APPLYING CUSTOMER CONSCIOUSNESS FOR BRAND CONSTRUCTION USING NLP

[1]TOMOHIRO KAWATA, [2]TAKUMI KATO, [3]KAZUHIKO TSUDA

Kawata Architects Office Japan, University of Tsukuba Japan, University of Tsukuba Japan
E-mail: mohi.tomohiro.kawata@gmail.com, s1540108@u.tsukuba.ac.jp, tsuda@gssm.otsuka.tsukuba.ac.jp

Abstract- Social-media are arising especially in teens and twenties and becoming more important to companies which tend to construct and develop their brand today. However, natural language processing which is necessary to analyse social-media are not effectively utilized yet for study of branding. In this paper, a method of applying natural language processing for branding was developed. To apply natural language processing, dictionaries which is used to verify target texts have certain attributes is created and target texts are analysed with morphological analysis and N-gram model. The authors inspected how accurate the dictionaries work with reading texts and judging the interpretation was collect and evaluated precise, although it could stand further improvement, especially in tagging sensibilities.

Index Term- Branding, Brand Equity, Natural language processing, Consumer Generated Media

I. INTRODUCTION

To understand and improve brand is important when building Corporate value. In world market capitalization 2016, as well as Apple, 4 companies out of the top 5 companies were also ranked in the top 5 of Forbes the World's Most Valuable Brands.[12] Although branding is so important, it is bad idea that developing a stereotypical attitude toward branding is valid. it is needed, rather, to define brand images inspiring consumers to have toward our brand or to define brand images which is wanted to appeal to consumers. In 2002, UNIQLO, clothing company, entered to online resale business of vegetables and left the market just for one year and a half with 2.6 billion yen deficit. UNIQLO thought they had "reasonable and high quality", on the other hand, consumers thought UNIQLO was "cheap and nasty". Because of this mismatch, entry to the perishable market in which "quality and safety" was very important, was end in failure. In these days, considering "what kind of brand" spreading to consumer and learning "how correctly the brand is captured" are required to success a business.
In this paper, we attempt to achieve a scheme to analyse and applying customer consciousness using natural language processing(NLP) .

II. RELATED WORK

In this section, as the related work, we mention existing study of brand and NLP for customer consciousness.

A. Study of Brand

The basis about study of brand was established [1]. He introduced 'Brand Equity'. After 'Brand Equity' introduced, study of brand came to be taken seriously. D. Aaker [2], also, introduced the concept of 'Brand Identity' which is a unique set of brand associations representing what the brand stands for and offers to customers an aspiring brand image. Keller [3] developed 'Brand Equity'. He defined 'Brand Equity' as 'the differential effect that brand knowledge has on consumer response to the marketing of the brand. and proposed 'Customer-Based Brand Equity' and indicated effectiveness about practice of 'Brand Knowledge'. The core of study about brand, including 'Brand Equity', 'Brand Identity', 'Brand Knowledge' As 'Brand Image' is invisible and ambiguous, how to measure it is matter of main focus. There is an idea similar to 'Brand Image'. Brand Personality is a set of human characteristics that are attributed to a brand and is a framework that helps a company or organization shape the way people feel about the brand. J. Aaker [4] evaluated brand personalities of 37 well-known brands by a questionnaire survey with 114 question. She suggests major factors associated with 'Brand Personality' called 'Big 5'. Tonegawa [5] introduced quantitatively approach to measurement of 'Brand Personality' using 'Big 5'.

B. NLP for Customer Consciousness

While J. Aaker [4] and Tonegawa [5] developing study of brand by a questionnaire survey, these days, social media grow increasingly important from the aspect of easiness to access and character that consumers post or spread with their own words. Okumura [6] summarize succinctly method about NLP for social media. Remaining issues in NLP for social media, especially CGM (Customer Generated Media) such as blogs or SNS, are precise understanding of them with treating unknown words or irregular syntax or grammar

II. STRUCTURING QUANTIFICATION OF SOCIAL MEDIA DATA

In this section, we describe the structure of quantification of social media data and illustrate environment developed.

[1] market capitalization date as of September 12, 2016. "Google" is stated as "Alphabet"
[2] The World's Most Variable Brands 2016

A. Overall Structure

Environment developed and processing flow are below

.

1) Environment developed

Figure 1 is the environment diagram we developed using Amazon Web Service(AWS). social media data is analysed using python and stored in Relational Database Service(RDS). In the Figure 1, considering redundancy and availability, Act-Standby configuration is employed, but this is optional.

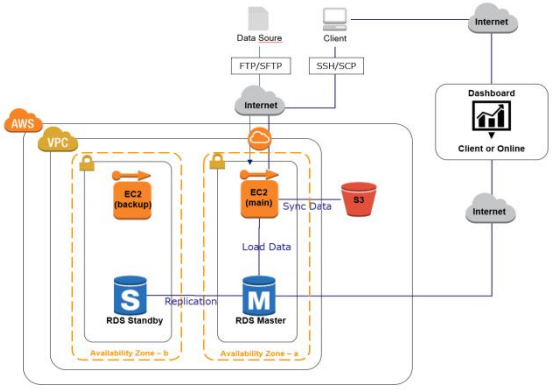


Figure 1 : Configuration Diagram

2) Processing Flow

This is processing flow of quantification of social media data

1. Define dictionaries and tags using analysis
2. Acquire data and standardize
3. Morphological analyse and tag data
4. Manage tags and restraint misinterpretation

B. Standardization of data specification

1) Diversification of media and Importance of cross-media analysis

As diversifying media, these days, what source consumers obtain information and where they utter their thoughts. Since the beginning of the 2000, not only existing media such as newspaper or television, Social Network Services(SNS) such as Facebook or Twitter is used actively, and in the new decade new sns services like Instagram, Line and Snow are arising.

According to research of Ministry of Internal Affairs and Communications in Japan [3], usage time of social media of teens and twenties in Japan is more than 50% of the media usage time in a day. Based on this situation, it is necessary to analyse various media including social media in addition to traditional media, especially social media is beneficial to understanding how consumers recognize each brand because the characteristic of social media that consumers transmit their own thoughts and, certainly they get information via them help us understanding of 'Brand Image'.

this is the reason social media is involved in this effort.

2) Categorization of media

In this approach, media are categorized as below

Table 1 : Types of media

Media Type	Summary
tv	Television
news	Newspaper
twitter	Twitter
blog	blog
sns	SNS except for Twitter
web	online news e.g. Yahoo News
forum	Bulletin board

defining media types and categorizing media into media types make it possible to analyse data with sorting out or weighting media. Each data source has their own format, so that it is difficult to handle them with uniform process. Defining common format can help uniform process. Table 2 is the sample common format

Table 2 : Example of common format

id	date	time	media_type	media	body
abc0001	2016-09-01	08:15:00	tv	CBS	xxxxxxxxx

3) Data Acquisition

The way to acquire social media data is pushed aside in this article. It will be using API, web crawling, or even purchase.

C. Define Dictionary and Tagging

1) Objective of Dictionary

In this article, the role of dictionary used to analyse and quantify the social media data is below.

- Append Vocabulary
- Normalization words
- Identification of Target Words

a) Appending Vocabularies

Open dictionary like IPAdic or UniDic is not insufficient in the specific industry or specific context. Thereby, it is necessary to append vocabularies, idioms, or even slangs.

b) Normalization of Words

If words written differently have same meaning or interpretation, these orthographical variants should be corrected to exactly perform the analysis. For instance, there are different words such as 'pleasure', 'enjoyment', 'entertainment' which similar to 'fun', but along in some purpose of analysis, treating all these words as 'fun' is required. Similarly, a word

[3] White Paper on telecommunications 2015(Japanese)

'enjoyment' has many notations such as 'enjoy', 'enjoyed', 'enjoys', 'enjoyable' should standardize.

c) Identification of Target Words

Identification of target words is useful when a word has different meaning as usual or a meaning a word have in the background is more important than one as usual. For instance, 'iPhone' is a product of 'Apple' and in the situation that a tweet should be interpreted as talking about 'apple', when the word 'iPhone' appeared in a tweet this technique is useful.

2) Types of Dictionary

9 types of dictionary are used to analyse social media data in this effort. Each dictionary except for 'Term Dictionary' is also used as 'XXX Tag' when it is tagged. Table 3 illustrates the types of dictionary.

Table 3 : Definition of Dictionaries

Dictionary Name	Tag Name	Purpose
Term Dictionary	-	appending Vocabularies and normalization of words. this dictionary is not used as tagging.
Company Dictionary	Company Tag	used to identify company name or product's name as subject of analysis
Topic Dictionary	Topic Tag	used to identify object or topic like 'where' or 'what'
Risk Dictionary	Risk Tag	used to identify topic which is undesirable or are considered to require immediate attention
Image Dictionary	Image Tag	used to identify image associated to the brand
Sensibility Dictionary	Sensibility Tag	used to identify sensibility which is wanted to belong with brand or emotion associated to the brand
Message Dictionary	Message Tag	used to identify terms other than above to which are concerned such as slogans or product characteristics
Screening Dictionary	Screening Tag	used to identify terms which are not concerned and should be marked as unrelated posts
Skip Dictionary	Skip Tag	used to prevent mis-extraction of terms. For instance, when a target term is 'app'(shorthand notation of 'application'),'apple' does not meet the purpose

3) Treatment of Dictionary

After tagging a post, the tags are combined with themselves which have different tag type. Figure 2-1 is valid v of tags (henceforth as tag-set). the post has Company tag, Topic tag and Image tag. Figure 2-2 is invalid because it doesn't have an image tag or a sensibility tag and it doesn't have information about the brand image of the company. On the other hand, Figure 2-3 doesn't have company tag so that the target company is vague. Figure 2-4 is seemed to be valid but it is suspected as advertisement which is not appropriate for the sake of analysis.[4]

[4] For sake of illustration, this examples are stated in English but actual texts are written in Japanese.

1. iPhone in the annual event is so cool.
2. The annual iPhone event is held.
3. The event is so cute
4. New iPhone is launched!! BUY NOW!! CLICK HERE!!

Figure 2: Sample texts

4) How to Tagging

There are two way tagging posts. One is morphological analysis and the other is N-gram. Sensibility tag is tagged through morphological analysis, since sensibility have attribute that the dictionary has few terminologies and an order of terms or a part of speech need to be taken in consider. The others are tagged thorough N-gram, since the dictionaries have many terminologies, and consideration about an order or a part of speech is not necessary.

a) Morphological Analysis and Acquisition of Sensibility

Acquisition of sensibility is reached through morphological analysis and rules. Applying rules toward results of morphological analysis enable complex sensibility kept in the background of the posts. In the Figure 3-1, the post written in Japanese means "New iPhone is so cute!!" and a sensibility 'Praise' is found. Similarly, in the Figure 3-2, the post means "I wish iPhone were cheaper!" and a sensibility 'Desire' is found.

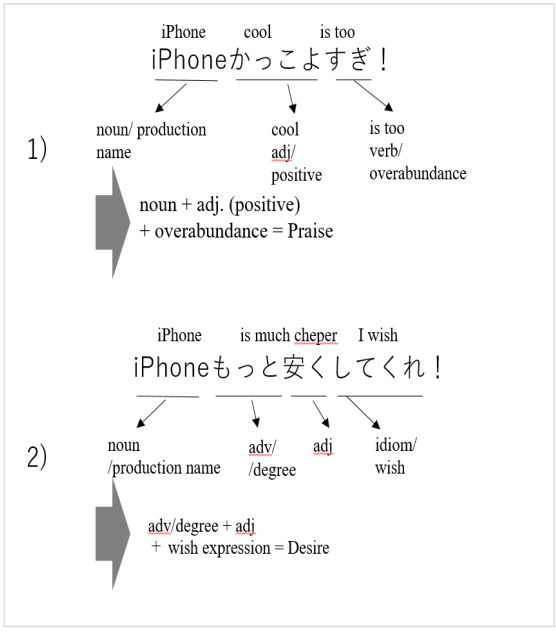


Figure 3: Example of sensibility analysis

b) Tagging by N-gram

By contrary, tagging of interpretation is reached through N-gram. N-gram is used for efficient approximate matching. Perfect match search is enabled by using inverted index which has a

contiguous sequence of n characters from a given sequence of text and appearing positions. Besides, preventing tag noises, Skip Dictionary is used. In the Figure 4-1, the post means' iPhone is cool!' is tagged Company Tag 'iPhone' and Image Tag 'Cool'. On the other hand, Figure 4-2 show the prevention of miss-extraction. The text means "I want to eat an applepie" which Company Tag 'Apple' is not appropriate is controlled by Skip Tag.

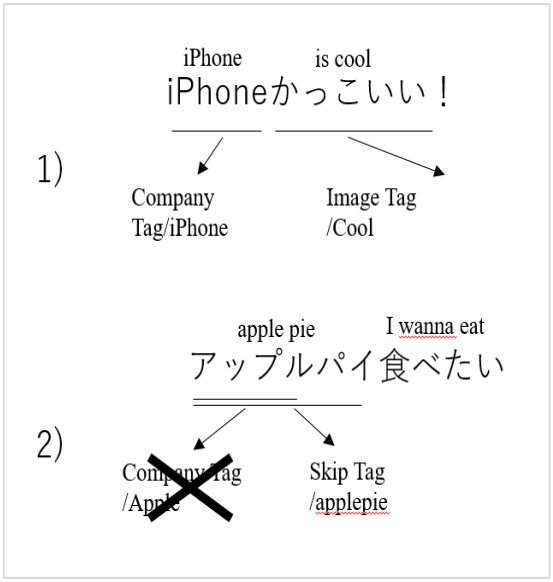


Figure 4: Example of tagging

III. ACCURACY AND EVALUATION OF TAGGING

In this section, evaluation of this efforts are stated. How accurate interpretation or sensitivity tagged is evaluated by man hands reading texts and judge the interpretation tagged. The target tags evaluated are below.

[0] Company Tag
[1] Topic Tag
[2] Image Tag
[3] Sensibility Tag
[4] Risk Tag

A. Process of Evaluation

Process of evaluation is below. Evaluation was executed twice. First time only Image Tag was evaluated and last time All target tags were evaluated by tag-sets.

1. Data Acquisition
2. Analyse Data
3. Combining tag-sets
4. Evaluate tag-sets
5. Calculating Precision Ratio

1) Data Set

Twitter and Blogs data were collected to evaluation. The Data was covered 3 months from Oct. 2015 to Dec. 2015 and it had 686,735 twitter posts and 122,240 blog posts, 808,975 posts as total.

2) Analysis

Analysis was executed in the way stated prior chapter. As the result of morphological analysis and N-gram analysis, valid texts were 111,459. Detailed results of each tag are below.

Table 4: Number of tags

Tag Name	Number of tags
Company Tag	192,707
Topic Tag	155,624
Risk Tag	140,951
Image Tag	135,427
Sensibility Tag	277,591

3) Preparing Tag Sets

After analysis, each tag was combined to tag-sets with determining validity like existence of 'Screening Tag'. In this evaluation, tag-sets below were treated as valid.

- Company Tag × Topic Tag × Image Tag
- Company Tag × Topic Tag × Sensibility Tag
- Company Tag × Risk Tag

4) Evaluating Tag Sets

Process for evaluation of accuracy of tag-sets are below.

1. Tag-sets are the same as human understanding.
2. If not, which tags are mismatch.
3. If not even all tags seemed correct, why and how tag-sets are mismatch.

5) Calculation of Precision Ratio

The way how to calculate precision ratio is below.

$$PR = 1 - f/t$$

PR: Precision Ratio
f: number of tags failed to accurate analysis
t: total number of tags in valid target texts

B. Results of tagging

The results of evaluation are stated below for each tag. Table 5 shows the detail.

1) Company Tag

Company Tag was evaluated in the last time. The number of evaluated tags were 1595 and the precision ratio was favorably 99.9%.

2) Topic Tag

Topic Tag was evaluated in the last time. The number of evaluated tags were 1595 and the precision ratio was favorably 98.9 % as well.

3) Image Tag

Image tag was evaluated twice. In the first time, tags were evaluated without considering tag-sets and, then

in the last time, tags were evaluated with considering Tag-sets. As the result, in the first time, the number of evaluated tags were 706 and he precision ratio was 62.3% and in the last time the number of evaluated tags were 628 and he precision ratio was 84.2%. points of improvement were mainly expanding Screening Dictionary and Skip Dictionary and adopting or rejecting the terms of Image Dictionary.

4) Sensibility Tag

Sensibility Tag was evaluated in the last time. The number of evaluated tags were 888 and the precision ratio was favorably 65.1%. The reasons of imprecision are mainly that sensibility found in the text didn't suit in the context considering companies and topics even though it suited in ordinary or modification relations were mismatch due to complexity of Japanese grammar and syntactic structure.

5) Risk Tag

Risk Tag was evaluated in the last time. The number of evaluated tags were 79 and the precision ratio was favorably 89.9%. Image Tag and Sensibility Tag were focused in preference to seemingly accurate Risk tags, thus the number of samples is a few.

Table 5: Result of evaluation

		1st exam	2nd exam			
		8/12/16	8/29/16	8/30/16	8/31/16	total
Company × Topic × Image	Number of precision	-	117	45	367	529
	Number of sample	-	133	71	424	628
	Precision ratio	-	88.0%	63.4%	86.6%	84.2%
Company × Topic ×Sensibility	Number of precision	-	16	488	51	555
	Number of sample	-	49	778	61	888
	Precision ratio	-	32.7%	62.7%	83.6%	62.5%
Company×Risk	Number of precision	-	71	-	-	71
	Number of sample	-	79	-	-	79
	Precision ratio	-	89.9%	-	-	89.9%
Company	Number of precision	-	261	847	485	1593
	Number of sample	-	261	849	485	1595
	Precision ratio	-	100.0%	99.8%	100.0%	99.9%
Topic	Number of precision	-	261	828	485	1574
	Number of sample	-	261	849	485	1595
	Precision ratio	-	100.0%	97.5%	100.0%	98.7%
Image	Number of precision	440	117	45	367	529
	Number of sample	706	133	71	424	628
	Precision ratio	62.3%	88.0%	63.4%	86.6%	84.2%
Sensibility	Number of precision	-	16	511	51	578
	Number of sample	-	49	778	61	888
	Precision ratio	-	32.7%	65.7%	83.6%	65.1%
Risk	Number of precision	-	71	-	-	71
	Number of sample	-	79	-	-	79
	Precision ratio	-	89.9%	-	-	89.9%

IV. CONCLUDING REMARKS

Investigating method of analysing social-media for branding is the purpose of this paper. As before questionnaire surveys are general way to understand 'Brand Image', we developed how to apply social media data including CGM like twitter or blogs for understanding 'Brand Image'. Further consideration will be needed to yield any finding about understand sensibilities with considering ambiguous expressions and contexts, even though understanding interpretation like companies or topics are proven throughout effort. This effort will be applied to quantification of 'Brand Image' or assessment of publicity activities from the branding viewpoint, or even early detection branding risks.

REFERENCES

[1] Aaker, D.A., “Managing Brand Equity”, Free Press, Sep. 1991

[2] Aaker, D.A., “Building Strong Brands”, Free Press, Dec. 1995

[3] Keller, K.L., “Strategic Brand Management”, Prentice Hall, Feb. 1998

[4] Jennifer L Aaker, “Dimensions of brand personality”, Journal of Marketing Research, vol.34 No.3, Aug. 1997.

[5] Tonegawa and Haku, “Proposing a quantitative methodology using brand personality”, Policy science, vol. 15 No.2, pp13-23, Feb. 2008 (Japanese)

[6] Okumura, Social Media Mining, IEICE Fundamentals Review Vol.6 No.4 Vol. 6, Apr. 2013 (Japanese)

★★★