Kotonush: Understanding Concepts Based on Values behind Social Media

Tatsuya Iwanari
University of Tokyo

Kohei Ohara
University of Tokyo

Naoki Yoshinaga
IIS, University of Tokyo

Nobuhiro Kaji
Yahoo Japan Corporation

Masashi Toyoda
IIS, University of Tokyo

Masaru Kitsuregawa
IIS, University of Tokyo
NII, Japan

Abstract

Kotonush, a system that clarifies people's values on various concepts on the basis of what they write about on social media, is presented. The values are represented by ordering sets of concepts (*e.g.*, *London*, *Berlin*, and *Rome*) in accordance with a common attribute intensity expressed by an adjective (*e.g.*, *entertaining*). We exploit social media text written by different demographics and at different times in order to induce specific orderings for comparison. The system combines a text-to-ordering module with an interactive querying interface enabled by massive hyponymy relations and provides mechanisms to compare the induced orderings from various viewpoints. We empirically evaluate Kotonush and present some case studies, featuring real-world concept orderings with different domains on Twitter, to demonstrate the usefulness of our system.

1 Introduction

When we want to investigate unfamiliar entities or concepts (*e.g.*, iPhone SE) as consumers, or inversely, intend to supply new concepts as vendors, we typically endeavor to understand the value of a given concept by comparing or ordering it with familiar concepts (*e.g.*, Xperia X or Galaxy S7) from various perspectives (*e.g.*, *user-friendliness*). At present, people often spend a substantial amount of time wading through massive social media text to get an overview of others' perceptions, or spend a lot of money to call for votes from experts in order to come up with a convincing ordering.

In this study, we present **Kotonush**, a system that induces people's values on given concepts from social media text as concept orderings on the basis of common attribute intensity expressed by an adjective. Our system enables users to interactively ask queries (concepts and an adjective) and compare the induced orderings for deeper understanding of the concepts. Assuming that a user has at least one target concept (or entity) in mind, our querying interface helps the user to interactively list similar entities using massive hyponymy relations (Sumida et al., 2008). Receiving a query, a text-to-ordering module (Iwanari et al., 2016) collects posts from social media text written by specific (gender, region) users and at a certain time of interest (say, *domain*) to induce concept orderings specific to the chosen domain. Our ordering visualizer then provides intelligent interfaces to compare orderings from various perspectives to gain a deeper insight into the domain-specific values of concepts.

Our system is beneficial not only in practical terms for understanding entities from others' values (orderings with related entities) to make correct decisions (*e.g.*, ordering smartphones in terms of *user-friendliness*) but also in sociological terms for inversely understanding common views shared by a certain demographic and/or from a certain period of time. We conclude this work with a handful of interesting case studies comparing concept orderings in different domains taken from our 4-year Twitter archive.

2 Related Work

There have been no attempts other than our own previous work (Iwanari et al., 2016) on ordering concepts on the basis of the intensity of their attributes. Although aspect-based sentiment analysis mines reviews

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: http://creativecommons.org/licenses/by/4.0/

Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations, pages 292–296, Osaka, Japan, December 11-17 2016.

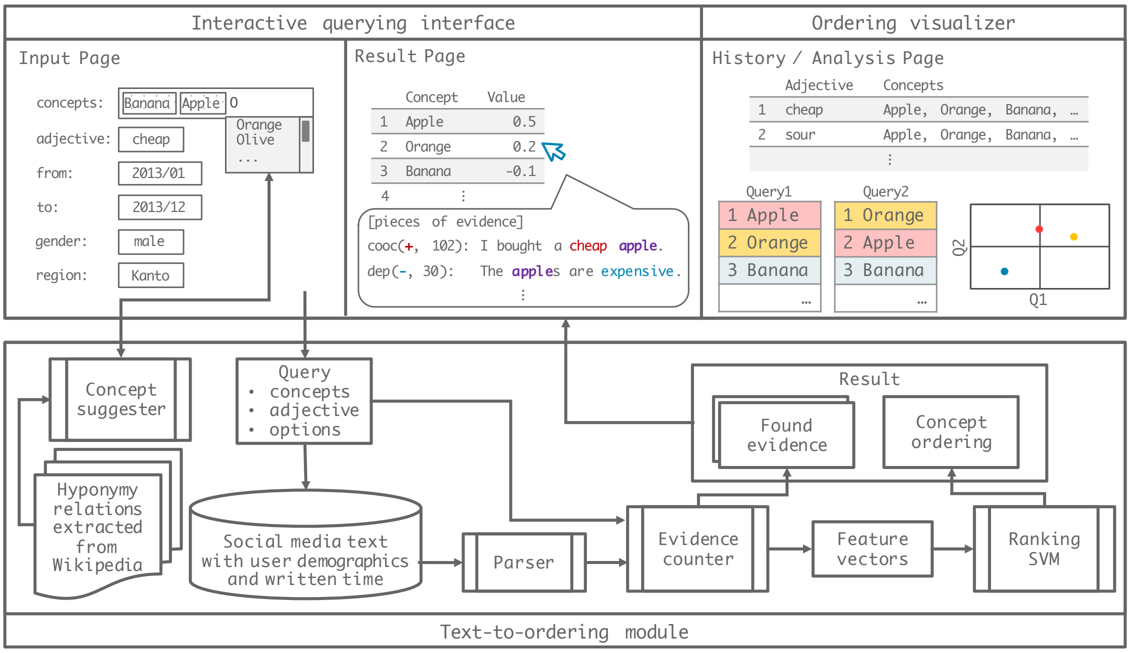


Figure 1: Overview of Kotonush, the system we developed to acquire values from social media text.

or other texts for opinions on entities (Pang and Lee, 2008), such analysis focuses on the polarity of specific aspects (i.e., whether the '*atmosphere*' of a restaurant is good or not), while our system supports not only the polarity but also the intensity of attributes (a restaurant is *cozy* or *lively*).

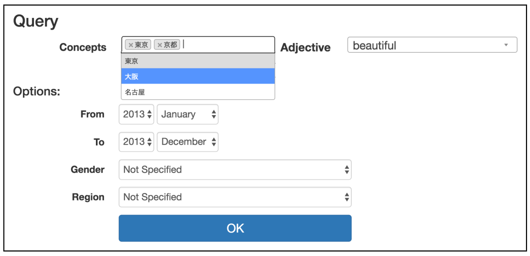
Iwanari et al. (2016) initiated the task of ordering concepts and proposed methods that order concepts by gathering various pieces of evidence from social media text and integrating them with a supervised learning. They confirmed that it is possible to obtain common views from the text people write. However, there are a couple of issues when it comes to using their method for our purpose of understanding the target concepts. First, it is not easy to conceive other concepts to compare with the target concept. Second, monolithic orderings induced from entire social media texts do not provide a deeper insight into the target concepts. To address these issues, we have built a system that suggests to users other concepts in the same category as the target concepts along with tools to understand the values in different domains.

3 System Architecture

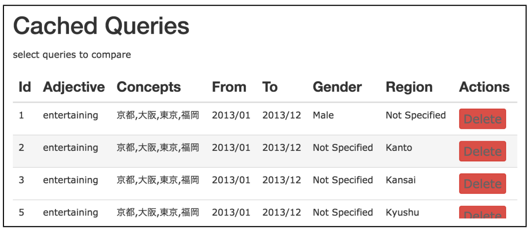
Our system consists of three parts: (1) an interactive querying interface, (2) a text-to-ordering module, and (3) an ordering visualizer (Figure 1). Our querying interface enables users to interactively input a set of concepts and an adjective as a query (Figure 2a) and then sends them to the text-to-ordering module. The querying interface accepts several options that specify domains, such as the gender and region of social media users as well as the time periods of interest. After receiving a query, the text-to-ordering module collects posts from social media text in the domain and returns a convincing ordering along with the pieces of evidence used (to justify the ordering). The system keeps track of the results of asked queries so that users can compare the (cached) results with other queries on our system's History / Analysis page (Figure 2b and 2c). This enables us to compare concepts from various viewpoints (adjectives) or to observe differences of ordering in each domain to see which factors affect orderings.

Note here that the domain analyses provide deeper and closer insight into not only target concepts but also target domains (*e.g.*, *women in Japan like Disney movies better than action movies*, as we will reveal in the following case studies). In the following, we describe the workflow of our system in more detail.

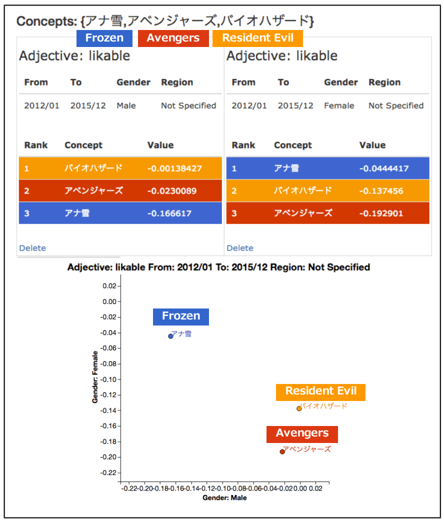
Preprocessing We assume a search engine to retrieve posts that include concepts and adjectives and have built a simple inverted index-based search engine for that purpose. This search engine can easily be replaced with other search engines such as the Twitter API (to obtain up-to-date orderings), since all the text analyses to collect evidence on concept ordering are done online.



(a) Interactive querying interface accepts a set of concepts, an adjective, and options. The system suggests concepts in the same category.



(b) History page keeps cached query results.



(c) Users can compare orderings with different settings. This example compares movies (Frozen, Avengers, and Resident Evil) in terms of 'likable' with two genders.

Figure 2: Snapshots of our system for ordering objects on the basis of common attribute intensity.

As with the indexing, we briefly identify the gender and location (prefecture) of social media users from their posts and profiles for domain analyses and then associate text with those attributes. Since this process is outside the focus of this study, here we just use existing methods based on bag-of-words.

Interactive querying Users input a query by adding concepts one by one and selecting an adjective from a (short) list that meets the users' practical demands. The list prompts users to compare concepts in different ways that might not come to mind on their own. Users can also specify domains (Figure 2a).

Although users can input any concepts they want, they may not conceive of concepts they might wish to compare. For example, when you browse rental movies at a shop, you may not be able to remember appropriate movies for comparison. The same applies here. To help such users, Kotonush suggests concepts related to given concepts. We exploit hyponymy relations extracted from Wikipedia (Sumida et al., 2008) to suggest concepts that share the same hypernym with the given concepts.

Concept ordering After receiving a query, the text-to-ordering module retrieves posts including one or more of the given concepts and the adjective from social media text in the specified domain. The posts are then online parsed with J.DepP, a state-of-the-art dependency parser for Japanese (Yoshinaga and Kitsuregawa, 2014), to process massive text online ($> 10,000$ sentences/s).

The parsed text is given to our implementation of Iwanari et al. (2016) to induce a concept ordering. The method uses four types of evidence to capture the common view on concepts from social media text: (1) co-occurrences of a concept and an adjective (*e.g.*, How large that whale is!), (2) dependencies from a concept to an adjective (*e.g.*, A whale is so big.), (3) similes (*e.g.*, He is brave as a lion.), and (4) comparative expressions (*e.g.*, Whales are larger than cats.). The first three implicitly suggest attribute intensity and can be understood as capturing the absolute intensity of the attribute that the concept has. The fourth directly captures the relative attribute intensity, which directly indicates the order of a subset of a concept set. The method encodes these four types of evidence as real-valued features by using pointwise mutual information (PMI) of the pairs of a concept and adjective for each piece of evidence and then performs an ordering based on ranking SVM. Finally, the text-to-ordering module returns the joint results of the outputs of found pieces of evidence so that users can know what social media users say about each item along with the ordering obtained by ranking SVM with scores computed for each item.

Ordering visualizer By keeping the results of past queries in our system, users can review and compare them on the History / Analyze page. This page provides complete sets of cached results as a table and tools to analyze queries with different settings such as bump charts (top of Figure 2c). With bump

Table 1: Correlations between Twitter user orderings and gold-standard orderings.

	Male	Female	All
Avg. ρ	0.681	0.674	0.661

Table 2: Spearman's ρ against gold-standard orderings.

	BASELINE		DOMAIN-UNAWARE		DOMAIN-AWARE	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
Avg. ρ	0.262	0.339	0.308	0.322	0.309	0.337

charts, users can, for example, determine the best season for each flower by varying periods.

In addition to bump charts, we implemented an interface of scatter plots on the page (bottom of Figure 2c). Although users can compare two or more queries at once with bump charts, scatter plots provide a more intuitive way of comparing two queries when a user wants to know the relative strength of the attribute intensity of each concept (*e.g*, *lemons are much more sour than apples and dorians*, *i.e.*, *lemon* $\gg$ *apple* $>$ *dorian*), compare orderings with different attributes (*e.g.*, *'cheap' and 'delicious' for restaurants*), or compare ordering in a different domain (*e.g.*, *male vs. female*).

4 Evaluation

We conducted experiments to evaluate Kotonush with our archive of 25 billion Twitter posts in terms of correlation between system-generated and gold-standard orderings. We used LIBLINEAR (`https://www.csie.ntu.edu.tw/~cjlin/liblinear/`) as an implementation of ranking SVM.

4.1 Settings

We prepared 28 queries with the same process in Iwanari et al. (2016), which used a word clustering-based method. They cover a wide variety of queries: from concepts (*e.g.*, '*car*') to instances (*e.g.*, '*Kinkaku-ji*', a shrine) and from objective adjectives (*e.g.*, '*fast*') to subjective ones (*e.g.*, '*likable*').

To prepare gold-standard orderings for training and testing Kotonush, we used a crowdsourcing service (`https://crowdworks.jp/`) to ask 53 Twitter users (workers) to answer (rank) each query. The users had various demographics: gender (24 males and 29 females), age (from 20s to 60s), location (29 out of 47 prefectures in Japan) and occupation (students, homemakers, office workers, etc.). We generated gold-standard orderings for each gender by choosing an ordering, in all permutations of concepts, that maximized the average of Spearman's rank correlation coefficient ρ against the orderings of the workers by gender, in addition to gold-standard orderings for all workers.

Table 1 shows the average correlations between the human and gold-standard orderings for three domains: all users, male users, and female users. The human-generated orderings have strong correlations and show higher correlations when we restrict workers in specific domains. By looking into these domain-specific orderings in detail, we can understand their values on concept orderings, *e.g.*, males' preferences regarding alcohol are quite different compared with those of females. The evaluation datasets will be available on `http://www.tkl.iis.u-tokyo.ac.jp/~nari/coling-16/`.

We have explored two different ways to train ranking SVM. Domain-unaware training uses the gold-standard orderings computed from the orderings given by all the workers, while domain-aware training uses the gold-standard orderings for individual domains (male and female). In domain-aware training, the number of training examples is multiplied by the number of domains (here, two) and the quality of the gold-standard orderings (correlations against human orderings) is higher than domain-unaware training, although it could suffer from a data sparseness problem. In testing, we input statistics collected from Twitter posts (Jan. 2012 - Dec. 2015) in each domain to obtain domain-specific orderings.

4.2 Results

Table 2 shows the experimental results obtained by leave-one-out cross-validation with the aforementioned datasets. We evaluated the system-generated orderings for each domain by computing Spearman's ρ against the gold-standard ordering in the domain. Here, BASELINE refers to the baseline

Table 3: Case studies in different settings.

(a) Flower (Beautiful): Three seasons

	Mar. - May	*Jun. - Aug.*	*Sep. - Nov.*
1	Cherry	Sunflower	Mum
2	Sunflower	Cherry	Sunflower
3	Mum	Mum	Cherry

(b) Disease (Fearful): Years

	2013	2014
1	Influenza	Dengue
2	Malaria	Influenza
3	Dengue	Malaria

(c) Fruit (Delicious): Regions

	Tohoku	*Shikoku*
1	Apple	Tangerine
2	Tangerine	Apple
3	Strawberry	Strawberry

method adopted in (Iwanari et al., 2016), which scores each concept on the basis of noun-adjective co-occurrences, i.e., the first evidence our system uses. The baseline method outperformed the proposed method in female domain because similes were hardly observed in posts written by female. The domain-aware training obtained better Spearman's ρ than the domain-unaware training.

5 Case Studies

This section presents four case studies that demonstrate the effectiveness of our system. We used the ranking SVM obtained by domain-unaware training along with statistics collected from Twitter posts. Even though we implemented the system to process all of the tasks in a single thread, it processes posts fast enough (about 10,000 posts in less than 5 sec), and they can be improved easily because all the tasks are perfectly parallel. We have hereafter translated the Japanese system outputs into English.

The first case study captures common views on movies in terms of gender (Figure 2c). In Japan, men tend to like action movies better than Disney movies and women vice versa. The next case compares three seasonal flowers – (Japanese) cherry, sunflower, and Chrysanthemum (mum) – in terms of beauty in different seasons (Table 3a). The results clearly show the blooming season (best time) of each flower. The third shows the time-series fearfulness of three diseases: Influenza, Malaria, and Dengue fever. The rise of Dengue fever from 2013 to 2014 reflects its spreading over Japan in 2014. Table 3c shows the region-parameterized results of '*Fruit (Delicious)*', which is reasonable for Japanese because the Tohoku (north) and Shikoku (south) areas are famous for the production of apples and tangerines, respectively.

6 Conclusion

We presented **Kotonush**, a system that acquires and compares orderings of concepts on the basis of intensity of their common attributes. Our system enables us to easily obtain concept orderings specific to a certain demographic and period from social media text. We empirically confirmed that our system outperformed the baseline based on noun-adjective co-occurrences, and we provided some case studies that compare concept orderings induced from a different domain in our 4-year Twitter archive.

We are now working to support languages other than Japanese, since a cross-lingual comparison between orderings obtained from text in different languages will reveal the differences of perception of different language speakers. We will release the codes of Kotonush for the academic and industrial communities under BSD License at `http://www.tkl.iis.u-tokyo.ac.jp/~nari/coling-16/`.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number 16K16109 and 16H02905.

References

T. Iwanari, N. Yoshinaga, N. Kaji, T. Nishina, M. Toyoda, and M. Kitsuregawa. 2016. Ordering concepts based on common attribute intensity. In *IJCAI-16*.

B. Pang and L. Lee. 2008. *Opinion Mining and Sentiment Analysis*. Now Publishers Inc.

A Sumida, N Yoshinaga, and K Torisawa. 2008. Boosting precision and recall of hyponymy relation acquisition from hierarchical layouts in wikipedia. In *LREC-08*.

N. Yoshinaga and M. Kitsuregawa. 2014. A self-adaptive classifier for efficient text-stream processing. In *COLING-14*.