An Affect-Based Text Mining System for Qualitative Analysis of Japanese Free Text

Makoto Sano

JUSTSYSTEM Corporation
Aoyama bldg. 1-2-3 Kita-Aoyama Minato-ku Tokyo 107-8640 Japan
makoto_sano@justsystem.co.jp

Abstract

Manufacturing companies are concerned about ordinary customers' opinions of their products. This paper presents an affect-based text mining system for Japanese that can aid in the analysis of customers' reviews of commercial products. From free-text product reviews, the system creates an *Adjective-Relation Database* containing relations extracted between adjective and noun phrases. Using a pre-existing lexicon of general affect words together with scaled *Positive/Negative* evaluation, the system provides a quantified evaluation of *modified-noun* phrases appearing in the product survey. In addition, using correspondence analysis techniques, an *Affect Map* reflecting the preferences of target customers is created displaying the relations between user profile data and positive and negative affect-laden words. In a practical experiment, we used the results of such correspondence analysis between high-frequency adjective phrases and customer profiles on a questionnaire-based survey of opinions on a brand-name handbag, to further modify our affect lexicon, improving correspondence results.

Introduction: The Commercial Challenge

In more and more Internet contexts, ordinary customers are expressing their opinions about commercial products (e.g., on consumer-related bulletin boards or in newsgroups). Increasingly, companies are monitoring "the voice of the customer" directly via questionnaires at their web sites. Such product reviews by customers play valuable roles for companies, from providing frank information about preferences to revealing attitudes about competitors. In cases where the *functional* needs of the customer might be satisfied by several alternative products, it is often the *emotional* response of the customer to the product that makes the critical difference. Thus, in the hope of understanding such attitudes, many users of our business systems are especially interested in a qualitative analysis of customers' free-text comments on products. More specifically, users want to know which elements of their products have emotion-invoking effects and, subsequently, how perceived emotional benefits may affect their customers' buying decisions. This paper describes a text mining system for the qualitative analysis of Japanese free text that attempts to address such issues.

Past work on the problem of qualitative analysis of text reflects a variety of approaches. Some researchers in Japan have developed methods to handle qualitative information as image search systems (Watanabe, Nakamura and Nagao 1993; Kiyoki, Kaneko, and Kitagawa 1996; Harada, Itoh, and Nakatani 1999; Mukunoki, Tanaka, and Ikeda 2001). Watanabe et al. (1993) present a search system for painting works with key subjective features extracted from the attached texts. They used the semantic features of all nouns and verbs found in a thesaurus. Harada et al. (1999) studied searching for pictures of chairs with free text queries including some adjectives. They represented subjective features in a parameter space specified for each commercial product, for example, chairs and televisions. They used the semantic differential method for constructing the feature parameter space. However, such methods are not practical for commercial systems because of the prohibitive cost of applying features to all the words used to describe commercial products. A great deal of effort would also be required if users wanted to modify features.

In previous work on subjectivity learning, Hatzivassiloglou and McKeown (1997) presented a method for automatically assigning semantic orientation labels to adjectives. Hatzivassiloglou and Wiebe (2000) described a method for classifying adjectives as *gradable* or *nongradable*. Wiebe (2000) applied the method to subjectivity tagging. Many of these learning methods require large corpora. However, it is difficult and impractical for users to prepare large corpora for each survey if the expected number of resultant reviews is too small to be used for learning.

Pang, Lee and Vaithyanathan (2002) and Turney (2002) have developed methods for classifying reviews. Their approach focused on document-based classification. However, many questionnaires include both "fixed-response" questions and free-text responses or comments on individual topics, designed to capture more detail about customers' opinions (e.g., positive or negative). So it is not always appropriate to classify whether a review (itself) is positive or negative at the document level.

In previous work on affect-based text mining, Subasic and Huettner (2000) used fast analysis and visualization of affect content for decision-making. This work is perhaps

the closest to ours. But they focused on *document-based* affect analysis, not individual survey items or product facets. In contrast, Tateishi, Ishiguro and Fukushima (2001) in their work on *opinion information retrieval* and Morinaga, Yamanishi, Tateishi and Fukushima (2002) in their work on *mining product reputations* focused on affect analysis of the products themselves.

Our users want an affect-based text mining system focused on the elements of their products, not on the products themselves for the purpose of making decisions about new product development or marketing strategies. Thus, to develop a practical commercial system, we focused on partial sentences that include typical adjectival expressions of customers' qualitative comments and focused on *phrase-relation-based* affect analysis. The principal goals of our affect-based text mining system are the following:

1. Minimize user's initial cost to prepare lexical resources;
2. Optimize functions for questionnaire analysis; and
3. Make it easy and efficient to modify lexical resources to accommodate various users' products and customers' preferences.

Creating an Affect Processing System

Definition of Phrases

The key features of our *phrase-relation-based* affect analysis system are *adjective phrases* and *noun phrases*.

Definition of Adjective Phrase. Okada (1985) defines the function of adjectives (in Japanese) as representing the concept of *difference*. We adopt this definition and use the following template to distinguish adjective phrases from other phrases, independently of the syntactic categories of the morphemes that are actually used in the phrase.

○○についていえば，Ａは Ｂより××

Evaluating ○○, A is more ×× (much ××-er) than B.

The phrase that appears as ×× represents an evaluation of the qualitative *difference* between A and B. We define such a phrase as an adjective phrase. For example, in the sentence:

"接客について言えば、A さんは B さんより「親切」だ

("As for receptionists, I'd say A-san is *kinder* than B-san.)",

we can treat *kind* as an adjective phrase.

Note again that in Japanese, as long as a phrase denotes a qualitative *difference* (between two things), it is considered to be an adjective phrase, even if the phrase does not contain an adjective morpheme (e.g., 品がない (*lacking in refinement*); ウキウキする (*cheerfully feel*); 使いづらい (*hard to use*)).

Definition of Modified-Noun Phrase. Okada (1985) observes that nouns, sometimes explicitly, sometimes tacitly, have contextual relations to adjectives. These relations are very important in expressing the qualitative features of objects. We address this in our work by specifically defining the noun phrase in relation to adjective phrase as a *modified-noun* phrase.

Adjective-Relation Database

Structure of Adjective-Relation Database. The *Adjective-Relation Database* has relational information about adjective phrases extracted from the free-text reviews. The database contains adjective phrases, *modified-noun* phrases in relation to the adjective phrases, the free texts as sources of extraction, and customers' individual profiles (if such exist). The individual profiles are used for modifying the affect lexicon. Table 1 presents the database structure and some example entries from the database. By using the database, users can extract adjective phrase relation patterns, parts of free text, and the customers' profiles.

Creating Adjective-Relation Database. The *Adjective-Relation Database* is built from the surface syntactic parser (Zhai, Tong, Milic-Frayling, and Evans 1997) used in the ConceptBase information management system (Fujita 1999), extracting adjective phrases and noun phrases from text after morphological processing. The database builder normalizes the surface forms of the adjective phrases and noun phrases and extracts relations between adjective phrases and *modified-noun* phrases with a dependency grammar parser. The grammar has 122 rules. The parser is similar to the Kudo et al.'s parser (2002) using a *Cascaded Chunking Model*. The builder extracts relations and counts the number of similar/identical patterns. Finally, the builder saves the results as a new *Adjective-Relation Database*. The database must be built in advance as a pre-process of *phrase-relation-based* affect analysis.

Table 1: *Adjective-Relation Database* Example Entries

Phrase Pattern		Free Text	Customer's Profile	
Modified-Noun	**Adjective**		**Sex/Gender**	**Age**
	Novel	Too novel.	Female	30-
	Novel	That's novel but overdone.	Male	40-
letter	Cute	Letters are cute.	Female	20-

Affect Lexicon

Structure of Affect Lexicon. The Affect Lexicon lists each adjective phrase along with a numeric value quantifying the *Positive/Negative* valence of the phrase. Each adjective is also given one of three levels (absolute values) of affect intensity: *high*, *medium*, or *low*. The numeric values are used to calculate the A*ffect Value* of *modified-nouns* in relation to the adjective phrases in the affect lexicon. By using the lexicon, users can get quantified evaluation of *modified-noun* phrases as elements of a product on a survey. Table 2 gives examples of entries in an affect lexicon.

Table 2: Example Affect Lexicon Entries

Adjective Phrase	Positive(+)/ Negative(–)	Intensity
よい (good)	+	middle
悪い (bad)	–	middle
完ぺき (perfect)	+	high
最悪 (worst)	–	high
安定 (stable)	+	low

Creating an Affect Lexicon. In developing the system, we also created a general purpose affect lexicon. This lexicon supports the analysis of *modified-noun* phrases as elements in the description of virtually any type of product. We believe flexibility and generality is very important in affect analysis. Once a lexicon is constructed, it is difficult to maintain it and to extend it to a new domain. Thus, we wanted to create a "lightweight" and flexible core affect lexicon that could serve as a basis for analysis in any domain-specific application.

We first selected general adjective phrases from users' comments on many product-purchase-related information message boards. We then identified the subset of phrases that independently indicate positive or negative evaluations. These are the phrases we included in the lexicon. A lexicographer in our group classified the selected adjectives into the *Positive/Negative* valences and the three levels of affect intensity. The resulting core affect lexicon has 1598 entries. By focusing on such general adjective phrases, we believe we have made a practical and extensible resource that users of our system can easily adapt to individual product surveys, principally by adding the special adjective phrases that may be appropriate to their products and markets.

The Text Mining System

The latest version of our system was released in December 2003 by Justsystem Corporation under the name "*CB Market Intelligence*." All functions of the affect processing system have been implemented in it. Figure 1 shows the system architecture. Users of the system perform affect analysis via a Web browser and the text mining server returns the result of the analysis in the browser. Results are presented visually via an SVG (Scalable Vector Graphics) viewer plugged into the browser. The outputs are generated as an SVG image file by an imaging module on the server.

The text mining server consists of five principal components. The first is the natural language processing (NLP) sub-system that processes free text, whose components include a morphological analysis module, a surface syntax analysis module, and a dependency structure analysis module. The NLP produces the second component, the *Adjective-Relation Database*s built from each product survey. The third component consists of the extraction process that identifies (quantifies) qualitative information related to the customer's profiles. The fourth

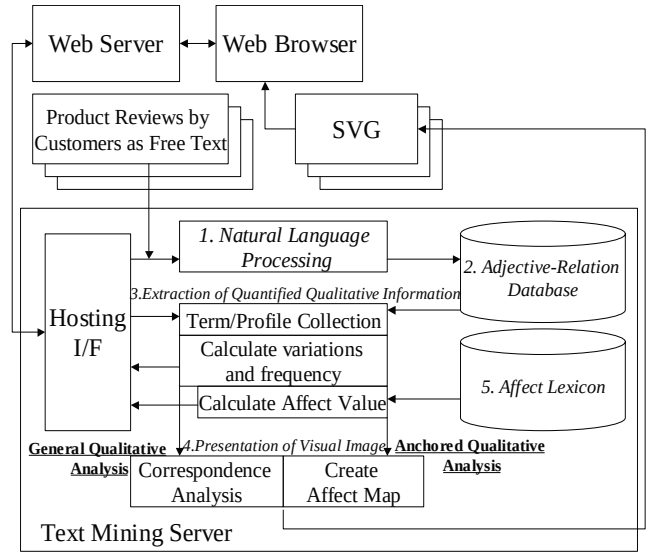

Figure 1: System Architecture

component consists of the module for visualizing outputs. Finally, the fifth component is the affect lexicon used by the extractor (third component).

Extraction of Quantified Qualitative Information

There are two ways to extract qualitative information from the *Adjective-Relation Database*. These are the key to the affect-based text mining. The first is based on *modified-noun* phrases for extracting positive elements or negative elements (of a product on survey). The second is based on adjective phrases for modifying the affect lexicon.

Anchored Qualitative Analysis. In the case of *modified-noun-based* extraction, there are three steps. First, the system lists *modified-noun* phrases and collects modifying adjective phrases from the *Adjective-Relation Database*. Second, the adjective phrase variations and their frequency are calculated for each *modified-noun* phrase. Third, the system retrieves collected adjective phrases from the affect lexicon. If the adjective phrases have numeric *Positive/Negative* values, these values are applied to the collected adjective phrases. The *Affect Values* of *modified-noun* phrases are given by the following formula:

$$A(n) = \frac{\sum_{i=1}^{t} (L(a_i) \times F(a_i))}{\sum_{i=1}^{t} F(a_i)} \qquad (1)$$

where n is a *modified-noun* phrase modified by the adjective phrases: a_1, a_2, a_3,…, a_i,…, and a_t, $L(a_i)$ is the numeric *Positive/Negative* value of a_i, and $F(a_i)$ is the frequency of a_i. We call the *modified-noun-based* extraction "*Anchored Qualitative Analysis*."

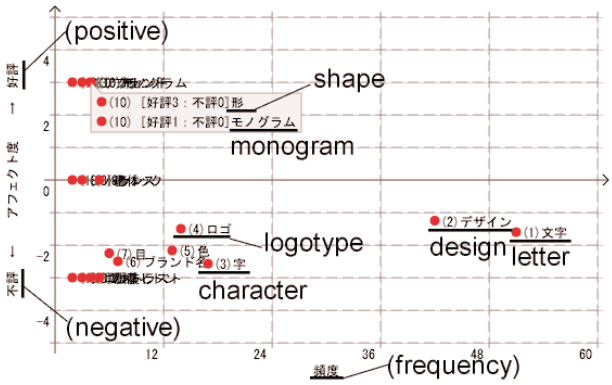

Figure 2: Affect Map with Basic Lexicon plus 32 Additional Entries for the Survey

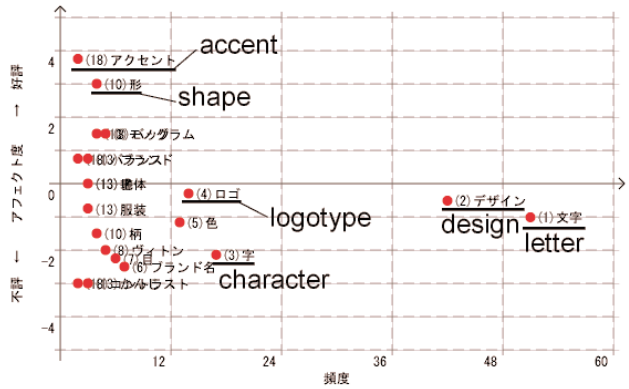

Figure 3: Affect Map with Modified (Extended) Affect Lexicon

General Qualitative Analysis. There are two steps in *adjective-based* extraction. First, the system lists adjective phrases and collects *modified-noun* phrases for each adjective phrase from the *Adjective-Relation Database*. Second, the number of individual profile variations and their frequency are calculated for each adjective phrase to assess the correspondence of adjective phrases and individual profiles. (See brief discussion of *Correspondence Analysis*, below.) We call the *adjective-based* extraction "*General Qualitative Analysis*."

Presentation of Visual Image

We use an *Affect Map* and *Correspondence Analysis* to visualize the relations among affects and attributes.

Affect Map. The *Affect Map* is a two-dimensional graph for visualizing relations (based on frequency) between *modified-noun* phrases and their *Affect Values* as resulting from *Anchored Qualitative Analysis* (as illustrated in Figures 2 and 3). Users can see what kinds of product features are frequently expressed as *modified-noun* phrases by customers. Users can also see which features have positive or negative evaluations.

The analyses in Figures 2 and 3 are based on identical texts with identical phrase extraction results. But the *Affect Values* shown are calculated using different affect lexicons. In particular, we can see the same frequencies for features on both maps' X (horizontal) axes, but the *Affect Values* on the Y (vertical) axes are different, reflecting the effect of using different affect lexicons.

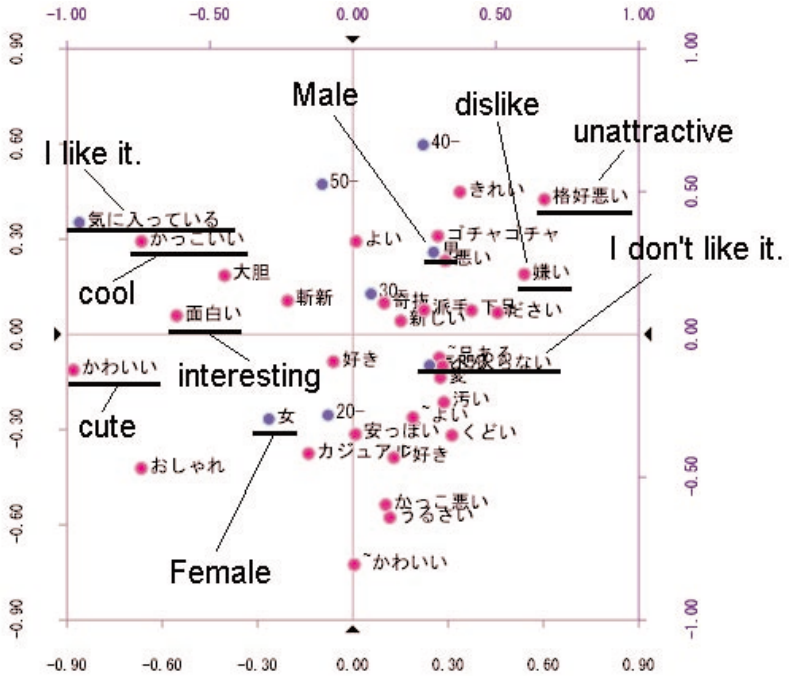

Figure 4: The Result of Correspondence Analysis Based on *General Qualitative Analysis*.

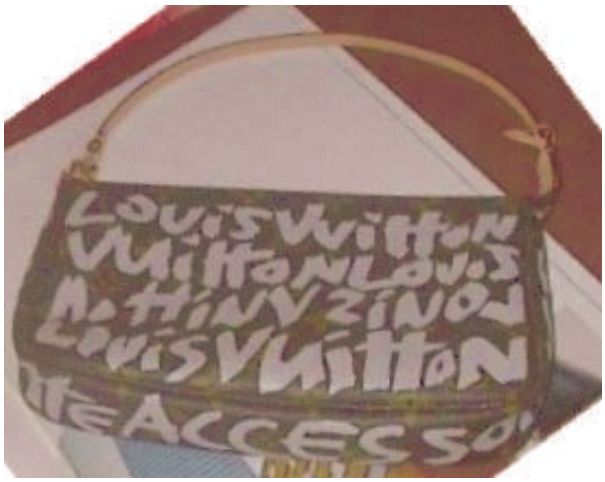

Figure 5: Louis Vuitton Graffiti Pochette

http://www.handbagdiva.com/louisvuitgra1.html

The picture shown here is similar, but not identical, to the one used on the questionnaire.

Correspondence Analysis. *Correspondence Analysis* (Clausen 1998) is a technique for multi-dimensional visualization. The system uses this to show relations of adjective phrases and individual profiles, as resulting from *General Qualitative Analysis*. The items that are grouped on either the left or right side (and on either the upper or lower side) tend to be associated.

Evaluation: Questionnaire Survey

As a concrete example, we present *CB Market Intelligence's* analysis of an actual questionnaire survey.

Questionnaire Design

We designed a survey on a Louis Vuitton handbag marketed in Japan in 2001 under the name "Graffiti." (The handbag is shown in Figure 5; the questionnaire in Figure 6.) We chose this particular product because we expected a variety of answers (expressing likes and dislikes) reflecting definite points of view. There were 512 subjects (293 males and 219 females); all were employees of Justsystem Corporation living in Japan.

Anchored Qualitative Analysis (I)

We analyzed answers to question 1-2 by collecting patterns of relations between adjective phrases and noun phrases. We can see the kinds of handbag features that were mentioned (evaluated) and whether they were regarded as positive or negative based on their association with high-frequency *modified-noun* phrases. Figure 2 is a map of *modified-noun* phrases as a result of *modified-noun-based* phrase extraction from the full-text answers to question 1-2. Table 3 is a list of *modified-noun* phrases from the same extraction results.

Question 1

Imagine that you are considering whether or not to buy the handbag.

1-1 What do you think about the design?

1-2 Why do you feel that way?

1-3 Do you like the handbag?

☐I like it. / ☐ I don't like it.

Question 2

Please let us know something about yourself.

2-1 Occupation : ☐sales department ☐R&D ☐general affairs ☐other

2-2 Service-Area : ☐Tokushima ☐Tokyo ☐Osaka-Nagoya-Fukuoka ☐other

2-3 Sex/Gender : ☐male ☐female

2-4 Age : ☐20- ☐30- ☐40- ☐50-

2-5 Marital Status : ☐married ☐single

Figure 6: Questionnaire

We have added 32 adjective phrases to the core affect lexicon for *Anchored Qualitative Analysis*. The new adjective phrases can be clearly distinguished by the two types of evaluations (*Positive/Negative*) regardless of customers' preferences (e.g., アホっぽい (stupid); 併せやすい (easy to coordinate)). The full answers have 132 adjective phrase variations that should be entries in the affect lexicon for the survey. After adding the 32 adjective phrases, the affect lexicon contains 98 adjective phrases out of the 132 phrases (thus, coverage is 74.2%).

The *modified-noun* phrase, *letter*, at the far right hand side of the map (Figure 2) was the most frequently mentioned property of the handbag, followed by the properties *design*, *character*, and *logotype*.

Quantified evaluation assigned *Affect Values* to the *modified-noun* phrases. The *modified-noun* phrase, *letter*, in the lower half of the map had a negative value. The other frequent phrases, *design*, *character*, *logotype*, were also negative. *Shape* and *monogram*, which ranked 10th in frequency, were among the few features with positive values. One subject's comments on *shape* were as follows: "Though the shape of the bag is tiny and cute, the logotype is a bold design. I have never seen such a design." The positive factor of *shape* seems to affect the subject's final positive evaluation: the answer to 1-3 was *I like it*.

Table 3: The List of Modified-Noun Phrases as a Result of Anchored Qualitative Analysis (I).

Modified-Noun	Frequency	Affect Value
文字 (letter)	51	-1.60
デザイン (design)	39	-1.25
字 (character)	17	-2.57
ロゴ (logotype)	14	-1.50
色 (color)	13	-2.16
ブランド名 (brand name)	7	-2.50
目 (eye)	6	-2.25
バック (bag)	5	0.00
ヴィトン (Vuitton)	5	-3.00
モノグラム (monogram)	4	3.00
形 (shape)	4	3.00
柄 (pattern)	4	-3.00

General Qualitative Analysis

We use correspondence analysis to obtain the associations between the adjective phrases and the individual profiles for the purpose of modifying the affect lexicon.

Figure 4 is an illustration of correspondence analysis. This figure represents an association between the elements of two sets. The first set consists of the top-30 most frequent adjective phrases derived from *adjective-based* phrase extraction over full-text responses to question 1-1. The second set is made up of answers to the questions 1-3 (*Do you like the handbag?)*, 2-2 (*Sex/Gender*) and 2-3 (*Age*). The X axis (horizontal dimension) explains 46.86% of the variance (inertia), and the Y axis (vertical dimension) explains 27.51% of the variance. The sum of the variance is 74.37%, which means most of the associated relations are expressed in the map.

On the left hand side of the plot, we find the label *I-like-it*, one of the answers to question 1-3. On the right hand side, we find *I-don't-like-it*. Thus, the X axis seems to represent the distribution of answers to question 1-3. The label *Female* is on the left hand side, while the label *Male* is on the right hand side. The distribution of answers to question 1-3 is associated with gender. In the upper region, we see the labels *50-* and *40-* for ages. In the lower region, we see the *20-* label. Thus, the Y axis seems to represent distributions of the ages of our subjects.

Examining the association of adjective phrases and axes, we also see that *cute*, *interesting*, and *cool* appear on the left hand side. Thus, there is a possibility that customers, especially *females*, answering question 1-3 with *I-like-it*, express their feeling with such adjective phrases as these on the left hand side. On the right hand side, we find *unattractive* and *dislike*. There is therefore the possibility that customers answering question 1-3 with *I-don't-like-it* express their feelings in such terms. Through such an analysis, we can predict customers' preferences by identification of the typical adjective phrases they use, even if we do not have the direct results of question 1-3.

Modifying Affect Lexicon with the Result of General Qualitative Analysis

When we know the preferences of a specific customer group (such as target customers for a new product), it would be helpful to be able to modify an affect lexicon to enrich our results and calibrate the analysis to the customers' expressed points of view. The result of the correspondence analysis in our survey has 46.86% of the variance on the X axis. That would appear to represent half of the tendencies expressed in the results. The X axis also seems to represent the distribution of answers to question 1-3. Thus, we can assign the scores of the adjective phrases appearing on the X axis to the adjective phrases themselves as modified numeric *Positive/Negative* values. In this way, we can modify an affect lexicon to align it with the observed preferences of our subjects. In our base analysis, we used the core affect lexicon with 32 additional adjective phrase entries. Now, we have added 7 new adjective phrases with numeric values derived from the correspondence analysis (e.g., カジュアル (*casual*): Positive; low). In addition, we have modified the intensity of 12 adjective phrases (e.g., かわいい (*cute*): middle to high; かっこいい (*cool*): middle to high; おしゃれ (*exquisite*): middle to high).

Anchored Qualitative Analysis (II)

Using the affect lexicon modified to the domain of discourse of the subjects talking about the handbag, we can re-analyze the features and themes associated with answers to question 1-2. Figure 3 is a map of *modified-noun* phrases resulting from *modified-noun-based* phrase extraction over the full text of responses to question 1-2. Table 4 is a list of *modified-noun* phrases based on the same result. Though many of the high-frequency *modified-noun* phrases still have negative valence, compared to the very negative values in Table 3, there is a clear shift in *Affect Values* toward the positive end of the scale. This is because the preferences of the respondents who answered question 1-3 with *I-like-it* have now been reflected in the affect lexicon. The positive factors in their feelings seem to affect the final evaluation, shifted toward the *modified-noun* phrases. The *modified-noun* phrase *bag* was given an overall positive score. (Previously, based on the analysis under the attributes and values as reflected in Table 3, *bag* had scored "neutral.")

Effect of Modifying of Affect Lexicon

To assess the effect of modifying the affect lexicon, we decided to test the accuracy of *Anchored Qualitative Analysis* under the two alternative affect lexicons—the affect lexicon with the additional 32 adjective phrases and the modified affect lexicon. For this purpose, we selected only reviews by female subjects who answered question 1-3 (*Do you like the handbag?*) affirmatively (*I like it*.). First, using *Anchored Qualitative Analysis*, we collected patterns of relations between adjective phrases and noun phrases

from the answers to question 1-2. Second, we counted up the number of adjective phrases that received *Positive* numeric values (as determined by *modified-noun* phrases) with each *modified-noun*. Third, we calculated the correlation between the valence of *modified-noun* phrases and the answers (of *I-like-it*) to question 1-3 (as the correct answer). This correlation is 0.440 for the analysis under the first lexicon (core + 32 additional entries)—a poor result—but 0.846 for the analysis under the second lexicon (consisting of the modified affects). The result proves that modifying the affect lexicon improves the accuracy of *Anchored Qualitative Analysis*.

Table 4: The List of Modified-Noun Phrases as a Result of Anchored Qualitative Analysis (II).

Modified-Noun	Frequency	Affect Value
文字 (letter)	51	-1.01
デザイン (design)	39	0.52
字 (character)	17	-2.14
ロゴ (logotype)	14	-0.30
色 (color)	13	-1.17
ブランド名 (brand name)	7	-2.50
目 (eye)	6	-2.25
バック (bag)	5	1.50
ヴィトン (Vuitton)	5	-2.00
モノグラム (monogram)	4	1.50
形 (shape)	4	3.00
柄 (pattern)	4	-1.50
かんじ (feeling)	3	-3.00
ブランド (brand)	3	0.75
書体 (style of type)	3	0.00
地 (texture)	3	0.00
服装 (clothes)	3	-0.75
アクセント (accent)	2	3.75
コントラスト (contrast)	2	-3.00

Note on Current Practice

The affect processing system described in this paper has been in commercial use since January 2003 as part of the *CB Market Intelligence* text mining system. To date, 13 companies have adopted the technology to analyze customer data. Examples of the commercial applications of affect analysis are illustrated in the following brief case studies.

Business Areas of the Commercial Users

Transportation & Telecom. The clients in this business area use affect-based text mining in their customer support departments for the purpose of improving service and customer satisfaction. The system supports the staff by helping analysts identify specific issues that can affect customer defection rates. The staff works to reduce customer defections by addressing the problems in their service.

Retail. The clients deploy our affect-based text mining in their management planning department to incorporate their customers' points of view to improve operations and services. The system helps managers develop more detailed understanding of their customers' needs.

Consumer Goods & Services. The clients in this business area use affect analysis in marketing—to quantify customer evaluations. The system helps them reduce customer defection by identifying their marketing strengths.

Manufacturing. The clients use our affect-based text mining in product design, especially to create new products that will reflect the preferences of customers as expressed in surveys. The system is used by people who are not expert in interpreting qualitative surveys.

Conclusions and Future Directions

This paper presents an affect-based text mining system. The system performs automatic extraction of structured customer preferences by using a general, core affect lexicon and *Adjective-Relation Database*s. In areas where competing products equally satisfy customers' functional needs, understanding and measuring customers' emotional responses to products becomes increasingly important. Our system addresses this problem by providing tools to extracting quantified information from qualitative analyses.

The system is very useful for fast analysis of customer data; it operates quickly with a fast database builder and the pre-existing affect lexicon. The system is very effective in processing customer comments as expressed in simple natural language. When customers' expressions become more complex, the system requires additional lexical resources to achieve good results. Thus, due to the difficulty of defining and constructing lexical resources for practical use in all potential areas of application, the general solution to the problem of quantifying qualitative information remains a challenge for us.

In our future work, we plan to handle more complex expressions of affect and go beyond surface analysis of phrases. In particular, we plan to work on tacit *modified-nouns*, nouns that are not explicitly expressed in customer comments, e.g., a comment on hotel rooms such as "is clean," without the *modified-noun* phrase "room." In addition, we will develop more effective and practical ways to construct qualitative feature spaces and to modify affect lexicons.

Acknowledgements

I want to thank Yan Qu, David A. Evans, David A. Hull, and Gregory Grefenstette for their invaluable comments and assistance in preparing this paper. Note that an earlier, shorter version of this paper has been submitted to Information Processing Society of Japan DBWeb2003 under the title *A Text Mining System with Affect Lexicon for Qualitative Analysis of Free Text*.

References

Clausen, S.T. 1998. *Applied Correspondence Analysis: An Introduction*, Sage University Papers Series.

Fujita, S. 1999. An Approach for Information Retrieval and Classification using Natural Language Processing, *Information Processing Society of Japan Magazine*, Vol. 40, No. 4, 352-357.

Harada, S.; Itoh, Y.; and Nakatani, H. 1999. On Constructing Shape Feature Space for Interpreting Subjective Expressions, *Journal of Information Processing Society of Japan*, Vol. 40, No. 5, 2356-2366.

Hatzivassiloglou, V., and McKeown, K.R. 1997. Predicting the semantic orientation of adjectives, *Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL*, 174-181.

Hatzivassiloglou, V., and Wiebe, J.M. 2000. Effects of adjective orientation and gradability on sentence subjectivity, *Proceedings of 18th International Conference on Computational Linguistics*.

Kiyoki, Y.; Kaneko, Y.; and Kitagawa, T. 1996. A Semantic Search Method and Its Learning Mechanism for Image Databases Based on a Mathematical Model of Meaning, *IEICE Transactions*, Vol. J79-D-II, No. 4, 509-519.

Kudo, T., and Matsumoto, Y. 2002. Japanese Dependency Analysis Using Cascaded Chunking, *Journal of Information Processing Society of Japan*, vol.43, No.6, 1834-1842.

Morinaga, S.; Yamanishi, K.; Tateishi, K.; and Fukushima, T. 2002. Mining product reputations on the Web, *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 341-349.

Mukunoki, M.; Tanaka, H.; and Ikeda, K. 2001. An Image Retrieval System Using the Feature Space Consisting of Antonymous Pairs of Adjectives, *Journal of Information Processing Society of Japan*, Vol. 42, No. 7, 1914-1921.

Okada, N. 1985. Conceptual Taxonomy of Attributes for Natural Language and Picture Pattern Understanding : Primitive Concepts in Adjectives, *Journal of Information Processing Society of Japan*, Vol. 26, No. 1, 25-31 .

Pang, B.; Lee, L.; and Vaithyanathan, S. 2002. Thumbs up? Sentiment classification using machine learning techniques, *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, 79-86.

Subasic, P., and Huettner, A. 2000. Affect Analysis of Text Using Fuzzy Semantic Typing, *Proceedings of the 9th IEEE International Conference on Fuzzy Systems*.

Tateishi, K.; Ishiguro, Y.; and Fukushima, T. 2001. Opinion Information Retrieval from the Internet, *SIGNL Information Processing Society of Japan*, NL-144-11, 75-82.

Turney, P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, *Proceedings of the 40th Annual Meeting of the A CL*, 417-424.

Watanabe, Y.; Nakamura, Y.; and Nagao, M. 1993. Extraction of objective and impression (KANSEI) information from explanation texts of pictures, *SIGCH Information Processing Society of Japan*, 93-CH-20, 13-20.

Wiebe, J.M. 2000. Learning subjective adjectives from corpora, *Proceedings of the 17th National Conference on Artificial Intelligence*.

Zhai, C.; Tong, X.; Milic-Frayling, N.; and Evans, D.A. 1997. Evaluation of syntactic phrase indexing- CLARIT NLP track report. *The 5th Text Retrieval Conference (TREC-5)*, 347-358.