SUMMARY

In this age of digital technology, the usage of online social networks is higher than ever before, which opens up new opportunities for analyzing high volumes of data. Today, people use social media for various purposes such as having conversations, sharing information, creating web content, marketing, and more. As the usage of social media platforms has increased, so has the amount of publicly available data, which can be analyzed and utilized for various purposes.

With the rise of many online job and employee experience sites, research have analyzed the large amounts of employee reviews available on these sites to better understand job satisfaction in relation to the performance of employees. For understanding these large amounts of textual unstructured data, topic modelling is a widely used technique, which given a collection of documents outputs a number of topics. These topics can then be further analyzed using methods such as correspondence analysis or sentiment analysis, which helps with identifying interesting patterns in the data.

In this study, the focus was set on analyzing large tech companies on the online employment experience platform Glassdoor. Reviews for 20 companies were scraped here, which went through a textual preprocessing phase resulting in a corpus to be used as input for topic modelling. In the next step, Latent Dirichlet Allocation (LDA) topic modelling was used on the corpus to generate a total of 9 workplace-related topics. As these topics provide information about what the employees speak of, the topics were further analyzed through a correspondence analysis for depicting which topics are more dominant in certain companies, and sentiment analysis using a lexicon sentiment analyzer for finding the best and worst topics in each company.

Findings reveal interesting facts about dominant topics across the large tech companies. Specifically the topic *Customer service* was more isolated compared to other topics, and the topic *Time and scheduling* was isolated as well, but more dominant in a few companies. The sentiment analysis discovered the topics *Work-life-balance* and *Pay and benefits* as being the most positively spoken of, and *Hiring and promotion* as being the most negatively spoken of. Future research can be taken into several directions, such as exploring companies of other industries or looking into other features of the review data that can potentially reveal some interesting facts about the employees.

ANALYSIS OF TOPICS DISCUSSED IN ONLINE EMPLOYEE REVIEWS

Markus H. Jeppesen
mjep17@student.aau.dk
Aalborg University
Aalborg, Denmark
Group: cs-22-hci-10-04

ABSTRACT

An increasing amount publicly available data on social media platforms opens up for more opportunities for learning about topics of discussion. In this study, 20 large tech companies were analyzed, utilizing the large amounts of employee experience reviews available on Glassdoor to get understanding of the topics employees talk about. This was achieved by firstly performing LDA topic modelling on a large corpus of 40.000 company reviews, which revealed 9 corporate topics. These topics were then analyzed in the individual companies using several analyses, including a correspondence analysis, and sentiment analysis. The findings from the analyses contributes with knowledge about the value of information to be found in employee reviews, which may help companies gain insight into the working life and satisfaction of employees.

1 KEYWORDS

Employee reviews, Text analysis, LDA topic modelling, Correspondence analysis, Sentiment analysis

2 INTRODUCTION

Using social media platforms today, people can find information about jobs through online reviews sharing experiences and opinions of working at a company. Online platforms like Glassdoor allows job seekers to find employee experience information in reviews from companies, which can help them in their decision of applying to a job [12].

Whether you are reading employee reviews or product reviews, reading the reviews can impact your decision of applying for a job or buying a product. Job information through word-of-mouth on social media can impact employer attractiveness and can therefore impact the job seeker's decision to apply for a job [3]. According to a survey from Dimensional research, 90 % of consumers who recalled reading online reviews claimed that positive reviews influenced their decision to buy. On the flip side, the negative reviews influenced 86 % of the consumer's buying decisions [27]. The online reviews can therefore be a great source of information not only for the customers, but also for the companies learning about the satisfaction of customers. Or in the case of Glassdoor employee reviews, knowledge about the satisfaction of employees working in a company can be extracted [18]. Previous work on the effect of social media on workplaces shows evidence of workplaces improving their practices and related disclosures after being reviewed on Glassdoor [10].

In this study, workplace-related topics are identified from employee experience information extracted from Glassdoor reviews with the purpose of drawing a picture of which workplace topic areas their employees talk about and maybe can learn from in order to improve their company. The study was conducted by scraping online reviews from the largest tech companies around the world according to market cap. Afterwards, topic modelling was performed on the reviews resulting in 9 identified topics. The topics will be referred to as "corporate topics" or just "topics".

Using the identified corporate topics as a basis, several analyses will be conducted on the companies, including a correspondence analysis and sentiment analyses. With this, the study will contribute with knowledge about what topics are discussed in the largest tech companies today, which in return will be beneficial for companies to know for finding improvements and stay competitive. Knowing how employees feel about certain topics can potentially help HR and business managers in finding areas to be improved on and in return improve employees job satisfaction.

3 RELATED WORK

Previous studies have utilized the increasing number of online textual reviews for trying to understand the satisfaction of the reviewers. A study has used online textual reviews for predicting the overall customer satisfaction by understanding how the linguistic characteristics, such as subjectivity and readability, influence customer ratings [30]. Another study identified job satisfaction factors using topic modelling on online reviews from IT companies in South Korea posted on jobplanet.co.kr [18]. In addition, the importance and sentiments (Negative, Neutral, Positive) of the identified factors were computed for the multiple companies, providing businesses with insights into the satisfaction of employees. Studies have even revealed that a relationship can be found between the satisfaction of employees and corporate performance [21] [16]. One study performed textual analysis on Glassdoor review data and discovered the corporate categories: safety, communication, and integrity to be negatively correlated with company performance [21].

As the company reviews give insights into employees' job satisfaction, the identified factors can be utilized to support decisions made by the business managers and improve the business in the end. Prior work in fields of HCI and CSCW has even introduced a design of tools for similar purposes. One study have introduced a tool called Enterprise Social Pulse (ESP), designed to support analysts' in understanding employee chatter in the form of social media content [25][2]. The tool was designed both for understanding what employees talk about and with what sentiment. The tool included visual UI-design features such as: an overview of the top 10 topics extracted, top positive topics, and top negative topics. Another study designed and developed an employee voice system as a Enterprise Social Network to facilitate discussion about workplace issues, which are often challenging to discuss openly at the workplace [2].

3.1 Text analysis trends

Using a text-mining approach like topic modelling seems to have become more and more popular for analyzing large amounts of unstructured data and finding interesting patterns [20] [5] [1]. One study in the field of HCI identified national AI policies using topic modelling on a set of 25 national AI documents from multiple countries. The study identified differences and similarities of discussed AI policies between countries, which contributed more knowledge for researchers to understand which areas to focus on within a country. Additionally, they extended their topic modelling analysis with a correspondence analysis to provide an overview of the differences between countries in terms of which topics were discussed. The same approach can be used in this study, to map out the differences between companies in terms of their corporate topics [5]. This will be useful for understanding how similar companies are in terms of their corporate topics. Topic modelling has previously been used to successfully extract job satisfaction factors such as (Salary, Vacation, Self-development) from online employee reviews from a Korean job site [18], which indicates the usefulness of using approach. This also makes it ideal for a comparison between these job satisfaction factors and the corporate topics identified in my study, as they should be slightly similar.

Sentiment analysis is a popular method used for understanding the feelings associated with words and texts as a whole. Through this analysis, the sentiment polarity (Positive, Neutral, and Negative) of sentences can be measured and used to understand how consumers feel about a certain product or how employees feel about their job and company. Previous HCI research have designed data analysis tools targeted data scientists that utilize sentiment analysis [25][29]. A recent study from 2020, developed an interactive system that enabled data scientist to get insights from reviews, including extracting the sentiments [29]. The study used a pre-trained state-of-art language model called BERT (Bidirectional Encoder Representation for Transformer), which has achieved great success in the field of AI [15]. But using an existing pre-trained model requires the data to be labelled with their ground truth sentiment, which is not the case for Glassdoor reviews used in this study.

Lexicon-based approaches to sentiment analysis have been used as an alternative for analyzing the sentiments of social media content. One study compared the machine learning approach with the lexicon-based, and the results indicated similar performance, but still the machine learning approach trained with manually classified data outperforms the lexicon-based [9]. My study explore the use of the lexicon-based approach, and its accuracy will be tested on the Glassdoor reviews, before it is chosen for the study.

Previous research show workplace satisfaction factors have been identified and analyzed in terms of importance and sentiments. This study will provide more knowledge about the corporate topics discussed in employee reviews of large tech companies, and what valuable knowledge about these topics can be extracted from company reviews. In addition, this study will analyze the corporate topics' sentiments in the different companies to get an understanding about the worst and best topics discussed by employees. Diverging from previous studies, this study will focus on extracting corporate topics from the largest tech companies in the world, as these companies have a massive amount of reviews available on online social media platforms, which may reveal interesting findings that can be utilized by the companies to gain insights into which areas to improve on.

Reliability of Glassdoor reviews

The data analyzed in this study will be employee social media data in the form of Glassdoor reviews, which will analyzed in terms of both related corporate topics and their sentiments. As Glassdoor wants to help people find a job and a company that they love, they try to be very transparent and honest about the data shared on the platform [12], which is a great indicator of how reliable the reviews are. Employees are allowed to write anonymously about both the good and bad things about their experiences working at the companies. As the reviews are anonymous, employees are more likely to tell their honest of opinion of working at a company [11]. In addition, they make sure to look into cases of fraudulent reviews, and enabling users to flag reviews if they have suspicion about fraudulent activity [13].

4 METHOD

4.1 Data collection

Review data was collected from 20 of the top 100 largest tech companies by market cap as of March 2022. To have a good sample size of reviews in the corpus used for topic modelling, 2000 reviews were scraped from each company on Glassdoor. I.e., a total of 40.000 reviews were included in the corpus. The reason for choosing 2.000 reviews from each of the 20 companies was due to the fact that not all of them had a sufficient amount of reviews to be able to include them in the study. While examining companies on Glassdoor, it was assessed that most of the companies had at least 2.000 reviews available, and therefore only companies who had at least 2.000 reviews were chosen.

To collect the reviews, the data was scraped in the order, starting from the most recent reviews. A review consist of a total of 3 text sections: pros, cons, and advice to management, as illustrated on figure 1. Every section of the review was scraped and included in the corpus.

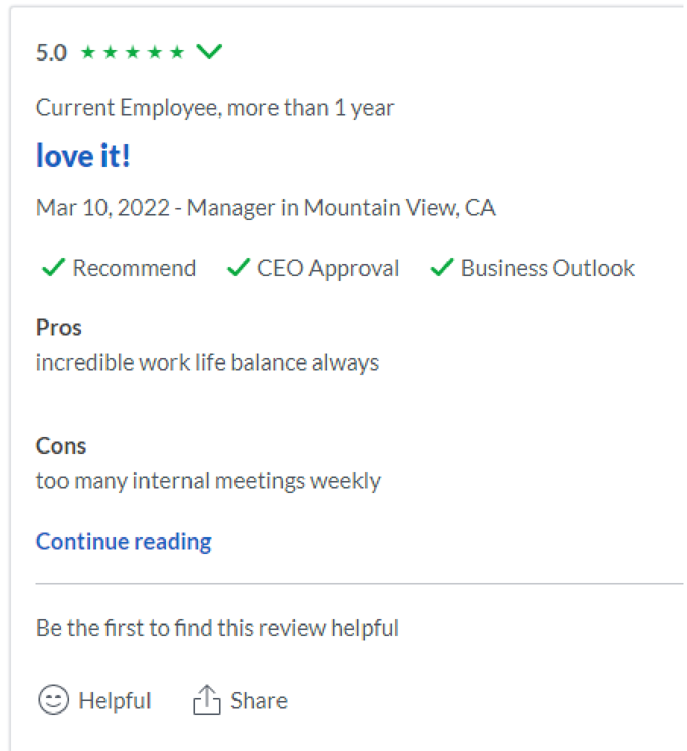

FIGURE 1. Example of Glassdoor review

4.2 Textual pre-processing

To prepare the corpus used for generating the topics as illustrated in figure 2, each review went through a textual pre-processing phase, converting the unstructured review data into structured data of tokens. This phase included the process of breaking down the sections of text into word-tokens and removing meaningless words in the form of stopwords and adjectives. Words were tagged according to their Part of Speech (POS), and only nouns, verbs, and adverbs were extracted, as other POS-categories don't provide much meaning to the sentences. In addition, lemmatization was performed on the word-tokens, reducing words to their base form (e.g. benefits $\rightarrow$ benefit). In the last step, frequent bigrams and trigrams were identified from the text. These word sequences consist of two consecutive words in the form of a bigram (e.g. fast pace), or three consecutive words in the form of a trigram (e.g. work life balance). After the bigrams and trigrams have been identified, an id-to-word dictionary and term-frequency corpus was created and used as input for topic modelling. The term-frequency corpus is used for looking up the number of occurrences of a term within a document/review and the id-to-word dictionary is used for looking up a word token associated with an integer ID [24].

5 RESULTS

5.1 Topic modelling

To perform topic modelling a Latent Dirichlet Allocation (LDA) topic modelling algorithm was used to extract corporate topics from Glassdoor reviews. By using topic modelling, topics that best describe a set of documents/reviews can be identified using probability distribution theory. The LDA model finds the topics based on the assumption that words in a document are related. The topic modelling algorithm chosen used Gibbs sampling, which is a slower but more precise method compared to the Variational Bayes sampling method [6]. Using Gibbs sampling will therefore be the better option for achieving topics with great coherence, which is a measure of the degree of similarity between the words of a single topic.

To find a good amount of topics, coherence scores were calculated for multiple topic models with a varying number of topics between 2 and 29 topics. The CV coherence metric was used ranging from 0 to 1, with a score closer to 1 indicating a better coherence between the words within a topic. The best model obtained a coherence score of 0.684 and generated a total of 9 topics. The topics can be seen in table 1 together with the word

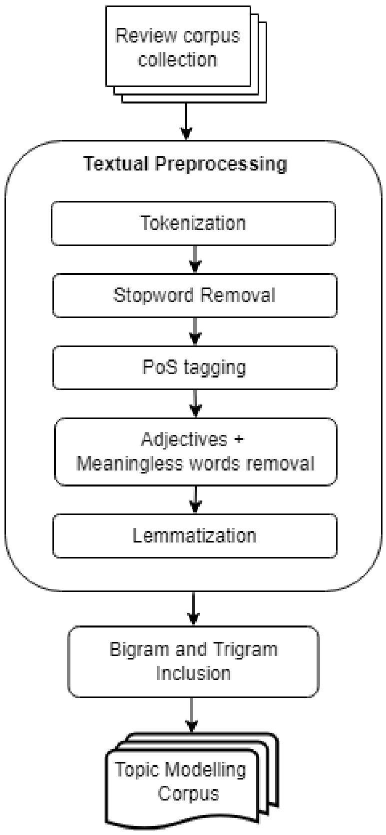

FIGURE 2. Overview of the textual preprocessing phase

tokens associated with the topics.

Each topic were labeled with an appropriate title according to the most salient terms, which are the words that are the most informative for the topic, thereby the most characteristic terms for defining a topic [8][22]. As an example, the most salient terms for topic number 6 on figure 1 are *Company*, *Opportunity*, *Growth*, *Learn*, and *Career*, which can with some human reasoning be aggregated to the title label *Development and growth*.

5.2 Correspondence analysis

Given the discovery of the corporate topics, a correspondence analysis between the topics and companies was conducted with the purpose of observing how the discussion of topics differs from one company to another. The same 20 large tech companies chosen for topic modelling were chosen for this analysis, which had a minimum of 2000 reviews. To access when a topic is being discussed, the frequencies of terms associated with a topic were counted, and lastly the sum of the term-frequencies for each topic was computed. This procedure was performed for each of the 20 companies and used in the correspondence analysis resulting in figure 3.

TABLE 1. The corporate topics generated with topic modelling together with the top 15 related word tokens

Topic	Title	Word tokens
0	Work-life-balance	Work, Life balance, Environment, Culture, People, Life, Love, Working, Colleague, Balance, Enjoy, Fast pace, Place, Travel, Load
1	Pay and benefits	Benefit, Pay, Salary, Compensation, Flexibility, Perk, Bonus, Culture, Coworker, Stock, Offer, Food, Location, Health, Plan
2	Time and scheduling	Time, Hour, Work, Job, Office, Home, Schedule, Week, Shift, Depend, Meeting, Expect, Require, Vacation, Break
3	Project management	Team, Management, Project, Support, Experience, Group, Communication, Lead, Resource, Engineer, Client, Improve, Build, Manage, System
4	Customer Service	Customer, Sale, Issue, Service, Feedback, line, Money, Share, Deal, Month, Provide, Hear, Call, Drive, Time
5	Employee management	People, Employee, Company, Care, Management, Training, Job, Treat, Listen, Program, Respect, Invest, Reward, Trust, Community
6	Business and innovation	Change, Product, Leadership, Focus, Business, Organization, Leader, Market, Politic, Culture, Talent, Impact, Decision, Term, Innovation
7	Hiring and promotion	Manager, Year, Management, Hire, Level, Promotion, Position, Promote, Department, Raise, HR, Performance, worker, base, Layoff
8	Development and growth	Company, Opportunity, Growth, Learn, Career, grow, Technology, Challenge, Role, Tech, Industry, Culture, Experience, Area, Provide

From observing the analysis, the companies seems to by spread widely amongst the 9 topics, but with a few patterns of interest. To the right side of the plot, the companies: *Apple*, *Tesla*, *Uber*, and *Amazon* are situated close to the topic *Time and scheduling*, which means this topic was frequently discussed in these companies. Another point of interest can be seen on the top middle part of the plot, where *Hiring and promotion* is

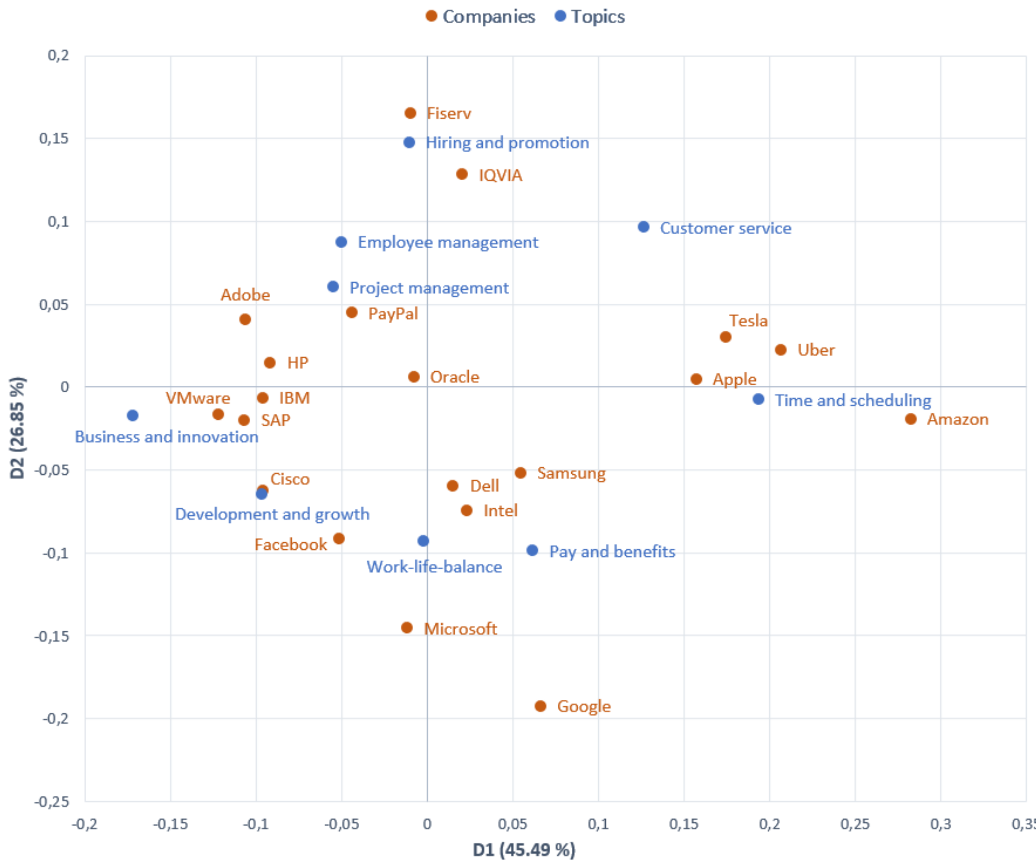

FIGURE 3. Correspondence analysis between corporate topics and companies

situated close to the companies: *Fiserv* and *IQVIA*. Companies like *Google* and *Microsoft* seems to be more isolated from topics compared to other companies, but their closest topics are *Work-life-balance* and *Pay and benefits*. Several other points of interest can be observed from the plot on figure 3, but these are some of the more noticeable ones worth mentioning.

5.3 Sentiment analysis

For extracting sentiments of the corporate topics, lexicon sentiment analyzers were tested on the Glassdoor reviews to give an indication of their accuracy on this kind of data.

As the Glassdoor reviews are already seperated into pros and cons, extracting positive and negative feelings and opinions about the data seems obvious. But by examining the individual review sentences, problems are revealed, e.g. it's not always the case that the positive statements are written in the pros section, and vice versa for the negative statements. Given the example sentences written in cons: ”This job has a high level of stress. Overall great experience though” tells both something negative about the job (high level of stress) and something positive about the job (great experience). If these sentences are written in the cons section of the review, both statements will be considered negative. Automatically processing the reviews one by one and assuming only the positive statements are mentioned in the pros and only the negative statements are written in the cons might therefore be a slightly inaccurate approach.

To deal with this problem, the option of using a sentiment analyzer was explored by testing two popular lexicon sentiment analyzers on 2.000 Glassdoor reviews. As the length of reviews varies, they were broken down into sentences by splitting at every punctuation resulting in a total of 5.075 sentences to be classified. Vader and Textblob are two popular textual lexicon sentiment analyzers chosen for this test, with Vader performing especially well on social media content [7]. For this test, an analyzer makes a correct prediction, if a sentence has been predicted with a polarity corresponding to the section to which it belongs to. For example, if the sentence was predicted with a negative polarity and belonged to the cons review section, then the prediction was correct.

Results of the tests in table 2 show improved accuracy with

TABLE 2. Lexicon sentiment analysis on 2000 Glassdoor reviews consisting of a total of 5.075 sentences

Analyzer	Accuracy	True positives	False positives	True negatives	False negatives
Vader	0.844	2413	726	1844	62
Textblob	0.808	2419	914	1656	56

the choice of the Vader lexicon sentiment analyzer compared to Textblob, which is also backed up by previous research [7]. Both analyzers perform poorly in predicting the cons and perform similarly in predicting the pros. With these results in mind, a lexicon analyzer will not be used for predicting the cons, as they are simply too inaccurate. Observing the number of false negatives, on the other hand, the accuracy was much better with a total of only 62 and 56 pros predicted with a negative sentiment. To access the amount of miss-placed pros (a con written in pros), the 62 sentences were manually checked to see if they belong in the pros. Out of the 62 wrongly predicted pros sentences, 14 of them did not belong in the pros. The other 48 sentences were either unclear as to which sentiment they should have, or simply just a wrong prediction by the analyzer. To handle the problem of miss-placed sentences it was decided that all the sentences in pros predicted with negative sentiment were discarded for sentiment analysis of the corporate topics. The Vader sentiment analyzer was therefore used exclusively on the positive review sections to filter off pros predicted with a negative sentiment

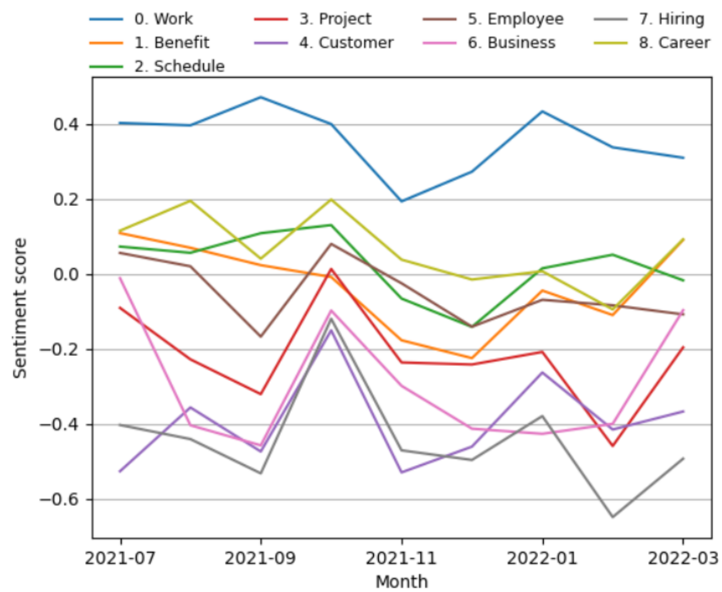

FIGURE 4. Time-line sentiment analysis of the 9 corporate topics for the company *IBM*

5.4 Topic sentiment analysis

Timeline analysis of the topics was conducted to showcase how the discussion of topics has changed in terms of sentiment over time in the companies. To calculate the sentiment of a topic, the following equation was used:

$$Sen(T_t, M_m) = \frac{\sum_{i=0}^{n}(P_{t,m,i} - C_{t,m,i})}{\sum_{i=0}^{n}(P_{t,m,i} + C_{t,m,i})} \tag{1}$$

where $Sen(T_t, M_m)$ is the sentiment score calculated for a topic T_t in a month M_n. $P_{t,m,i}$ is the number of pros associated with the t-th topic in the m'th month, in the i-th review-sentence, and $C_{t,m,i}$ is the number of cons associated with t-th topic in the m-th month, in the i-th review-sentence. The result of equation 1 shows that the sentiment of a topic will be closer to 1 the more positive it is, and closer to -1 the more negative it is. The results of mapping the topic sentiments for the company *IBM* can be seen on the line-chart of figure 4. By observing the chart, topic 0 about *Work-life-balance*, generally seems to be a positively loaded topic in this company, and on the other end topics 4 and 7 about *customer service*, and *hiring and promotion* seems to be more negatively loaded.

To assess whether the feelings about the corporate topics are similar for large companies in general, the same 20 large tech companies used for topic modelling were used for the topic sentiment analysis. The results of this analysis showing in figure 5 looks quite similar to figure 4, with the major difference being topic 1 about *pay and benefits* rising to the top in positivity alongside topic 0. This indicates a high average of employees speaking positively about *pay and benefits* in large tech companies through the past year. Comparing with the graph in figure 4 of the *IBM* company, *pay and benefits* seems to be a corporate topic to be improved on in this specific company, as the topic is significantly below the average sentiment score for this topic.

These graphs also seem to agree with with generally the best and worst topics in terms of sentiment for the 20 companies, as showcased in figure 6. Here 11 out of the 20 companies had *work-life-balance* as the topic with the highest sentiment and the

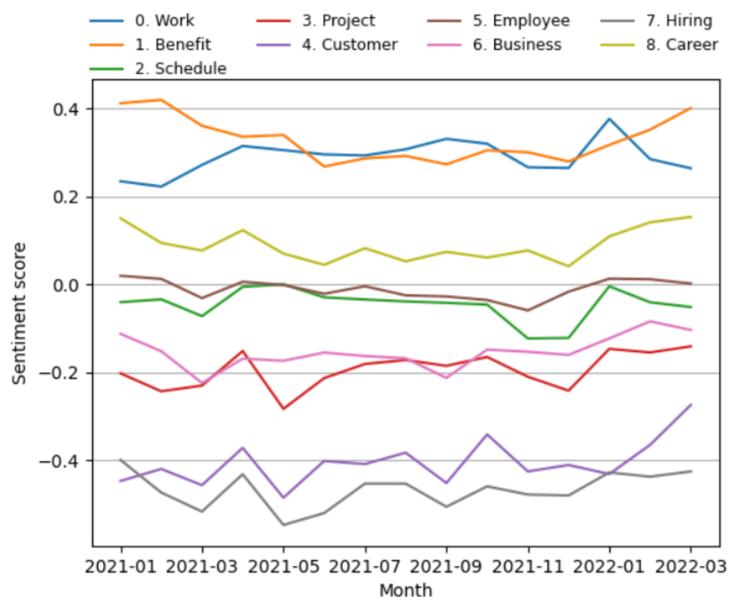

FIGURE 5. Time-line sentiment analysis of the 9 corporate topics for 20 large tech companies

other 9 companies had *Pay and benefits* as the topic with the highest sentiment. For the worst topics in terms of negative sentiment, *Hiring and promotion* was undoubtedly the most negative of them with 18 out of 20 companies scoring the lowest on this topic, followed by topic 4 *Customer service* scoring the lowest in 2 companies.

The results of the sentiment analysis seem to comply with both positive and negative topics identified from employee reviews in previous studies. A study identified the topic *Compensation/benefits* as being the most dominant from the positive feedback, which matches the high positivity of topic *Pay and Benefits* [26]. In addition, *recruitment* was an identified topic from the negative feedback, which explains why the similar topic of *Hiring and Promotion* has a negative sentiment in figures 4 and 5. This seems to be a trend in other text analysis studies as well [18].

6 DISCUSSION

Through multiple analyses of employee experience reviews on the Glassdoor platform, several findings have been discovered. Hidden corporate topics were identified in a large set of unstructured review data using LDA topic modelling. These topics were then utilized for analyses on a company level to reveal the dominance and sentiment of topics in individual companies.

6.1 Topic assessment

The corporate topics identified through topic modelling provide us with information about what kind of topics are discussed in the largest tech companies today. By comparing the corporate topics with topics found in existing Human Resource Management (HRM) related literature, some resemblance was found, which may indicate that these are common topics concerning the majority of employees [18][26][23]. Compared to a previous study identifying job 30 satisfaction factors, the corporate topics are somewhat similar to the factors identified [18]. Examining the factors closely revealed an overlap between the satisfaction factors and the topics. A factor like *Human resource* was also related to promotion and position, just like topic 7 *Hiring and promotion* in table 1, and the factor *Self-development* was very similar to the topic *Development and growth*, as it also included the keywords advancement, opportunity, experience, and growth. Another study studying employee topics in online reviews also find similar topics, except for a few topics such as *Hiring and promotion* [26]. The inconsistency between the number of topics and factors identified can be due to several parameters, such as the chosen topic modelling algorithm, the parameters used for running the algorithm, or the collection of review data.

The major differences in the topics identified in this study were:

topic 5: *Employee management*
topic 6: *Business and innovation*

These two topics were not clearly present in previous studies, as their associated keywords didn't match well with any topics. Taking a closer look into topic 5 of *Employee management*, compared to other topics, it appears to be very broad topic concerning several subcategories related to managing employees in a workplace, e.g. *employee care, training, trust* and *respect*. Looking into topic 6 of *Business and innovation*, the keywords also seem to be covering a wide range of smaller topics, such as innovation, leadership, business, and more, which made it difficult labelling this topic with a meaningful title. Choosing another topic modelling algorithm or choosing other parameters for the algorithm might have revealed those categories as topics. Even though some topics could be improved on, the topics identified covers many areas concerning the employee, and even extending with additional topics not found in previous studies.

6.2 Job satisfaction and topic sentiment

Observing the sentiment of the topics revealed a pattern across the 20 companies. Topics *0. Work-life-balance* and *1. Pay and benefits* scored the highest in terms of sentiment across all companies. Analyzing the results from a job satisfaction point of view, a study has revealed that fringe benefits (nonwage payment) are significant and positive determinants of job satisfaction, which could be a reason why *Pay and benefits* is such a hot

FIGURE 6. Best and worst topics in terms of sentiment for the 20 large tech companies

topic in the positive reviews [4]. Pay, on the other hand, a study has found to also bear a positive, but quite modest, relationship to job satisfaction [17]. For topic *0. Work-life-balance* a study has been conducted that shows the important role that work-life balance plays in promoting better life and job satisfaction for employees across different cultures, which can be a reason for its positive sentiment amongst the employees [14].

Looking at the worst topics, topic *7. Hiring and promotion* scored the lowest in terms of sentiment across all the 20 companies. This topic consist of frequent terms such as managers and management, which most likely refers broadly to the people controlling the operations of the company or the HR team responsible for recruiting, developing, and maintaining employees in the company. It makes sense why this topic scored the lowest across all companies, as the negative comments about a company are mostly directed towards the management responsible for running the company. In addition, it is the HR management team's job to keep employees engaged in their jobs, which studies have also shown to be positively related to job satisfaction [28] [19].

6.3 Implications

The results of this study are valuable for companies if they have any interests in understanding the topics employees speak about in their companies today, and maybe even utilize the data available from multiple online platforms, in addition to Glassdoor.

The knowledge about the topics might be useful for HCI researchers when designing interactive systems utilizing employee reviews, such as the tool Enterprise Social Pulse highlighted earlier [25]. Knowledge about the highly positive and negative topics, might find relevance in the process of designing interactive systems targeted large tech companies. HCI designers should consider what the employees value in a job, when designing interfaces, and put more emphasis on these values/topics, which this research might help them getting insight into. Looking from the employee's perspective, being aware of which topics are mentioned positively and negatively in a company can help guide them in the right direction when seeking a company to work at. Results of this study can therefore assist job applicants in finding a job.

The results might also find relevance in other types of research, including business and economics for learning about what topics employees care about and what business managers should be concerned about in their business. Knowledge about sentiment of the corporate topics in companies can provide business managers with an overview of positive and negative topics in their company, which can help them in findings areas of improvement. Measuring the change in sentiments of the topics over time as showcased in figures 4 and 5 by conducting a monthly or quarterly employee evaluation, could be an valuable procedure for companies to implement in order to identify areas of improvement.

6.4 Limitations and future work

This study was limited to analyzing large tech companies, as these companies had the most public available data. Other studies could focus on companies in other industries than the tech industry, which might reveal other topics that the employees speak of, which are particularly relevant for those industries. A problem with a lack of data might arise, which could potentially be solved by extracting data from other employment experience platforms, besides Glassdoor.

In this study a lexicon sentiment analyzer was used on the reviews in the pros section to filter off miss-placed cons. In future research, time could be spent on labelling the reviews with their ground truth sentiment and train a supervised machine learning language model on reviews for achieving better accuracy on both the pros and cons sections of the reviews.

Looking at other features of the review data, such as the role of the employee in the company writing the review, could be an interesting direction in a future study. In this case of analyzing tech companies, the majority of employees are software engineers followed by program managers, so investing what kind topics these different roles speak about could be interesting and valuable for businesses to know.

7 CONCLUSION

Through this study, more knowledge about workplace-related topics discussed by employees has been obtained. The topics identified from conducting topic modelling on reviews from tech companies have shown to be similar to topics identified in previous studies analyzing reviews from other companies on other employee and job experience platforms. These topics can therefore be valuable parameters for measuring employees' satisfaction in companies in general, which can potentially be useful knowledge both for companies and employees/job-seekers. In addition, findings from the sentiment analysis reveal that the topics about *work-life-balance* and *pay and benefits* were standing out in terms of positivity, and the topic *hiring and promotion* was standing out in terms of negativity. In future work, other features of the review data can be explored or companies from other industries can be explored and analyzed to find potential differences in the topics identified and the sentiment polarity of the topics.

References

[1] Ashraf Abdul et al. "Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–18. ISBN: 9781450356206. URL: https://doi.org/10.1145/3173574.3174156.

[2] Dinislam Abdulgalimov et al. "Designing for Employee Voice". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–13. ISBN: 9781450367080. URL: https://doi.org/10.1145/3313831.3376284.

[3] Faiz Ahamad. "Impact of word-of-mouth, job attributes and relationship strength on employer attractiveness". In: *Management Research Review* (2019).

[4] Benjamin Artz. "Fringe benefits and job satisfaction". In: *International journal of manpower* (2010).

[5] Niels van Berkel et al. "A systematic assessment of national artificial intelligence policies: Perspectives from the Nordics and beyond". In: *Proceedings of the 11th nordic conference on human-computer interaction: shaping experiences, Shaping Society*. 2020, pp. 1–12.

[6] David M Blei and Michael I Jordan. "Variational inference for Dirichlet process mixtures". In: *Bayesian analysis* 1.1 (2006), pp. 121–143.

[7] Venkateswarlu Bonta and Nandhini Kumaresh2and N Janardhan. "A comprehensive study on lexicon based approaches for sentiment analysis". In: *Asian Journal of Computer Science and Technology* 8.S2 (2019), pp. 1–6.

[8] Jason Chuang, Christopher D Manning, and Jeffrey Heer. "Termite: Visualization techniques for assessing textual topic models". In: *Proceedings of the international working conference on advanced visual interfaces*. 2012, pp. 74–77.

[9] Chedia Dhaoui, Cynthia M Webster, and Lay Peng Tan. "Social media sentiment analysis: lexicon versus machine learning". In: *Journal of Consumer Marketing* (2017).

[10] Svenja Dube and Chenqi Zhu. "The Disciplinary Effect of Social Media: Evidence from Firms' Responses to Glassdoor Reviews". In: *Journal of Accounting Research* 59.5 (2021), pp. 1783–1825.

[11] Robert S Feldman, James A Forrest, and Benjamin R Happ. "Self-presentation and verbal deception: Do self-presenters lie more?" In: *Basic and applied social psychology* 24.2 (2002), pp. 163–170.

[12] Glassdoor. *About us*. URL: https://www.glassdoor.com/about-us/ (visited on 04/11/2022).

[13] Glassdoor. *Fraudulent reviews*. Oct. 26, 2021. URL: https://help.glassdoor.com/s/article/Fraudulent-reviews?language=en_US (visited on 04/11/2022).

[14] Jarrod M Haar et al. "Outcomes of work–life balance on job satisfaction, life satisfaction and mental health: A study across seven cultures". In: *Journal of Vocational Behavior* 85.3 (2014), pp. 361–373.

[15] Xu Han et al. "Pre-trained models: Past, present and future". In: *AI Open* 2 (2021), pp. 225–250. ISSN: 2666-6510. DOI: https://doi.org/10.1016/j.aiopen.2021.08.002. URL: https://www.sciencedirect.com/science/article/pii/S2666651021000231.

[16] Minjie Huang et al. "Family firms, employee satisfaction, and corporate performance". In: *Journal of Corporate Finance* 34 (2015), pp. 108–127.

[17] Timothy A Judge et al. "The relationship between pay and job satisfaction: A meta-analysis of the literature". In: *Journal of vocational behavior* 77.2 (2010), pp. 157–167.

[18] Yeonjae Jung and Yongmoo Suh. "Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews". In: *Decision Support Systems* 123 (2019), p. 113074.

[19] Soonhee Kim. "Participative management and job satisfaction: Lessons for management leadership". In: *Public administration review* 62.2 (2002), pp. 231–241.

[20] Jongseo Lee and Juyoung Kang. "A study on job satisfaction factors in retention and turnover groups using dominance analysis and lda topic modeling with employee reviews on glassdoor. com". In: (2017).

[21] Ning Luo, Yilu Zhou, and John Shon. "Employee satisfaction and corporate performance: Mining employee reviews on glassdoor. com". In: (2016).

[22] Daniel Maier et al. "Applying LDA topic modeling in communication research: Toward a valid and reliable methodology". In: *Communication Methods and Measures* 12.2-3 (2018), pp. 93–118.

[23] Andy Moniz and Franciska de Jong. "Sentiment Analysis and the Impact of Employee Satisfaction on Firm Earnings". In: *Advances in Information Retrieval*. Ed. by Maarten de Rijke et al. Cham: Springer International Publishing, 2014, pp. 519–527. ISBN: 978-3-319-06028-6.

[24] Selva Prabhakaran. *Topic Modeling with Gensim (Python)*. Mar. 26, 2018. URL: https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python (visited on 03/14/2022).

[25] N Sadat Shami et al. "Understanding employee social media chatter with enterprise social pulse". In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 2014, pp. 379–392.

[26] Efthymia Symitsi et al. "The informational value of employee online reviews". In: *European Journal of Operational Research* 288.2 (2021), pp. 605–619.

[27] Verizon. *90% of Consumers Say Online Reviews Impact Buying Decisions . . . And Other Hot Topics*. Mar. 7, 2022. URL: https://www.verizon.com/business/small-business-essentials/resources/90-consumers-online-reviews-impact-buying-decisions-other-211526956/ (visited on 03/07/2022).

[28] Anton Vorina, Miro Simonič, and Maria Vlasova. "An analysis of the relationship between job satisfaction and employee engagement". In: *Economic themes* 55.2 (2017), pp. 243–262.

[29] Xiong Zhang et al. "Teddy: A System for Interactive Review Analysis". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–13. ISBN: 9781450367080. URL: https://doi.org/10.1145/3313831.3376235.

[30] Yabing Zhao, Xun Xu, and Mingshu Wang. "Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews". In: *International Journal of Hospitality Management* 76 (2019), pp. 111–121.