Journal of Organizational and End User Computing
Volume 33 • Issue 4 • July-August 2021

Assessing Public Opinions of Products Through Sentiment Analysis:

Product Satisfaction Assessment by Sentiment Analysis

C. Y. Ng, Lee Shau Kee School of Business and Administration, The Open University of Hong Kong, Hong Kong

Kris M. Y. Law, School of Engineering, Deakin University, Australia

https://orcid.org/0000-0003-3659-0033

Andrew W. H. Ip, Department of Mechanical Engineering, University of Saskatchewan, Canada

https://orcid.org/0000-0001-6609-0713

ABSTRACT

In the world of social networking, consumers tend to refer to expert comments or product reviews before making buying decisions. There is much useful information available on many social networking sites for consumers to make product comparisons. Sentiment analysis is considered appropriate for summarising the opinions. However, the sentences posted online are generally short, which sometimes contains both positive and negative word in the same post. Thus, it may not be sufficient to determine the sentiment polarity of a post by merely counting the number of sentiment words, summing up or averaging the associated scores of sentiment words. In this paper, an unsupervised learning technique, k-means, in conjunction with sentiment analysis, is proposed for assessing public opinions. The proposed approach offers the product designers a tool to promptly determine the critical design criteria for new product planning in the process of new product development by evaluating the user-generated content. The case implementation proves the applicability of the proposed approach.

KEYWORDS

Opinion Mining, Product Development, Sentiment Analysis, Social Networking Sites

1. INTRODUCTION

Social networking sites are internet-based applications supporting communications for social and business purposes. These sites enable an individual user to interact with others to efficiently share personal interest, ideas, thoughts, or activities. One unique commonality to the existing social networking sites is that the user-generated content, in different forms such as photos, videos, blogs, emoticons, or text posts, is openly shared. Text posts like comments or reviews of a target product are embedded with sentiment words that can be extracted for further analyse for making purchase decisions (Goldsmith & Horowitz, 2006). The analysis on opinion strengths would be very useful to product review references because these comments are directly from the consumers (Hu and Liu, 2004; Kim and Moon, 2011; Yoo et al., 2018) and can be utilised to support product design evaluations. The number of user-generated contents in the social networking sites is increasing drastically, the

DOI: 10.4018/JOEUC.20210701.oa6

sentiment analysis is emerging as a topic among researchers, regarding the capturing or summarising the text posts (Cambria et al., 2013).

Sentiment analysis, which focuses on the processing of the text for the identification of opinionated information, can handle large volumes of text posts (Mali et al., 2016). It can be used for the determination of the contextual polarity as well as the measurement of opinion strengths by searching the sentiment words in a set of text posts. Many applications using sentiment word analyses to summarise customer text posts have been successfully carried out for different product categories including digital cameras, laptops, cell phones, books, and health care products (Hu and Liu, 2004; Bucur, 2015; Kim et al., 2018).

The SentiWordNet (Guerini et al., 2013) is one of the commonly used for the determination of polarity and opinion strength. It is done by counting the number of sentiment words or summing up the sentiment scores. However, it may not be sufficient to classify a comment to be positive or negative by merely counting the number of sentiment words or determining the sentiment scores. Thus, an algorithm for categorising the comments into different polarities to support decision making is needed.

K-means (MacQueen, 1967) is a simplified approach to perform cluster analyses for multiple dimensional data. It aims to classify several data into *k* clusters. With its advantages for grouping the unlabeled data efficiently, the use of *k*-means for clustering the text posts is proposed. The text posts can be classified into three different groups, i.e. positivity, negativity, and objectivity, using sentiment analysis with the *k*-means algorithm. K-means can also be employed to facilitate the classification of various comments into corresponding design criteria.

The approach proposed has two distinct features. First, it offers an immediately applicable instrument for the evaluation of sentiment scores to present the results of sentiment analysis. Second, it helps to identify the critical design criteria and opinion strengths based on the user-generated content without reading all the text posts. Also, it offers a practical and prompt means for collecting feedback from the customers' perspective. The results are valuable for decision-makers to perform product analysis, especially for generating new design alternatives or revised models at the initial product development stage. The subsequent sections of the paper are organised as follows: Section 2 describes the related work of sentiment analysis and *k*-means for product evaluations. Section 3 outlines the procedure of the approach. Section 4 demonstrates the applicability of the method approach using a case application. Section 5 presents the results and conclusion.

2. BACKGROUND

2.1. Consumer Reviews and Product Development

In the new product development processes of consumer products, the product design stage is the most challenging, to gather customer concerns to support the decision-making on product design (He et al., 2015; Chang et al., 2018; Ng and Law, 2019), by collecting feedback from consumers (Liu et al., 2019).

The traditional product evaluation based on customer survey incurs time lag and significant resources for data processing (Pournarakis et al., 2017). It requires the pro-active participation of the users in the survey; thereby, those studies are conducted on a relatively smaller scale (Wang et al., 2018). Besides, the questions are set from the experts' point of view before conducting the survey rather than from the customers' perspectives (Hsiao et al., 2017). Thus, the interviewees are only able to provide their opinions in a specific context. In contrast, social networking sites enable customers to provide feedback and concerns about the products with relatively fewer restrictions. The product can gather constructive feedback through maintaining product pages for consumers to post product reviews. The consumers post their comments on a product concerning a specific product feature based on their qualitative judgement. These reviews, written by consumers or products end-users, reveal their expectations of the products (Li et al., 2014). Manufacturers can, therefore, obtain some reflection for the redesign of the product according to consumer's feedbacks (Gallaugher & Ransbotham, 2010;

Helander and Khalid, 2006). Hence, gathering opinions from consumers contributes significantly to the core processes of product design and development, which are critical in the value chain of the consumer product. The analysis on the consumer opinions is useful for identifying product life cycle criteria, to support product innovation (Muninger et al., 2019; Suseno et al., 2018) and new product development (Bashir et al., 2017; Poecze et al., 2018). The consumer opinions can be done by analysing the online reviews and ranking the options available (Liu et al., 2017). The analysis of the online review is to identify product features by assessing the sentiment strengths of user-generated content. The ranking of options helps interpret the results after carrying out the analysis of the online reviews. The text mining approach can then be applied for analysing the user-generated content using natural language processing and machine learning (Wang et al. 2017). The sentiment analysis involves the search, extraction, and evaluation of the unstructured text written by the writers to understand the writers' attitudes (Yadollahi et al., 2017).

2.2. Sentiment Analysis

Sentiment analysis is well known for summarising the public opinion. Opinionated information can be captured using corpus-based methods, machine learning-based methods, and hybrid (Tang et al., 2009; Liu, 2010; Yan et al., 2017; Basili et al., 2017; Tang et al., 2018; Yoo et al., 2018). The machine learning approaches often require a significant amount of training documents for text classification (Medhat et al., 2014). In contrast, the corpus-based approach can simply begin with a set of opinion words collected and then expanded the set of words by searching the synonyms and antonyms, according to the thesaurus. It starts with a list of sentiment words and then searches for the additional sentiment words with similar meaning to build a large corpus.

While previous studies have attempted to address this problem, they mainly rely on traditional sentiment classification methods, including lexicon-based methods [6], which are economical, expandable, and straightforward. The limitations in traditional sentiment classification stem from its dependence on human effort in labelling documents, time-intensive activities, low coverage, and limited effectiveness resulting from the ordinary and unstructured text in tweets [7], [8], [9]. Many researchers claim that employing a mix of lexicon-based and machine learning methods can produce improved results [10].

For the public opinion on a target topic, several automatic text classification techniques are introduced to classify the subjective words into different sentiment polarities (Dave, Lawrence, & Pennock, 2003; Pang & Lee, 2008; Yang and Lin, 2018) and to summarise the public opinions for the user reviews from social networking sites regarding their experiences on movies, electronic products, restaurants, or hotels (Hu and Liu, 2004; Thet et al., 2008; Yan et al., 2015; Ali et al., 2016; Tjahyanto and Sisephaputra, 2017). Numerous studies have attempted to apply the lexicon-based approach for conducting the sentiment analysis because of the simple and expendable (Naseem et al., 2020). However, the lexicon-based approach requires extensive use of resources in tagging the words for the unstructured content posted (Saeed et al., 2018). This limitation can be remedy by the adoption of the integration of the lexicon-based approach with learning algorithm (da Silva et al., 2014; Naseem et al., 2020).

2.2.1. Classifying and Polaritizing

The mechanism of classifying subjective words into different polarities is important in the sentiment analysis. The polarities of the subjective or emotional words are determined, while the sentiment strengths of the text are assigned with linguistic terms of preferences (Wilson et al., 2004). The lexical databases such as SentiWordNet (Guerini et al., 2013), SenticNet (Cambria et al., 2010) and Opinion Lexicon (Hu and Liu, 2004) are developed to support sentiment words searching. Liu and Hu Opinion Lexicon has over 6800 positive and negative opinion words (Hu and Liu, 2004) to facilitate the sentiment analysis. The opinion words are categorised into either positive or negative polarity to support the analysis of customer reviews. Another tool for sentiment analysis called SentiWords

(Guerini et al., 2013) has been developed recently. This lexical tool contains about 155,000 words. Each opinion word is classified as positivity, negativity, or objectivity. The opinion strength is associated with a sentiment score ranging from -1 to 1. For example, the word "bad" is associated with a sentiment score -0.625 in negative polarity with 0.375 in objective polarity (1-|-0.625|=0.375). The sentiment scores enable a practical approach to support sentiment word searches for summarising customer reviews. Many case applications using SentiWords have been reported including opinion mining in tourism products (Bucur, 2015), development of intercity safe travelling plans (Ali et al., 2017), and extraction of public opinion in financial services (Ravi et al., 2017).

SentiWordNet (Guerini et al., 2013) is capable of supporting the analysis of sentiment strengths. SentiWordNet assigns each word with three values, called triplets, ranging from -1 to 1 for indicating a sentiment strength of a subjective word and the sum of the absolute values of the triplets should be 1.0. SentiWordNet has been employed for many case applications (Zhao and Li, 2009; Jahiruddin et al., 2009; Dalal and Zaveri, 2013; Bucur, 2015; Ali et al., 2016). Extending the existing SentiWordNet lexical database with the triplets is a valuable approach for performing sentence-based analyses.

However, one of the disadvantages is that scores obtained from objective sentiment are often ignored when determining the sentiment polarity (Gonçalves et al., 2013). It is because merely summing up or taking averages on the sentiment scores for concluding the sentiment polarity of a review, the score of objectivity is always very high that affects the polarity determination, even the polarity of the text post can be classified to positivity or negativity. The use of an unsupervised learning approach for determining the polarity of the text posts based on sentiment scores thus remains open for discussions.

2.2.2. K-Means For Vector Quantisation

K-means is a type of unsupervised learning technique for vector quantisation to partition unlabeled data for data mining. The term *k*-means was first applied for analysing multivariate observations (MacQueen, 1967), and the standard algorithm was proposed to support the analysis of pulse-code modulation (Lloyd, 1982). It is used to group many observations *n* into groups or clusters *k* that each observation is partitioned to a cluster with the nearest mean. The centroids of the *k* non-intersection clusters, as well as the groupings of observations, can be determined by iterating the algorithm to search for the minimum value of the squared Euclidean differences (between the centroids of clusters and the corresponding clustered observations). The number of "*k*" clusters can be pre-determined or obtained by searching the "elbow point" (Ketchen & Shook, 1996). Once the value of *k* is determined, the initial centroids of clusters can be either generated or selected randomly from the observations.

The results of the *k*-means++ algorithm are found more reliable than the standard *k*-means algorithm (Bahmani et al., 2012; Öztürk et al., 2015). The k-means++ approach can overcome the deficiencies associated with determining the initial centroids for *k*-means. A study on different dataset has been conducted to demonstrate that *k*-means++ is capable of providing more reliable results, even the number of clusters is increased (Shindler, 2008).

2.2.3. Integration of K-means and Sentiment Analysis

Though *k*-means is straightforward with some lexical resources for performing sentiment analysis available, the combined *k*-means with sentiment analysis for summarising public opinions of product reviews has not been well explored until recent years. The combined approach has been employed for the evaluations of emotional signals based on emoticon or text posts obtained from Twitter (Hu et al., 2013), and this is probably more suitable for summarising the crowd responses on an incidence. It has been used for analysis on the impact of the economic costs of violence (Pejić Bach et al., 2018). Besides, it has also been applied for opinion mining (AL-Sharueea et al. 2018), where the online reviews are crawled from an Australia consumer opinion website, for processing a vast amount of data with no training or manual work is involved.

This combined approach can categorise the reviews into different rating for supporting purchase decisions (Riaz et al., 2017). However, specifically for evaluating online review for a target product or service, the assessment criteria often are not with equal importance, as well as the weights of the criteria. Result simply based on the majority of voting of polarities or the summation of sentiment scores is not sufficient to conclude a design evaluation of a target product or service. Hence, a new approach is proposed, and the details are discussed in the subsequent sections.

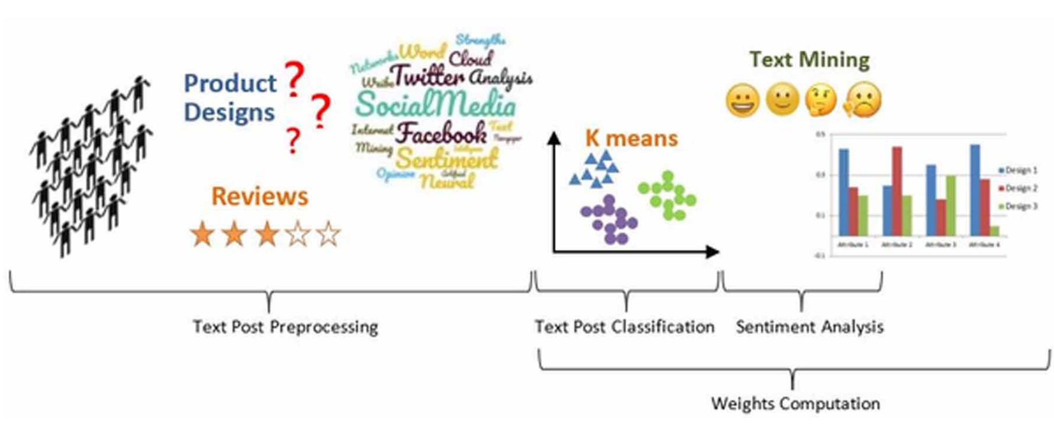
3. THE PROPOSED APPROACH

Customers may have different perceptions of the product models. Some of them would post their comments or reviews about the product models on the corresponding fans' pages. These comments or reviews posted on social networking sites help the decision-makers to evaluate different product models and identify the importance among different attributes by extracting the ideas from the crowd. The development of text mining approach supports automatically evaluate and summarise a vast amount of comments.

The proposed opinion mining approach starts with the determination of the assessment criteria of the target products (Figure 1). It includes the extraction of text posts from social networking sites, identifying keywords to form a corpus for the product category, the adoption of K-means for classifying the posts according to different design attributes, and evaluating the posts using sentiment analysis.

The decision-makers firstly determine the assessment criteria and the number of criteria (i.e. clusters) for the implementation of the approach. The text posts are crawled from the corresponding social networking sites for further processing. The unnecessary information of the text posts such as the post identification numbers, writers' nicknames, date information, and query times are removed. The stop words are deleted to reduce the searching time. The words of the posts are stored in an array through tokenisation to support the search of sentiment words. In the second stage, the key phrases related to the assessment criteria are tagged manually and extracted into the corresponding arrays to facilitate criteria classification processes.

Figure 1.



A search process is then carried out for counting the number of occurrences of key phrases. For example, three assessment criteria related to mobile devices, namely "Display", "Specifications", and "Camera" are identified. The associated key phrases are pre-determined, with a text post "*Nice screen but the fingerprint sensor is useless and plz solve the overheating problem"* extracted. The extracted post contains three key phrases that can be associated with the assessment criteria. They are “display”, “fingerprint”, and “overheat”. Further details of employing *k*-means to categorise the text posts into assessment criteria will be presented in the subsequent section.

In the third stage, sentiment analysis would be conducted using SentiWordsNet (Guerini et al., 2013). Suppose a sentiment word “superb” is found in a text post, the corresponding sentiment scores are (P: 0.875, N: 0, O: 0.125) and a three-dimensional vector can represent the scores. The *k*-means iteration then can be carried out to categorise the text posts into the three polarities of sentiment analysis (i.e. positivity, negativity, and objectivity). The fourth stage is to finalise the weights for ranking the product models of the target product. The priority weights of polarities can be determined through normalisation. The overall priority weights can be obtained by multiplying the corresponding criteria weights with the sum of sentiment scores. A higher total priority weight of a product model scored indicates that it is preferred than the other assessed models.

4. CASE IMPLEMENTATION

4.1. The Case

The application of the proposed approach is illustrated with a case. Three smartphone models from three different manufacturers are chosen. The case implementation first begins with crawling the comments from the individual models’ Facebook fan pages. The fan pages enable the public to post comments. Total 2,412 comments posted within six months are extracted to illustrate the applicability of the proposed approach. The details of extraction and tokenisation of the posts for opinion mining can be found here (Ng and Law, 2019). The proposed algorithm helps summarise public opinions to collect feedback from customers. The results are valuable for decision-makers to perform product analyses and product redesigns. The user-generated comments of three mobile phone models, denoted as M1, M2, and M3, in the relevant FaceBook pages are extracted for the illustration of the applicability of the proposed work. The unnecessary punctuations, stopwords, and information are removed. The posts are then tokenised for performing the word matches.

4.2. The Process

4.2.1 Select The Assessment Criteria and Determine The Number of Criteria (The Value of K)

Six assessment criteria of the mobile phone devices namely Specifications (A1), Display (A2), Camera and Storage (A3), Battery (A4), Design (A5), and Software (A6) are identified. These criteria often employed for the evaluations of mobile phones performances in related websites or magazines for consumers. Thus, the value, k=6, is assumed. As the objective of this paper is to showcase the use of *k*-means for summarising user-generated content from the social networking sites, the details of selecting the assessment criteria are not given here. The three mobile phone models will be assessed based on these six criteria.

4.2.2 Identify The Key Phrases of The Selected Assessment Criteria

A set of key phrases concerning each assessment criterion are identified by tagging the phrases based on the posts extracted from the fan pages by manual. The key phrases are identified by counting the number of occurrences, and then tag the phrases, which are with a higher number of occurrences, for grouping them into the corresponding assessment criteria. For example, a text post is read as “Its’ “camera” “has” “a” “wider” “aperture”. The key phrases “camera” and “aperture” are included in the array of key phrases under A3. Over 500 phrases are categorised into the six criteria using the method of word count. Table 1 shows some of the key phrases used in this case application. The next step is to search the text posts using the key phrases in the arrays. The results are represented by a vector defined as:

$p_i = \{A_j\}$ i=1,2,…, n and j=1,2,…, k=6 (3)

where p and A are denoted as the text posts and assessment criteria respectively, n is the number of text posts for evaluation, and k is the number of clusters. Using the case example, while the words “camera” and “aperture” are the key phrases of A3, the text post, p_{ex}, can then be represented by a vector, p_{ex}={ 0, 0, 2, 0, 0, 0}.

Table 1. Assessment criteria and their corresponding key phrases

Assessment Criteria	Related areas of assessment	Key phrases
Specifications, A1	Processing power, CPU and GPU performance, rom, ram, sim card slots, wireless connectivity, sound and speaker quality	quad, octa, processor, qualcomm, snapdragon, LTE, Bluetooth, DAC, sim, B&O, speaker, fingerprint, jack, heat, overheat, feature, spec, hardware, core, ram, rom, 2GB, 4GB, 6GB, IP68, etc.
Display, A2	Resolution, screen size, durability, glass type, surface, hardness,	720p, 1080p, 4k, 18.5:9, ppi, 5”, 6” 720p, amoled, oled, IPS, breakable, broke, display, screen, curve, edge, brightness, brittle, crack, fragile, scratch, glass, gorilla, hard, protector etc.
Camera and Storage, A3	Aperture value, photo quality, flash, optical zoom, expandable memory, sd card size,	32gb, 64gb, 128gb, 256gb, SD, sdcard, microsd, aperture, megapixel, 16mp, 5mp, fps, hdr, camera, cam, capture, blue, exposure, flash, motion, optic, optical, pics, photo, pictures, portrait, zoom, shoot, redeye, slow-motion, stabilisation, shot, selfies, movie, etc.
Battery, A4	Battery endurance, charging time,	battery, charger, charging, detachable, mah, explode, bomb, etc.
Design, A5	Design, materials, size, weight, shape, colour	Bezel, borderless, colour, plastic, metal, metallic, stylus, aesthetic, design, shape, materials, size, weight, square, pretty, comfortable, heavy, light, etc.
Software, A6	Firmware update, software, user interface, Apps of the hardware modules	App, software, firmware, android, app, bug, bootloader, oreo, os, setting, version, root, interface, etc.

4.2.3 Use The K-Means++ Algorithm To Determine The Centroids and Obtain Criteria Weights

The k-means algorithm supplemented by k-means++ is employed to categorise the extracted posts into various clusters for each product model. As mentioned in section 2.2.2, the results generated by the standard k-means algorithm are sensitive to the initial values of centroids assigned, and the proposed approach uses k-means++ for determining the initial seeds to improve the reliability of the results. The centroids for clusters A1, A2, …, A6 can be determined using k-means++. The detailed steps of calculations can be found in the Appendix section.

The adoption of k-means algorithms can be used to categorise the data points into corresponding clusters by minimising the sum of squared differences within the clusters. It can be done by calculating the Euclidean distance between a data point and the six centroids. The optimum clustering solution can be determined by selecting the data point with the minimum value of the sum of the Euclidean distances. The optimum values of centroids can be calculated using the equation (2) if a new minimum distance can be found in that particular iteration. The iterations will be terminated if no further minimum value is found. After looping the equations (1) and (2) for the data points, the posts are then categorised into the six assessment criteria, as shown in Table 2.

Table 2. Summarised results after the k-means iterations of assessment criteria categorisation for the three models

Assessment Criteria	M1		M2		M3	
	No. of Post	Normalised	No. of Post	Normalised	No. of Post	Normalised
A1	205	0.1902	134	0.2306	192	0.2465
A2	85	0.0788	119	0.2048	62	0.0796
A3	158	0.1466	61	0.1050	122	0.1566
A4	91	0.0844	97	0.1670	34	0.0436
A5	478	0.4434	123	0.2117	325	0.4172
A6	61	0.0566	47	0.0809	44	0.0565

4.2.4 Use The k-Means/ k-Means++ Algorithms To Categorize Sentiment Polarities

The sentiment analysis on the extracted text posts is conducted by searching the sentiment words using the lexical resource databases. In our case, we use SentiWordNet (Guerini et al., 2013) for illustrating the applicability of the proposed approach. Some other lexical-based tools for determining the sentiment scores such as Sentic Net (Cambria et al., 2017), TextBlob (Subirats et al., 2018), and Valence Aware Dictionary and sEntiment Reasoner (2014) can also be applied in the proposed *k*-means approach. By the adoption of these tools, the sentiment strengths of the words are represented by sentiment polarities and scores. In our proposed approach, a vector can be defined, given in equation (4), for representing the sentiment strengths of each post,

$$s_i = \{P_i, N_i, \mathrm{O}_i\} \; i=1,2,3\ldots, n \tag{4}$$

where s_i is a 3-dimensional vector for representing sentiment strengths, P, N, O are denoted as the sum of positivity, negativity, and objectivity sentiment strengths respectively; n is the number of text posts extracted of a target product model.

The posts of each product model are categorised into three polarities by iterating the *k*-means and *k*-means++ algorithms. The normalised weights are shown in Table 3 and Table 4. The ranking of the product models can be prioritised by multiplying the criteria weights with the weights of sentiment polarities. The ranking of the models remains unchanged when the value of f_p is increased by 0.01 within [0, 1] (Figure 2).

5. DISCUSSIONS OF THE PROPOSED K-MEANS AND SENTIMENT ANALYSIS APPROACH

5.1 Rapid Production Evalution

The adoption of the *k-means* algorithm in summarising the user-generated contents extracted from social media sites to support a fast-track product evaluation is illustrated in the case example in Section 4.

Results, as shown in Table 4, show that the majority of the posts are related to the Design (A5) and Specifications (A1). The next critical criterion is Display (A2) because many positive comments

Table 3. Summarised results of sentiment analysis for the three models

Assessment Criteria		M1			M2			M3		
		Positivity	Negativity	Objectivity	Positivity	Negativity	Objectivity	Positivity	Negativity	Objectivity
A1	No. of posts	68	112	25	64	43	29	104	63	25
	Normalised	0.3317	0.5463	0.122	0.4706	0.3162	0.2132	0.5417	0.3281	0.1302
A2	No. of posts	35	36	14	50	33	36	19	18	25
	Normalised	0.4118	0.4235	0.1647	0.4202	0.2773	0.3025	0.3065	0.2903	0.4032
A3	No. of posts	92	45	21	15	32	13	49	49	24
	Normalised	0.5823	0.2848	0.1329	0.25	0.5333	0.2167	0.4016	0.4016	0.1967
A4	No. of posts	26	43	22	62	25	10	12	17	5
	Normalised	0.2857	0.4725	0.2418	0.6392	0.2577	0.1031	0.3529	0.50	0.1471
A5	No. of posts	247	74	157	36	68	18	91	195	39
	Normalised	0.5167	0.1548	0.3285	0.2951	0.5574	0.1475	0.2800	0.60	0.120
A6	No. of posts	30	19	12	13	22	12	23	14	7
	Normalized	0.4918	0.3115	0.1967	0.2766	0.4681	0.2553	0.5227	0.3182	0.1591

Table 4. A summary of weights for the product models (f_p = 0.5)

Model Criteria	M1	M2	M3
A1	-0.0088	0.0424	0.0424
A2	0.0061	0.0456	0.0167
A3	0.0315	-0.0035	0.0154
A4	0.0023	0.0404	6.06E-6
A5	0.1531	-0.0121	-0.0417
A6	0.0107	0.00258	0.01026
Overall weights	0.1948	0.1154	0.0430

are posted concerning its high-density display of the model M2. In contrast, the model M1 is ranked as the best among the candidates.

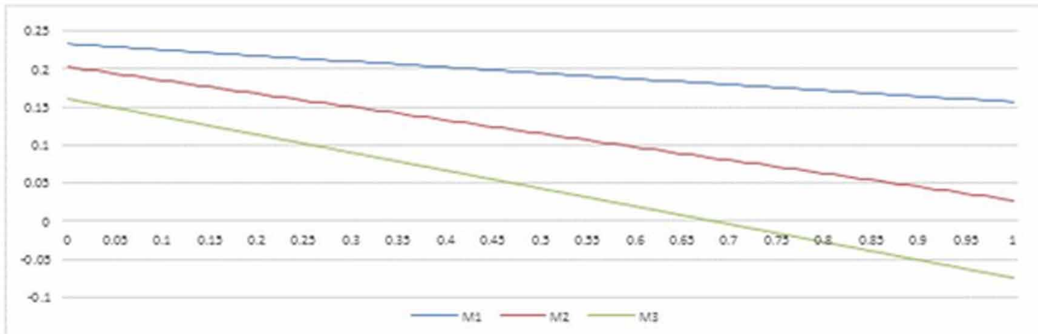
The factor, f_p, is determined intuitively by decision-makers for combining the weights of assessment criteria with the weights of sentiment polarities. A sensitivity analysis is conducted to verify that the final ranking would not be affected by changing the values of vector *f,* as shown in Figure 3.

Our results show that Design (A5) and Specifications (A1) are the most critical assessment criteria, implying that users are concerned more about these two criteria. Consumers pay more attention to the aesthetics of the products such as colour and finishing. Manufacturers may need to put more efforts into developing new phones with better specifications such as the processing power of microprocessors capacity of RAM/ ROM, and sound quality. Some product models emphasised on the image quality, however, the amount of comments obtained is relatively less than the other criteria, and this may imply the 'image quality' is not the foremost concern of consumers. Putting more efforts in enhancing the performances in these criteria may not be leading to inc increase the overall attractiveness of these product models.

5.2 Linking The Sentiment Analysis and product Development

The adoption of sentiment analysis on the user-generated comment to support the evaluation of product design, from a customer perspective, can help to collect the feedback and public views efficiently. This proposed approach is a new initiative on the topic of sentiment analysis and product design evaluation.

Figure 2.



In the proposed approach, the sentiment polarities of the text posts, which are categorised by *k*-means, eliminate the drawbacks related to majority votes or summation of the sentiment scores. Each text post is represented as a data point and categorised into a cluster, the problem related to unreasonably high sentiment score of a particular polarity, which is boosted by posts that contains a large number of sentiment words, can be eliminated. Thus, the confidence in the results of sentiment analysis is enhanced.

Launching a new product to the market is a crucial driver to support a company's long term growth and success (Battistoni et al., 2013). The new product development process consists of a series of the identification on the necessary changes or improvement on existing products, products' idea generation, design of the products, and making the products real and beneficial. The new product development is, therefore,e crucial to product success (Suharyanti et al., 2017).

The proposed approach supports the product designers to promptly determine the critical design criteria for new product planning in the process of new product development by evaluating the user-generated content. Improvement areas of the existing product model can be identified from the end-users opinions without traditional product surveys, and this is a significant contribution to the rapid product development.

6. CONCLUSION

The proposed approach provides a practical way to perform product evaluations by considering the sentiment words of the text posts uploaded to the social networking sites. It offers the product designers a tool to promptly determine the important design criteria for new product planning in the process of new product development by evaluating the user-generated content. The more the comments on a specific product model posted online, the more attention it attracts from the public. However, counting the number of sentiment words and summing up the corresponding sentiment scores are not sufficient to support product evaluation as the design criteria are very often not with equal importance. The use of *k*-means to identify the significance among different design criteria is a novelty of the proposed approach. Improvement areas of the existing product model can be identified from the end-users' opinions without traditional product surveys, and this is a significant contribution to the rapid product development.

6.1 Contributions

The approach combining the *k*-means algorithm with the lexical resource database offers a logical and practical solution to summarise the customer-generated comments without reviewing thousands or more text posts, which is certainly a time and resource-consuming process. Another contribution is the integration of text mining using sentiment analysis and new product development. The use of *k*-means for clustering the product design criteria along with the SentiWordNet for summarising the user-generated content is a new attempt in the field of opinion mining.

6.2 Limitation and Future Works

The proposed approach has some limitations. It requires users to select the product models or alternatives for comparison purposes, and the selection of suitable referencing product models may affect the overall results. In the case implementation, three android mobiles phones with more or less the same price and specifications are selected. In contrast, phones driven by iOS are not chosen because of different user-interface. Therefore, expert judgement is needed when selecting appropriate product models for product comparison. Furthermore, the sentiment scores are given by the lexical resource sometimes may not be sufficient or precise enough because the sentiment words chosen to represent opinions vary among different persons. Therefore, the use of uncertainty analysis in conjunction with sentiment analysis and clustering algorithms are the possible direction for the future works.

REFERENCES

Al-Sharuee, M. T., Liu, F., & Pratama, M. (2018). Sentiment analysis: An automatic contextual analysis and ensemble clustering approach and comparison. *Data & Knowledge Engineering*, *115*, 194–213. doi:10.1016/j.datak.2018.04.001

Ali, F., Daehan, K., Pervez, K., Riazul, S. M., Kim, K. H., & Kwak, K. S. (2017). Fuzzy Ontology-based Sentiment Analysis of Transportation and City Feature Reviews for Safe Traveling. *Transportation Research Part C, Emerging Technologies*, *77*, 33–48. doi:10.1016/j.trc.2017.01.014

Arthur, D., & Vassilvitskii, S. (2007). k-means++: the advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 1027–1035.

Bahmani, B., Moseley, B., Vattani, A., Kumar, R., & Vassilvitskii, S. (2012). Scalable k-means++. *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*, *5*(7), 622–633. doi:10.14778/2180912.2180915

Basili, R., Croce, D., & Castellucci, G. (2017). Dynamic polarity lexicon acquisition for advanced Social Media analytics. *International Journal of Engineering Business Management*, *9*, 1847979017744916. doi:10.1177/1847979017744916

Battistoni, E., Fronzetti Colladon, A., Scarabotti, L., & Schiraldi, M. M. (2013). Analytic hierarchy process for new product development. *International Journal of Engineering Business Management, 5*, 5-42.

Bucur, C. (2015). Using Opinion Mining Techniques in Tourism. *Procedia Economics and Finance*, *23*, 1666–1673. doi:10.1016/S2212-5671(15)00471-2

Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2017). Sentiment analysis is a big suitcase. *IEEE Intelligent Systems*, *32*(6), 74–80. doi:10.1109/MIS.2017.4531228

Cambria, E., Schuller, B., Xia, Y., Havasi, C. (2013, Mar.). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Computer Society*, 15-21.

Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010) SenticNet: A publicly available semantic resource for opinion mining In *Commonsense Knowledge: Papers from the AAAI Fall Symposium*. AAAI Press.

da Silva, N. F., Hruschka, E. R., & Hruschka, E. R. Jr. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, *66*(C), 170–179. doi:10.1016/j.dss.2014.07.003

Goldsmith, R. E., & Horowitz, D. (2006). Measuring motivations for online opinion seeking. *Journal of Interactive Advertising*, *6*(2), 2–14. doi:10.1080/15252019.2006.10722114

Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013). Comparing and combining sentiment analysis methods. *COSN '13 Proceedings of the first ACM conference on Online social networks*, 27-38.

Hsiao, Y. H., Chen, M. C., & Liao, W. C. (2017). Logistics service design for cross-border e-commerce using Kansei engineering with text-mining-based online content analysis. *Telematics and Informatics*, *34*(4), 284–302. doi:10.1016/j.tele.2016.08.002

Hu, M., & Liu, B. (2004). Mining and summarising customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 168-177.

Hu, X., Tang, J., Gao, H., & Liu, H. (2013). Unsupervised sentiment analysis with emotional signals. *Proceedings of the 22nd international conference on World Wide Web*, 607-618. doi:10.1145/2488388.2488442

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International AAAI Conference on Weblogs and Social Media.*

Ketchen, D. J. Jr, & Shook, C. L. (1996). The application of cluster analysis in Strategic Management Research: An analysis and critique. *Strategic Management Journal*, *17*(6), 441–458. doi:10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G

Kim, A. R., Lee, S., & Song, M. (2018). Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis. *Information Sciences*, *454-455*, 200–215. doi:10.1016/j.ins.2018.04.079

Kim, Y., & Moon, I. (2011). A study on algorithm for selection priority of contents in social network service. *Proceedings of the Korean Institute of Information and Communication Sciences Conference.*

Kreutzer, J., & Witte, N. (2013). *Opinion Mining Using SentiWordNet*. Uppsala University. http://stp.lingfil.uu.se/~santinim/sais/Ass1_Essays/Neele_Julia_SentiWordNet_V01

Liu, Y., Bi, J. W., & Fan, Z. P. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion*, *36*, 149–161. doi:10.1016/j.inffus.2016.11.012

Lloyd, S. P. (1982). Least squares quantisation in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. doi:10.1109/TIT.1982.1056489

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 281–297.

Mali, D., Abhyankar, M., Bhavarthi, P., Gaidhar, K., & Bangare, M. (2016). Sentiment Analysis of Product Reviews for E-commerce Recommendation. *International Journal of Management and Applied Science, 2*(1).

Medagoda, N., Shanmuganathan, S., & Whalley, J. (2015). Sentiment Lexicon Construction Using SentiWordNet 3.0. *11th International Conference on Natural Computing*. doi:10.1109/ICNC.2015.7378094

Medhat, W., Hassan, A., & Korashy, H. (2014). *Sentiment analysis algorithms and applications: a survey*. Ain Shams Eng. Journal.

Naseem, U., Razzak, K., Musial, K., & Imran, N. (2020). Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis. *Future Generation Computer Systems*.

Ng, C. Y., & Law, K. M. (2019). Investigating Consumer Preferences on Product Designs by Analysing Opinions from Social Networks using Evidential Reasoning. *Computers & Industrial Engineering*, 106180.

Öztürk, M. M., Cavusoglu, U., & Zengin, A. (2015). A novel defect prediction method for web pages using k-means++. *Expert Systems with Applications*, *42*(19), 6496–6506. doi:10.1016/j.eswa.2015.03.013

Pejić Bach, M., Dumičić, K., Jaković, B., Nikolić, H., & Žmuk, B. (2018). Exploring impact of economic cost of violence on internationalisation: Cluster analysis approach. *International Journal of Engineering Business Management*, *10*, 1847979018771244. doi:10.1177/1847979018771244

Ravi, K., Ravi, V., & Krishna, P. (2017). Fuzzy formal concept analysis based opinion mining for CRM in financial services, Research article. *Applied Soft Computing*, *60*, 786–807. doi:10.1016/j.asoc.2017.05.028

Riaz, S., Fatima, M., & Kamran, M. (2017). *Cluster Computing*. Advance online publication. doi:10.1007/s10586-017-1077-z

Saeed, Z., Abbasi, R. A., Sadaf, A., Razzak, M. I., & Xu, G. (2018). Text stream to temporal network-a dynamic heartbeat graph to detect emerging events on Twitter. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer.

Shindler, M. (2008). *Approximation algorithms for the metric k-median problem. Written Qualifying Exam Paper*. University of California.

Silva, N. F. F. D., Coletta, L. F., & Hruschka, E. R. (2016). Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning. *ACM Computing Surveys*, *49*(1), 49–15. doi:10.1145/2932708

Subirats, L., Reguera, N., Bañón, A., Gómez-Zúñiga, B., Minguillón, J., & Armayones, M. (2018). Mining facebook data of people with rare diseases: A content-based and temporal analysis. *International Journal of Environmental Research and Public Health*, *15*(9), 1877. doi:10.3390/ijerph15091877 PMID:30200209

Suharyanti, Y., Subagyo, , Masruroh, N. A., & Bastian, I. (2017). A multiple triangulation analysis on the role of product development activities on product success. *International Journal of Engineering Business Management*, *9*, 1847979017723544. doi:10.1177/1847979017723544

Thet, T., Na, J., & Khoo, C. (2008). Sentiment classification of movie reviews using multiple perspectives. *Proceedings of the international conference on Asian digital libraries (ICADL)*, 184–193. doi:10.1007/978-3-540-89533-6_19

Wang, W. M., Li, Z., Tian, Z. G., Wang, J. W., & Cheng, M. N. (2018). Extracting and summarizing affective features and responses from online product descriptions and reviews: A Kansei text mining approach. *Engineering Applications of Artificial Intelligence*, *73*, 149–162. doi:10.1016/j.engappai.2018.05.005

Yadollahi, A., Shahraki, A. G., & Zaiane, O. R. (2017). Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys*, *50*(2), 1–33. doi:10.1145/3057270

Yoo, S. Y., Song, J. I., & Jeong, O. R. (2018, September). Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications*, *105*(1), 102–111. doi:10.1016/j.eswa.2018.03.055

APPENDIX A

Use The k-Means++ Algorithm To Determine The Centroids

The *k*-means algorithm supplemented by *k*-means++ is employed to categorise the extracted posts into various clusters for each product model. As mentioned in section 2.2.2, the results generated by the standard *k*-means algorithm are sensitive to the initial values of centroids assigned, and the proposed approach uses *k*-means++ for determining the initial seeds to improve the reliability of the results.

Suppose we have 10 data points (d_1, d_2, d_3,…, d_{10}) for the model M1 and each data point can be represented by a vector with 6 dimensions as shown in Table 5. Firstly, d_3 = {0, 0, 2, 0, 0, 0} is randomly picked as the first centroid among the 10 data points. Second, for each data point, calculate the squared distances between the data and centroid. The results are listed in the column "Dist(d_3)^2". Third, a new centroid is selected randomly according to the cumulative probability distribution of "Dist(d_3)^2". Here, d_8 is selected based on a random number, 95, which is within an interval 0 to 129. The squared distances and the cumulative probability distribution are then calculated and listed in the column "Dist(d_8)^2". To exclude those data points which have already been selected as centroids in the previous rounds, the minimum values between "Dist(d_3)^2" and "Dist(d_8)^2" is chosen for calculating the column "Cum. Dist(d_8) ^{2}". The next step is to generate new random numbers for determining the next centroids and calculate the squared distances as well as compute the cumulative probability distributions until the initial values for all the *k* centroids are determined. The centroids for clusters A1, A2, …, A6 determined by *k-means*++ are d_{10}, d_6, d_3, d_2, d_7, and d_8 respectively.

APPENDIX B

Apply The K-Means Algorithm For Obtaining The Criteria Weights

The adoption of *k*-means algorithms can be used to categorise the data points into corresponding clusters by minimising the sum of squared differences within the clusters. Using d_1 as an example, the Euclidean distance between d_1 and the six centroids are first calculated, and the corresponding distances are 5.099, 1.0, 4.583, 5.0, 5.657, and 4.583. The minimum value is 1.0, and therefore, d_1 is initially categorised into cluster A2. The remaining data points can be partitioned into the clusters by repeating this procedure. The initial solution can be obtained with all the data points clustered into the criteria. The next step is, to sum up, the minimum distances of all data points and look up the optimum clustering solution based on the objective of minimising the sum of distances. The values of centroids would then be updated immediately using the equation (2) if a new minimum distance can be found in that particular iteration. The iterations would be terminated until no further minimum value can be obtained. After looping the equations (1) and (2) for the 10 data points, the solution can be obtained {A1: d_9 and d_{10}; A2: d_1 and d_6; A3: d_3; A4: d_2; A5: d_4 and d_7; A6: d_5 and d_8}. Once the iterations of the *k*-means algorithms for each product model (M1, M2, and M3) are completed, the posts are then categorised into the six assessment criteria. The results can be found in Table 2 under section 4.2.3.

Calculate The Sentiment Scores of Text posts and use The k-Means/ k-Means++ Algorithms To Categorise Sentiment Polarities

The sentiment analysis on the extracted text posts is conducted by searching the sentiment words using the lexical resource databases. The SentiWordNet (Guerini et al., 2013) is used here. A post extracted as "I" "love" "the" "performance" "specially" "its" "photo" "quality", the positive sentiment words are "love", "performance", and "quality" and the positive, negative, and objective sentiment strengths

Table 5. Dataset for the illustration of the procedure of k-means++

Data_ no. (M1)	A1	A2	A3	A4	A5	A6	Dist(d_3)^2	Cum. Dist(d_3)^2	Dist(d_8)^2	Cum. Dist(d_8)^2
d_1	1	4	0	0	0	0	21	21	21	21
d_2	1	0	0	3	0	0	14	35	14	35
d_3	0	0	2	0	0	0	0	35	8	35
d_4	0	0	0	0	3	0	13	48	13	48
d_5	1	0	0	0	0	2	9	57	1	49
d_6	1	3	0	0	0	0	14	71	14	63
d_7	1	0	0	0	4	0	21	92	21	84
d_8	0	0	0	0	0	2	8	100	0	84
d_9	2	0	0	0	0	0	8	108	8	92
d_{10}	4	0	0	0	1	0	21	129	21	113

are 0.375, 0, and 2.625 respectively. A vector can be defined, given in equation (4), for representing the sentiment strengths of each post,

$$s_i = \{P_i, N_i, O_i\} \text{ i=1,2,3…, n} \qquad (4)$$

where s_i is a 3-dimensional vector for representing sentiment strengths, P, N, O are denoted as the sum of positivity, negativity, and objectivity sentiment strengths respectively; n is the number of text posts extracted of a target product model.

After the determination of the centroids, the posts of each product model are categorised into three polarities by iterating the *k*-means and *k*-means++ algorithms. Thus, the posts can then be categorised. The normalised weights have been presented in Table 3.

Calculate The Overall Weights and Prioritise Products

The final step is to rank the product models by combining the weights obtained. It can be done simply by multiplying the criteria weights with the weights of sentiment polarities. For multiplying the values, a 3-dimensional vector, *f*, is introduced: f = {*fp*, *fn*, *fo*}, where *fp* = *fn* (–1), and *fp* = 1–*fo*, where *fp* is a within [0, 1] determined by the decision-maker subjectively. Taking M_1 as an example, the normalized weights obtained are, A_1=0.1902 and s_{A1} = {0.3317, 0.5463, 0.122}. Assuming *fp* = 0.5, the factor, *f*, would then be equal to {0.5, –0.5, 0.5}, and we can obtain -0.0088 for the criterion A_1. By repeating the multiplication processes, the overall weights of all assessment criteria are 0.1948, 0.1154, and 0.043 for the three models, respectively. That means, based on the result summarised from user-generated comments, M_1 is considered as the best model, then followed by M_2 and M_3. Table 4 summarises the weights for the target product models. Besides, a sensitivity analysis is conducted to verify whether the final ranking is sensitive to the value of *fp*.

C. Y. Ng received his Ph.D. in Systems Engineering and Engineering Management. He managed many new product development projects before joining the academia. His current research interests include logistic management, meta-heuristic search algorithms, design for environment, and product life cycle assessment. He is a member of the Institution of Engineering Designers and the Society of Environmental Engineers. He is also a Chartered Environmentalist.

Kris M. Y. Law is currently an Associate Professor at the School of Engineering, Deakin University, Australia. Prior to her joining Deakin University, she was a Lecturer at the Department of Industrial and Systems Engineering, Hong Kong Polytechnic University. She currently also holds a Docentship (Adjunct Professorship) in the Department of Industrial Engineering and Management, Oulu University in Finland. She undertook a post-doctoral research scholarship and was a Visiting Researcher at the Graduate Institute of Industrial Engineering, National Taiwan University in 2009–2011. Her expertise lies in organizational learning and development, technology and innovation management, technology-based entrepreneurship, project management and engineering education. She has published in journals and edited several books in the areas of performance management, organizational learning, technology and knowledge management.