Understanding Brands with Visualization and Keywords from eWOM using Distributed Representation

Hoseong Yang
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul, Korea
hoseong@dm.snu.ac.kr

Sungzoon Cho
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul, Korea
zoon@snu.ac.kr

ABSTRACT

We propose a method that transforms brands into vectors using distributed representation for visualizing relationships of brands and extracting keywords. Our proposed model, Brand2Vec, can be applied to both calculating brand similarities and extracting keywords because brand information and text data are used at the same time when training vectors using neural network. Brand2Vec is also an objective and repeatable model because this model has been observed to be robust to parameters, and does not require many preprocessing steps. The relationships of brands can be visualized using hierarchical clustering and t-SNE using trained brand vectors. In addition, it is able to extract discriminative keywords using trained distributed representation of brands. In this paper, we demonstrate the case of famous desktop brands such as Apple and Microsoft, and we validate it qualitatively. We predict that this method can be expanded to entertainers, politicians, places and et cetera if we are able to use data from social media websites such as Facebook, Twitter and Youtube.

Keywords

eWOM, distributed representation, brand management, perceptual map

1. INTRODUCTION

Customer perception is important concept in marketing. Particularly, if there are various products with similar specifications, the concept positioned in the consumer's mind has a big influence on purchasing products. Thus, many companies consider it important to understand consumer perception of own brand and that of competitors.

Perceptual map is a visualization tool to understand the consumer perception of brands. Also called "positioning map", it shows relative relationships of brands which are positioned in consumers' mind.

The prevalent method of drawing perceptual map is collecting data from a survey and visualizing the relationship of brands using MDS (Multidimensional Scaling) or correspondence Analysis [3, 10]. Survey-based method is easy to analyze and understand. However, it has weakness in grasping the consumer perception objectively because human intervention is involved in and the number of sample is limited due to the constraints of time, space and money.

To overcome above shortcomings, many researchers and managers have tried to understand consumer perception using electronic word of mouth (eWOM) [2, 8]. eWOM is defined as any statements consumers share via the Internet (e.g., web sites, social networks, instant messages, online reviews) about an event, product, service, brand or company [9].

We can collect eWOM data with low cost from various kinds of online communities to extract live opinions about product or service as it is voluntarily, spontaneously and continuously updated. Also, eWOM contains a variety of information such as written time, rating, user location and shared count. Above all, eWOM can exert strong influence on the brand image because it has higher credibility, empathy and persuasive on customers than marketer-created source [1, 7].

Previous studies have utilized the number of times that brand is mentioned in eWOM [5, 17]. However, it is difficult to calculate similarities between the brands using the counting information. Furthermore, we cannot utilize live and abundant information which is given by consumer opinion. Also, in this counting based method, we have to build each model separately because approaches for visualizing perceptual map and extracting keywords are completely different in general.

In this paper, we suggest the Brand2Vec model which transforms brands into distributed vectors and its applications. Our approach is inspired by Word2Vec [15] and Distributed Memory model [12]. These are the algorithms that represent a word or document as a distributed vector with neural network. While existing eWOM applications use the number of times that brand is mentioned in corpus, the Brand2Vec method utilizes every context information in corpus. And it is possible to calculate similarities between brands objectively because this method does not require many preprocessing steps before training vectors.

By using these properties, we observe that visualizing the brand relationships is possible using hierarchical clustering and t-SNE. Also, we propose keyword extraction method by calculating similarities between brands and words for understanding characteristics of brands. In this research, we apply our method to brands such as Apple and Microsoft and validate the result qualitatively.

The main contributions of this paper are as follows: (1) By using the property which is able to calculate similarities between brands with less preprocessing steps, visualizing hierarchical relationships and positioning of brands is possible. (2) It provides interpretability by automatically extracting keywords using distance measure. (3) With parameter search experiments on 125 combinations of parameter sets, we show that this model is robust to parameters. (4) To best of our knowledge, our work is the first approach that addresses distributed representation on the application of brand management. (5) This methods can be easily extended to other objects such as a movie, entertainer and politician.

This paper is organized as follows. Next section provides a brief literature review on the visualizing brands using eWOM and distributed representation. In Section 3, we propose our Brand2Vec method for brand representation and visualization methods. In Section 4, we describe the dataset and provide experiment results of parameter search, visualization and keyword extraction. We conclude in Section 5 with some discussion and directions for future work.

2. RELATED WORK

2.1 Perceptual Map using eWOM

eWOM is easily collected compared to the survey data and it reflects live opinions about brands. Due to these advantages, there have been many attempts to use this data in the business field.

A methodology for visualizing luxury wine brands was described in [17], based on the number of times that brand was mentioned. They used correspondence analysis for dimension reduction. Another research collected the number of brand was mentioned in news website and Twitter, and visualized brand positioning using MDS (Multidimensional scaling) and Minimum spanning tree [5].

Both research considered only the number of times brands were mentioned in eWOM. That is, they do not provide interpretability. Besides, it is not reasonable and objective method because they use the similarity measure as the number of times brand is mentioned in eWOM, which lose a lot of information.

Another approach is to use topic modeling with Latent Dirichlet Allocation (LDA) [20]. This research visualized brands using word distribution for each topic which was extracted by LDA. While this approach use all the text in the corpus, it is inappropriate for training large amounts of data because of high computational complexity. Also, LDA model is difficult to reproduce because it involves a lot of subjectivity when extracting topics and is sensitive to parameters. Moreover, this method is limited in understanding the context because extracting keywords is difficult.

However, the Brand2Vec can handle large size corpus and calculate similarities between brands with customer's live opinions mentioned in eWOM. Through this method, we are also able to calculate similarities between words and brands, so we can extract keywords of brands. During the entire process, we can secure repeatability and objectivity because human intervention is minimized.

2.2 Distributed Representation

To apply text data into machine learning algorithms, first thing to do is transforming text data into numbers. We can classify transforming methods into two groups; Discrete and Distributed. Representative method in the discrete representation category is one-hot encoding which represents vectors as combinations of a single one and all the others zero. But this method do not contain semantic and syntactic information of words and cannot calculate similarities between words.

To overcome those limitations, word distributed representation was proposed [18]. Unlike discrete representation, distributed representation can make vectors with containing semantic and syntactic information. For example, the syntactic relationship between "biggest" and "big" is similar to that of "smallest" and "small". So, we can express this relationship with simple algebra as follows.

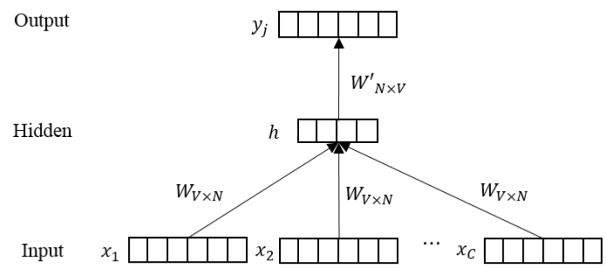

Figure 1: The structure of CBOW model

$$
\begin{aligned}
&vector(\text{"}biggest\text{"}) - vector(\text{"}big\text{"}) \\
&\qquad\approx vector(\text{"}smallest\text{"}) - vector(\text{"}small\text{"})
\end{aligned}
$$

Thanks to those advantages, several approaches have been tried to make appropriate distributed representation of words [4, 16]. However, those approaches are inappropriate for training large dataset because of high computational cost. In 2013, Word2Vec model which was able to train billions of words to vectors with simple neural network was proposed [15]. Word2Vec do not require many preprocessing steps if data size is sufficient and the vectors learned by Word2Vec model encode linguistic regularities and patterns. Due to these advantages, Word2Vec model has been actively applied to various fields of NLP tasks such as sentiment analysis and machine translation [6, 11].

Word2Vec model is either Continuous Bag-of-Word (CBOW) model which predicts next words with a given word or Skip-gram model which predicts next word with given context size words. Figure 1 shows brief structure of CBOW model.

As shown in Figure 1, CBOW model is composed of input nodes C and one hidden layer without activation function. Each of the input nodes x_i is one-hot encoding vector and V dimension. Hidden layer h is N dimension, equivalent to word dimension.

$W_{V\times N}$ is a weight matrix between input layer and hidden layer, $W'_{N\times V}$ is a weight matrix between hidden and output layer. Objective function of the Word2Vec is to maximize the probability of predicting next word given context words as equation 1.

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p(w_t|w_{t+j}) \tag{1}$$

where T is total number of words, $2c$ is equal to window size, w_t is word vector and $p(w_t|w_{t+j})$ is calculated by softmax function.

Word2Vec is unsupervised algorithm to represent words as vectors. Furthermore, representing paragraphs or documents as vectors, called Doc2Vec, is proposed [12]. Previously, TF-IDF [19] is prevalent method to represent documents as vectors. However, TF-IDF has a curse of dimension issue when the vocabulary of corpus is large.

To overcome such problems, Doc2Vec that can represent vectors with less dimension than TF-IDF is proved to give

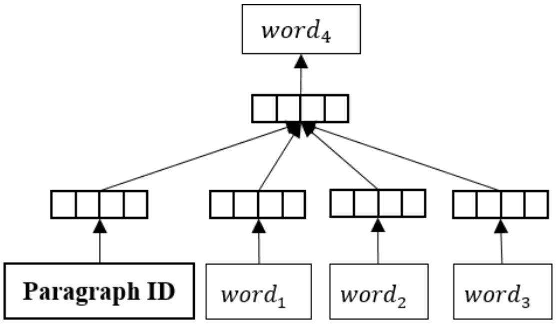

Figure 2: The structure of PV-DM

better performance for sentiment classification than Bag-of-words such as TF-IDF. Doc2Vec model trains vectors of words and paragraph simultaneously as show on Figure 2. Not only high performance in sentiment classification, Doc2Vec shows good performance compared to bag-of-words model in the information retrieval task.

3. PROPOSED METHODS

3.1 Brand2Vec

In this paper, we propose the Brand2Vec model for representing brands as distributed vectors and suggest business applications such as visualizing brand relationships and extracting keywords distinguishing from competitors. The structure of the proposed Brand2Vec model is as shown on Figure 3. Word vectors and brand vectors are simultaneously trained to maximize probability of predicting next words given brand information and context words.

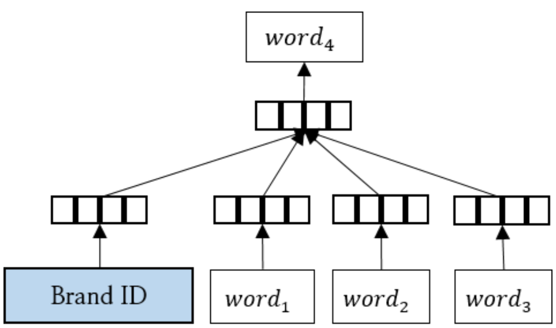

Figure 3: The structure of Brand2Vec

The objective function of Brand2Vec model is as follows:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\leq j\leq c, j\neq 0}\log p(w_t|w_{t+j}) + \frac{1}{T_b}\sum_{i=1}^{T_b}\sum_{k=1}^{T_k}\log p(b_i|w_k) \tag{2}$$

Same as the Word2Vec, weight matrix is trained by backpropagation. T_b is total number of brands and T_k denotes the number of words corresponding to brand b_i. While the left part of objective function is equivalent to that of CBOW model, the right part is added to train brand vectors. The meaning of the right part is conditional probability of the brand b_i given a word w_k mentioned in the corpus. We can learn the vectors of words mentioned in corpus and brands at the same time by backpropagation.

3.2 Visualization method : Hierarchical clustering and t-SNE

Brands can be classified into hierarchical structure. For example, brands can be classified into large group such as manufacturing, service and construction industry. Furthermore, manufacturing group can be divided into cars, furniture, textiles, etc. In this paper, we choose hierarchical clustering for representing such hierarchical structure of brands. There are two main ways to do hierarchical clustering; agglomerative and divisive method. Agglomerative hierarchical clustering is a bottom-up approach which starts clustering by defining each data point as a cluster and combine existing clusters at each step. On the other hand, divisive method is a top-down approach which starts with all data points as one cluster and splits into smaller groups. Generally, divisive methods is not mainly used because the number of divisive cases increases exponentially.

So, we use agglomerative hierarchical clustering method instead of divisive method. Widely used distance measure between two sets of groups are single, complete and average linkage method. In this paper, we use ward's variance minimization algorithm [22]. At each step, ward's method finds the pair of clusters that leads to minimum increase in total within-cluster variance after merging the clusters.

While hierarchical relationship of brands can be visualized by hierarchical clustering, it is difficult to understand relative relationships among several brands. Thus, we employed t-distributed stochastic neighbor embedding (t-SNE) [21] for understanding similarities among many brands. t-SNE is a machine learning algorithm for dimensionality reduction that is particularly well suited for embedding high-dimensional data into a space of two or three dimensions. t-SNE is appropriate for the proposed method because we also use high dimensional vector of brands or documents.

4. RESULTS

4.1 Dataset Description

In this paper, review dataset of amazon.com which is the largest Internet-based retailer in the United States has been used [14]. Attributes of dataset are review text, reviewer ID, written time, price, rating, brand and so on. We can minimize preprocessing step which requires lots of human intervention and time because proposed model is based on Word2Vec. For preprocessing text data, we just have changed to lower case, eliminated special character and parsed by space.

For the parameter search experiments, we select 3,000 brands for each categories (Electronics, Clothing and Beauty). To avoid class imbalance problem, randomly selected 500,000 reviews for each categories are used for training the model.

For the business applications, only the reviews of Electronics category brands are selected which have been written after 2012. About 3 million reviews corresponded to 5,079 brands are used for training the model. Total number of tokens is 21,828,099 and that of unique tokens is 886,768. A document is composed of 74.12 tokens on the average. Figure 4 shows the brands in the Electronics category sorted by the number of written reviews.

4.2 Parameter Search

We evaluate the model parameters by two experiments. In the first experiment, we classify the category of brands using

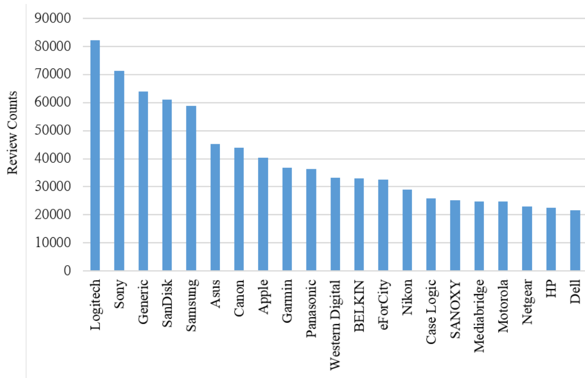

Figure 4: Electronics Brands

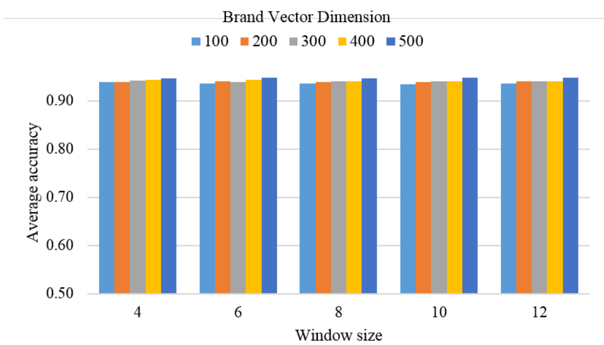

Figure 5: Classification accuracy by parameters

brand vectors which is learned by Brand2Vec. We evaluate the performance, changing window size, vector dimension and training epoch (total 125 combinations). Table 1 shows the 10-fold cross validation result of the first experiment by changing window size and vector dimension when training epoch is 10. Figure 5 shows the same results visualized by bar graph.

Table 1: Classification accuracy by parameters

		Brand vector dimension				
		100	200	300	400	500
Window Size	4	0.9397	0.9390	0.9421	0.9440	0.9462
	6	0.9364	0.9405	0.9397	0.9443	0.9484
	8	0.9357	0.9399	0.9405	0.9402	0.9467
	10	0.9341	0.9389	0.9413	0.9407	0.9486
	12	0.9362	0.9404	0.9411	0.9406	**0.9490**

This result shows that even if there is a little variance from 93 to 95%, the proposed model has robustness in parameters. Moreover, even though the experiment is 3-class classification problem, accuracy is higher than 95%, which means characteristics of each brands are well reflected on the distributed vectors.

Parameter setting when window size is 12 and dimension is 500 shows the highest accuracy. In this case, accuracy variation by the changing training epoch is as shown on the Figure 6. Performance improved by increasing the number of epoch and six time is enough to train vectors.

In the second experiment, we represent each reviews as vectors using Doc2Vec for evaluating the vector quality in a document level. We use 10-fold cross validation for checking the performance of the model, changing three parameters (total 125 combinations). Table 2 and Figure 7 shows the results of the experiment 2, changing window size and vector dimension.

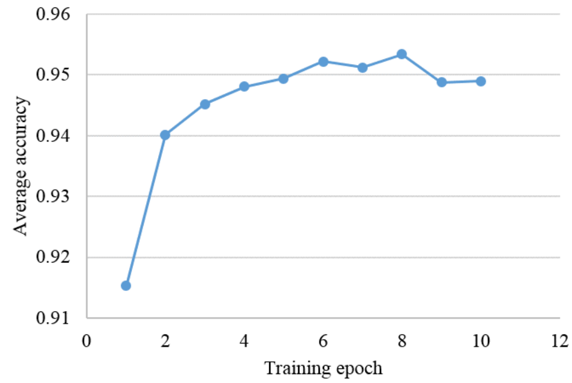

Figure 6: Change of accuracy by the training epoch

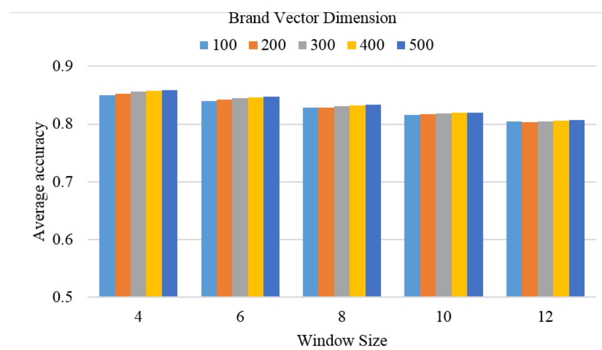

Figure 7: Classification accuracy by parameters

Table 2: Classification accuracy by parameters

		Document vector dimension				
		100	200	300	400	500
Window Size	4	0.8503	0.8527	0.8558	0.8577	**0.8589**
	6	0.8400	0.8420	0.8445	0.8455	0.8473
	8	0.8287	0.8290	0.8307	0.8327	0.8335
	10	0.8155	0.8165	0.8178	0.8193	0.8202
	12	0.8047	0.8030	0.8046	0.8061	0.8074

Similar to the first experiment, the accuracy variation of the second experiment also low from 0.8 to 0.86. While the performance increases as the dimension larger like the first experiment, the smaller window size shows the better accuracy in the second experiment. The best parameter set in the second experiment is window size is 4 and dimension is 500. In this case, accuracy change by training epoch is as shown on Figure 8.

Judging by the results of two experiments, documents and brands are both well represented when the dimension is large. On the other hand, the better performance when window size is large if each review is represented as a vector. But the better performance when window size is smaller if a brand is represented as a vector.

Considering both cases, we choose the dimension as 500,

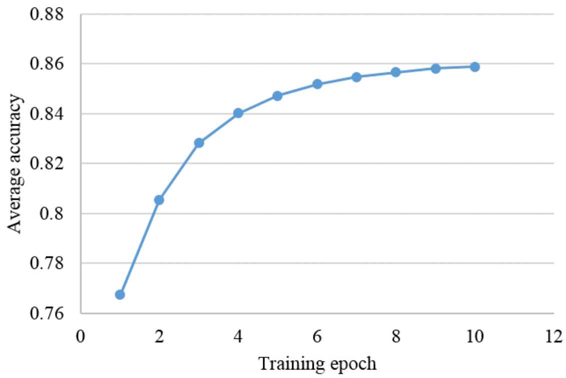

Figure 8: Change of accuracy by training epoch

window size as 8 which is the medium of 4 and 12. We select training epoch as 10 because 10 epochs is enough to train vectors in both cases.

4.3 Brand visualization

Based on the parameter search result, we train Brand2Vec model using about 3 million reviews in Electronics category products. Calculating similarities between brands is possible because each of brands is represented as a distributed vector which contains the characteristics of the brand. For example, Table 3 lists similar brand vector of Samsung and Canon.

Table 3: Similar brands with Samsung and Canon vector

Samsung		Canon	
Brand Vector	Similarities	Brand Vector	Similarities
Acer	0.5991	Nikon	0.8371
Toshiba	0.5541	Focus Camera	0.7949
Proscan	0.5025	Fujifilm	0.7776
Lenovo	0.4961	Sigma	0.7640
Sharp	0.4939	Pentax	0.7629
HP	0.4935	Tamron	0.7615
TCL	0.4666	Tokina	0.7094
Dell	0.4654	Rokinon	0.7041
VIZIO	0.4636	SSE	0.6905
Kocaso	0.4568	Vivitar	0.6746

Through the Table 3, similar vectors with Samsung are Acer and Toshiba, which produce similar products with Samsung. Also, similar brands with Canon are Nikon, Focus Camera, Fujifilm, etc. So, understanding brands relationships is possible through calculating cosine similarities between two brand vectors.

Additionally, the proposed method, Brand2Vec, is able to calculate similarities between words and brands because both information is included when training the model. For example, Table 4 lists the similar brand vectors with the average vector of *computer* and *desktop*.

Similarly, Table 5 shows similar brands with each of the average vector of *earphone* and *headphone* and the average vector of *camera* and *cameras*. The brands in (a), Sennheiser and Monster, are making earphone and headphone and the brands in (b), Canon and Nikon, are camera manufacture brand. Through these results, brands vector contains the property of products and calculating similarities between brands and words is possible.

Table 4: Similar brands with $(computer + desktop)/2$

Brands	Similarities
Dell	0.1418
StarTech	0.1229
SIB	0.1040
HP	0.1035
Cooler Master	0.0996

Table 5: Similar vectors with the average vector of words

Brands	Similarities
Sennheiser	0.2117
Monster	0.1875
JVC	0.1456
Monoprice	0.1346
Bose	0.1233

(a) Similar brands with $(earphone{+}headphone)/2$

Brands	Similarities
Canon	0.1991
Nikon	0.1873
Neewer	0.1410
Case Logic	0.0779
Panasonic	0.0748

(b) Similar brands with $(camera{+}cameras)/2$

As you can see, distributed representation of brands and words learned by the Brand2Vec can calculate similarities not only among brands but also brands and words. Thus, using this property, for identifying relative relationships of more brands, we use hierarchical clustering with ward's method and t-SNE using the vectors of top reviewed 50 brands. For hierarchical clustering, we calculate pair-similarity matrix using cosine similarity first. And the result of clustering is shown on the Figure 9.

On the upper left side of Figure 9, cable manufacture companies, BlueRigger, MediaBridge, Monoprice, are located nearby. Cheetah, VideoSecu are brands which is selling TV mounts also close. So, we named each clusters as `Cables`, `Mounts`. As you can see, dendrogram drawn by brand vectors shows similarities between brands. Moreover, similarities between clusters are also visualized in the dendrogram. For example, if we cluster desktop brands such as "Dell", "HP", etc. and named it as `Desktop` and "Microsoft" and "Logitech" as `PC accessories`, two clusters are located nearby. Therefore, not only similarities between brands but also hierarchical relationships of brands can be visualized if you use brand vectors.

For understanding more relationships of brands, we visualize the same brands into 2 dimension using t-SNE. Like the dendrogram, Figure 10 shows that brands which are producing similar products are located nearby. Especially, brands which are specialized in selling specific products are located closely. For example, TV mounts brands (Cheetah, VideoSecu), Storage brands (Western Digital, Seagate) and Car Audio brands (Pyle, Pioneer) are far from the other brands. Also, it is interesting that Apple is more similar to cell phone accessory brands like BELKIN, Generic, SANOXY than desktop brands. This means that words mentioned in Apple reviews are more similar to accessory brands than those of Samsung. So, you can draw objective perceptual map which contains the characteristics of brands using dendrogram and t-SNE.

4.4 Brand keyword extraction

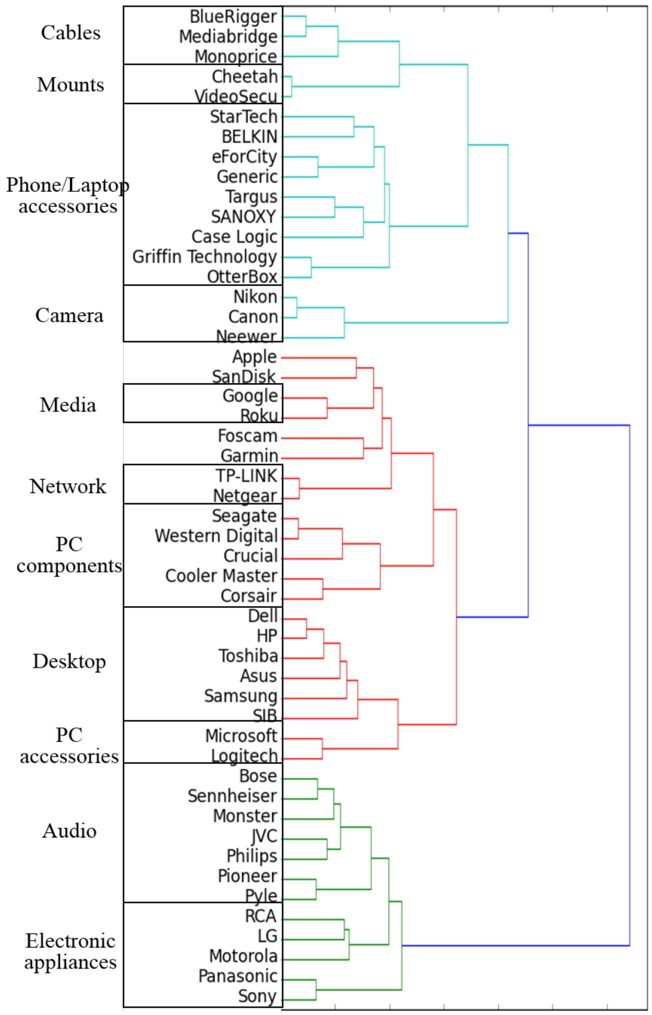

Figure 9: The dendrogram of 50 brand vectors

As we mentioned in the previous chapter, the proposed method, Brand2Vec, is able to calculate similarities between brands and words because they are represented in the same space. Table 6 lists top 10 words which are similar to each vector of Samsung and Canon. The most similar words with Samsung brand vector are “samsung”, “samsung’s”, “smart’, etc. In the case of Canon, “canon”, “zoom” are similar to the Canon brand vector. Through these results, it can be seen that the word which is similar to brand vector is indicating a feature of the brand.

In the actual marketing, the important thing is to find the relative features of the brand compared to its competitors. Therefore, we propose relative keyword extraction method using the above property that calculating similarities between a brand and a word is possible.

For extracting keywords from reviews, we first select 500 adjectives from randomly selected 30,000 reviews. Table 7 shows the examples of extracted adjectives. We select nine famous desktop brands as shown on Table 8.

We introduce a distance measure for calculating relative distance between each of adjectives and brands. First, we denote set of adjective vectors as $A \in \{vector(great), \ldots\}$. And we use $B \in \{vector(Samsung), \ldots\}$ to denote a brand vector set. We define distance measure between a brand and a word as the equation 3.

$$P(B_i|A_j) = \frac{\exp(B_i \cdot A_j)}{\sum_{brands} \exp(B_{brand} \cdot A_j)} \tag{3}$$

The value of the proposed distance measure is between 0 and 1. And the word is more similar to the brand as the value of distance measure is close to one.

For example, the keywords extracted from the Apple and Microsoft reviews are shown on Table 9. We extract keywords that distance measure is close to one. Third column of the Table 9a is frequency which means the number of mentioned in each reviews. Compared to the other brands, the words which is similar to Apple brand are “air”, “classic”, “cute” and so on among 500 adjectives. Extracted keywords, “air”, “classic”, are related to the name of Apple products. And we can extract key opinions about Apple products such as “compatible”, “handy”, “sturdy” and “heavier”. Also, the proposed method can extract keywords even if the frequency of words is low such as “magnetic”, “popular” and “stronger”.

For the case of Microsoft, adjectives about mouse and keyboards such as “vertical”, “stiff” and “ergonomic” are extracted from the Microsoft reviews. This is because we use only the reviews in the Electronics category. Through this results, we can understand that keyboards and mouse is distinguishing products of Microsoft in the Electronic category.

For verifying the extracted keywords, we extracted co-occur words using PMI (Pointwise Mutual Information) [13] and corresponding original reviews. Table 10 shows mean-

Table 6: Words similar to Samsung and Canon brand vectors

Samsung		Canon	
Word Vector	Similarities	Word vector	Similarities
samsung	0.3117	canon	0.4292
samsung’s	0.2840	zoom	0.4077
smart	0.2519	dslr	0.4064
google	0.2493	telephoto	0.3893
ativ	0.2316	nikon	0.3769
lg	0.2297	slr	0.3765
series	0.2272	bodies	0.3626
plasma	0.2257	megapixel	0.3525
smarttv	0.2242	macro	0.3479
viera	0.2187	point-and-shoot	0.3393

Table 7: Top 10 words from 500 adjectives

	Words	Frequency
1	great	9,071
2	good	7,610
3	easy	3,794
4	new	2,863
5	cable	2,434
6	nice	2,143
7	small	2,100
8	first	2,022
9	old	2,020
10	best	1,818

Table 8: Target 9 desktop brands

	Brands	Count
1	Samsung	58,852
2	Asus	45,222
3	Apple	40,282
4	HP	22,582
5	Dell	21,727
6	Microsoft	18,437
7	Toshiba	12,645
8	Acer	9,589
9	Lenovo	9,908

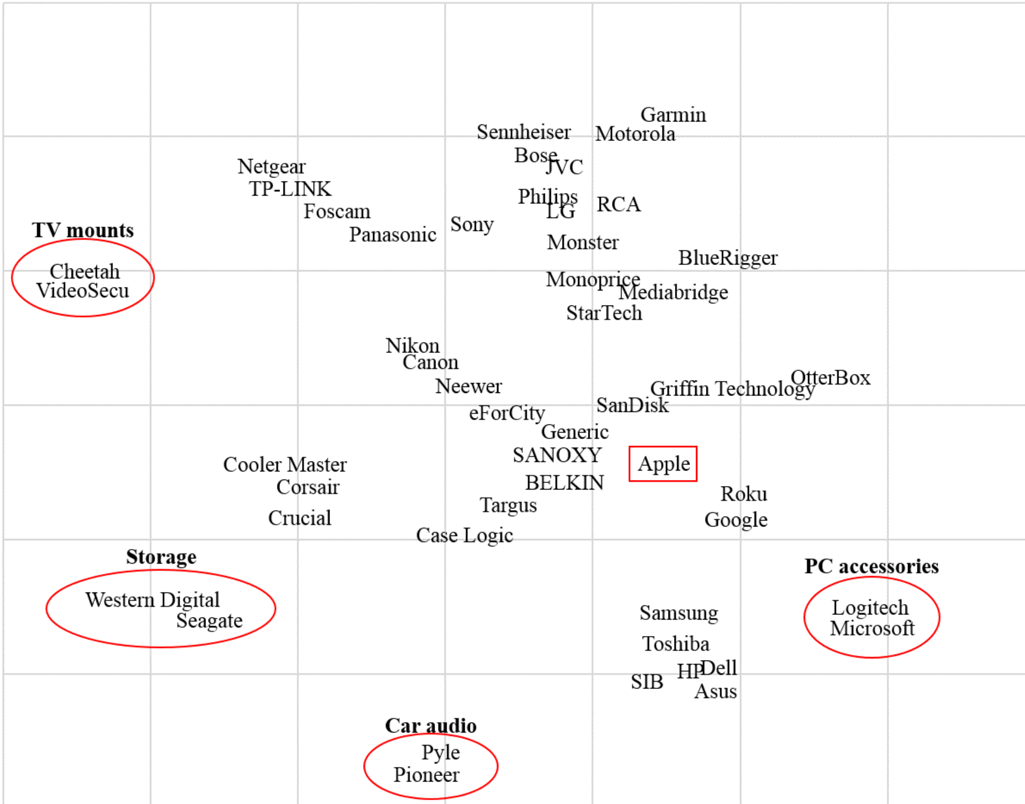

Figure 10: Brand visualization using t-SNE

Table 9: Extracted keywords using the proposed method

Apple		
Words	Measure	Freq
air	0.99	2,144
classic	0.99	706
cute	0.99	118
magnetic	0.99	89
proprietary	0.99	121
popular	0.99	80
compatible	0.99	400
magic	0.99	392
versatile	0.99	95
similar	0.99	384
white	0.99	587
handy	0.99	088
short	0.99	524
sturdy	0.99	221
stronger	0.99	60
substantial	0.99	38
heavier	0.99	126
impossible	0.99	143
protective	0.99	158

(a) Apple Keywords

Microsoft		
Words	Measure	Freq
vertical	0.99	99
stiff	0.99	165
closer	0.99	112
comfortable	0.99	2,034
traditional	0.99	195
key	0.99	2,669
natural	0.99	786
mechanical	0.99	225
couch	0.99	139
ergonomic	0.99	1,403
responsive	0.99	558
soft	0.99	360
harder	0.99	158
smooth	0.99	581
sensitive	0.99	308
uncomfortable	0.99	235
regular	0.99	558
love	0.99	2,991
easier	0.99	508

(b) Microsoft Keywords

ingful neighboring words and original reviews corresponding to the extracted keywords.

You can easily understand that words such as "air", "classic" are keywords of Apple because those are included in the name of Apple products. "compatible" is frequently mentioned with "cable" and "not" in the Apple reviews. Looking at the original reviews, people have written reviews about the compatibility issue many times.

Although "sturdy" and "protective" have different dictionary definition, they are related to the durability issue. In the context, metallic materials of iPad or iPhone give an impression of "sturdy". On the other hand, protective cover or case is required for avoid damage.

In the case of Microsoft, Table 11 shows the keywords, corresponding co-occur words and original reviews. The function of mouse such as "Vertical scrolling" is mentioned, and the keyboard related keywords such as "stiff", "ergonomic" are extracted. Looking at the original reviews, while "stiff" is used in negative reviews, "ergonomic" is used in positive context.

5. CONCLUSIONS

In this paper, we have proposed that Brand2Vec represents brands as distributed vectors. First, we have shown the robustness of Brand2Vec by the two parameter search experiments.

The dendrogram and t-SNE have been used to visualize the relationships of brands. We have visualized the hierarchical clustering process using the dendrogram, showing that not only are the "Dell" and "HP" brands similar, but also that the Desktop cluster, which includes Dell and HP, is similar to the PC accessories cluster which composed of Logitech and Microsoft. The t-SNE method helps to understand the positioning of most brands. Unlike the existing methods, the proposed methods can draw an objective perceptual map using all the text information as mentioned in eWOM.

Brand2Vec is also able to calculate similarities between brands and words because they are represented in the same space. Using such properties, we have proposed a keywords extraction method. We extracted keywords "air", "compatible", "sturdy" from the reviews of Apple, as well as "vertical", "scroll", "stiff" from the reviews of Microsoft, and checked the validity of these keywords through the reading of original reviews.

The limitations of this research is as follows: firstly, the Brand2Vec method is restricted to analysis of various dimensions of brands because it represents a brand as a single vector. Generally, consumers view brands as multidimensional, and people can have different opinions about the price, quality and service of the same brand. This problem can be solved by topic modelling. Secondly, we did not utilize sentiment information when training the model. Review data contains the rating that shows whether the review is positive or negative. We believe that negative and positive information can be visualized if we train the model with rating information.

In the future research, in order to identify the change of consumer perception, dynamic analysis will be possible if such documents are trained by the different time. Data from social media sites such as Facebook and Youtube can be used for training the model. If we use eWOM from social media, we believe that a movie, entertainer, politician, or athlete who is also frequently mentioned in eWOM can be represented as vectors and therefore be used in making decisions.

6. ACKNOWLEDGMENTS

This work was supported by the BK21 Plus Program(Center for Sustainable and Innovative Industrial Systems, Dept. of Industrial Engineering, Seoul National University) funded by the Ministry of Education, Korea (No. 21A20130012638), the National Research Foundation(NRF) grant funded by the Korea government(MSIP) (No. 2011-0030814), and the Institute for Industrial Systems Innovation of SNU.

7. REFERENCES

[1] B. Bickart and R. M. Schindler. Internet forums as influential sources of consumer information. *Journal of interactive marketing*, 15(3):31–40, 2001.

[2] A. S. Cantallops and F. Salvi. New consumer behavior: A review of research on ewom and hotels. *International Journal of Hospitality Management*, 36:41–51, 2014.

[3] I.-P. Chiang, C.-Y. Lin, and K. M. Wang. Building online brand perceptual map. *CyberPsychology & Behavior*, 11(5):607–610, 2008.

[4] R. Collobert and J. Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167. ACM, 2008.

[5] P. Dwyer. Inferring brand proximities from user-generated content. *Journal of Brand Management*, 19(6):467–483, 2012.

[6] A. Frome, G. S. Corrado, J. Shlens, S. Bengio, J. Dean, and T. Mikolov. Devise: A deep visual-semantic embedding model. In *Advances in Neural Information Processing Systems*, pages 2121–2129, 2013.

[7] R. E. Goldsmith and D. Horowitz. Measuring motivations for online opinion seeking. *Journal of interactive advertising*, 6(2):2–14, 2006.

[8] E. Jeong and S. S. Jang. Restaurant experiences triggering positive electronic word-of-mouth (ewom) motivations. *International Journal of Hospitality Management*, 30(2):356–366, 2011.

[9] J. Kietzmann and A. Canhoto. Bittersweet! understanding and managing electronic word of mouth. *Journal of Public Affairs*, 13(2):146–159, 2013.

[10] D. J. Kim, W. G. Kim, and J. S. Han. A perceptual mapping of online travel agencies and preference attributes. *Tourism management*, 28(2):591–603, 2007.

[11] R. Kiros, Y. Zhu, R. R. Salakhutdinov, R. Zemel, R. Urtasun, A. Torralba, and S. Fidler. Skip-thought vectors. In *Advances in Neural Information Processing Systems*, pages 3276–3284, 2015.

[12] Q. V. Le and T. Mikolov. Distributed representations of sentences and documents. *arXiv preprint arXiv:1405.4053*, 2014.

[13] C. D. Manning and H. Schütze. *Foundations of statistical natural language processing*. MIT press, 1999.

[14] J. McAuley, C. Targett, Q. Shi, and A. van den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International*

Table 10: Apple keywords and original reviews

Keywords	Neighbor words	Original reviews
air	macbook, ipad, pro	“it has proven to be a great purchase the macbook **air**” “Being able to see whatever is on my iPad with **air** play can also access all movies music”
compatible	cable, not	”USB connection port is not **compatible** with the connector in my car.” “Apple TV is not **compatible** with amazon.com”
sturdy	feels, seems	“Apple uses aluminum for the back cover of their tablet instead of plastic. It feels **sturdy**.” “The device is incredibly thin and light though it still feels very **sturdy**.”
protective	case, cover	“You’ll want a **protective** case and some screen protectors to keep your iPad looking fresh.” “Make sure you get a **protective** cover for it. so it doesn’t get scratched up.”

Table 11: Microsoft keywords and original reviews

Keywords	Neighbor words	Original reviews
vertical	horizontal, scrolling	“This mouse really helps windows become more usable especially switching through apps **vertical** and horizontal scroll” “I think it’s because the ratio of **vertical** movement to horizontal movement is different from my current mouse”
stiff	bar, buttons	“I just returned a microsoft arc mouse because the buttons were too **stiff**” “I am unable to type with this keyboard because the space bar is so **stiff**”
ergonomic	keyboard, shape	“**Ergonomic** shape is all what I was concerned and I managed to meet my needs” “I’ve always wanted to try the **ergonomic** style keyboards and the price was right”

ACM SIGIR Conference on Research and Development in Information Retrieval, pages 43–52. ACM, 2015.

[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[16] T. Mikolov, S. Kombrink, L. Burget, J. H. Cernocky, and S. Khudanpur. Extensions of recurrent neural network language model. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 5528–5531. IEEE, 2011.

[17] M. Reyneke, L. Pitt, and P. R. Berthon. Luxury wine brand visibility in social media: an exploratory study. *International Journal of Wine Business Research*, 23(1):21–35, 2011.

[18] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Cognitive modeling*, 5:3, 1988.

[19] G. Salton. Automatic text processing: The transformation, analysis, and retrieval of. *Reading: Addison-Wesley*, 1989.

[20] S. Tirunillai and G. J. Tellis. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4):463–479, 2014.

[21] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.

[22] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.