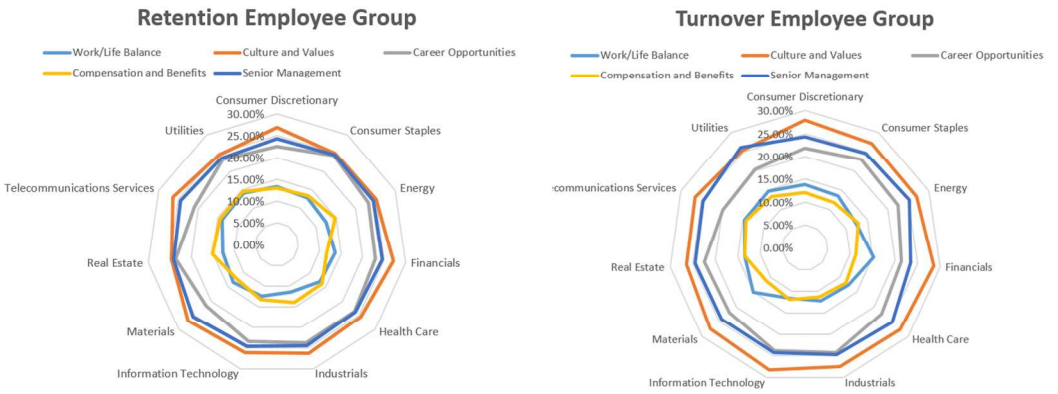

Figure 6. Relative Importance of Job Satisfaction Factors by Sector (Retention Employee Group VS Turnover Employee Group)

Conclusion

This study aims to provide a new direction for analysis of job satisfaction in retention/turnover employee groups. In order to present a talent management strategy for employee retention based on data analysis, we collect employee reviews from Glassdoor.com, which is a representative employee review site. Using LDA topic modeling and dominance analysis, we propose a novel model to analyze satisfaction in retention/turnover employee groups. Our study makes two contributions. First, we apply inductive reasoning to organizational research by using big data, and perform various analyses for diverse groups using text mining and dominance analysis. Second, through our analyses, we identify key points to help improve employee retention. Corporations should focus on creating a corporate culture that meets employees needs thereby encouraging them to identify with it and adapt easily to it. The Senior Management factor is also relatively important in both groups indicating that employees relationship with management is crucial. Upper management should ensure that they communicate frequently with their employees in an approachable and inclusive manner, through diverse avenues, so that employees feel valued and safe enough to air their grievances as well as make positive contributions. In the future, we plan to extend our study in several directions. Through EDA, we also found additional differences in job satisfaction factors for IT workers in jobs with titles such as IT Analyst and Software Developer. Through analysis we plan to investigate differences between groups with different job functions across the same sector. In addition, the influence of online reviews on job seekers' choices may warrant additional study. Finally, we hope to develop a predictive model to detect tendencies indicating possible future employee turnover at a preemptive stage. Such a predictive model would help management to react more timely and proactively to the impending loss of key human resources.

References

Bersin, J. 2016. "Analytics: The New Center of Expertise in HR," Deloitte, Available at: http://www.slideshare.net/jbersin/people-analytics-state-of-the-market-top-ten-list/4-Copyright_2015_Deloitte_Development_LLC, (Accessed March 29, 2017).

Beugelsdijk, S. 2008. "Strategic human resource practices and product innovation," *Organization Studies* (29:6), pp 821-847.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent dirichlet allocation," *Journal of machine Learning research* (3:Jan), pp 993-1022.

Budescu, D. V. 1993. "Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression," *Psychological bulletin* (114:3), p 542.

Cho, V., and Huang, X. 2012. "Professional commitment, organizational commitment, and the intention to leave for professional advancement: An empirical study on IT professionals," *Information Technology & People* (25:1), pp 31-54.

Dickinger, A., Lalicic, L., and Mazanec, J. 2017. "Exploring the generalizability of discriminant word items and latent topics in online tourist reviews," *International Journal of Contemporary Hospitality Management* (29:2), pp 803-816.

Gildea, D., and Hofmann, T. 1999. "Topic-based language models using EM," Sixth European Conference on Speech Communication and Technology.

Grimmer, J. 2010. "A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases," *Political Analysis* (18:1), pp 1-35.

Guo, S., Ding, L., Luo, H., and Jiang, X. 2016. "A Big-Data-based platform of workers' behavior: Observations from the field," *Accident Analysis & Prevention* (93), pp 299-309.

Hattori, Y., and Nadamoto, A. 2013. "Tip information from social media based on topic detection," *International Journal of Web Information Systems* (9:1), pp 83-94.

Ji, Y., Rozenbaum, O., and Welch, K. T. 2017. "Corporate Culture and Financial Reporting Risk: Looking Through the Glassdoor," SSRN, Available at: https://ssrn.com/abstract=2945745, (Accessed September, 3, 2017).

Kath, L. M., Stichler, J. F., Ehrhart, M. G., and Sievers, A. 2013. "Predictors of nurse manager stress: a dominance analysis of potential work environment stressors," *International journal of nursing studies* (50:11), pp 1474-1480.

Lakin, T. 2015. "Why you shouldn't ignore Glassdoor," *HRmagazine*).

Lawler III, E. E., and Porter, L. W. 1976. "The effect of performance on job satisfaction," in *Job satisfaction—A reader*, Springer, pp. 207-217.

LeBreton, J. M., Binning, J. F., Adorno, A. J., and Melcher, K. M. 2004. "Importance of personality and job-specific affect for predicting job attitudes and withdrawal behavior," *Organizational Research Methods* (7:3), pp 300-325.

Locke, E. A. 1976. "The nature and causes of job satisfaction," *Handbook of industrial and organizational psychology* (1), pp 1297-1343.

Locke, E. A. 2007. "The case for inductive theory building," *Journal of Management* (33:6), pp 867-890.

Locke, E. A., and Latham, G. P. 2004. "What should we do about motivation theory? Six recommendations for the twenty-first century," *Academy of management review* (29:3), pp 388-403.

Luo, N., Zhou, Y., and Shon, J. 2016. "Employee Satisfaction and Corporate Performance: Mining Employee Reviews on Glassdoor. com," in *Thirty Seventh International Conference on Information Systems, Dublin 2016 (ICIS 2016)*, Association for Information Systems: Dublin, Ireland.

MacKay, D. J., and Peto, L. C. B. 1995. "A hierarchical Dirichlet language model," *Natural language engineering* (1:3), pp 289-308.

McAbee, S. T., Landis, R. S., and Burke, M. I. 2017. "Inductive reasoning: The promise of big data," *Human Resource Management Review* (27:2), pp 277-290.

Moqbel, M., Nevo, S., and Kock, N. 2013. "Organizational members' use of social networking sites and job performance: An exploratory study," *Information Technology & People* (26:3), pp 240-264.

Obalola, M., and Adelopo, I. 2012. "Measuring the perceived importance of ethics and social responsibility in financial services: a narrative□inductive approach," *Social Responsibility Journal* (8:3), pp 418-432.

Okpara, J. O. 2004. "Personal characteristics as predictors of job satisfaction: An exploratory study of IT managers in a developing economy," *Information Technology & People* (17:3), pp 327-338.

Provost, F., and Fawcett, T. 2013. "Data science and its relationship to big data and data-driven decision making," *Big Data* (1:1), pp 51-59.

Rice, R. W., Gentile, D. A., and McFarlin, D. B. 1991. "Facet importance and job satisfaction," *Journal of Applied Psychology* (76:1), p 31.

Schmiedel, T., Müller, O., Debortoli, S., and vom Brocke, J. Year. "Identifying and quantifying Cultural Factors that Matter to the IT workforce: an Approach based on Automated Content Analysis," ECIS2016, p. ResearchPaper177.

Sen, S. 2015. "How Analytics Are Becoming Important For HR," startup.focus., Available at: http://startupfocus.saphana.com/how-analytics-are-becoming-important-for-hr/, (Accessed March 29, 2017).

Smart, M. C., Andrew;. 2016. "Why Do Workers Quit? The Factors that Predict Employee Turnover," Glassdoor RESEARCH REPORT, Available at: https://research-content.glassdoor.com/app/uploads/sites/2/2017/02/WhyDoWorkersQuit_Glassdoor.pdf, (Accessed September, 3, 2017).

Spector, P. E., Rogelberg, S. G., Ryan, A. M., Schmitt, N., and Zedeck, S. 2014. "Moving the pendulum back to the middle: Reflections on and introduction to the inductive research special issue of Journal of Business and Psychology," *Journal of Business and Psychology* (29:4), pp 499-502.

Stahl, G. K., Björkman, I., Farndale, E., Morris, S. S., Paauwe, J., Stiles, P., Trevor, J., and Wright, P. M. 2012. "Six principles of effective global talent management," *MIT Sloan Management Review* (53:2), pp 25-32.

Tett, R. P., and Meyer, J. P. 1993. "Job satisfaction, organizational commitment, turnover intention, and turnover: path analyses based on meta□analytic findings," *Personnel psychology* (46:2), pp 259-293.

Vroom, V. H. 1982. *Work and motivation*, (Robert E. Krieger Publishing Company.

Wallach, H. M. 2006. "Topic modeling: beyond bag-of-words," Proceedings of the 23rd international conference on Machine learning, ACM2006, pp. 977-984.

Wanous, J. P., and Lawler, E. E. 1972. "Measurement and meaning of job satisfaction," *Journal of applied psychology* (56:2), p 95.

Whanger, J. 2002. "The application of multiple regression dominance analysis to organizational behavior variables," JM LeBreton & JW Johnson (Chairs), Application of relative importance methodologies to organizational research. Symposium presented at the 17th annual conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., and Li, X. Year. "Comparing twitter and traditional media using topic models," European Conference on Information Retrieval, Springer2011, pp. 338-349.