Graph Analysis of the Semantic Interface of the Top Cosmetic Brands by SVD: Use of $\sigma_i \boldsymbol{v}_i$

○MATSUDA, Noriyuki, MACHIDA, Akiko, MIZUNO, Kei
Graduate School of Systems and Information Engineering
University of Tsukuba

1. Introduction

Manufacturers actively employ web pages to appeal to consumers who, in turn, use the pages to gain information to satisfy their concerns or curiosity. That is, they serve as a vital interface between businesses and consumers. Moreover, rival companies and market watchers try to gain information from those web for business purposes. Whether the interface works for B-to-C or B-to-B, its analysis in terms of semantic interface is expected to provide valuable findings to marketers and to stimulate researchers in various fields.

The present analysis is a sequel to Machida, Mizuno and Matsuda' work (2007b) that makes use of $\sigma_i \boldsymbol{u}_i$ resulting from *SVD* (Singular Value Decomposition): $\boldsymbol{A}=\boldsymbol{U\Sigma V}^{\mathrm{T}}$. And, the two together complement our earlier work (Machida, Mizuno and Matsuda, 2007a) in which graphs were constructed on the basis of the cosine similarities computed from the reduced matrices $\boldsymbol{A}^*$ via SVD.

In these studies, we focused on the skincare products, since their presentation in the web and other media, like other cosmetics, illustrate an artistic balance between legal constraints and business interests including aesthetic appeals (see, Machida, Mizuno and Matsuda, 2007a, for more detailed explanation).

Before proceeding further, a brief note on SVD and web seems in order. SVD decomposes any matrix into the product form as $\boldsymbol{A}_{mxn}=\boldsymbol{U}_{mxn}\boldsymbol{\Sigma}_{nxn}\ \boldsymbol{V}_{nxn}^{\mathrm{T}}$ where $\boldsymbol{U}$ and $\boldsymbol{V}$ are ortho-normal bases, and $\boldsymbol{\Sigma}$ is a diagonal matrix containing singular values ordered from the largest to the smallest ($\sigma_1 \geq \sigma_2 \geq \ldots$). By the nature of these matrices, we obtain the following important relationships:

$$\boldsymbol{A}\boldsymbol{v}_i=\sigma_i \boldsymbol{u}_i \text{ and } \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_i=\sigma_i \boldsymbol{v}_i$$

They allow column- and row-space analysis, respectively. In our series of studies, matrix $\boldsymbol{A}$ consisted of frequencies of verbal terms (rows) in various contexts (columns) as explained below.

2. Method

Data—The text data were adopted from the web sites for ELIXIR (SHISEIDO) and SEKKISEI (KOSÉ) during the first week of April, '07. We extracted total 204 and 245 terms across 10 and 12 product items, respectively. The contexts were the direct product of [items]x[slogan, ingredient, benefit, misc]. The frequency matrix $\boldsymbol{A}$ (term by context) for each brand were adjusted by *tf-idf* prior to SVD, all done by software package *R* (see, Machida, Mizuno and Matsuda, 2007a, for more detailed explanation).

ELIXIR **雪肌精**

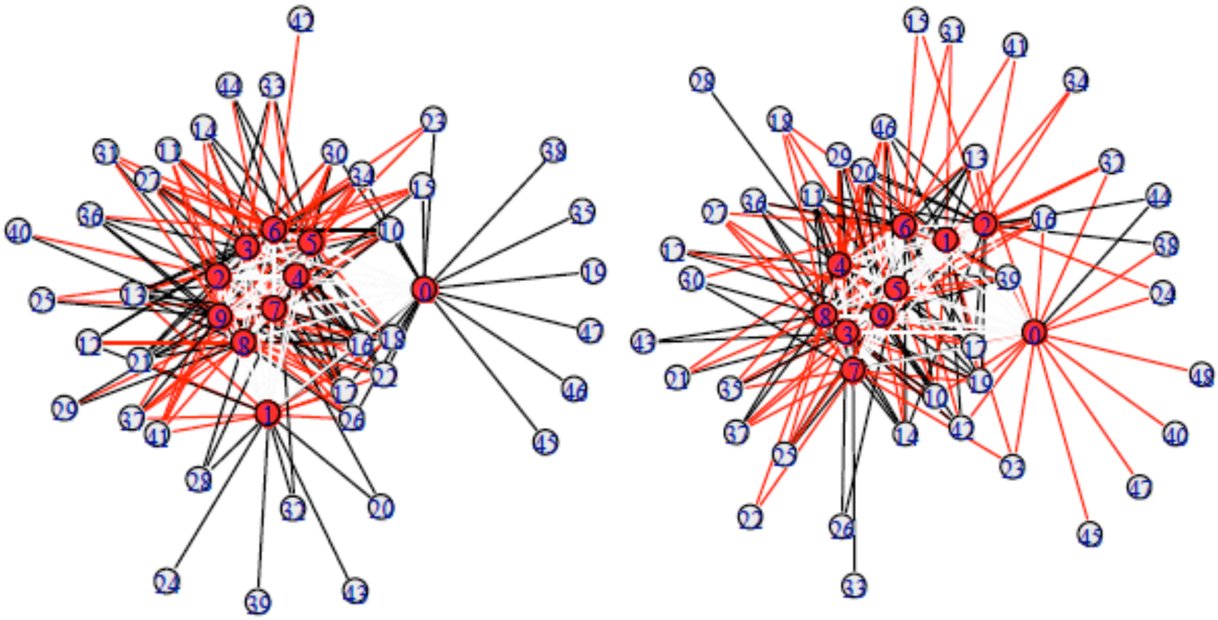
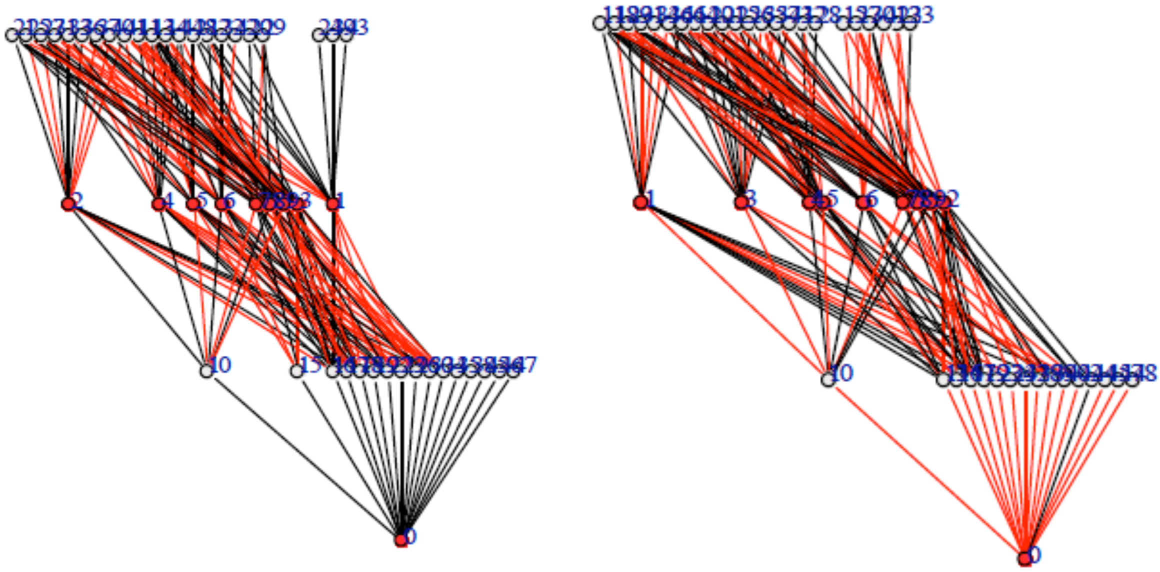


Figure 1. Networks of contexts for ELIXIR (left) and SEKKISEI (right) laid out by Kamada-Kawai's (1989) algorithm



ELIXIR **雪肌精**

Figure 2. Networks of contexts for ELIXIR (left) and SEKKISEI (right) laid out by Reingold-Tilford's (1991) algorithm

Graph construction and analysis—For the sake of simplicity, we derived a matrix $\boldsymbol{B}=\sigma_i \boldsymbol{v}_i$ for the largest 10 singular values. The rows of ***B*** pertain to the contexts. Each column of ***B*** contains "loadings" of the contexts that are analogous to principal component analysis. Again, to reserve simplicity, another matrix ***B**** was derived from ***B***, by selecting contexts whose loadings fell in the high and low ends of the loading distribution in each column—20% each. Hence, ***B**** contains these high- and low-end contexts instead of numbers.

Contexts in a column were, then, treated as vertices of a (sub)graph linked to component C_i (i=1,..,10). We represented the entire structure of the contexts by linking 10 components in order of the magnitude of σ_i, i.e., C_1-C_2-...-C_{10}.

As an aid to visual inspection of the structure, the two algorithms were employed for displaying graphs—Kamada and Kawai's (1981) and Reigngold and Tilford's (1989). The former is an improved spring-model, suitable for an bird'-eye view, while the latter is intended to give a hierarchical vertical one.

Finally, the spin-glass community algorithm (Reichardt and Bornholdt, 2006) was applied in order to detect local communities within which vertexes would be more densely connected among each other than to those lying outside. Note that the results remain non-deterministic, since the algorithm starts with randomly set initial states to minimize the Hamiltonian function. Hence, we repeated the computation several times.

3. Results

The graphs for the two brands laid out by two algorithms are displayed in Figures 1 and 2. The brands were similar in the number of vertices (48—49 for ELIXIR and SEKKISEI), edges (160--160) and density (.142--.136) which is the ratio of the number of edges of the realized graph to that of the possible complete graph. The degree distributions, ranged from 1 to 9, were similar with respect to the mean and variance (4.2—4.1, 7.6—7.1), but somehow differed in mode (1.0—2.0) and median (4.0—3.0). Additionally, they were similar with respect to the contexts at the high and low ends of the distributions: Mask and cream related contexts had the highest degrees, whereas essence and emulsion related ones had the lowest degrees.

The brands did not greatly differ in term of the community indexes. There were four communities with comparable sizes that ranged between 7 to 17. The modularity of ELIXIR tended to be slightly smaller than SEKKISEI's (approximately .220--.230).

Although, the indices were similar between the brands, close inspections of Figure 2 led to interesting contrasts between them. Substantive analysis will be reported at the meeting.

4. Summary

Graph analysis via SVD now seems to be a viable means to studying the semantic interface of commercial products in the web as well as other media where verbal explanations often do not comply with sentence structures. Close examinations of our related works will reveal interesting similarities and differences between the two brands.

References

Kamada, T., & Kawai, S. (1989). An Algorithm for Drawing General Undirected Graphs, *Information Processing Letters*, 31, 7–15, 1989.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.

Machida, A., Mizuno, K., & Matsuda, N. (2007a). Diagnoses of the Semantic Interface of Top Cosmetic Brands on the Web by SVD (Singular Value Decomposition). A paper presented at HCI, International 2007. Beijin**.**

Machida, A., Mizuno, K., & Matsuda, N. (2007b). Graph Inspection of the Semantic Interface of the Top Cosmetic Brands by SVD: Use of $\sigma_i \boldsymbol{u}_i$. A Paper to be presented at the Behavior Metric Society of Japan. Kyoto.

Reingold, E., & Tilford, J (1981). Tidier drawing of trees. IEEE Transactions on Software Engineering, 7(2), 223–228.

Reichardt, J., & Bornholdt, S. (2006). :Statistical Mechanics of Community Detection, Physical Review, E, 74,016110, http://arxiv.org/abs/cond-mat/0603718

Web Sites and Software

Csardi, G. (2006). *IGraph Library*. http://cneurocvs.rmki.kfki.hu/igraph/.

ELIXIR, SHISEIDO: http://www.shiseido.co.jp/eis/index.htm

FDA. (2007). *Cosmetics*. Center for Food Safety and Applied Nutrition. U.S. Food and Drug Administration.: http://www.cfsan.fda.gov/~dms/cos-toc.html

MHLW. (2007). *Drug Legislation*. Ministry of Health, labour and Welfare, Japan: http://wwwhourei.mhlw.go.jp/hourei/index.html

R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

SEKKISEI, KOSÉ: http://www.sekkisei.com/jp/