Social Media-Based Opinion Retrieval for Product Analysis Using Multi-Task Deep Neural Networks

Necip Gozuacik [a*], C. Okan Sakar [a], Sercan Ozcan [b, c, d]

[a] Department of Computer Engineering, Bahcesehir University, Istanbul, Turkey

[b] Portsmouth Business School, University of Portsmouth, Portsmouth, UK

[c] Department of Engineering Management, Bahcesehir University, Istanbul, Turkey

[d] National Research University Higher School of Economics (Russian Federation), Moscow, Russia

e-mails: {necip.gozuacik@bahcesehir.edu.tr; okan.sakar@eng.bau.edu.tr;sercan.ozcan@port.ac.uk}

Abstract

Social media platforms are considered one of the most effective intermediaries for companies to interact with consumers. Social media-based decision support systems for the marketing domain are highly developed, but product development and innovation-oriented studies remain limited. This study offers a novel approach which utilises opinion retrieval theme along with sentiment analysis to support the decision-making process for product analysis and development. To achieve this aim, we propose an end-to-end social media-based opinion retrieval system and utilise machine learning and natural language processing techniques. Google Glass is chosen as a use-case as this product was unable to achieve its commercial targets despite its superior technological offerings. We design a multi-task deep neural network architecture for the training of sentiment prediction and opinion detection tasks. We first divide the tweets containing certain useful opinions and suggestions into two categories based on their sentiment labels. The negative tweets are analysed to identify product-related concerns, whereas the positive and neutral tweets are used to extract innovative ideas and identify new use cases for product development. We visualise and interpret the clusters of keywords extracted from each sentiment label group . Apart from methodological contributions, this study offers practical contributions for the next generations of smart glasses.

Keywords: deep learning, feedback retrieval, natural language processing, opinion mining, sentiment analysis, text analytics

1. Introduction

For innovations and new product development (NPD) approaches, customer involvement is either at the heart of the developments or their involvement is considered a supportive mechanism at different phases of various decision-making processes. Some academics believe that customers are the key for a technology to be accepted (Chang & Taylor, 2016), whereas some believe customer opinion is not so critical, especially for radical/discontinuous innovations or where there is an asymmetry between company and customer knowledge (Trott, 2001). There are a number of reasons for the negative opinion towards customer-centric decision support systems in NPD. Firstly, it is costly, especially if third party based surveys and product analysis are considered. Secondly, it is believed that customers may not possess adequate information especially in high tech environments. Finally, due to the traditional customer-centric product analysis (i.e. surveys, focus groups, product trials), the gathered information is considered to be too static for fast-paced environments.

Innovative methods using social media data in customer-centric product analysis approaches minimise or eliminate the above mentioned problems. Social-media data are used in a wide range of areas from digital marketing to customer satisfaction analysis as social media-based operations are low-cost, dynamic, targeted and informative due to the new possibilities presented by the 4Vs of big data and advanced machine learning approaches (Saura, 2020). Many failed innovations also indicate the need for advanced product analysis systems.

In recent studies, social media data have been used for product analysis for different reasons, such as the examination of customer reaction to the launch of new products or technologies (Lipizzi, 2015; Nuortimo & Ha¨rk¨onen, 2018), the assessment of a product's competitive advantages (Liu et al., 2019; Jiang et al., 2019) and assistance with the development of next-generation products (Li et al., 2014; Hou et al., 2019; Mirtalaie et al., 2017). The majority of these studies build multi-stage models using different combinations of sentiment analysis, topic modelling, natural language processing, named entity recognition, and term extraction techniques to analyse user generated content (Saura & Bennett, 2019). However, these approaches are limited in their ability to offer an end-to-end social media-based feedback mechanism for decision support systems in innovations and NPDs. Examining the literature specific to the product development and ideation steps, there are only a few studies in which social media data are used as a part of the NPD process.

The major weakness in the literature is the design of the methodology and its linkage to the product development process. The current literature fails to offer strong examples of models that can retrieve the product development and innovation-oriented opinion of customers. Many studies in the literature are designed to retrieve customer sentiment regarding the product in general or existing features of the product (Hasson et al., 2019; Ibrahim & Wang, 2019; Liu et al., 2019). Studies where only sentiment analysis is performed for product analysis fail to provide detailed information to understand why the customers do or do not like the product. However, the studies that perform aspect-based analysis for feature-based product analysis do not offer innovative ideas that can be used for product development. There are only a few studies that aim to generate innovative ideas related to the specific features of the product via sentiment analysis of the customer comments shared on social media (Li et al., 2014; Mirtalaie et al., 2017). Such studies propose frameworks to identify sentiment and customer opinion to improve the next version of products or to introduce new types of products based on specific customer suggestions that focus on groups of product features.

Considering this notable gap in the literature, in this study we propose a framework that uses social media data to reveal the reasons for a failed innovative product from the customer perspective and to suggest new use cases and innovative ideas for product development. We selected the case of Google Glass as this product had a failed launch despite the technological offerings and the level of innovations it featured. Google Glass, a smart glass brand, is an optical head-mounted device that can be controlled via the voice and motions of its user and can assist the user by displaying information on its screen. It is regarded as a useful case that offers rich data considering both positive and negative customer feedback. As our aim is to analyse product-related concerns and extract innovative ideas for product development rather than performing a quantitative customer satisfaction analysis, we integrated an opinion detection module into our system to clean the dataset by removing the user generated content which does not include useful feedback or suggestion but only expresses satisfaction/dissatisfaction about the reference product. To achieve our objectives, we propose the design of a multi-task deep neural networks (DNN)-based framework that learns sentiment analysis and opinion detection tasks jointly. The main motivation of designing a multi-task model is to improve the detection ability of the learning tasks and retrieve sentiment-based customer opinions with a single model. The practical and methodological objectives of our study are as follows:

- Propose a general framework that analyses product feedback
- Utilise various NLP techniques for effective text representation
- Propose a DNN-based model for the supervised learning tasks required for opinion retrieval
- Provide suggestions for the next generations of smart glasses

The key contributions of this study are: 1) the proposal of an end-to-end product-related opinion retrieval and analysis system, 2) the design of a multi-task DNN architecture predicting two outputs simultaneously for better generalisation of the machine learning model, and 3) the illustration of a proposed system for product analysis on the Google Glass case with both negative and positive feedback.

The remainder of this paper is organised as follows. We review the relevant literature in Section 2. In Section 3, we provide the general framework of the methodology, the details of the used methods and the DNN architectures designed to construct the overall. Section 4 presents the experimental and practical results of our study, and we discuss the results in Section 5. Finally, in Section 6 we conclude the study by presenting its implications, contributions and limitations.

2. Related Works

Rapid advancements in Internet of Services, Web 2.0, and social media have led to potential applications and better interactions with consumers. Product development and innovation activities can now be supported by knowledge retrieval applications. Consumers share their opinions in public domains, and their input can be used as the source of ideas for new generation products, services and business models. Two innovative approaches can be implemented to retrieve information from the crowd: 1) crowdsourcing models, where consumers are led to a certain task and information is retrieved (Djelassi & Decoopman, 2013; Schemmann et al., 2016), and 2) idea or opinion mining, where information is retrieved from generic public data with a specific approach (Lipizzi, 2015; Li et al., 2014). Both approaches may have advantages and disadvantages. Crowdsourcing is a better approach if the required information is scarce and lacking in detail, in cases where a company requires feedback for a certain stage of product development (i.e. prototyping). However, crowdsourcing-based approaches could be costly, and continuously implementing such approaches is more difficult. For idea/opinion mining approaches, the optimal approach is to retrieve information from big data to ensure the continuity and efficiency of the process.

Table 1 summarises the review of the relevant studies, considering their research focus and methodological approaches. Accordingly, the studies are grouped as: 1) company, product, or sector-specific customer feedback analysis using social media, 2) product review analysis using product-specific online marketplace or product forums, and 3) product review analysis for product development from social media data. The analysis of the literature studies summarised in Table 1 shows context and methodology-specific gaps and weaknesses. Most of the relevant studies performed a general sentimental analysis based on the sentiments of the content generated by the users without presenting specific feedback related to the product features. For example, a customer comment may consist of an overall positive sentiment (i.e. "I love this product") but may lack product- or feature-specific useful feedback (i.e. "I love this product's camera quality"). For product development purposes, it is more valuable to identify customer feedback that has an opinion to improve the current offering or has suggestions for new types of products. Table 1 shows that there are some studies that utilise social media data for product development. These studies focus on finding new features that can be integrated to the next version of the products. However, to the best of our knowledge, there is no prior study that proposes an end-to-end framework that uses social media data to reveal the reasons for a failed innovative product from the customer perspective and to suggest new use cases and innovative ideas for product development. This gap in the literature can be addressed by implementing sentiment analysis and opinion detection modules together to identify customer comments with product-specific feedback and suggestions. Therefore, in this study we focus on developing a multi-task learning-based opinion retrieval system that can analyse social media data for product analysis and development.

Previous studies illustrate the application of text mining in different phases of decision-making in product and technology development. From these, it is evident that social media data are used for to examine emerging technologies, product competitiveness or life cycle analysis and overall opinion gathering. As the increasing amount of social media data provides opportunities for building decision support systems regarding product or service improvement, some of these studies focus on processing the social media content for product-related feedback analysis (Botchway et al., 2019; Hasson et al., 2019; Jiang et al., 2011; Mirtalaie et al., 2017; Saura & Bennett, 2019).

Micro-blogging services are one of the main data sources for companies to gather feedback from potential customers and improve their services and products in the light of this analysis (Araque et al., 2017). Twitter is an extremely popular micro-blogging service and has been used widely by academics and companies as an important source of customer feedback. Hasson et al. (2019) concluded that Twitter data is a less costly alternative to customer satisfaction surveys. They used Twitter data belonging to a large biotechnology company and compared the customer feedback gathered from social media with the survey responses in terms of content and value. For this purpose, they first applied pre-processing operations on the Twitter data, used another already labelled sentiment dataset to train a classifier, applied the obtained sentiment analysis model on their original dataset to classify each tweet with a sentiment label, and constructed a hash-tag co-occurrence matrix for the top six products and/or services of the company. Finally, they retrieved a sample of survey data from the company and comparatively analysed the sentiment analysis and survey results. The findings showed that some products/services that are not the focus of the company were discussed more than some of their ground-breaking work in both negative and positive tweets, showing that the company and customers may differ in their interests in the products and services. Their results also revealed that analysis on social media data can provide

more dynamic information and new insights about customer satisfaction compared to customer satisfaction surveys.
In a recent study, Ibrahim and Wang (2019) analysed Twitter data related to five leading UK online retailers by combining topic modelling, sentiment analysis and network analysis techniques, with the aim of identifying the main concerns of the customers. Different from our study, they did not experiment with different DNN-based models for sentiment analysis but instead used a sentiment analysis tool to identify the negative tweets. They then specifically highlighted the services and products perceived negatively by their customers. They used Latent Dirichlet Allocation (LDA) for topic modelling and listed the most important eight topics from the negative tweets of the customers. Rathan et al. (2018) proposed a Twitter-specific sentiment analysis model using features such as emoji detection, emoticon detection and spelling correction. The model was applied to the "Smartphone" domain. The authors presented an aspect-based analysis considering different attributes of smartphones such as battery, camera and display.

Liu et al. (2019) applied, similar to our study, a two-step supervised learning approach to analyse social media text data. The main goal of the study was to assess competing products from the perspective of customers. The first step of their approach was sentiment analysis performed using a domain-specific sentiment lexicon. They subsequently built a classification model to detect the comparative user-generated content. Finally, they presented the most important advantages of the target product compared to its competitors. The results indicated that sentiment analysis plays a key role in analysing customer feedback from different perspectives. Sun et al. (2019) proposed a machine learning-based framework to identify the degree of a review's informativeness, using data from an online electronic marketplace in China. As stated in this study, sentiment analysis of user-generated content has an important role in determining the informativeness of a review. Therefore, in addition to the binary classification model built to determine whether the tweet includes a feedback or suggestion, we incorporate sentiment analysis into our framework to analyse positive and negative reviews for aspect-based opinion retrieval and analysis.

In another study, Rane and Kumar (2018) analysed 14,640 tweets associated with six major US airlines. They first manually labelled all tweets. Subsequently, after text pre-processing operations, they used a word embedding technique, Doc2vec, and seven different classifiers to classify the tweets as negative, neutral or positive. They presented a comparative analysis of the classifiers' performances and identified the common terms which appeared in the negative feedback of the customers. Similarly, Botchway et al. (2019) applied sentiment analysis techniques to Twitter data to analyse the customer feedback related to the products and services of one of the largest bank in Europe. They used a rule-based sentiment analysis tool specifically designed to process social media data. They also presented top hashtags which appeared in the customers' tweets. Basiri et al. (2020) performed sentiment analysis in the medical domain by applying various deep learning models to the drug reviews shared in Drugs.com. They compared their proposed deep fusion models with the existing traditional and deep learning-based techniques.

Table 1: Summary of the related works

Group of Studies	Authors	Research Aims	Methods	In comparison with the previous studies
Company, product or sector- specific customer feedback analysis using social media	Hasson et al. (2019) Rane and Kumar (2018) Botchway et al. (2019) Ibrahim and Wang (2019), Mai and Le (2020), Lipizzi et al. (2015)	To perform sentiment analysis and determine the most common customer complaints To identify customers' primary topics of concern To compare customer satisfaction surveys with social media-based customer feedback To perform sentiment analysis on tweets related to a specific company and determine the most popular hashtags about the products and brands of the related company To analyse early reactions of customers for new products	Supervised learning techniques combined with natural language processing, topic modelling and network analysis modules	Do not offer a detailed sentiment-based analysis for product development with a word network that can be utilised by decision makers, but mostly present only sentiment category count related to each product/service of the company Do not have an opinion detection module to clean the dataset by removing the entries that do not contain useful feedback Mostly focus on only analysing the negative tweets to determine the products and services that are perceived negatively by the customers
Product review analysis using a product-specific online marketplace, or product forums	Liu et al. (2019) Sun et al. (2019) Basiri et al. (2020), Eldin et al. (2020), Jiang et al. (2019)	To analyse product-specific user comments retrieved from a product specific platform To assess the competing products from the customer perspective To identify the degree of a review's informativeness	Traditional supervised learning and deep learning techniques combined with domain-specific lexicon generation approaches and named entity recognition methods	Less general solutions are offered as customer feedback or opinion retrieved from product or service specific online marketplace or product forums are analysed Mostly domain-specific features such as membership status of the user, availability of a public user image, etc. are used in addition to the review text Do not offer innovation-oriented ideas for product development Do not offer an automated end-to-end approach from data retrieval to the extraction of word network of customer feedback that can be utilised for specific product development
Product review analysis for product development from social media data	Mirtalaie et al. (2017), Li et al. (2014)	To propose a decision support framework to retrieve product-specific innovative ideas from social media	Similarity metrics, rule-based approach or a knowledge base for sentiment analysis, word clouds, various metrics to evaluate the reviews (e.g. influence score, expertise score, review rating)	Machine learning/deep learning techniques are not used Performs cross-domain analysis to extract innovative ideas for the reference product

Mai and Le (2020) used deep learning techniques to examine user comments regarding smartphone products shared on a social media platform. Similar to our study, they built a multi-task learning framework to analyse user comments. Specifically, while we perform multi-task learning for opinion detection and sentiment analysis tasks, they train sentence-level and aspect-level sentiment analysis tasks jointly to design a more generalisable sentiment analysis model. Different from our study, their framework presents the proportion of positive and negative comments for specific attributes of a product rather than generating innovative ideas and new use cases. Similar to the main goal of our study, Mirtalaie et al. (2017) aimed to generate innovative ideas from user reviews retrieved from social media. They proposed a framework consisting of three stages. The proposed framework identifies related products with the reference product with a cross-domain analysis and identifies new features that can be integrated into the future versions of the reference product. The studies applying sentiment analysis in various domains show that is a key technique for building managerial decision-making tools for product and service improvement.

Having reviewed the literature and demonstrated the relevance of our study in Table 1, it is evident that there are limited studies where such approaches are being implemented for the purpose of retrieving feedback and opinions for product development or innovation-oriented processes. Most studies are limited to offering direct solutions that can be used for product analysis or development but mostly present an overall quantitative assessment about the reference product. There are also methodological gaps in terms of combining deep learning and advanced word embedding techniques with the aim of retrieving useful product and service-related suggestions and feedback by benefiting from user generated content in social media.

3. Methodology

The primary motivation of this research is to introduce an end-to-end social media-based opinion retrieval system. In this section, we first outline the general framework of the system. Subsequently, we describe each step of the proposed framework and the dataset used in the experiments.

3.1. General Framework

The general framework of the proposed system consists of the following steps:

- Data Pre-processing
- Text Representation with Bag-of-Words (BoW) and Word Embedding Techniques
- Modelling for Opinion Retrieval
- Performance Evaluation and Prediction
- Co-occurrence Matrix-Based Visualisation

The methods and operations applied to implement each step of the general framework are shown in Figure 1. We provide the details for these methods in the following subsections.

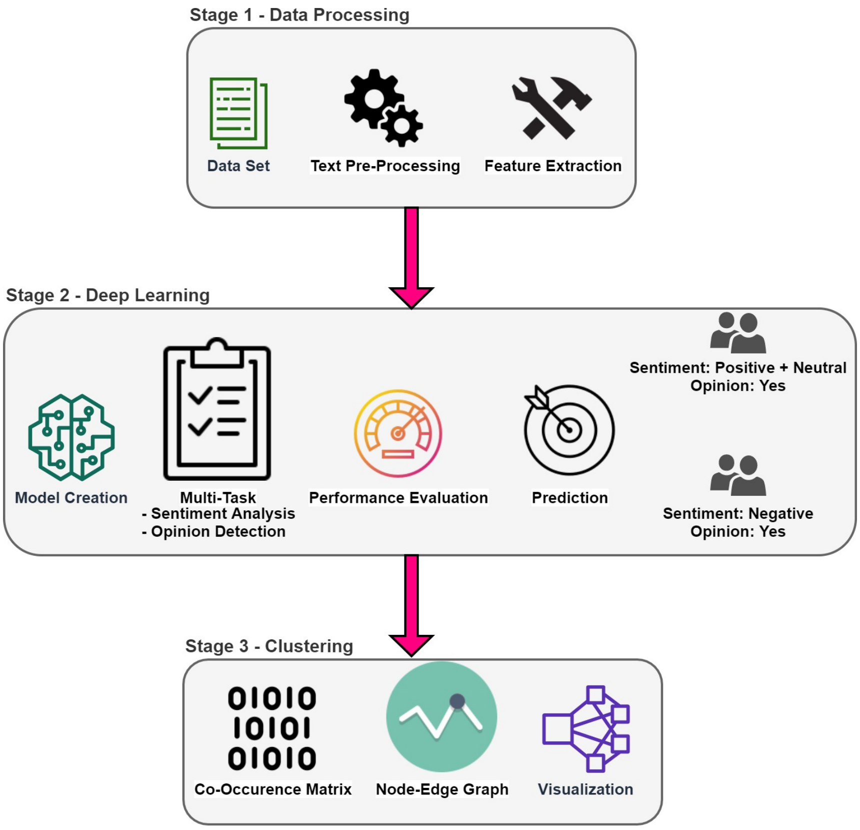

Figure 1: Methods and operations applied to implement the end-to-end opinion retrieval system

3.1.1. Data Pre-processing

For this study, we used one of the most popular micro-blogging services, Twitter, as the data source. As given in Section 2, Twitter data have recently been used in various academic studies with sentiment analysis for various purposes. After collecting the dataset consisting of the targeted tweets, we applied several pre-processing NLP operations. First, as tweets may include the hashtag character (#), which defines the topic on the Twitter platform, we removed it to expose the real meaning of the words. We also removed the addresses of the websites, numeric forms and account names starting with characters, followed by stop words that do not contribute to the semantic meaning of the tweets. After these operations, we manually labelled some of the tweets to be used in the training of the supervised learning algorithms required for opinion retrieval task. We provide detail of the data collection and labelling steps, along with the description of the dataset used in the experiments, in Section 3.2.

3.1.2. Text Representation

In this step, the aim was to represent the text data in a structured form to be processed by the DNN architectures built for sentiment prediction and opinion detection tasks. Initially, we created a vocabulary from the content of the tweets. We then represented each word with various traditional BoW and recent word embedding techniques.

The traditional BoW techniques used as baseline methods in our study are based on term frequency

(TF) and term frequency-inverse document frequency (TF-IDF) metrics (Skansi, 2018). In BoW representation, the text is described in terms of the word occurrences of each word in the vocabulary. We used TF and TF-IDF metrics to represent the occurrence of each word. In TF-based representation, each word is represented with its frequency in the given text, whereas TF-IDF-based representation not only uses the frequency of the word in the given text but also to what extent that word is common in the entire corpus (Igual & Segu´ı, 2017). The main drawback of BoW representation is that it disregards the semantic meaning and word orders and hence poorly represents the text.

Word embedding is an advanced text representation technique where words from the vocabulary are converted to vectors of real values in a low-dimensional space, depending on the size of the original vocabulary. This representation technique can capture the semantic and syntactic similarity between the words, and hence can represent their context text better (Goyal, 2018). These models are generally designed based on the use of neural network architectures. In this study, we used two word embedding techniques, Word2Vec and GloVe, to represent the tweets in our dataset.

The Word2Vec model is one of most utilised and transformed models in various NLP tasks (Ergen & Kozat, 2017; Kim et al., 2020; Li & Shah, 2017; Ren et al., 2019). The main aim in this method is to learn the word associations or vector representation of the words in a given set of texts. In the Word2Vec model, with the use of a two-layered neural network architecture, the words that have similar semantic meanings are expected to have word vectors with high similarity. The Word2Vec approach can be implemented using one of the two model architectures which are known as continuous BoW (CBoW) and skip-gram. Whereas in CBoW-based architecture the current word is predicted from a window of surrounding context words, in skip-gram-based architecture the surrounding window of context words are predicted from the current word. The main advantage of skip-gram is to incorporate the order of context words into the word embedding process, hence resulting in the better representation of infrequent words (Bhoir et al., 2017).

GloVe is another word embedding model proposed by Pennington et al. (2014). The underlying approach in GloVe differs from Word2Vec by giving more importance to the frequency of co-occurrences of the words in the given text. The principal aim of GloVe is to provide word representations by utilising the advantages of both co-occurrence matrix-based statistics and predictive models. The word vectors obtained after the training process are related to the probability of the co-occurrences of the words.

In this study, we used Word2Vec and GloVe to obtain word vectors for the sentiment prediction and opinion detection tasks. We used both CBoW and skip-gram implementations of Word2Vec with and without transfer learning in our experiments; the GloVe model is used only as transfer learning. In transfer learning, the main goal is to utilise a pre-trained model already built on a huge dataset for a different but similar task. In our study, the pre-trained Google News corpus word vector model consisting of 3 million 300-dimensional English word vectors and a pre-trained Twitter dataset were used with the Word2Vec and GloVe models (Rezaeinia et al., 2019).

3.1.3. Modelling for Opinion Retrieval

After representing the Twitter messages with the text representation methods given in Section

3.1.2, we constructed DNN-based architectures to apply the classification tasks to the extracted features. The opinion retrieval system built in this study is based on the use of two outputs for each Twitter message. The first output is sentiment label, which can be obtained by addressing a multi-class classification task including three classes which are negative, neutral and positive labels. The second output is a binary class value representing the existence of a feedback/suggestion/opinion in the related tweet.

In the model creation step, we followed two main DNN-based approaches for the abovementioned sentiment analysis and opinion detection tasks. In the first approach, we created two independent models for each of the learning tasks to obtain the sentiment and opinion outputs separately. The DNN architectures of the single-task learning models are shown in Figure 2. The input layer consists of the features obtained using the various word representation techniques described in Section 3.1.2. As is evident, each architecture has two hidden layers and two regularisation techniques, drop-out and batch normalisation, used after each hidden layer. In dropout, some of the randomly selected neurons are discarded in particular epochs during the learning process to improve the generalisation ability of the network (Srivastava et al., 2014). In batch normalisation, the normalisation process is applied for each batch as well as in the hidden layers. The main goal is to prevent the weights and outputs from getting extreme values during the learning process (Ioffe & Szegedy, 2015).

In the second approach, we designed a multi-task learning scheme to solve the sentiment prediction and opinion detection tasks simultaneously. Figure 3 shows one of the multi-task DNN architectures designed and tested in our study. We followed a hard-parameter sharing approach in which all hidden layers are shared between the two NLP tasks. Although the data source was the same for the tasks, jointly learning them with shared layers helped to reduce the risk of overfitting as well as the training time (Li et al., 2017; Park et al., 2019; Parwez et al., 2019). This way of learning, in which the network is forced to find (Li et al., 2017) a shared representation that can predict both tasks, can be considered another regularisation mechanism. As seen in Figure 3, first, we concatenated features from two different text representation techniques into a single vector. We then fed the obtained representation into a dense layer which has a lower number of hidden neurons than the concatenated input layer. The dense layer was followed by dropout and normalisation operations; the obtained representation was then fed into the second hidden layer which has a lower number of hidden neurons. After the second round of regularisation, we mapped the representation to the specific tasks.

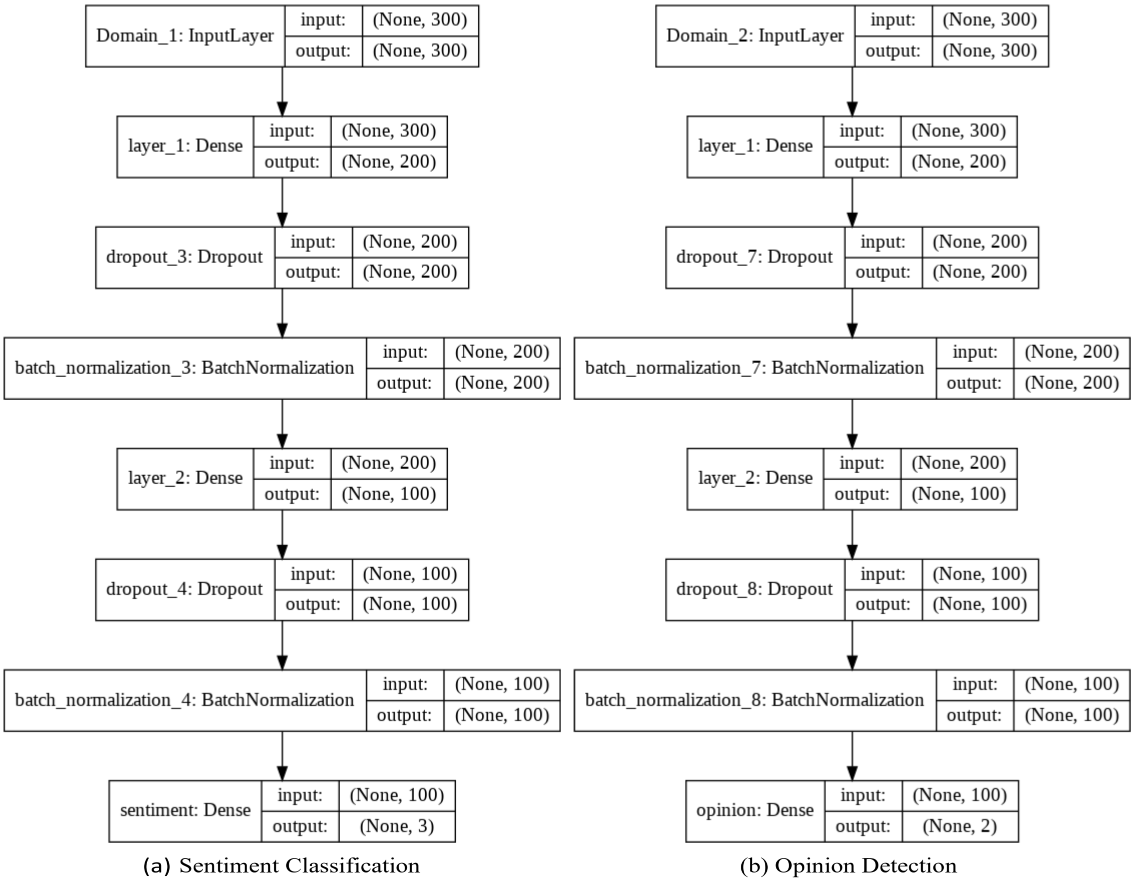

Figure 2: Single-task DNN architecture for (a) sentiment target (b) opinion target

3.1.4. Performance Evaluation and Prediction

We trained the single and multi-task learning architectures using the input space obtained with various word representation techniques. For model construction and evaluation, we initially partitioned the labelled dataset into training and test sets with 80%–20% ratios, respectively. and then divided the left-out training set into training and validation sets for hyper-parameter optimisation. The best model on the validation set was applied to the left-out test set. We repeated this procedure 10 times and performed statistical tests to validate the statistical significance of the comparative results. The results were evaluated in terms of F1 Score and accuracy metrics. Finally, we applied the best model to all the dataset including unlabelled tweets and obtained the sentiment and opinion predictions for use in the ultimate sentiment-based opinion retrieval task.

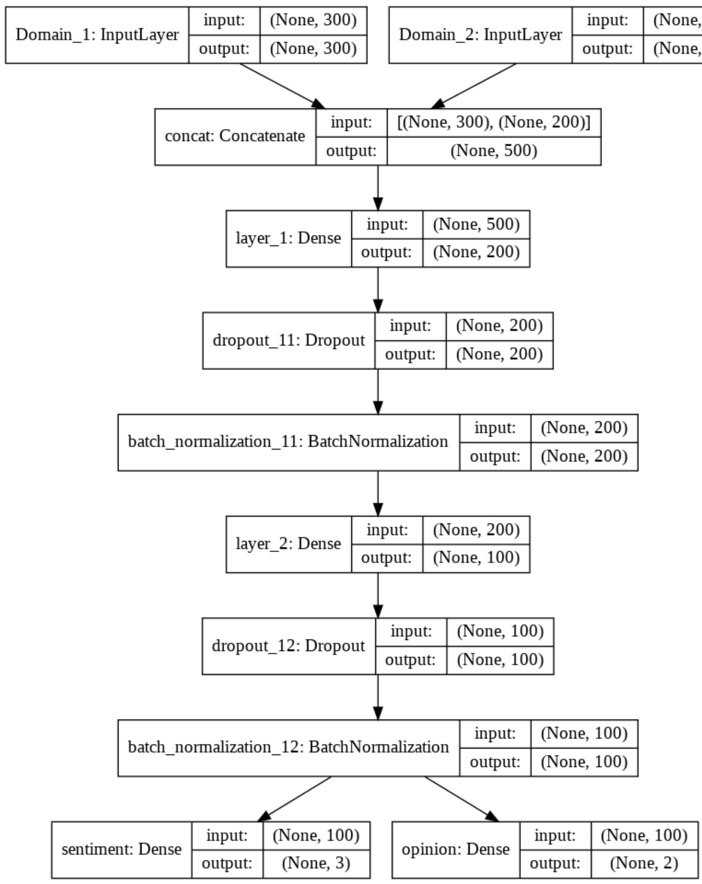

Figure 3: Multi-task DNN architecture using Word2Vec + GloVe features together

3.1.5. Co-occurrence Matrix-Based Visualisation

After obtaining the predictions for all tweets in the dataset, first, we eliminated all the tweets that did not contain any product-related feedback/suggestion/opinions as they did not contain any useful information for the opinion retrieval task. This step is a data cleaning operation which enables a better representation of opinions shared in the social media. We then used the predicted sentiment labels to divide the remaining tweets into two groups. The first group contained the tweets labelled as negative by the sentiment analysis model, representing the complaints, deficiencies, and concerns regarding the product. The second group contained the remaining positive and neutral tweets representing the customer satisfaction and expectations about the product. Accordingly, we constructed vocabulary sets for each group according to their average TF-IDF scores; a vocabulary consisting of 1-Gram and 2-Gram terms was created for this purpose. Subsequently, we created a co-occurrence matrix showing to what extent each pair of the words in the vocabulary are related for each group. Each entry of the co-occurrence matrix represents the number of times the corresponding word pair appears in the same tweet.

Finally, using each co-occurrence matrix as a complete graph, we generated and visualised the word clusters for sentiment-based opinion analysis using the force-atlas algorithm (Jacomy et al., 2014). Cosine similarity was used for clustering. For visualisation, we applied a pre-processing operation known as edge-filtering (Jia et al., 2008). Related to this, a threshold value for similarity was determined and the edges above this threshold were eliminated to keep only the edges that represent strong linkage between the corresponding word pairs. We investigated the practical implications based on the obtained word clusters.

3.2. Dataset

In this study, the data source was the Twitter social media platform. We collected the data using the Twitter package/tool officially published by Python Package Index (pypi) platform. To identify the relevant data, we used a lexical search strategy, specifically, Google Glass relevant terms and hashtags. We only retrieved English tweets posted between July 2016 and July 2019. The query resulted in 4,956 unique tweets. Each tweet consists of various data fields such as username, date, year, number of retweets, text, mentions, hashtags and the original link. We used only the text content of the tweets to design a generic opinion retrieval system that can be applied to data from different social media platforms.

After collecting the tweet dataset, we performed the manual labelling step, labelling each tweet with sentiment and opinion labels. For the labelling step, three researchers labelled 1,000 tweets separately. All labels were combined in columns side by side, and the contradicting labels were discussed and confirmed. Following the approaches listed in Table 1, we placed the sentiment target into three categories – negative, neutral, or positive – whereas the opinion target is a binary variable. A tweet text entry with the label "Opinion No" indicates that the user made mention of "Google Glass" but did not give any functional feedback or suggestion. Table 2 shows the distribution of 1,000 tweets, manually labelled for the single and multi-task supervised learning problems designed to construct the opinion retrieval system. Table 3 shows some exemplary tweets along with their label information . Subsequently, we applied the DNN models built using the labelled dataset to the remaining unlabelled tweets, and both sentiment and opinion labels were obtained for all tweets in the dataset.

4. Results

In this section, we first present the experimental setup and the results obtained with various word embedding techniques and DNN architectures. Subsequently, we discuss the practical results.

Table 2: Class distribution of labelled tweets

Sentiment	Opinion	# of Tweets
Negative	No	45
Negative	Yes	47
Neutral	No	165
Neutral	Yes	307
Positive	No	65
Positive	Yes	371

Table 3: Exemplary labelled tweets

Text	Sentiment	Opinion
Looked like #googleglass failed.	Negative	No
Google Glass Dangerous For Drivers	Negative	Yes
What Happens to #Brain When You Use #GoogleGlass	Neutral	No
Imagine a future where #GoogleGlass knows how you are feeling	Neutral	Yes
Loved seeing #GoogleGlass	Positive	No
#GoogleGlass Improves #Productivity Of #Boeing #Workers	Positive	Yes

4.1 Experimental Setup

To perform the experiments, we used the Google Colab platform. We used Keras library with Tensorflow backend to build DNN models and Gensim library (Řehůřek & Sojka, 2010) to obtain Word2Vec and GloVe representations of the tweets. For hyper-parameter optimisation, we tested various values for the hyper-parameters of feature extraction and tested model training algorithms. For this purpose, first we split the labelled tweet dataset consisting of 1,000 samples into training and test sets at 80% and 20%, respectively. Next, we used 20% of the training set as a validation set to find the optimal values of hyper-parameters for each algorithm. The best model obtained on the validation set was applied on the test set for performance evaluation. This procedure was applied 10 times and the Wilcoxon test was applied to evaluate the statistical significance of the results.

Regarding feature extraction, we tested different values of vocabulary size from 50 to 500 to construct the feature matrices for the TF, TF-IDF, Word2Vec and GloVe methods, using both CBoW and Skip-Gram implementations of Word2Vec. For transfer learning, we also used the pre-trained word vectors of Word2Vec and GloVe. For DNN-based model creation, we tested two activation functions, tanh and ReLU, and various drop-out rates between 0.2 and 0.8. The number of hidden units in the dense layers were tested with [(600,300),(200,100),(100,50)] values. We performed the experiments for 14 different configurations of single/multi-task DNN architectures and various word representation methods.

4.2. Experimental Results

As stated in Section 3, the opinion retrieval system proposed in this study is based on two learning tasks: sentiment prediction and opinion detection. Table 4 shows the test set performance results for the sentiment prediction task with various combinations of DNN architectures and word representation methods. As evident, the best performance, with a 0.63 F1 score and 0.66 accuracy, was obtained with a multi-task DNN architecture using Word2Vec for word representation. The second best result, with an F1 score of 0.62 and accuracy of 0.66, was also obtained with a multi-task architecture using Word2Vec and GloVe features together. However, we should note that the difference between these two models is not significant (p-value > 0.05). The third highest F1 score of 0.60 and accuracy of 0.65 were achieved by a single-task DNN architecture with Word2Vec features. The Wilcoxon test indicated that the difference between F1 scores of multi-task + Word2Vec and single-task + Word2Vec models is significant (p-value < 0.05). The results also show that, in general, the DNN models based on word embedding features yield better performance compared to the models using BoW features as input.

Table 4: Average performance results on the test set for the sentiment prediction task

Target	Model	F1 score	Accuracy
Multi-task	Word2Vec	0.632	0.663
Multi-task	Word2Vec + GloVe	0.618	0.663
Single-task	Word2Vec	0.597	0.654
Multi-task	GloVe	0.575	0.641
Single-task	GloVe	0.558	0.646
Multi-task	TF-IDF + Word2Vec	0.507	0.605
Multi-task	TF-IDF + GloVe	0.446	0.563
Multi-task	TF + Word2Vec	0.431	0.536
Multi-task	TF + GloVe	0.413	0.529
Single-task	TF-IDF	0.376	0.475
Single-task	TF	0.370	0.470
Multi-task	TF + TF-IDF	0.366	0.466
Multi-task	TF-IDF	0.288	0.381
Multi-task	TF	0.280	0.398

The performance results regarding the opinion detection task can be seen in Table 5. Similar to the results obtained for the sentiment prediction task, the best two performances with an F1 score of 0.68 were yielded by the Multi-task + Word2Vec and Multi-task + Word2Vec + GloVe models. The third highest F1 score was also obtained with a multi-task DNN architecture using GloVe features as input, and the performance differences among these top three models are not statistically significant (p-value >0.05). The best performing single-task DNN architecture for the opinion detection task yielded an F1 score of 0.65 and accuracy of 0.72 with Word2Vec features. The statistical test revealed that the F1 scores of the best performing multi-task models, Multi-task + Word2Vec and Multi-task + Word2Vec + GloVe, are significantly higher than that of the best single-task model. In parallel to the results obtained for the sentiment prediction task, the word embedding features performed better than the BoW features.

Table 5: Average performance results on the test set for the opinion detection task

Target	Model	F1 score	Accuracy
Multi-task	Word2Vec	0.677	0.719
Multi-task	Word2Vec + GloVe	0.676	0.727
Multi-task	GloVe	0.657	0.717
Single-task	Word2Vec	0.652	0.723
Single-task	GloVe	0.602	0.698
Multi-task	TF-IDF + Word2Vec	0.583	0.700
Multi-task	TF-IDF + GloVe	0.542	0.676
Multi-task	TF-IDF	0.484	0.610
Multi-task	TF + GloVe	0.474	0.661
Multi-task	TF + Word2Vec	0.472	0.668
Multi-task	TF	0.471	0.641
Single-task	TF-IDF	0.470	0.632

Multi-task	TF + TF-IDF	0.443	0.641
Single-task	TF	0.426	0.645

Regarding the results seen in Table 4 and 5, the Multi-Task + Word2Vec model yielded the highest F1 scores for both the sentiment prediction and opinion detection tasks. Therefore, we applied this model to the remaining unlabelled 3,956 tweets. The distribution of the obtained predictions with respect to the class labels are shown in Table 6.

Table 6: Class distribution of the tweets labelled with the best multi-task model

Sentiment	Opinion	# of Tweets
Negative	No	332
Negative	Yes	227
Neutral	No	881
Neutral	Yes	1385
Positive	No	401
Positive	Yes	1730

4.3. Practical Results

The keyword network map for each category obtained using the predictions of the best DNN model following the methodology detailed in Section 3.1.5 can be seen in Figures 4 and 5. As shown in Figure 4, there are five identified major clusters which summarise negative opinions for the Google Glass. Cluster 1 (C1) shows overall problems regarding the market positioning of the Google Glass. Some consumers believe it is better to position this product for scientists and companies rather than generic consumer usage. C2 illustrates the privacy concerns for this product, as users could potentially record others without their consent. C3 reveals that many customers require a more appealing design and a better user experience. C4 illustrates some of the consumer concerns are related to the potential for brain cancer due to the wireless radiation in the device. Finally, C5 depicts the safety concerns of consumers such as the potential of the glass to limit the vision of or distract drivers, cyclists, or pedestrians in particular. Table 7 shows the sample predicted tweets for negative opinion clusters.

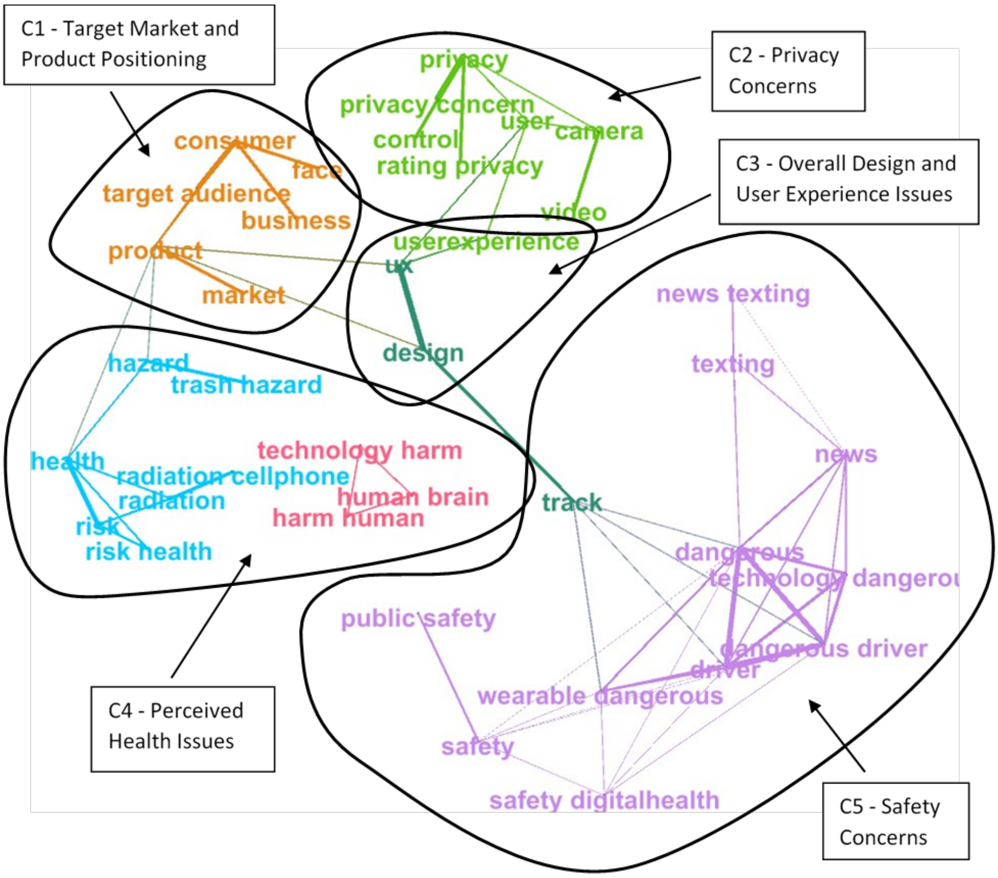

Figure 4: Word cluster visualisation of the tweets that includes opinion and is labelled as negative by the sentiment prediction model

Figure 5 shows the clustering visualisation of tweets predicted as positive or neutral by the sentiment prediction model. There are seven major clusters of the related keywords. C1 illustrates the application areas of Autism and Brain Disorder: there are various applications of this to help patients socialise and interact. C2 shows the various hardware- or software-oriented capabilities that consumers like. C3 illustrates the many application areas of Google Glass in factory conditions to mainly improve productivity and support with work tool-related applications. C4 demonstrates the application areas of this solution for medical students, such as surgical simulations. C5 groups disability-related user cases, such as supporting visually impaired individuals. C6 shows the applications of the product in surgical cases as a collaborative tool. Finally, C7 reveals the many e-health application-related suggestions and feedback to improve doctor–patient interaction. Table 8 lists the sample predicted tweets for the positive and neutral opinion clusters.

Table 7: Sample tweets from each cluster belonging to the negative class

Cluster	Tweet
C1 – Target Market and Product Positioning	#googleglass was not a failure, it was marketed incorrectly. There would be something like #GoogleGlass for consumers but more money doing it for #businesses
C2 – Privacy Concerns	After failures of #GoogleGlass, due to #privacy issue, [ANON] plans to launch voice powered glass #Failure of #Googleglass has shown that it will never take off. Because it need to have a camera
C3 – Overall Design and User Experience Issues	UX least requires better interaction design and perhaps a different form factor #GoogleGlass #Googleglass failed because it forgot about human #Design
C4 – Perceived Health Issues	#GoogleGlass technology to harm human brain #GoogleGlass Alert Potential health risks from wireless radiation
C5 – Safety Concerns	Amazing sunset tonight, but was driving. Makes me miss #GoogleGlass, but used the drone for good shot It's dangerous and not very comfortable to play game but it would be ok if #GoogleGlass supported

Table 8: Sample tweets from each cluster belonging to the positive and neutral classes

Cluster	Tweet
C1 Autism and Brain Disorder Use Cases	#GoogleGlass Could Be Gamechanger For Kids on the Autism Spectrum #GoogleGlass can help diagnose brain disorders early
C2 – Hardware and Software Capabilities	University is working to power the lenses. Next generation #googleglass How about applying AI scanning technology into #googleglass?
C3 – Workplace, Productivity and Support Use Cases	It's designed to improve the productivity of factory operators #GoogleGlass Question: #GoogleGlass supports Bluetooth pairing, will it pair directly with my arm band
C4 – Use Cases for Medical School Students	Exciting! The new and improved #GoogleGlass could one day transform medical training #googleglass gets another chance at life remote monitoring #clinical students
C5 – Disability Use Cases	What if we used google glass to audio translate what it sees to assist visually impaired people How technology such as #GoogleGlass can help deaf and/or disabled people?
C6 – Surgical Use Cases	Study shows surgeons using #GoogleGlass perform markedly better #GoogleGlass I would like to be a very detailed and reference for organ surgery.
C7 – E-health Use Cases	Wow! Ideas include color changing #anxiety meter, wearable #googleglass for #health Can #googleglass increase efficiency for #healthcare?

Considering the results in both Figure 4 and 5, it is evident that Google Glass needs to resolve a number of issues and position itself better for commercial success. Accordingly, the company should follow two different strategies based on different market targets and product positioning. Firstly, the next generation of Google Glass should have different versions for professionals such as for those scientists, doctors and workplace conditions. This targeted approach can help eliminate negative feedback such as privacy, safety and design issues. The second strategy can be for direct consumers by enhancing its hardware and software capabilities such as enhanced AI-based support, better design and the resolution of privacy and safety concerns. Examining both strategies, a niche market or targeted approach for professional use cases is a better strategy in the short or medium term for the next generations of this product. In the longer term, Google could then work towards an enhanced design for the general public whilst attempting to address the public's concerns. To ensure vast technological adoption, legal issues may be the key factor for this product.

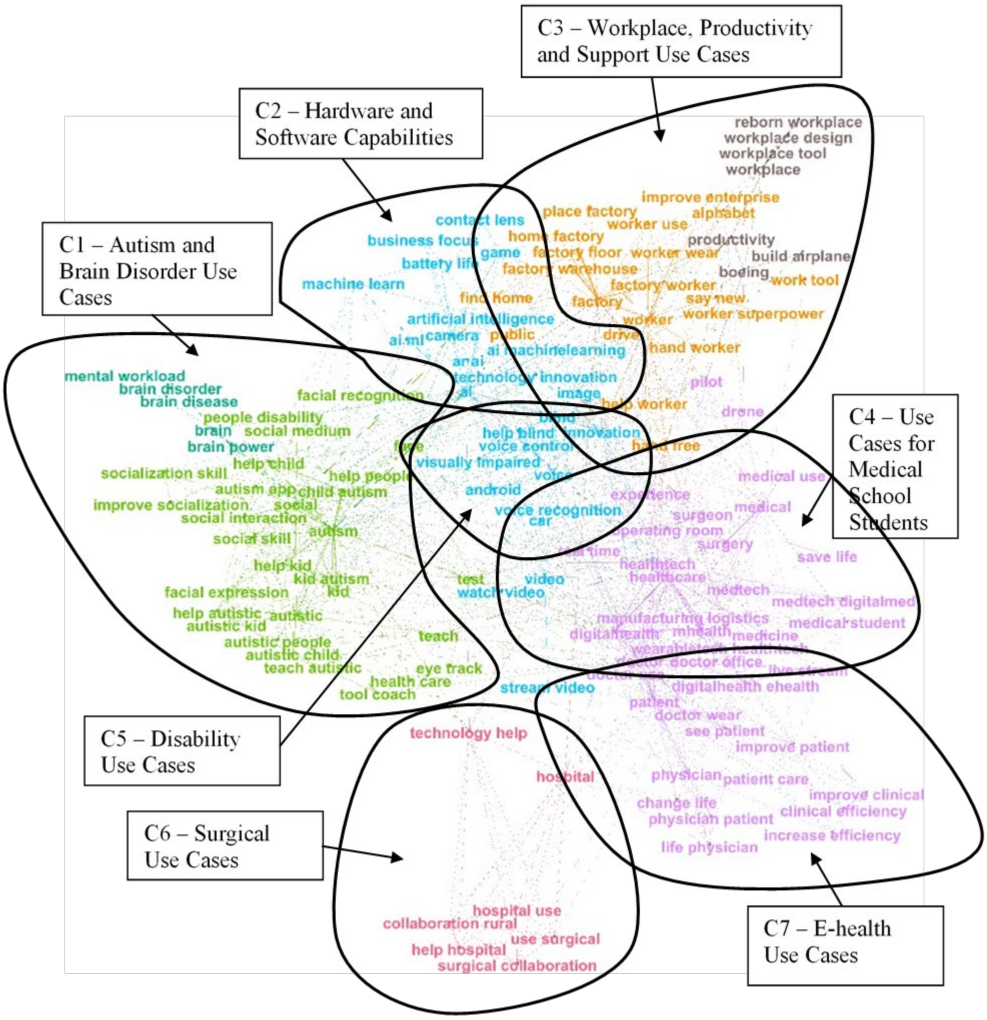

Figure 5: Word cluster visualisation of the tweets that includes opinion labelled as positive or neutralby the sentiment prediction model

5. Discussion

As illustrated in Table 1, sentiment analysis and NLP approaches have been proven to be effective techniques for product or technology reviews and for retrieving opinion from the relevant community. These approaches are even more effective when the relevant datasets are difficult to examine due to time, budget or resource constraints. Our practical and methodological findings show that the proposed method in this study is an effective one considering the performance metrics and the specific case of Google Glass. Following studies that aim to utilise social media data to retrieve novel ideas for product development (Li et al., 2014; Mirtalaie et al., 2017), we extend the current knowledge and relevant applications considering the proposed model – "social media-based opinion retrieval using multi-task deep neural networks" – as well as the application area of the model (product analysis), in order to identify new use cases for an innovative product and support product development-oriented innovation activities.

The interpretations of the results can be further extended by focusing on both academic and methodological aspects. Starting from the academic perspective, previous literature has focused on product analysis with different approaches; Lipizzi (2015) and Nuortimo and Harkonen (2018) examined it from a product launch perspective, Rane and Kumar (2018) and Ibrahim and Wang (2019) examined negative feedback, and some authors have examined sentiments for companies (Botchway et al., 2019) or products (Basiri et al., 2020). In terms of new product development and innovations, many of these are appropriate for incremental innovations where products are improved upon the previous offering but may not be highly suitable for radical innovations (Holahan et al., 2014). However, relative to previous approaches, our results show that our model is highly suitable for both incremental and radical innovations as we identify radical (game changing) ideas as well as incremental ones (improvements based on the previous offerings).

To extend further the discussion on the suitability of our model for new product development and innovation practices, we should consider the criticism of Trott (2001) that consumer opinion is not suitable, especially for radical innovations, particularly where there is asymmetric knowledge between the companies and consumers. In our case, we found that when knowledge is being retrieved from a large database with a systematic approach, it is possible to organise the relevant information to be beneficial, even for high-tech products. Furthermore, it is possible to tap into the hidden information or market positioning opportunities (Google Glass may be better suited for industrial or scientific actors instead of direct consumers). For example, Mirtalaie et al. (2017) successfully retrieved product feature-specific information and performed cross-domain analysis to identify novel ideas that can be integrated into future generations of the reference product (See Table 1). We extend this information by uncovering new product application areas, product improvement opportunities and next generation innovations.

We also make methodological contributions to the related literature. As illustrated in Table 1, a detailed product analysis from unstructured data consisting of user comments requires the design of a multi-stage framework. The proposed opinion retrieval system in this study is also based on two machine learning tasks which address sentiment analysis and opinion detection problems. We used the opinion detection module to identify the user comments that include useful feedback or suggestions about the related product. Thus, we removed the comments which did not include useful feedback but only expressed sentiment regarding the reference product. Combining the outputs of the models allowed us to conduct a detailed sentiment-based analysis of the user opinions with a word network map consisting of keyword clusters appearing in the related tweets. By training these tasks with a multi-task DNN architecture, we improved the generalisation ability of the resulting system.

Finally, we should also the mention transferability and further application areas of our model. Although we applied it to the consumer electronics area, others can implement it in other areas by training the model with specific data and labelling steps. The proposed model could also be implemented for service innovation or business model innovation activities.

6. Conclusions

In this study, we proposed a framework that uses social media data to reveal the reasons for a failed innovative product from the customer perspective and to suggest new use cases and innovative ideas for product development. To do so, we applied various single and multi-task

DNN models with different word representation techniques to perform opinion retrieval based on social media. As a case study, Google Glass was selected using the Twitter platform. For the main opinion retrieval task, we identified two supervised learning tasks as sentiment prediction and opinion detection. Sentiment prediction was handled as a multi-class problem with negative, neutral and positive class labels representing the sentiment of the user's opinion about the related product. We designed the opinion detection module as a binary-class problem representing the existence of a product-related opinion in the related content.

Our experimental results show that for both sentiment prediction and opinion detection tasks, multi-task DNN models yielded significantly higher F1 scores than single-task DNN models. From the point of feature input set, the best results were obtained using Word2Vec features individually and Word2Vec and GloVe features together. The best performing model, which yielded the highest F1 scores for both tasks, was the multi-task DNN architecture with Word2Vec features. We applied this model to the unlabelled tweets to obtain the predictions for both target variables and used the predicted labels to eliminate the tweets that do not contain any feedback or suggestion. We then identified and analysed the remaining negative tweets and positive + neutral tweets separately, then visualised the obtained results as a word network map of the related terms.

6.1. *Implications and Contributions*

This study offers both practical and methodological contributions to the literature. The main practical contribution is the resulting opinion retrieval framework, which can be used as a decision support system for product development approaches. It can be integrated into different phases of an NPD process for both incremental and radical innovations. The implementation of this approach to incremental innovations could make it easier to identify problems that can be eliminated in the next generation of products. Radical innovation may present more difficulties as it would require a complete redesign of products considering the feedback received at a holistic level. For the case of Google Glass, the findings suggest a redesign and radical innovation are required based on both the negative and positive opinions retrieved.

The main methodological contribution of this study is the design of a multi-task DNN architecture that learns two supervised learning tasks simultaneously which are required for a sentiment-based detailed opinion retrieval system. The results confirm that a multi-task DNN model offers better generalisation ability for both sentiment prediction and opinion detection tasks when compared to two single-task DNN models trained independently for each task. Another contribution is the proposal of an end-to-end system for product-related sentiment-based opinion retrieval from social media. The system, from data collection to the sentiment-based visualisation of the keyword maps, offers a detailed product analysis framework for decision makers.

Our study contributes to the previous literature as shown below:

- Existing studies mostly analyse only the tweets labelled as negative by the sentiment analysis module to present the products and services about which the customers complain most. Our proposed system use the opinion detection module to also detect and analyse the positive and neutral comments that may contain useful feedback or suggestion that can be utilised for product development, innovation-oriented decisions and finding new use cases for the related product,

- Existing studies mostly use domain-specific features or lexicons for data representation, whereas we propose a generalisable model using state-of-the-art word embedding techniques that can be applied to any social media-based text data,
- Most of the existing studies apply multiple phases for social-media analysis. Our findings show that the overall accuracy of the system can be improved by training these tasks simultaneously with a multi-task learning approach,
- Most of the existing studies present some statistical results, such as the frequencies of sentiment labels of the user reviews about the related products or services, whereas the proposed system in this study presents word network maps to summarise the sentiment-based opinions that can be directly utilised for product development or to determine new use-cases,
- Our proposed system clusters and identifies customer opinion considering the sentimental classifications specific to the product development and innovation opportunities.

Our results have implications for relevant practitioners such as product development specialists and also have implications specific to the smart glass industry. Our proposed model calls for the relevant practitioners to implement smart or advanced approaches into product development practices to minimise failure rates in commercialisation activities. Our selected case highlights how technological advancements alone are not adequate for successful innovations. In the case of Google Glass, it was apparent that other factors, such as legality, safety and usability aspects, were some of the key reasons why such a great product had lower adoption or diffusion rates. Specific to the smart glass industry, relevant companies need to work on an enhanced design for the general public that considers design, hardware and legal issues. However, the public's privacy concerns may be the most difficult challenge to overcome – the results illustrate that smart glass development strategies appear to be more suitable for specific markets (i.e. workplace and productivity use cases). Considering industry-specific application areas of smart glasses, the involvement of key stakeholders may be the most crucial aspect for technological acceptance considering the required software applications as well as industry specific know-how. To resolve this issue, companies such as Google may need to collaborate with industry-specific partners at a global scale to accelerate the diffusion of this technology.

6.2. Limitations and Future Research

Future research can further enhance the proposed opinion retrieval system using bidirectional transformers for word representation such as BERT and XLNet. In addition, unlabelled tweets can be utilised during training by a semi-supervised learning approach with the aim of improving the accuracy of the overall system and also reducing the required number of labelled samples for a robust model. Furthermore, the proposed model can be tested in different areas such as low-tech environments to analyse its generalisability and performance. Other researchers can implement our model for service innovations or business model innovations. Finally, further opinion retrieval systems can be designed for different phases of the product development process.

References

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications, 77, 236–246.

Basiri, M. E., Abdar, M., Cifci, M. A., Nemati, S., & Acharya, U. R. (2020). A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. Knowledge-Based Systems, 198, 105949.

Bhoir, S., Ghorpade, T., & Mane, V. (2017). Comparative analysis of different word embedding models. In 2017 International conference on advances in computing, communication and Control (ICAC3) (pp. 1–4).

Botchway, R. K., Jibril, A. B., Kwarteng, M. A., Chovancova, M., & Oplatková, Z. K. (2019). A review of social media posts from unicredit bank in europe: A sentiment analysis approach. In Proceedings of the 3rd International conference on business and information Management (pp. 74–79).

Chang, W., & Taylor, S. A. (2016). The effectiveness of customer participation in new product development: A meta-analysis. Journal of Marketing, 80, 47–64.

Djelassi, S., & Decoopman, I. (2013). Customers' participation in product development through crowd-sourcing: Issues and implications. Industrial Marketing Management, 42, 683–692.

Ergen, T., & Kozat, S. S. (2017). Online training of LSTM networks in distributed systems for variable length data sequences. IEEE Transactions on Neural Networks and Learning Systems, 29, 5159–5165.

Goyal, P. (2018). Deep learning for natural language processing: Creating neural networks with Python. Berkeley, CA: Apress.

Hasson, S. G., Piorkowski, J., & McCulloh, I. (2019). Social media as a main source of customer feed-back: alternative to customer satisfaction surveys. In Proceedings of the 2019 IEEE/ACM International conference on advances in social networks analysis and mining (pp. 829–832).

Holahan, P. J., Sullivan, Z. Z., & Markham, S. K. (2014). Product development as core competence: How formal product development practices differ for radical, more innovative, and incremental product innovations. Journal of Product Innovation Management, 31, 329–345.

Hou, T., Yannou, B., Leroy, Y., & Poirson, E. (2019). Mining customer product reviews for product development: A summarization process. Expert Systems with Applications, 132, 141–150.

Ibrahim, N. F., & Wang, X. (2019). A text analytics approach for online retailing service improvement: Evidence from twitter. Decision Support Systems, 121, 37–50.

Igual, L., & Segu´ı, S. (2017). Introduction to data science. New York, NY: Springer Berlin Heidelberg.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International conference on machine learning (ICML) (pp. 448–456).

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. PloS One, 9, e98679.

Jia, Y., Hoberock, J., Garland, M., & Hart, J. (2008). On the visualization of social and other scale-free networks. IEEE Transactions on Visualization and Computer Graphics, 14, 1285–92.

Jiang, H., Kwong, C., Kremer, G. O., & Park, W.-Y. (2019). Dynamic modelling of customer preferences for product design using denfis and opinion mining. Advanced Engineering Informatics, 42, 100969.

Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). Target-dependent twitter sentiment classification. InProceedings of the 49th annual meeting of the association for computational linguistics: human language technologies (pp. 151–160). Portland, Oregon, USA: Association for Computational Linguistics.

Kim, S., Park, H., & Lee, J. (2020). Word2vec-based latent semantic analysis (w2v-lsa) for topic modeling: a study on blockchain technology trend analysis. Expert Systems with Applications, (p. 113401).

Li, B., Lin, Y., & Zhang, S. (2017). Multi-task learning for intrusion detection on web logs. Journal of Systems Architecture, 81, 92 – 100.

Li, Q., & Shah, S. (2017). Learning stock market sentiment lexicon and sentiment-oriented word vector from StockTwits. In Proceedings of the 21st conference on computational natural language learning (CoNLL 2017) (pp. 301–310). Vancouver, Canada: Association for Computational Linguistics.

Li, Y.-M., Chen, H.-M., Liou, J.-H., & Lin, L.-F. (2014). Creating social intelligence for product portfolio design. Decision Support Systems, 66, 123 – 134.

Lipizzi, C. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using twitter streams. International Journal of Information Management, 35, 490–503.

Liu, Y., Jiang, C., & Zhao, H. (2019). Assessing product competitive advantages from the perspective of

customers by mining user-generated content on social media. Decision Support Systems, 123, 113079.
Mai, L., & Le, B. (2020). Joint sentence and aspect-level sentiment analysis of product comments. Annals of Operations Research, (pp. 1–21).
Mirtalaie, M. A., Hussain, O. K., Chang, E., & Hussain, F. K. (2017). A decision support framework for identifying novel ideas in new product development from cross-domain analysis. Information Systems, 69, 59–80.
Nuortimo, K., & Härkönen, J. (2018). Opinion mining approach to study media-image of energy production. implications to public acceptance and market deployment. Renewable and Sustainable Energy Reviews, 96, 210–217.
Park, H., Bharadhwaj, H., & Lim, B. Y. (2019). Hierarchical multi-task learning for healthy drink classifi-cation. In 2019 International joint conference on neural networks (IJCNN) (pp. 1–8).
Parwez, M. A., Abulaish, M. et al. (2019). Multi-label classification of microblogging texts using convolution neural network. IEEE Access, 7, 68678–68691.
Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In empirical methods in natural language processing (EMNLP) (pp. 1532–1543).
Rane, A., & Kumar, A. (2018). Sentiment classification system of twitter data for us airline service analysis. In IEEE 42nd Annual computer software and applications conference (COMPSAC) (pp. 769–773).
Rathan, M., Hulipalled, V. R., Venugopal, K., & Patnaik, L. (2018). Consumer insight mining: Aspect based twitter opinion mining of mobile phone reviews. Applied Soft Computing, 68, 765–773.
Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In Proceed-ings of the LREC 2010 workshop on new challenges for NLP frameworks (pp. 45–50). Valletta, Malta:ELRA.
Ren, J., Long, J., & Xu, Z. (2019). Financial news recommendation based on graph embeddings. Decision Support Systems, 125, 113115.
Rezaeinia, S. M., Rahmani, R., Ghodsi, A., & Veisi, H. (2019). Sentiment analysis based on improved pre-trained word embeddings. Expert Systems with Applications, 117, 139–147.
Saura, J. R. (2020). Using data sciences in digital marketing: Framework, methods, and performance metrics. Journal of Innovation & Knowledge.
Saura, J. R., & Bennett, D. R. (2019). A three-stage method for data text mining: Using UGC in business intelligence analysis. Symmetry, 1, 519.
Schemmann, B., Herrmann, A. M., Chappin, M. M., & Heimeriks, G. J. (2016). Crowdsourcing ideas: Involving ordinary users in the ideation phase of new product development. Research Policy, 45, 1145– 1154.
Skansi, S. (2018). Introduction to deep learning. New York, NY: Springer Berlin Heidelberg.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15, 1929–1958.
Sun, X., Han, M., & Feng, J. (2019). Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products. Decision Support Systems, 124,113099.
Trott, P. (2001). The role of market research in the development of discontinuous new products. European Journal of Innovation Management, 4, 117–126.