

## Clustering Corporate Brands Based on Social Metrics: A Case Study of the Cosmetic Brands

Ha-Na Kang<sup>\*</sup>, Hye-Ryeon Yong<sup>\*</sup>, Hyun-Seok Hwang<sup>\*\*</sup>

Graduate School of Interaction Design, Hallym Univ.<sup>\*</sup>, Division of Business,  
Hallym Univ.<sup>\*\*</sup>  
{khnnn0607, yong-}@naver.com, hshwang@hallym.ac.kr

**Abstract.** Since the Internet provides a space for expressing and sharing Internet users' ideas or minds, corporate marketers can acquire measurable and actionable insights from the web archives like SNS (Social Network Service). As SNS is widely used to record people's daily activities and thoughts, researcher can gather and read people's mind on SNS through opinion mining or sentiment analysis. In this study we suggest a framework for clustering brand names of cosmetic products using social metrics gathered on the social media - Microblog, Twitter. To cluster the brand names, we calculate the distance of paired brand names based on the total number paired brand names mentioned together. These distances are transformed to project the brand names onto a 2-dimensional and a 3-dimensional space using MDS (Multi-Dimensional Scaling). After projecting the brand names, we found the clusters of the brand names using k-means clustering and identified the characteristics of each cluster.

**Keywords:** Brand clustering, Social Media, Cosmetic products, Multidimensional Scaling, Cluster analysis

### 1 Introduction

Many people record and share their daily minds and opinions on news, products and brands in SNS. Therefore social media such as Facebook, Twitter, and LinkedIn play a crucial role for collecting, transmitting, and sharing opinions of people. Many firms are eager to identify these collective opinions about firms' products or brands. Since customers are likely to share ideas pertaining to firms' products or brand frankly on the web - especially on social media, gathering and mining opinions has become a critical factor for marketers who are trying to identify customer preferences.

In this paper we suggest a framework for clustering corporate brands based on social metrics - the number of brand names mentioned together on microblog and on Twitter. After reviewing the related works of brand positioning and clustering in section 2, we propose a framework for clustering corporate brands - seven steps. To verify the suggested framework, we conduct a case study of the cosmetic products. Social metrics of cosmetic brand names are extracted and clustered using the suggested framework.

## 2 Related Works

The previous research can be classified into two major approaches. The first approach aims to reveal the effect of social media on brand images. The second approach focuses on providing a market positioning framework to support marketing decision.

### 2.1 Brand and Social Media

The De Vries et al. (2012) analyzed the effects of social media marketing using the brand posts - video clip, messages, quizzes and information about the brand - found on 11 international brand fan pages [2]. Their findings explain that many different drivers influence the number of "Like" and the number of comments. However this research focused on the effect of social data on the drivers for social media users' evaluation results.

Bifet & Frank (2011) suggested a real-time data mining technique for analyzing sentiment words in Twitter data streams [3]. Khan et al. (2014) suggested a framework for mining Twitter opinion using three hybrid classification schemes [4]. Nagy & Stamberger (2012) proposed a method to extract sentiment words in disaster micro-blogs [5]. These research mined social opinions and sentiments rather brand positioning.

### 2.2 Brand Clustering

Punj & Moon (2002) suggested a psychological categorization framework to support marketing managers in making major positioning decisions [1]. They developed five sets of propositions connecting positioning options for product with the processing of brand-level information. Chintagunta (1994) proposed a heterogeneous logit model for branding positioning. The suggested model is used to analyze data regarding the purchases of liquid laundry detergents [6].

## 3 Research Framework

To cluster brand names using SNS data, we propose the following procedure depicted in Fig. 1 as a research framework. The basic assumptions of this research are:

- i) The more frequent two brand names are mentioned, the closer two brand names are perceived by consumers
- ii) The distance between two brand names is in inverse proportion to the frequency of co-mention

The first step is choosing an industry and brand names. It would be better to choose an industry having dozens of brand names to cluster brand names properly. Small number of brand names may result in too trivial clusters and excessive number of brand names may also cause difficulty in understanding the result and practical uses. Step 2 is creating a co-mention matrix based on SNS data. If two brand names are

mentioned frequently, people perceive these brands as similar. Hence a co-mention matrix becomes a base table to calculate the similarity (or distance) between brand names. The next step is calculating distance between all pairs of brand names using a co-mention matrix. In Step 4 we project brand names onto 2-dimensional or 3-dimensional space. MDS(Multi-Dimensional Scaling), a statistical method to project instances (brand names) onto a multi-dimensional space. In this research we choose 2-dimensional or 3-dimensional space to interpret projection results intuitively. The fifth step is grouping brand names using a cluster analysis - minimizing inter-cluster distances while maximizing intra-cluster distances. After finding out the clusters, we are needed to identify the characteristics of each cluster for better understanding of the clusters. Final step will be building a brand positioning strategy to deploy marketing actions.

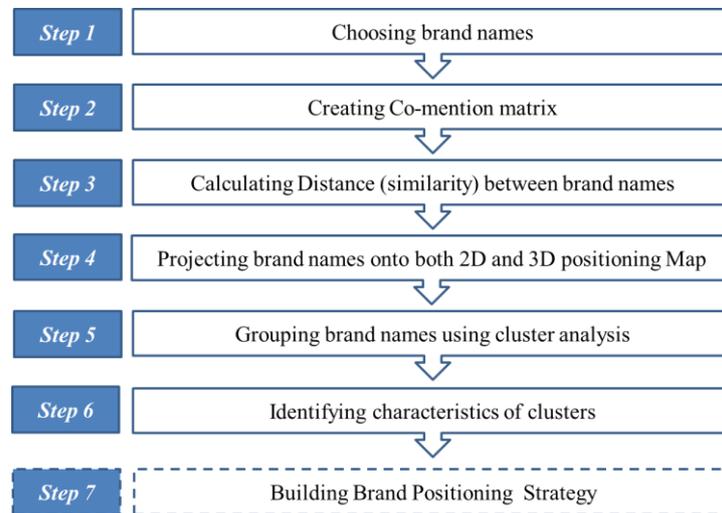


Fig. 1. Research framework

## 4 Case Study

To verify the feasibility of the suggested framework, the cosmetic products are chosen. We sampled the most famous 30 cosmetic brand names. The social metric data from microblogs articles and twitter mentions are gathered for 60 days. In case of no brand name, the posted articles or the twitter mentions are excluded.

If a certain brand name is mentioned with other brand names, we count the numbers of times that the two brand names are mentioned together. Let  $C_{ij}$  be the element of a matrix  $C$ , co-mention matrix, denoting the sum of co-mention number that  $i$ th brand name is mentioned with  $j$ th brand name. Table 1 shows some part of co-mention matrix  $C$ .

**Table 1.** Co-mention matrix C

Brand names	Innisfree	Missha	Etude	Thefaceshop	Mac	...	Nars
Innisfree	-	406	1137	391	195	...	
Missha	105	-	836	247	185	...	
Etude	1137	837	-	630	499	...	323
Thefaceshop	391	246	631	-	116	...	
Mac	0	0	369	0	-	...	608
...	...	...	...	...	...	...	...
Nars	89	97	244		607	...	-

And then the distance matrix D, denoting the distance between two brand names, is can be created from C matrix. Figure 2 shows a formula to create elements,  $d_{ij}$ , in D matrix from the elements in matrix C.

$$d_{ij} = \begin{cases} 1/(c_{ij}+1), & \text{where } i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

**Fig. 2.** Matrix Conversion formula

Table 2 denotes the results of converting the co-mention matrix C, distance matrix D

**Table 2.** Distance matrix D

	Innisfree	Missha	Etude	Theface shop	Mac	...	Nars
Innisfree	0.000	0.002	0.001	0.003	0.005	...	1.000
Missha	0.009	0.000	0.001	0.004	0.005	...	1.000
Etude	0.001	0.001	0.000	0.002	0.002	...	0.003
Thefaceshop	0.003	0.004	0.002	0.000	0.009	...	1.000
Mac	1.000	1.000	0.003	1.000	0.000	...	0.002
...	...	...	...	...	...	...	...
Nars	0.011	0.010	0.004	1.000	0.002	...	0.000

We use MDS (Multidimensional Scaling) to project brand names onto both 2-dimensional and 3-dimensional space. Figure 3 and Figure 4 show the projected results of brand names onto 2-dimensional and 3-dimensional space respectively. The brand names on the space depicted by the same shape belong to the same cluster.

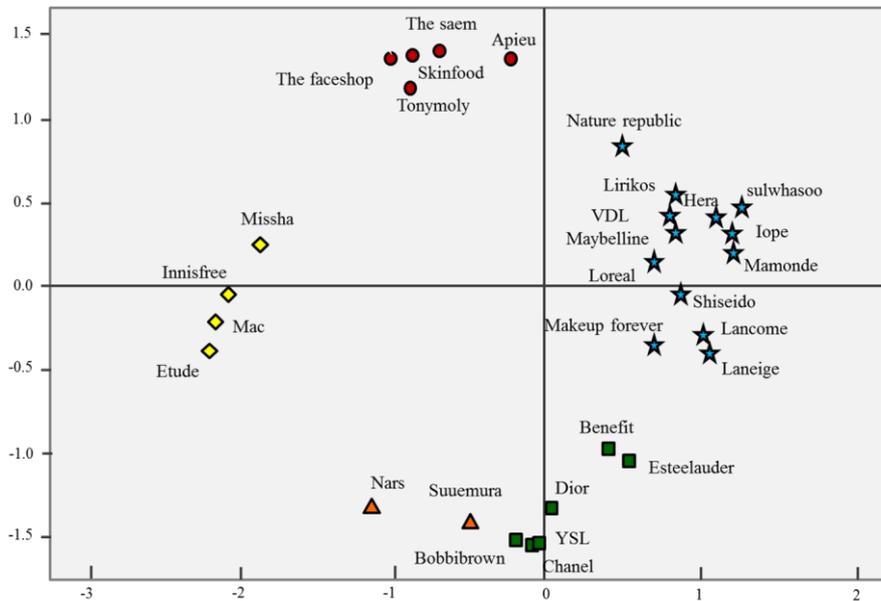


Fig. 3. Projection result onto 2-Dimensional Space

## 5 Conclusion and Future Research Plan

In this study, we used social media data to suggest a framework for identifying the brand name cluster of the cosmetic products using social media data. We expect this research will assist marketers to develop additional insight into customer preferences. However, as described below, our study has some limitations and further research needs to be performed: i) The co-mention matrix C needs to be elaborated. One of the possible compliments is considering sentiment words on the social media as shown in Fig. 1. ii) To satisfy the minimum level of stress value in MDS analysis we need to gather much social data.

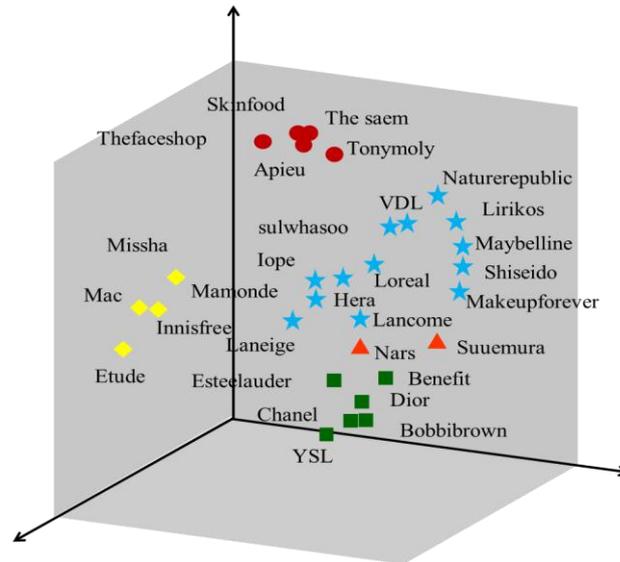


Fig. 4. Space Projection result onto 3-Dimensional Space

## References

1. Punj, G. and Moon: Positioning options for achieving brand association: A psychological categorization framework. *J. of Business Research*. 55, 275--283 (2002)
2. de Vries, L. Gensler, S. and Leeflang, P.S.H.: Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *J. interactive marketing*. 26, 83--91 (2012)
3. Bifet, A., Holmes, G., Pfahringer B.: MOA-Tweet Reader: real-time analysis in twitter Streaming data. *LNCS*, Vol. 6926, pp 46--60. (2011)
4. Khan, F.H., Bashir, S. Qamar, U.: TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision support systems*. 57, 245--257 (2014)
5. Nagy, A., Stamberger J.: Crowd sentiment detection during disasters and crises: Proceedings of the 9th International ISCRAM Conference, pp. 1--9, Vancouver, Canada (2012)
6. Chintagunta, Pradeep K.: Heterogeneous Logit Model Implications for Brand Positioning. *J. marketing research*. 31, 304--311 (2013)