

MASTER'S THESIS

**Learning to Order Concepts for Inferring Values  
of Social Media Users**

(ソーシャルメディアユーザの価値観推定を  
目的とした概念語の順序付け学習)



February 3, 2017

Supervisor  
Professor Masaru KITSUREGAWA

Department of Information and Communication Engineering  
Graduate School of Information Science and Technology  
The University of Tokyo

48-156408    Tatsuya IWANARI

# Abstract

We make decisions every day by ordering concepts such as political parties to vote or things to do in order of priority, and these orderings reflect our values on the things. It is beneficial for us to infer human values shared in various domains (*e.g.*, women, teenagers, people living in the same area). For example, a dressmaker can improve their products by knowing how female customers think about the company’s dresses comparing to those of other companies. Our research exploits social media text and tries to order concepts from various perspectives to understand writers’ values in various domains. Solving this task is not only beneficial in the practical sense but also interesting from a sociological perspective. For the research goal, we firstly target the whole social media text and propose statistical ordering methods that integrate heterogeneous evidence extracted from text on concept ordering. This first study aims at deriving collective wisdom on concept ordering from social media text. Experiments on real-world concepts reveal a strong correlation between orderings obtained by our methods and gold-standard orderings. Extending the study, we then propose a method to obtain domain-specific values by classifying the domains (*e.g.*, genders, regions they live in) of users to gain deeper insight into the values in specific domains. As an implementation of the method inducing domain-specific values, we develop Kotonush, a system that clarifies people’s values in specific domains by targeting social media text written by different demographics and at different times. The system combines a text-to-ordering module with an interactive querying interface enable 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.

**Keyword:** natural language processing, social media, learning to rank, concept ordering

# Acknowledgements

I would first like to express my profound gratitude to my advisor Professor Masaru Kitsuregawa, who has been a mentor and great inspiration to me. He has provided the tremendous circumstances for research from various viewpoints and gave valuable advice.

I would like to thank Associate Professor Masashi Toyoda, Associate Professor Naoki Yoshinaga, Project Assistant Professor Daisuke Yokoyama, Project Associate Professor Masahiko Itoh, and those who accompanied me in the presentation practice, who made helpful comments during the weekly meetings, who joined me for regular discussions. They spent much time to support my research, sometimes cooperated with as co-authors, and encouraged me to make study forward. In particular, Associate Professor Yoshinaga often brought we students to amazing places when we go to conferences, which truly motivated me and I had a wonderful time. I also would like to thank Dr. Nobuhiro Kaji (currently, he works at Yahoo! JAPAN Research) for his valuable advice.

I would like to appreciate all members of Kitsuregawa, Toyoda Lab. They were very kind and made me feel at home. Thanks my fellow students for their feedback. I have learned so much from them about not only my field but also their research. Shonosuke Ishiwatari, a doctoral student, always made us laugh by his really funny jokes. I spent valuable time with Shoetsu Sato and Mika Koizumi encouraging each other in hard time. Kohei Ohara helped me as a co-author by developing a system together. Satoshi Akasaki often gave useful advice on my research and presentations.

Special thanks to my family, who have been extremely supportive even in rough times. My gratitude to my parents is cannot be described in words. Thank you all of my friends who supported me and had a good time with me during my college life. They always made me happy and positive.

February 3, 2017

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Task Setting of Concept Ordering: Two Types of Target Users</b>	<b>5</b>
2.1 Concept Ordering Targeting Whole Social Media Users . . . . .	5
Input . . . . .	5
Output . . . . .	6
Gold Standard . . . . .	6
2.2 Concept Ordering Targeting Social Media Users in Specific Domains	6
Input . . . . .	6
Output . . . . .	6
Gold Standard . . . . .	6
<b>3 Related Work</b>	<b>8</b>
<b>4 Methods for Concept Ordering</b>	<b>10</b>
4.1 Inducing General Values on Concept Ordering . . . . .	10
4.1.1 Evidence on Concept Ordering . . . . .	10
4.1.2 Two Approaches to Ordering Concepts . . . . .	12

4.1.2.1	Ordering Concepts with Ranking SVM . . . . .	12
4.1.2.2	Ordering Concepts with SVR . . . . .	13
4.2	Inducing Common Values in Specific Domains . . . . .	15
4.2.1	Classify Domains of Users . . . . .	15
4.2.2	Training with Domain-Specific Data . . . . .	16
<b>5</b>	<b>Evaluation</b>	<b>18</b>
5.1	General Values on Concept Ordering . . . . .	18
5.1.1	Datasets . . . . .	18
5.1.2	Results . . . . .	20
5.2	Common Values in Domains . . . . .	25
5.2.1	Datasets . . . . .	26
5.2.1.1	Evaluation Datasets . . . . .	26
5.2.1.2	Twitter Datasets . . . . .	27
5.2.2	Results . . . . .	30
<b>6</b>	<b>A System for Concept Ordering</b>	<b>32</b>
6.1	System Workflow . . . . .	33
6.1.1	Preprocessing . . . . .	33
6.1.2	Interactive Querying . . . . .	34
6.1.3	Concept Ordering . . . . .	35
6.1.4	Ordering Visualizer . . . . .	36
6.2	Case Studies . . . . .	37
<b>7</b>	<b>Conclusion</b>	<b>40</b>
<b>A</b>	<b>Domain-Specific Gold-Standard Orderings and System-Generated Orderings</b>	<b>42</b>
<b>B</b>	<b>Annotation Manual for Crowdsourcing (in Japanese)</b>	<b>47</b>
	<b>Bibliography</b>	<b>52</b>
	<b>Publications</b>	<b>56</b>

# List of Figures

4.1	Two ways to train ranking SVM (Domain-aware/unaware). The yellow numbers show the procedures. . . . .	17
5.1	Correlation between the human orderings and the gold-standard ordering. . . . .	21
5.2	Correlations of the gold-standard ordering and the proposal methods. X-axis shows the average Spearman's $\rho$ s of gold-standards against human orderings and Y-axis shows the Spearman's $\rho$ s of proposal methods. . . . .	23
6.1	Overview of Kotonush, the system we developed to acquire values from social media text: interactive querying interface, text-to-ordering module and ordering visualizer. . . . .	33
6.2	Interactive querying interface accepts a set of concepts, an adjective and options. . . . .	34
6.3	The system suggests concepts in the same category by using hypernym relations. . . . .	35
6.4	History page keeps cached query results. . . . .	37
6.5	Users can compare orderings with different settings. This example compares movies (Frozen, Avengers, and Resident Evil) in terms of 'likable' with two genders. . . . .	38

# List of Tables

5.1	Evaluation dataset for acquiring general values. . . . .	19
5.2	Results on ordering concepts: Spearman’s $\rho$ against gold-standard ordering. . . . .	22
5.3	Ablation test for ranking SVM. . . . .	23
5.4	Examples of system-generated orderings. . . . .	24
5.5	Evaluation dataset for acquiring common values in specific domains. This table shows the gold-standard ordering of all the workers (ALL). . . . .	27
5.6	Correlation between human orderings. For each specific domain, five most different correlations compared to their ALL domain correlations are double-underlined. . . . .	28
5.7	Differences between some gold-standard examples from female (F, FEMALE) / male domains (M, MALE). Notable points are underlined. . . . .	29
5.8	Brief evaluation on gender classifications. . . . .	30
5.9	Spearman’s $\rho$ against domain-specific gold-standard orderings. . . . .	31
6.1	Case studies in different settings. . . . .	39
A.1	The gold-standard ordering of all the female (F, FEMALE) / male workers (M, MALE). . . . .	43
A.2	The baseline system-generated ordering of all the female (F, FEMALE) / male workers (M, MALE). . . . .	44
A.3	The domain-unaware system-generated ordering of all the female (F, FEMALE) / male workers (M, MALE). . . . .	45
A.4	The domain-aware system-generated ordering of all the female (F, FEMALE) / male workers (M, MALE). . . . .	46



# Chapter 1

## Introduction

We make decisions every day by ordering two or more concepts on the basis of common knowledge or common sense to which we are privy. For example, imagine a situation in which we buy fruit juice. If we want something sweet, we choose apple juice rather than lemon juice because we know that apples are generally sweeter than lemons.

The main objective of this study is to examine whether we can derive such common views on concept ordering from social media text, which reflects the values of the writers. Answering this question is not only interesting from a sociological perspective but also practically beneficial to those who want to order unfamiliar entities in terms of subjective attributes that are hard to quantify (*e.g.*, ordering smartphones in terms of *user-friendliness* or ordering tourist areas in terms of *fashionableness*) in order to make a correct decision. To come up with convincing ordering, we are presently forced to spend a substantial amount of time reading massive amounts of text to sum up people’s perceptions or call for votes from domain experts.

Considering these situations in mind, we tackle a task of ordering nominal concepts in accordance with the intensity of their common attributes as specified by adjectives, which was proposed by Nishina et al. [1] (in Japanese). A set of nominal concepts (*e.g.*, {*elephant*, *whale*, *dog*, *mouse*}) is provided in the task, along with

an adjective (*e.g.*, *large*) that represents an attribute shared by all members of the set. Given these two inputs, our goal is to output an ordered list of the items. The expected output in this example is *whale*  $\succ$  *elephant*  $\succ$  *dog*  $\succ$  *mouse*, where *whale* is the largest, *elephant* is the second largest, and so forth.

An issue to be addressed in performing this task is how to define a gold-standard, or goal ordering. Since a concept could refer to various instances that have different attribute intensities, and some attributes are subjective, it is inherently difficult to define an absolutely correct ordering agreed on by all. We, therefore, asked multiple volunteers to order a given set of concepts and then use the ordering that achieved the best average Spearman [2]’s rank correlation coefficient,  $\rho$ , against the human orderings as a gold-standard ordering. The resulting ordering can be seen as a representative ordering that sums up the general human perception, and is thereby meaningful as a goal in our task.

We present a method of ordering concepts to solve this task on the basis of textual evidence obtainable from massive amounts of social media text. The issues we address are twofold: (1) what kind of textual evidence to exploit and (2) how to integrate multiple kinds of evidence to obtain an appropriate ordering. We exploit heterogeneous textual evidence to address the first issue that indicates a possible ordering and then integrate the evidence to obtain an appropriate ordered list of the items. The types of evidence we used include noun-adjective (*i.e.*, concept-attribute) co-occurrences, noun-adjective dependencies, similes, and comparative expressions on nouns. The first three indicate the absolute strength of the attribute intensity, while the last captures the relative strength among the attributes of concepts.

We explore two approaches to integrating the heterogeneous evidence to address the second issue. The first uses a pairwise learning-to-rank framework, specifically, a ranking support vector machine (SVM) [3], while the second directly estimates Spearman’s  $\rho$  for each candidate ordering to output the ordering with the highest estimated Spearman’s  $\rho$  as the most likely ordering using a support vector regression (SVR) [4].

We performed experiments to evaluate our methods in terms of correlation between the system-generated and the gold-standard orderings for real-world concepts obtained from blog text. The results demonstrated that both our methods outperformed a dependency-based baseline (Nishina et al. [1]) that was inspired by Turney [5]’s work.

As an extension to the methods to induce the general values from whole social media users, we propose investigating domain-specific values of social media users by classifying the domains of users. Here, *domain* refers to a user group sharing the same attributes (*e.g.*, demographics, regions they live in). We develop Kotonush, a system that induces people’s values from indexed social media text as an implementation of the method. 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 [6]. Receiving a query, a text-to-ordering module collects posts from social media text written by specific domain users and/or at a certain time of interest to order concepts specific to the chosen domain. Our ordering visualizer then provides intelligent interfaces to compare orderings from various perspectives (*e.g.*, comparing transportation from different perspectives like ‘fastness’ and ‘expensiveness’) to gain a deeper insight into the domain-specific values. We empirically evaluate the system with a handful of interesting case studies, comparing concept orderings in different domains taken from our 4-year Twitter archive.

The remainder of this paper is organized as follows.

- **Chapter 2** formally defines our task of ordering concepts on the basis of common attribute’s intensity.
- **Chapter 3** introduces work related to our study.

- **Chapter 4** presents methods of ordering concepts using heterogeneous types of evidence that directly or indirectly suggest the attribute's intensity.
- **Chapter 5** explains how we evaluated our methods.
- **Chapter 6** shows a system that we developed as an implementation of our methods.
- **Chapter 7** concludes this study and address future work.

## Chapter 2

# Task Setting of Concept Ordering: Two Types of Target Users

In this chapter, we formally define the tasks of concept ordering.

### 2.1 Concept Ordering Targeting Whole Social Media Users

We exploit whole social media text (I.E., text in the most general domain) to infer the common values of all social media users.

**Input** A set of nominal concepts (*e.g.*,  $\{elephant, whale, dog, mouse\}$ ) is provided in the task, along with an adjective (*e.g.*, *large*<sup>1</sup>) that represents an attribute shared by all members of the set.

---

<sup>1</sup> We provide an antonym of a given adjective (*e.g.*, *small*) if any exists to reduce the ambiguity of the adjective.

## 2.2 Concept Ordering Targeting Social Media Users in Specific Domains

---

**Output** Given these two inputs, our goal is to output an ordered list of the items. The expected output in this example is *whale*  $\succ$  *elephant*  $\succ$  *dog*  $\succ$  *mouse*, where *whale* is the largest, *elephant* is the second largest, and so forth.

**Gold Standard** We ask multiple volunteers to order a given set of concepts and then use the ordering that achieved the best average Spearman [2]’s rank correlation coefficient,  $\rho$ , against the human orderings as a gold-standard ordering. The resulting ordering can be seen as a representative ordering that sums up the general human perception, and is thereby meaningful as a goal in our task.

## 2.2 Concept Ordering Targeting Social Media Users in Specific Domains

We exploit social media text in specific domains to induce the common values shared by the users in the domains. The domains of users are identified in advance (Section 5.2.1.2).

**Input** Along with the input described in Paragraph 2.1, our method accepts the domain information (*e.g.*, *women*, *living in Kanto region*).

**Output** The output format for this task is the same with Paragraph 2.1 but the output ordering reflects the common values included from the specified domain.

**Gold Standard** We also ask many volunteers to order a given set of concepts from various viewpoints (adjectives) and provide their domain information (*e.g.*, age, gender, prefecture they live in, SNS they use). We then generate the gold orderings for a domain that maximize the average Spearman’s  $\rho$  against the orderings

## 2.2 Concept Ordering Targeting Social Media Users in Specific Domains

---

of users in the domain. The resulting ordering can be seen as common values shared in a domain and provides a closer analysis of the target domain.

# Chapter 3

## Related Work

To the best of our knowledge, there have been no attempts to order concepts on the basis of the intensity of their attributes other than Nishina et al. [1] (in Japanese). Related tasks are discussed firstly in this chapter.

Question answering systems extract answers to factual questions (*e.g.*, ‘*What is the average temperature in Tokyo?*’) from text [7]. Similarly, some researchers have attempted to extract objects’ attributes and their values from the Web [8–13]. These studies can partly help us to perform our task, particularly when we order concepts in terms of the intensity of objective and numerical attributes (*e.g.*, *largeness*, *heaviness*, and *expensiveness*).

Aspect-based sentiment analysis mines reviews or other texts for opinions on entities (*e.g.*, products or movies) [14]. Some of these studies have handled statements comparing multiple items (*e.g.*, ‘*car x is two feet longer than car y*’ [15]). Kurashima et al. [16] proposed aggregating such statements to rank products in accordance with their popularity. This sort of information is also used with our method but is integrated with other evidence to obtain orderings for concepts that are not directly compared in texts. This strategy distinguishes our method from those proposed for aspect-based sentiment analysis.



In contrast to these studies, our task is more general in that it handles not only objective attributes (with numerical intensity, *e.g.*, *size* [13]) but also subjective attributes. Further, it handles not only entities (with specific values for attributes) but also concepts (with a range of values for attributes).

There has been a range of studies on aggregating pairwise comparisons (partial orderings) to a single consensus ordering [17–21]. These studies assume pairwise comparisons that are prepared (*e.g.*, search aggregation in meta-search or student evaluations via peer grading) or available from crowd-sourcing, while we do not assume them in our task setting to increase the applicability of the method.

Note that pairwise comparisons for combinations of specific concepts (entities) in comparative expressions are sparse and are not so helpful to solve our task. We thus integrate pieces of evidence in addition to pairwise comparisons to solve the task.

Făgărășan et al. [22] proposed a method of inducing feature norms [23] for a concept from text, which we use to perceive the concept. The features include adjectives (*e.g.*, ‘is\_sweet’ for sugar) but exclude ordering expressions (*e.g.*, ‘is\_larger\_than\_a\_pencil’ for dog) so their task is complementary to our task, which helps us to suggest possible attributes for given concepts.

Nishina et al. [1] (in Japanese) initiated the task that we tackled in this paper and proposed a method that orders concepts on the basis of the point-wise mutual information (PMI) of noun-adjective dependencies inspired by Turney [5]’s work. We also use the information but we combine it with other evidence as features in the framework of supervised learning.

# Chapter 4

## Methods for Concept Ordering

### 4.1 Inducing General Values on Concept Ordering

We resort to massive amounts of social media text to collect textual evidence that validates our perception on concept ordering (Section [4.1.1](#)) by assuming that our general values on concept ordering implicitly or explicitly affect the text we write. We then integrate that evidence to obtain a complete order of the concepts in the framework of supervised learning (Section [4.1.2](#)).

#### 4.1.1 Evidence on Concept Ordering

We exploit four types of evidence in this study to enable effective concept ordering. The first three implicitly suggest the absolute intensity of the attribute that the concept has. The fourth directly indicates an order of concepts, which captures the relative attribute intensity.

## 4.1 Inducing General Values on Concept Ordering

---

All pieces of evidence are represented by contexts where one or more concepts appear with a given adjective. The evidence is language-independent and we can tailor corresponding clue expressions for our target language, although our dataset is in Japanese. The example sentences that follow are provided in English to increase the applicability of our method and help reader understanding.

**Noun-adjective co-occurrence** If the intensity of the attribute of a concept is strong, we are likely to mention the attribute intensity along with the concept, which results in more sentences that include both the concept (noun) and the attribute (adjective).

- *Look how large that elephant is!*

**Noun-adjective dependency** A dependency relation between a nominative concept (noun) and an attribute (adjective) directly indicates the attribute intensity.

- *Elephants are so big.*

We used J.DepP [24–26],<sup>1</sup> a state-of-the-art dependency parser, to extract such dependency relations. This evidence is less frequent but provides stronger evidence than co-occurrences, since co-occurrences do not always indicate the intensity of the attribute (e.g., ‘*Ants are so small that elephants cannot harm them*’).

**Simile** If the intensity of the attribute of a concept is salient, we refer to the concept in a simile.

- *He is as brave as a lion.*

We use a couple of lexico-syntactic patterns tailored to detect similes, ‘ADJ as NP’ and ‘ADJ like NP’ (‘*marude* NP *no* you *ni* ADJ’ in Japanese).

**Comparative** This evidence directly indicates an order of a subset of a concept set. For example, the sentence below indicates *elephant*  $\succ$  *dog*.

- *Elephants are larger than dogs.*

---

<sup>1</sup><http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp/>

---

## 4.1 Inducing General Values on Concept Ordering

We use dependency relations for comparative adjectives with clues studied in the literature [15] (a post particle ‘*gori*’ in Japanese).

### 4.1.2 Two Approaches to Ordering Concepts

We explore two approaches to integrating the different types of heterogeneous evidence described in Section 4.1.1 to order concepts. The first uses a ranking support vector machine (svm) [3] to obtain an order and then induces concept-wise features from the evidence. The second uses a support vector regression (svr) [4], which learns a function that directly maps a given order to Spearman’s  $\rho$  against gold-standard ordering and induces order-wise features for candidate ordering.

#### 4.1.2.1 Ordering Concepts with Ranking svm

In ranking SVM, we represent each item (here, concept) with concept-wise features and perform pairwise training that generates  $\mathcal{O}(n^2)$  training examples (feature vectors) from all the pairs of  $n$  concepts in the gold-standard order. The ranking SVM then solves an optimization problem that involves minimizing the number of incorrect partial orderings. Testing of the ranking SVM involves the dot product between feature and weight vectors for the input  $n$  concepts, and hence only requires  $\mathcal{O}(n)$  time.

We encode the first three types of evidence described in Section 4.1.1 as real-valued features so that they directly indicate the attribute intensity of the concept. Generalizing Turney [5]’s work that used the point-wise mutual information (PMI) of a pair of a word and ‘excellent’ (and ‘poor’) to compute the semantic orientation of the word, we measure the polarity of words in terms of attributes other than ‘goodness.’ We compute feature value,  $\phi(\mathbf{x})_{coc}$ , for the noun-adjective co-occurrence of

## 4.1 Inducing General Values on Concept Ordering

a noun-adjective pair,  $\mathbf{x} = (\textit{noun}, \textit{adj})$ , as:

$$\begin{aligned}\phi(\mathbf{x})_{\textit{cooc}} &= SO_{\textit{cooc}}^{\textit{adj}}(\textit{noun}) \\ &= \text{PMI}(\textit{noun}, \textit{adj}) - \text{PMI}(\textit{noun}, \overline{\textit{adj}}) \\ &= \log \frac{p(\textit{noun}, \textit{adj})p(\overline{\textit{adj}})}{p(\textit{noun}, \overline{\textit{adj}})p(\textit{adj})}\end{aligned}\tag{4.1}$$

Here,  $\overline{\textit{adj}}$  refers to the antonym adjective or the adjective with negation. The feature values for dependency and simile are analogously computed, using co-occurrence counts based on dependency and simile.

We then encode the comparative expressions as two real-valued features so that they directly indicate the attribute intensity of the concept. For concept  $c$  and the adjective (here, *large*), we examine whether a comparative expression, ‘ $c$  is larger than  $c'$ ,’ can be found for any other concept  $c' (\neq c)$  in the given set of concepts. We encode the number of  $c'$  found in the expression divided by the number of items in the given concept set to obtain one real-valued feature, which indicates the positive intensity of the attribute. We analogously compute another real-valued feature by looking at a comparative expression, ‘ $c'$  is larger than  $c$ ,’ to express the negative intensity of the attribute.

### 4.1.2.2 Ordering Concepts with svr

In SVR, we represent each possible ordering by order-wise features and then directly map it to a real-valued measure. The SVR can incorporate the evaluation measure into the ordering process and directly optimize an ordering by outputting the ordering with the maximum estimated measure. We use Spearman’s  $\rho$  against the gold-standard order as a target variable for regression.

An issue to be addressed here is how candidate orderings for training and testing SVR are generated. There are factorial orders of candidate orderings,  $n!$ , for a given set of  $n$  concepts. As we do not need to use all the possible orderings in training,

## 4.1 Inducing General Values on Concept Ordering

---

we adopted a Monte Carlo method that randomly samples only a tractable number of orderings from all possible permutations of the given concept set. We extract the same number of candidate orderings from each concept set to obtain balanced training data, and that number is set to  $n_{min}!$ , where  $n_{min}$  is the smallest number of members of all the concept sets in the training data. There were up to eight concepts for ordering in the experiments that follow, considering the annotation cost to order concepts and a typical application scenario where users order concepts that are chosen (or narrowed down) from candidates *a priori*. We thus consider all possible candidate orderings in testing to understand the maximum performance of this approach. We can greedily add one remaining concept to the (ordered) list by choosing the best insertion position in accordance with the SVR value of the resulting ordering in testing with a set of a large number of concepts, starting from an empty list. This requires  $\mathcal{O}(n^2)$  time.

We induce real-valued features from the first three types of evidence described in Section 4.1.1 for the candidate ordering. We count the number<sup>2</sup> of ordered pairs  $(c, c')$  for which the  $SO_{cooc}^{adj}(c)$  is larger than  $SO_{cooc}^{adj}(c')$ , of all the ordered pairs in the candidate ordering. This is used to examine the extent to which pairwise orderings in the candidate ordering conform to the ordering specified by  $SO_{cooc}^{adj}$ . The feature values for dependency and simile are analogously computed using co-occurrence counts based on dependency and simile.

We then induce two real-valued features from comparative expressions by counting the number<sup>2</sup> of ordered pairs  $(c, c')$  in which  $c$ 's intensity is larger (or smaller) than  $c'$ 's, in all the ordered pairs in the candidate ordering for the given adjective. Here, we assume the attribute intensity of a concept is given by the corresponding feature value in the ranking SVM. A larger feature value means that the candidate ordering satisfies (or dissatisfies) more partial orders found in the text.

---

<sup>2</sup>These numbers are normalized by dividing them by the number of all the ordered pairs in the candidate ordering.

---

## 4.2 Inducing Common Values in Specific Domains

Here we say ‘ $x$  is larger than  $y$ ’ is positive for  $(x, y)$  and ‘ $y$  is larger than  $x$ ’ negative, we induce the positive evidence in the same way mentioned above, and the negative evidence with reverse check.

We then induce two real-valued features from comparative expressions by counting the number<sup>2</sup> of ordered pairs  $(c, x)$  in which comparative expression ‘ $c$  is larger than  $x$ ’ (or ‘ $x$  is smaller than  $c$ ’) can be found, in all the ordered pairs generated from the candidate ordering for the given adjective (here, *large*). A larger feature value means that the candidate ordering satisfies more partial orders found in the text.

## 4.2 Inducing Common Values in Specific Domains

Extending the methods to induce general values of whole social media users, we propose to induce common values shared by users in a specific domain.

### 4.2.1 Classify Domains of Users

In this study, we firstly target Twitter users and try to classify the genders and locations of them from their posts and profiles because these domains are relatively easy to acquire.

For gender, we adopted a simple heuristic that determines the gender according to the number of clue expressions (in their posts) indicating either gender; the clue expressions include first-person pronouns and sentence-ending particles specific to each gender [27].

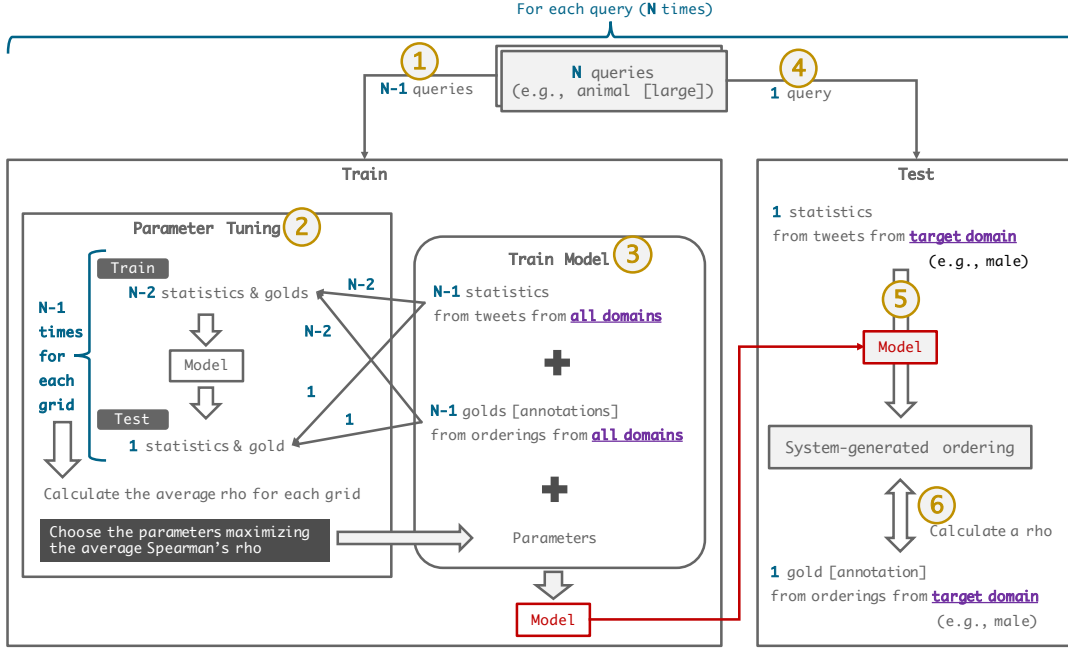
For location, since the precision of geo-locating posts is low, where the best median error distance is around 30km [28], we exploited the user profiles to annotate the location (living prefecture) of users. We extracted common locations specified by the users in their profiles by sorting the locations according to their frequency. We then manually assigned the common locations to an appropriate prefecture.

### 4.2.2 Training with Domain-Specific Data

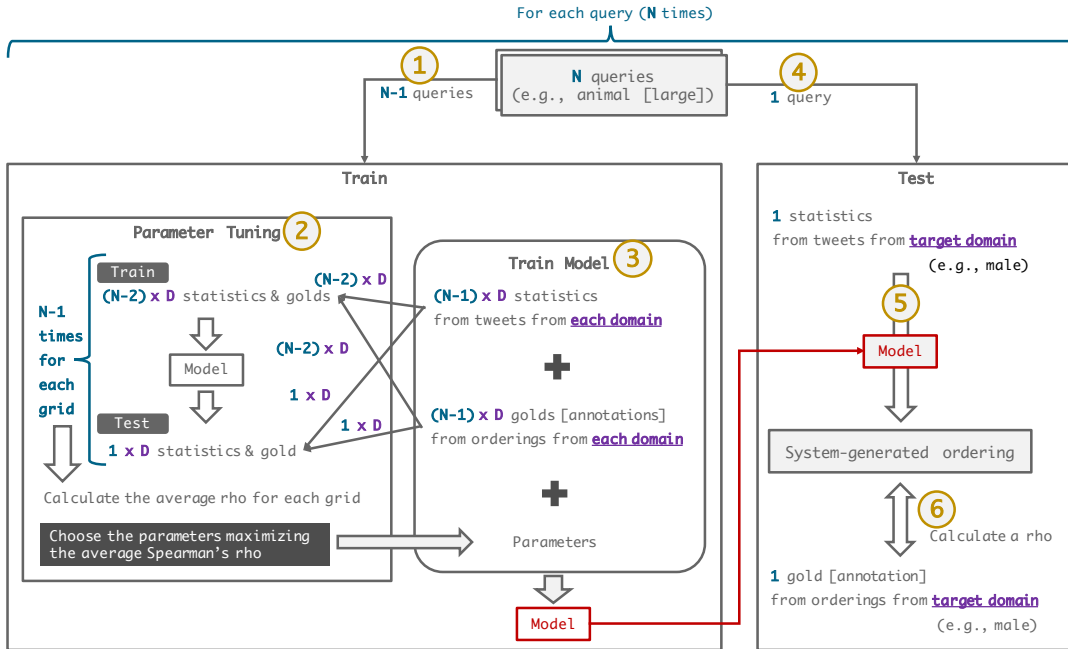
We have explored two different ways to train ranking SVM. Domain-unaware training uses the gold-standard orderings computed from the orderings given by all volunteers, while domain-aware training uses the gold-standard orderings for individual domains (*e.g.*, gender, region). In domain-aware training, the number of training examples is multiplied by the number of domains. Figure 4.1 briefly shows the procedures of the two training methods.



## 4.2 Inducing Common Values in Specific Domains



(A) Domain Unaware Training.



(B) Domain Aware Training.

FIGURE 4.1: Two ways to train ranking SVM (Domain-aware/unaware). The yellow numbers show the procedures.

# Chapter 5

## Evaluation

### 5.1 General Values on Concept Ordering

We performed experiments to evaluate our methods with open-domain datasets in terms of correlation between the system-generated and gold-standard orderings. We used LIBLINEAR [29]<sup>1</sup> as implementations of ranking SVM and SVR (with all hyper-parameters respectively tuned by cross-validation on training data).

#### 5.1.1 Datasets

We used around 260 million Japanese blog articles, which we have crawled since 2006, to build a dataset for evaluation and obtain the statistics for our methods. The blog articles consist of around two billion sentences written by more than a million users.

We built an evaluation dataset for this task from the blog articles to include real-world concepts that are often absent in handcrafted ontologies such as WordNet.

---

<sup>1</sup><https://www.csie.ntu.edu.tw/~cjlin/liblinear/>

## 5.1 General Values on Concept Ordering

TABLE 5.1: Evaluation dataset for acquiring general values.

Category	Adjective	Gold-standard ordering
flower	beautiful	sakura, rose, lily, lavender, platycodon, sunflower, camellia, daisy
jewel	elegant	sapphire, emerald, pearl, ruby, amethyst, opal, tourmaline, turquoise
alcohol	delicious	beer, wine, champagne, shōchū, chūhai, highball, tequila, makgeolli
sports	entertaining	football, table tennis, basketball, sumo, tennis, volleyball, baseball, professional wrestling
mammal	clever	dog, whale, cat, elephant, mouse, horse
mammal	large	whale, elephant, horse, dog, cat, mouse
conveyance	comfortable	Shinkansen, taxi, airplane, bicycle, bus, train
conveyance	fast	airplane, Shinkansen, train, taxi, bus, bicycle
food	yummy	steak, ramen, pasta, curry, pizza, fried rice, hamburger
instrument	soothing	flute, cello, clarinet, organ, trumpet, guitar, harmonica, drum
programming	easy	Ruby, Python, Perl, Java, JavaScript, Lisp, Scala, Haskell
programming	slow	Ruby, Perl, Python, JavaScript, Lisp, Haskell, Scala, Java
animal	lovely	squirrel, rabbit, dog, penguin, panda, horse, lizard, lion
vegetable	tasty	spinach, onion, pumpkin, eggplant, broccoli, napa cabbage, cucumber, sprout
fruit	sweet	melon, peach, apple, cherry, strawberry, tangerine, apricot
fruit	small	cherry, strawberry, apricot, tangerine, peach, apple, melon
appliance	useful	smartphone, PC, digital camera, car navigation system, printer, camera, speaker
flesh	preferable	chicken, beef, pork, lamb, brawn, venison, horseflesh
bird	cute	penguin, owl, quail, sparrow, swan, chicken, pheasant, eagle
weather	unpleasant	yellow sand, rain, thunder, gale, mist, snow, frost, fine
country	safe	UK, Thailand, Spain, India, Russia
country	warm	India, Thailand, Spain, UK, Russia
temple	famous	Kinkaku-ji, Ginkaku-ji, Hōryū-ji, Yakushi-ji, Zenkō-ji, Chūson-ji, Tō-ji, Zōjō-ji
temple	old	Hōryū-ji, Zenkō-ji, Yakushi-ji, Tō-ji, Chūson-ji, Zōjō-ji, Kinkaku-ji, Ginkaku-ji
cartoon	amusing	Gundam, Dragon Ball, One Piece, Vagabond, Kochikame, Gatchaman, Yatterman, Oishinbo
manufacturer	famous	Sony, Panasonic, Toshiba, NEC, Hitachi, Fujitsu, Canon, Seiko Epson
MLB team	famous	NY Yankees, SEA, BOS, LAD, NY Mets, CWS, BAL, CLE
fast-food chain	tasty	MOS Burger, Freshness Burger, KFC, Mister Donut, Burger King, McDonald's
automaker	healthy	Toyota, Honda, Yamaha, Mazda, Daihatsu
corner store	useful	7-Eleven, Lawson, FamilyMart, Seicomart, Ministop
corner store	numerous	7-Eleven, Lawson, FamilyMart, Ministop, Seicomart
browser	friendly	Chrome, Firefox, Safari, Opera, Sleipnir
city	safe	London, Berlin, Paris, Hong Kong, Chicago, Rome, Moscow
coffee shop	likable	Starbucks, Saint Marc, Tully's, Pronto, Doutor, Excelsior, Ginza Renoir
town	fashionable	Aoyama, Shibuya, Shinjuku, Shinagawa, Nakano, Ikebukuro, Ueno, Asakusa

We first applied word clustering [30] to one tenth of the blog articles in 2009 and obtained word clusters. We then looked into each cluster to manually collect nominal concepts in the same semantic category. Next, we associated each set of concepts with one or two adjectives that represented common attributes by examining the average PMI between the adjective and each concept, ultimately obtaining 35 pairs of a set of concepts and adjective (Table 5.1). There were 7.0 concepts per set on average and 28 unique sets of concepts (seven sets of concepts are associated with two adjectives). The resulting dataset included general to specific concepts (or instances) in various open-domain categories, and objective to subjective attributes for ordering that varied from one category to another.

## 5.1 General Values on Concept Ordering

We then asked seven volunteers (three graduate students, three researchers including the second author, and one system engineer) to provide an ordering for each pair of concepts and attribute. We regarded an ordering, in all permutations of concepts, that maximized the average of Spearman [2]’s rank correlation coefficient,  $\rho$ ,<sup>2</sup> against the seven human orderings, to be a gold-standard ordering. The gold-standard ordering we obtained is listed in Table 5.1.

We computed Spearman’s  $\rho$  between an ordering created by each of the seven subjects and the gold-standard ordering to find the agreement between each human ordering and the gold-standard ordering. The results are summarized in a box-and-whisker diagram (Figure 5.1). We should state that the average Spearman’s  $\rho$  between each human ordering and the gold-standard ordering was 0.75, which indicates a strong correlation.

We can see that the average Spearman’s  $\rho$  between each human ordering and the gold-standard ordering is larger than 0.4 for all the pairs of concepts and adjective, larger than 0.6 for 30 pairs except for ‘*sports (entertaining)*,’ ‘*mammal (clever)*,’ ‘*instrument (soothing)*,’ ‘*vegetable (tasty)*,’ and ‘*coffee shop (likable)*,’ and even larger than 0.8 for 15 pairs that included subjective ones such as ‘*programming (easy)*,’ ‘*bird (cute)*,’ and ‘*temple (famous)*,’ which confirms that the perception of the relative intensities of common attributes is largely shared by human subjects. This demonstrates that the gold-standard ordering is appropriate as general values, shared by volunteers, on ordering for each pair.

### 5.1.2 Results

We conducted leave-one-out cross-validation using the evaluation dataset described in Section 5.1.1. The appropriateness of the system-generated orderings was then measured by computing Spearman’s  $\rho$  between the system-generated and gold-standard orderings. The experimental results are listed in Table 5.2. Here, HUMAN

---

<sup>2</sup>Spearman’s  $\rho$  measures the strength of the correlation between two ordered lists. It ranges from  $-1$  to  $1$ . The negative value indicates an inverse correlation.

## 5.1 General Values on Concept Ordering

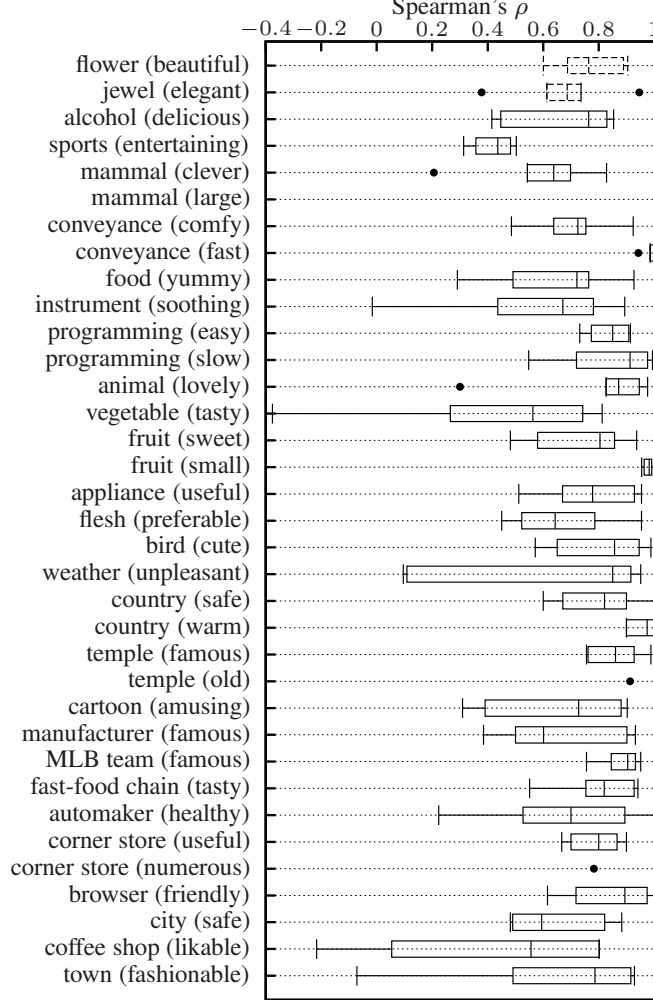


FIGURE 5.1: Correlation between the human orderings and the gold-standard ordering.

refers to the average Spearman's  $\rho$  between each human ordering and the gold-standard ordering, while BASELINE refers to a method that scores each concept by using the PMIs of noun-adjective dependencies for the given adjective and its antonym ( $SO_{depend}^{adj}$ ), as was done in Nishina et al. [1]'s work. SVM and SVR refer to our methods based on ranking SVM and SVR (Section 4.1.2), respectively. The results revealed that our methods overwhelmed the baseline, which indicated the effectiveness of integrating heterogeneous evidence in a supervised manner. Ranking SVM achieved the best average performance and indicated a positive correlation for

## 5.1 General Values on Concept Ordering

TABLE 5.2: Results on ordering concepts: Spearman’s  $\rho$  against gold-standard ordering.

Category (adjective)	HUMAN	BASE- LINE	SVM	SVR
flower (beautiful)	0.767	0.190	<u>0.357</u>	0.167
jewel (elegant)	0.682	0.299	0.524	<u>0.548</u>
alcohol (delicious)	0.663	−0.048	<u>0.762</u>	<u>0.810</u>
sports (entertaining)	0.422	0.238	<u>0.381</u>	−0.095
mammal (clever)	0.598	−0.20	<u>0.143</u>	0.029
mammal (large)	1.000	<u>0.943</u>	0.771	0.886
conveyance (comfy)	0.712	0.371	<u>0.486</u>	0.257
conveyance (fast)	0.986	0.257	0.543	<u>0.771</u>
food (yummy)	0.639	0.000	<u>0.607</u>	0.464
instrument (soothing)	0.583	−0.048	<u>0.310</u>	0.238
programming (easy)	0.845	0.476	0.619	<u>0.643</u>
programming (slow)	0.840	0.192	<u>0.381</u>	0.238
animal (lovely)	0.806	<u>0.738</u>	0.548	0.595
vegetable (tasty)	0.462	0.238	<u>0.524</u>	0.476
fruit (sweet)	0.729	<u>0.964</u>	0.607	0.607
fruit (small)	0.979	0.286	<u>0.607</u>	0.536
appliance (useful)	0.772	<u>0.536</u>	0.393	0.500
flesh (preferable)	0.662	−0.429	<u>0.143</u>	−0.286
bird (cute)	0.819	0.881	<u>0.929</u>	<u>0.929</u>
weather (unpleasant)	0.664	<u>0.762</u>	0.524	0.143
country (safe)	0.804	−0.500	−0.200	<u>0.000</u>
country (warm)	0.961	<u>0.900</u>	0.700	0.700
temple (famous)	0.861	0.168	0.524	<u>0.619</u>
temple (old)	0.988	0.500	0.595	<u>0.667</u>
cartoon (amusing)	0.648	−0.167	0.429	<u>0.476</u>
manufacturer (famous)	0.659	0.381	<u>0.619</u>	0.286
MLB team (famous)	0.885	0.952	0.905	<u>0.976</u>
fast-food chain (tasty)	0.807	0.543	<u>0.886</u>	0.543
automaker (healthy)	0.665	−0.900	<u>−0.700</u>	−0.900
corner store (useful)	0.791	<u>0.400</u>	0.100	0.100
corner store (numerous)	0.969	0.300	<u>0.500</u>	<u>0.500</u>
browser (friendly)	0.856	<u>−0.200</u>	−0.600	−0.600
city (safe)	0.655	0.000	<u>0.357</u>	0.250
coffee shop (likable)	0.405	0.071	<u>0.786</u>	0.464
town (fashionable)	0.673	<u>0.512</u>	0.381	0.262
average	0.750	0.275	<u>0.441</u>	0.366

all pairs other than ‘*country (safe)*,’ ‘*automaker (healthy)*,’ and ‘*browser (friendly)*.’ Figure 5.2 shows the correlations between the Spearman’s  $\rho$ s of the gold-standard orderings and proposal methods. This figure implies that they have correlations,

## 5.1 General Values on Concept Ordering

TABLE 5.3: Ablation test for ranking SVM.

Method	Spearman’s $\rho$
SVM (all)	0.441
–co-occurrence	0.391
–dependency	0.407
–simile	0.292
–comparative	0.424

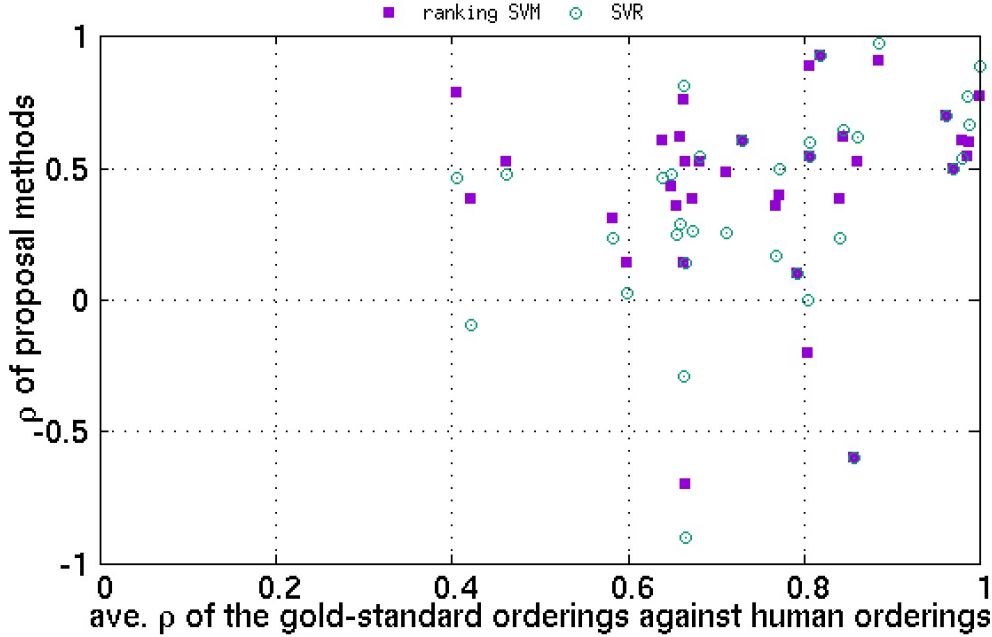


FIGURE 5.2: Correlations of the gold-standard ordering and the proposal methods. X-axis shows the average Spearman’s  $\rho$ s of gold-standards against human orderings and Y-axis shows the Spearman’s  $\rho$ s of proposal methods.

i.e. our methods generate better results for the gold-standard orderings that have higher average Spearman’s  $\rho$  against human orderings.

Table 5.3 shows an ablation test that evaluates the impact of individual pieces of evidence in ranking SVM. All the evidence contributed to improving the correlation against the gold-standard orderings. The simile most significantly affects the performance, followed by noun-adjective co-occurrences and dependencies. This conforms to our expectation since similes are useful for finding concepts with strongest

## 5.1 General Values on Concept Ordering

TABLE 5.4: Examples of system-generated orderings.

	GOLD	BASELINE	SVM	SVR
<b>food (yummy)</b>				
1	steak	steak	steak	curry
2	ramen	pizza	curry	steak
3	pasta	fried rice	pizza	ramen
4	curry	curry	ramen	pizza
5	pizza	hamburger	fried rice	fried rice
6	fried rice	pasta	pasta	hamburger
7	hamburger	ramen	humburger	pasta
<b>alcohol (delicious)</b>				
1	beer	makgeolli	beer	wine
2	wine	beer	wine	beer
3	champagne	highball	shōchū	champagne
4	shōchū	champagne	champagne	shōchū
5	chūhai	shōchū	makgeolli	makgeolli
6	highball	chūhai	highball	chūhai
7	tequila	wine	tequila	tequila
8	makgeolli	tequila	chūhai	highball
<b>fruit (small)</b>				
1	cherry	apricot	cherry	cherry
2	strawberry	peach	apricot	apricot
3	apricot	tangerine	tangerine	apple
4	tangerine	cherry	apple	tangerine
5	peach	apple	peach	peach
6	apple	strawberry	strawberry	strawberry
7	melon	melon	melon	melon
<b>programming (easy)</b>				
1	Ruby	Ruby	Ruby	Ruby
2	Python	Scala	Perl	Java
3	Perl	Perl	Java	Perl
4	Java	JavaScript	Scala	JavaScript
5	JavaScript	Java	JavaScript	Scala
6	Lisp	Python	Python	Python
7	Scala	Lisp	Haskell	Haskell
8	Haskell	Haskell	Lisp	Lisp

(or weakest) attribute intensity, which greatly contributed to the improvement of Spearman’s  $\rho$  against the gold-standard orderings. The comparative expressions are rarely observed in the text, so that had the least impact.

We manually investigated the system-generated orderings to identify which evidence



had contributed to the orderings. Table 5.4 lists examples of system-generated orderings. Noun-adjective co-occurrences between concept (noun) and *yummy* in the ordering of ‘*food (yummy)*’ raised the ordering of *ramen*, so the resulting orderings of our systems correlated with the gold-standard ordering. Similes in the ordering of ‘*fruit (small)*’ greatly contributed to improving the ordering of *cherry* (to top-1), which again increased the performance of our systems. Comparative expressions in the ordering of ‘*programming (easy)*’ contributed to raising the ordering of *Java*, which again increased the performance of our systems. These changes suggest that the use of heterogeneous evidence increases the chances of reaching ordering that has a higher correlation with the gold-standard ordering, which justifies our approach to integrating the pieces of evidence in a supervised manner.

The low correlation in ‘*country (safe)*’ and ‘*automaker (healthy)*’ was caused by their status changes over our time-series text. As for ‘*browser (friendly)*,’ similes and comparative expressions were not observed, and this data sparseness reduced the correlation. There were also some difficult cases in gathering reliable evidence. For example, we sometimes mention unusual and surprising things such as “*The mice around here are large,*” which have a negative impact on ordering concepts.

## 5.2 Common Values in Domains

We next conducted experiments to evaluate our method to infer values in specific domains (Section 4.2), with our archive of 25 billion Twitter posts, as we evaluated our methods with blog articles in Section 5.1, 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.

### 5.2.1 Datasets

#### 5.2.1.1 Evaluation Datasets

We prepared 28 queries (pairs of a set of concepts and an adjective) with the same process described in Section 5.1.1, which used a word clustering-based method. They cover wider variety of queries than the previous dataset: 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, we used a crowdsourcing service (<https://crowdworks.jp/>) to ask 53 Japanese Twitter users to answer (rank) each query and provide their attributes (genders, prefectures they live in, ages, social networking services they use, occupations). 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 domain (*e.g.*, male Twitter users) by choosing an ordering, in all permutations of concepts, that maximized the average of Spearman [2]’s rank correlation coefficient  $\rho$  against the orderings of the workers by domain. The evaluation datasets and correlations of some domains are shown in Table 5.5 and 5.6.

Here, ALL refers to the average Spearman’s  $\rho$  between all crowd worker orderings, while FEMALE and MALE refers to the average  $\rho$  between female and male crowd workers respectively. In addition to them, the average  $\rho$ s were calculated for data with KANTO tag that were gathered only from users living in Kanto region. The result shows strong correlations between human orderings (more than 0.65) and more specific domain has higher average correlation, that is to say, people in more specific domain agree more with their gold-standard orderings. Therefore, looking into the correlates, we can see the differences between domains.

## 5.2 Common Values in Domains

TABLE 5.5: Evaluation dataset for acquiring common values in specific domains.  
This table shows the gold-standard ordering of all the workers (ALL).

Category	Adjective	Gold-standard ordering
animal	large	whale, elephant, bear, cow, horse, dog, cat, mouse
flower	beautiful	sakura, rose, lily, sunflower, camellia, lavender, platycodon, daisy
jewel	elegant	pearl, sapphire, emerald, ruby, amethyst, opal
alcohol	delicious	champagne, wine, chuhai, beer, highball, whisky, shochu
sports	entertaining	football, baseball, tennis, volleyball, table tennis, basketball, sumo, professional wrestling
mammal	clever	dog, elephant, cat, whale, horse, bear, cow, mouse
conveyance	comfortable	Shinkansen, airplane, taxi, train, bicycle, bus, boat
food	yummy	steak, ramen, curry, pasta, pizza, hamburger, fried rice
instrument	soothing	piano, flute, cello, guitar, harmonica, trumpet, drum
bird	lovely	squirrel, penguin, rabbit, dog, panda, horse, lion, lizard
vegetable	tasty	pumpkin, onion, eggplant, spinach, cucumber, napa cabbage, broccoli, sprout
fruit	sweet	peach, melon, strawberry, grape, apple, cherry, tangerine, persimmon
appliance	useful	PC, smartphone, car navigation system, digital camera, printer, speaker
flesh	preferable	chicken, beef, pork, lamb
bird	cute	penguin, sparrow, owl, quail, swan, eagle, pheasant, chicken
weather	likable	fine, snow, mist, rain, frost, gale, thunder, yellow sand
country	safe	UK, France, Spain, America, Thailand, India, Russia, China
conveyance	fast	airplane, Shinkansen, train, taxi, boat, bus, bicycle
café	likable	Starbucks, Doutor, Tully 's, Saint Marc
corner store	useful	7-Eleven, Lawson, FamilyMart, Ministop
automaker	healthy	Toyota, Honda, Daihatsu, Mazda, Yamaha
fast-food chain	tasty	MOS Burger, KFC, Mister Donut, McDonald's
manufacturer	famous	Sony, Panasonic, Toshiba, Canon, Hitachi, Fujitsu, Seiko Epson
temple	famous	Kinkaku-ji, Horyu-ji, Ginkaku-ji, Yakushi-ji, Zenko-ji, Chuson-ji, To-ji, Zojo-ji
fruit	large	melon, apple, peach, persimmon, tangerine, grape, strawberry, cherry
country	warm	India, Thailand, Spain, America, China, France, UK, Russia
coffee shop	numerous	7-Eleven, Lawson, FamilyMart, Ministop
actress	cute	Ki Kitano, Yuka, Kazue Fukishi, Akina Minami, Yumiko Shaku, Rinka, Tomochika

To show these differences more clearly, we manually investigated the gold-standard ordering of female (FEMALE) and male (MALE) domains, and summarized them in Table 5.7. In the table, some notable points are underlined. For example, as for alcohol (delicious), women have much stronger correlation than men have.

### 5.2.1.2 Twitter Datasets

We targeted Twitter as data source for classifying domains by using Twitter users' posts and profiles. We have crawled for more than five years' worth of Twitter posts using Twitter API since March 11, 2011. We started crawling timelines from 30 famous Japanese users, and then repeatedly expanded the set of users by following retweets and mentions appeared in the timelines while tracking their timelines. Our

## 5.2 Common Values in Domains

TABLE 5.6: Correlation between human orderings. For each specific domain, five most different correlations compared to their ALL domain correlations are double-underlined.

Category (adjective)				KANTO		
	ALL	FEMALE	MALE	ALL	FEMALE	MALE
mammal (large)	0.974	0.975	0.978	0.971	0.967	0.983
flower (beautiful)	0.621	0.606	0.632	0.631	0.666	0.644
jewel (elegant)	0.411	<b><u>0.552</u></b>	0.443	0.539	<b><u>0.653</u></b>	0.619
alcohol (delicious)	0.348	<b><u>0.541</u></b>	<b><u>0.151</u></b>	0.392	<b><u>0.582</u></b>	<b><u>0.134</u></b>
sports (entertaining)	0.448	<b><u>0.558</u></b>	<b><u>0.407</u></b>	0.491	<b><u>0.585</u></b>	0.477
mammal (clever)	0.612	0.645	0.626	0.625	0.648	0.629
conveyance (comfortable)	0.619	0.592	0.652	0.597	0.583	0.632
food (yummy)	0.374	0.406	0.395	0.341	0.414	0.256
instrument (soothing)	0.582	0.652	0.574	0.645	0.654	0.651
animal (lovely)	0.662	0.679	0.721	0.673	0.654	0.785
vegetable (tasty)	0.310	0.283	<b><u>0.448</u></b>	0.383	<b><u>0.301</u></b>	<b><u>0.541</u></b>
fruit (sweet)	0.602	0.643	0.601	0.628	0.604	0.688
appliance (useful)	0.804	0.817	0.793	0.809	0.821	0.789
flesh (preferable)	0.489	0.492	0.620	0.516	0.460	<b><u>0.647</u></b>
bird (cute)	0.715	0.700	0.795	0.750	0.757	0.841
weather (pleasant)	0.756	0.811	0.724	0.749	0.794	0.710
country (safe)	0.605	<b><u>0.526</u></b>	<b><u>0.707</u></b>	0.578	0.502	<b><u>0.732</u></b>
conveyance (fast)	0.922	0.928	0.924	0.914	0.904	0.957
coffee shop (likable)	0.517	0.522	0.509	0.421	0.469	0.531
corner store (useful)	0.738	0.813	0.720	0.750	0.793	0.712
automaker (healthy)	0.628	0.636	0.590	0.642	0.648	0.644
fast-food chain (tasty)	0.570	<b><u>0.463</u></b>	<b><u>0.673</u></b>	0.586	<b><u>0.461</u></b>	<b><u>0.799</u></b>
manufactur (famous)	0.690	0.632	0.734	0.724	0.707	0.782
temple (famous)	0.839	0.847	0.830	0.853	0.870	0.835
fruit (large)	0.886	0.880	0.896	0.890	0.881	0.916
country (warm)	0.882	0.879	0.891	0.869	0.854	0.895
corner store (numerous)	0.808	0.748	0.892	0.892	0.884	0.925
actress (beautiful)	0.641	0.595	0.669	0.675	0.611	0.788
average	0.645	0.658	0.664	0.662	0.669	0.698

archive has more than 2 million users and 25 billion tweets. For the time being, we have used posts written in Japanese since we need to develop various NLP tools to analyze gender and location of Twitter users from their posts in addition to fast part-of-speech tagger and dependency parser to extract textual evidence on concept

## 5.2 Common Values in Domains

TABLE 5.7: Differences between some gold-standard examples from female (F, FEMALE) / male domains (M, MALE). Notable points are underlined.

Query	Ave. $\rho$	Gold-standard ordering
flower (beautiful)	0.606 0.632	F sakura, rose, lily, camellia, sunflower, platycodon, lavender, daisy M sakura, rose, lily, sunflower, lavender, camellia, platycodon, daisy
jewelry (elegant)	0.552 0.443	F pearl, sapphire, emerald, ruby, amethyst, opal M emerald, sapphire, ruby, pearl, amethyst, opal
alcohol (delicious)	0.541 0.151	F champagne, <u>chuhai</u> , wine, highball, beer, whisky, shochu M champagne, wine, beer, highball, <u>chuhai</u> , whisky, shochu
sport (entertaining)	0.558 0.407	F football, volleyball, tennis, baseball, table tennis, basketball, sumo, pro wrestling M <u>baseball</u> , football, tennis, volleyball, basketball, table tennis, pro wrestling, sumo
food (yummy)	0.406 0.395	F steak, pasta, pizza, curry, <u>ramen</u> , hamburger, fried rice M steak, <u>ramen</u> , curry, pizza, pasta, fried rice, hamburger
animal (lovely)	0.679 0.721	F dog, squirrel, penguin, rabbit, panda, lion, horse, lizard M squirrel, rabbit, penguin, dog, panda, horse, lion, lizard
vegetable (tasty)	0.283 0.448	F pumpkin, onion, spinach, napacabbage, eggplant, cucumber, broccoli, sprout M pumpkin, onion, eggplant, cucumber, spinach, napacabbage, sprout, broccoli
appliance (useful)	0.817 0.793	F smartphone, PC, car navigation system, digital camera, printer, speaker M <u>PC</u> , smartphone, car navigation system, digital camera, printer, speaker
flesh (preferable)	0.492 0.62	F chicken, pork, beef, lamb M beef, chicken, pork, lamb
fast-food (tasty)	0.463 0.673	F MOS Burger, KFC, Mister Donut, McDonald's M same
fruit (large)	0.88 0.896	F melon, apple, peach, persimmon, <u>tangerine</u> , <u>grape</u> , strawberry, cherry M melon, apple, peach, persimmon, <u>grape</u> , <u>tangerine</u> , strawberry, cherry
actress (cute)	0.595 0.669	F Ki Kitano, Yuka, KazueFukishi, AkinaMinami, Yumiko Shaku, Rinka, Tomochika M same

orderings.

As described in Section 4.2.1, we briefly analyzed the gender and location of Twitter users from their posts and profiles in order to annotate posts with user attributes. The gender classifier detected 311 thousand males and 345 thousand females (Japanese users), and the region classifier detected 201 thousand Japanese users.

We shortly evaluated the precision of the simple classification method for gender with a classification method based on user profiles which uses pattern matching (*e.g.*, matches *housewife*) and can be considered as a more precise method, but classifies a smaller number of users because of the lack of information. Table 5.8 shows the result. We used NAIVE method for gender classification and PROFILE-BASE is the more precise method, which exploits user profiles. According to the

## 5.2 Common Values in Domains

TABLE 5.8: Brief evaluation on gender classifications.

The number of users	PROFILE-BASE	NAIVE
12774	F	F
1554	F	M
2439	M	F
5468	M	M

result, NAIVE achieved the precision of 77.9% for classifying men and 84.0% for women.

Because the total amount of tweets was huge, we narrowed down them by filtering out the tweets which don not have a pair of a concept and an adjective in the given query set and roughly estimated the PMI with these tweets.

### 5.2.2 Results

We used two methods to train ranking SVM described in Section 4.2.2. We evaluated our methods with two domains (male and female). While domain-unaware training exploited the gold-standard orderings computed from the orderings of all the workers, domain-aware training used the gold-standard orderings of the workers in individual domains (here, two domains). In testing, we inputted statistics collected from Twitter posts (Jan. 2012 - Dec. 2015) in each domain to obtain domain-specific orderings.

Table 5.9 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  $\rho$  against the gold-standard ordering in the domain. Here, BASELINE refers to the baseline method which scores each concept on the basis of noun-adjective dependencies (Nishina et al. [1]). Both of our methods overwhelmed the baseline and the domain-aware training obtained better Spearman’s  $\rho$  than the domain-unaware training, which showed the effectiveness of considering the domains while training ranking SVM.

## 5.2 Common Values in Domains

TABLE 5.9: Spearman’s  $\rho$  against domain-specific gold-standard orderings.

Category (adjective)	BASELINE		DOMAIN-UNAWARE		DOMAIN-AWARE	
	FEMALE	MALE	FEMALE	MALE	FEMALE	MALE
animal (large)	-0.095	0.143	0.214	0.619	0.214	0.619
flower (beautiful)	0.810	0.714	0.571	0.571	0.619	0.571
jewel (elegant)	0.371	-0.714	0.886	-0.657	0.714	-0.657
alcohol (delicious)	-0.107	-0.179	-0.107	0.321	-0.143	0.571
sports (entertaining)	-0.143	-0.310	-0.143	-0.548	-0.095	-0.405
mammal (clever)	0.548	0.190	0.619	0.238	0.619	0.262
conveyance (comfortable)	0.464	0.179	0.571	0.179	0.571	0.179
food (yummy)	0.464	0.286	0.250	0.714	0.250	0.714
instrument (soothing)	-0.071	-0.179	-0.036	-0.357	-0.036	-0.357
bird (lovely)	0.048	0.024	0.310	-0.048	0.333	0.000
vegetable (tasty)	0.619	0.500	0.619	0.500	0.571	0.619
fruit (sweet)	-0.667	-0.143	0.167	0.405	0.167	0.405
appliance (useful)	-0.543	0.143	0.371	0.143	0.371	0.143
flesh (preferable)	0.000	0.200	0.400	0.800	0.600	0.200
bird (cute)	0.810	0.690	0.881	0.786	0.833	0.667
weather (likable)	0.810	0.690	0.619	0.714	0.667	0.881
country (safe)	0.690	-0.333	0.643	-0.667	0.524	-0.643
conveyance (fast)	0.464	0.393	-0.107	0.214	0.071	0.286
café (likable)	-0.400	0.200	-0.800	0.200	-0.800	0.200
corner store (useful)	0.800	0.800	0.400	0.400	0.400	0.400
automaker (healthy)	-0.800	-0.100	0.500	0.200	0.500	0.200
fast-food ch (tasty)	0.400	0.400	0.400	0.400	0.400	0.400
manufacturer (famous)	0.071	-0.036	0.679	0.643	0.679	0.643
temple (famous)	0.405	0.119	-0.048	0.714	0.048	0.714
fruit (large)	-0.333	-0.238	0.643	0.214	0.762	0.333
country (warm)	-0.833	-0.833	-0.381	-0.595	-0.381	-0.595
coffee shop (numerous)	0.000	-0.600	0.800	0.400	0.800	0.400
actress (cute)	0.214	0.536	0.214	0.393	0.357	0.393
Avg. $\rho$	0.143	0.091	0.326	0.246	0.343	0.255

## Chapter 6

# A System for Concept Ordering

We build Kotonush, a demonstration system which can derive orderings from segmented text written in different periods (in 2014 vs. in 2015), in different geographical areas (Kanto vs. Kansai), or by different demographics (males vs. females), to observe changes in common values over time or place or by different demographics.

Our system consists of three parts: (1) an interactive querying interface, (2) a text-to-ordering module, and (3) an ordering visualizer (Figure 6.1). Our querying interface enables users to interactively input a set of concepts and an adjective as a query (Figure 6.2) and then sends them to the text-to-ordering module. The querying interface accepts several options that specify domains such as genders and regions where social media users live in 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 6.4 and Figure 6.5). 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.



## 6.1 System Workflow

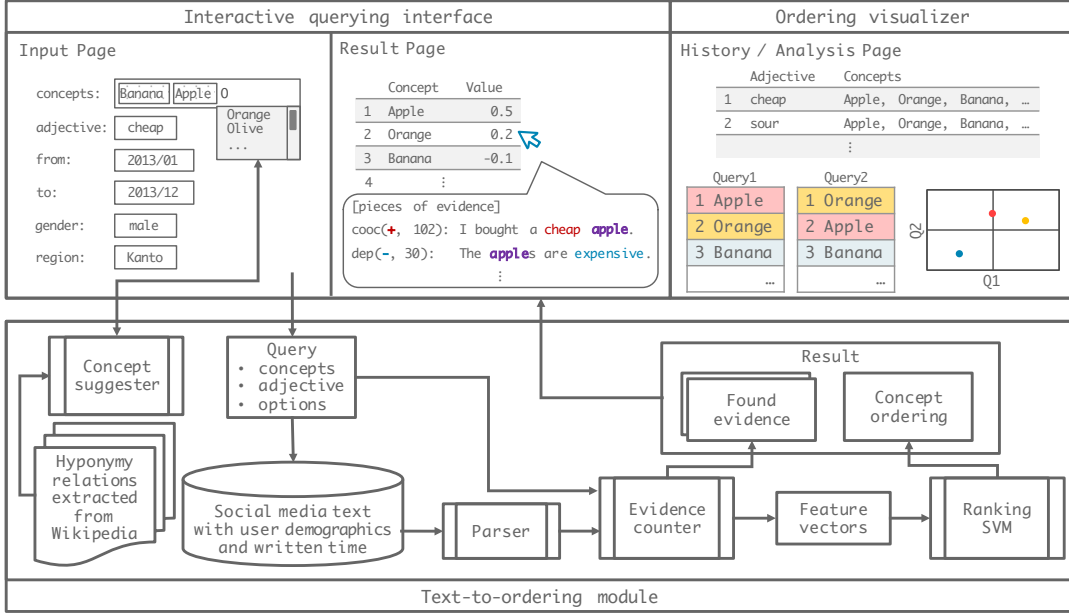


FIGURE 6.1: Overview of Kotonush, the system we developed to acquire values from social media text: interactive querying interface, text-to-ordering module and ordering visualizer.

Note that the domain analyses provide deeper and closer insight into not only target concepts but also target domains (*e.g., men in Japan like action movies much better than Disney movies*, as we will reveal in the following case studies).

## 6.1 System Workflow

In the following, we describe the workflow of our demonstration system in more detail.

### 6.1.1 Preprocessing

Our system needs a mechanism that retrieves useful posts fast from social media text because we assume using a large amount of text as archive to gather pieces of

FIGURE 6.2: Interactive querying interface accepts a set of concepts, an adjective and options.

evidence. In our system, the process is done dynamically to avoid keeping statistics about concepts that takes much space of disks. For the purpose, we assume a search engine to retrieve posts that include concepts and adjectives and have built a simple inverted index-based search engine. 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 orderings are done online.

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 (Section 5.2.1.2). Since this process is outside the focus of this study, here we just use existing methods based on bag-of-words.

### 6.1.2 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 6.2).

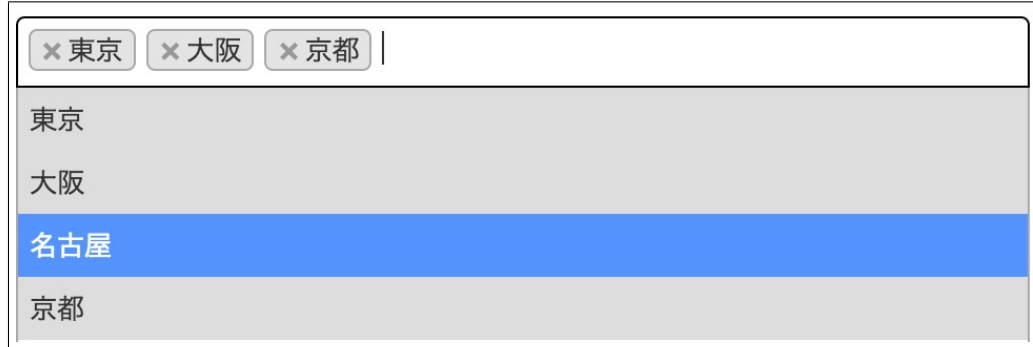


FIGURE 6.3: The system suggests concepts in the same category by using hypernym relations.

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 (Figure 6.3). We exploit hyponymy relations extracted from Wikipedia [6] to suggest concepts that share the same hypernym with the given concepts. Although handcrafted ontologies such as WordNet are widely used, as ontologies often lack real-world concepts, we, therefore, used hyponymy extracted from Wikipedia.

### 6.1.3 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 [24–26],<sup>1</sup> that is just as fast as the fastest front-end part-of-speech tagger<sup>2</sup> (> 15,000 posts/s on an Intel® Xeon® E7-4830 2.13GHz CPU server).

<sup>1</sup>J.DepP: <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/jdepp>

<sup>2</sup>MeCab: <http://taku910.github.io/mecab/>

The parsed text is given to our implementation of our method to induce a concept ordering, which we explained in Chapter 4 as the text-to-ordering module. The method encodes the four types of evidence as real-valued features so that they directly indicate the attribute intensity of the concept by using point-wise 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 or SVR. We choose ranking SVM for our demo system since we showed ranking SVM outperformed SVR. As we mentioned, our method (Section 4.1) requires an evidence counter and a supervised learning method for gathering and combining the evidence. Our evidence counter counts pieces of evidence about given concepts and adjectives in the parsed posts and outputs feature vectors for supervised learning along with all found pieces of evidence. Receiving the output of the evidence counter, the ranking SVM solves an optimization problem that tries to minimize the number of incorrect partial orderings. Receiving the output of our evidence counter program, the ranking SVM generates the best ordering. Finally, the text-to-ordering module returns the joint results of the best ordering and found pieces of evidence so that users can know what social media users say about each item.

### 6.1.4 Ordering Visualizer

By keeping the results of past queries in our system, users can review and compare them on the History / Analyze page (Figure 6.4 and Figure 6.5). This page provides complete sets of cached results as a table and tools to analyze queries with the same concepts and different settings such as bump charts (top of Figure 6.5). With bump charts, users can, for example, determine the best season for each flower by varying periods (*e.g.*, four seasons).

In addition to bump charts, we implemented an interface of scatter plots on the page (bottom of Figure 6.5). 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

Cached Queries							
select queries to compare							
Id	Adjective	Concepts	From	To	Gender	Region	Actions
1	entertaining	京都,大阪,東京,福岡	2013/01	2013/12	Male	Not Specified	Delete
2	entertaining	京都,大阪,東京,福岡	2013/01	2013/12	Not Specified	Kanto	Delete
3	entertaining	京都,大阪,東京,福岡	2013/01	2013/12	Not Specified	Kansai	Delete
5	entertaining	京都,大阪,東京,福岡	2013/01	2013/12	Not Specified	Kyushu	Delete

FIGURE 6.4: History page keeps cached query results.

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*).

## 6.2 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 (Section 5.2) 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.

**Movie (Likable)** The first case study captures common values on movies in terms of gender (Figure 6.5). In Japan, men tend to like action movies much better than Disney movies but women don’t like action movies so much.

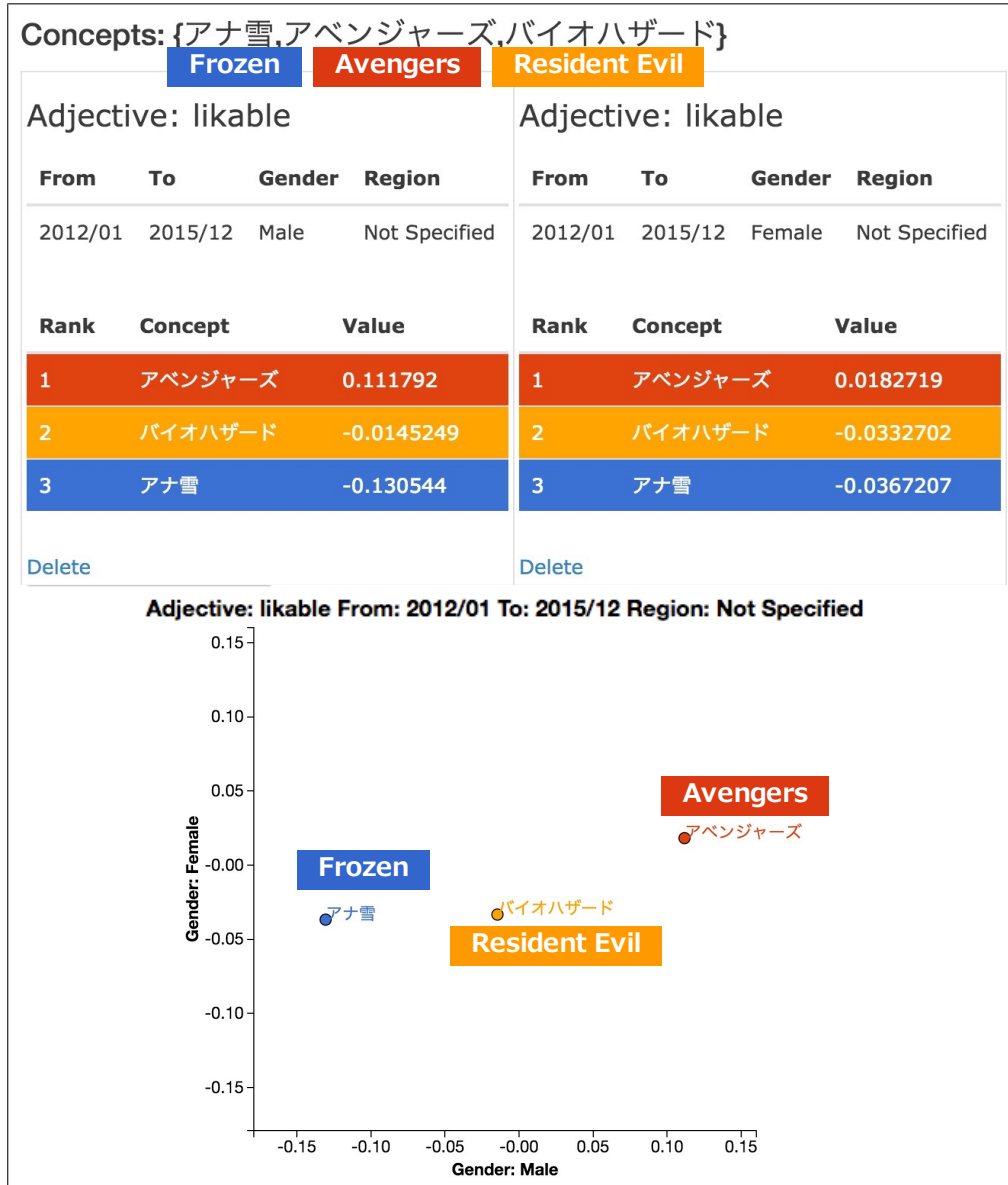


FIGURE 6.5: Users can compare orderings with different settings. This example compares movies (Frozen, Avengers, and Resident Evil) in terms of ‘likable’ with two genders.

**Flower (Beautiful)** The next case compares three seasonal flowers – (Japanese) cherry, sunflower, and Chrysanthemum (mum) – in terms of beauty in different seasons (Table 6.1a). The results clearly show the blooming season (best time) of each flower.

TABLE 6.1: Case studies in different settings.

(A) Flower (Beautiful): Three seasons

	<i>Mar. - May</i>	<i>Jun. - Aug.</i>	<i>Sep. - Nov.</i>
1	Cherry	Sunflower	Mum
2	Mum	Cherry	Sunflower
3	Sunflower	Mum	Cherry

(B) Disease (Fearful): Years

	2013	2014
1	Influenza	Dengue
2	Malaria	Influenza
3	Dengue	Malaria

(C) Fruit (Delicious): Regions

	<i>Tohoku</i>	<i>Shikoku</i>
1	Apple	Tangerine
2	Strawberry	Apple
3	Tangerine	Strawberry

**Disease (Fearful)** 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.

**Fruit (Delicious)** Table 6.1c 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.

# Chapter 7

## Conclusion

We have addressed a task of ordering concepts on the basis of the intensity of their common attributes and proposed methods of ordering a given set of concepts by aggregating heterogeneous pieces of evidence obtained from massive amounts of social media text. As the experimental results obtained from real-world concepts revealed a strong correlation between the system-generated and gold-standard orderings, we could induce general values on ordering concepts from text written by social media users.

Extending the methods to infer the general values from whole social media users, we proposed a method to obtain domain-specific values by classifying the domains (*e.g.*, genders, regions they live in) of users to gain deeper insight into the values in specific domains. As an implementation of the method inducing domain-specific values, we developed Kotonush, a system that enables users to interactively ask queries (concepts and an adjective) and compare the induced orderings for deeper understanding of the concepts. We empirically evaluated the system with a handful of interesting case studies, comparing concept orderings in different domains taken from our 4-year Twitter archive.



These results are not only interesting from the sociological perspective but are also beneficial in practice since they would help us make decisions from among alternative concepts in daily life.

This research has involved various directions for productive future research. For example, one of the closest analyses by using our method is to know the ordering-based values of each user and this also can be used to user segmentation for marketing. Marketers generally have segmented users based on diverse standards such as demographics (*e.g.*, people in their 30's vs. 50's), and our method can be applied to cluster users based on the values as a new segmentation.

To exploit our method for specific domains or each person, we should tackle some difficulties, and the most difficult one is the data sparseness problem because of the size of data from more specific domains. To solve this, we plan to gather evidence aggressively, for example, we will count relevant adjectives (*e.g.*, *heavy* relates to *large*), and consider hypernym/hyponym for each concept.

We have released the evaluation datasets with human orderings and the experimental codes for the academic and industrial communities at <http://www.tkl.iis.u-tokyo.ac.jp/~nari/> to facilitate the reproducibility of our results and their use in various application contexts.

# Appendix A

## Domain-Specific Gold-Standard Orderings and System-Generated Orderings

In this appendix, we show some tables that list the full set of the domain-specific gold-standard orderings and system-generated orderings mentioned in Section [5.2.1.1](#). They use the same notations to refer to the domains (FEMALE, MALE) which are used in Table [5.6](#).

## Appendix A Domain-Specific Gold-Standard Orderings and System-Generated Orderings

TABLE A.1: The gold-standard ordering of all the female (F, FEMALE) / male workers (M, MALE).

Query	Ave. $\rho$	Gold-standard ordering
animal	0.973	F whale,elephant,bear,cow,horse,dog,cat,mouse
(large)	0.976	M whale,elephant,bear,cow,horse,dog,cat,mouse
flower	0.650	F sakura,rose,lily,camellia,sunflower,platycodon,lavender,daisy
(beautiful)	0.596	M sakura,rose,lily,sunflower,lavender,camellia,platycodon,daisy
jewel	0.575	F pearl,sapphire,emerald,ruby,amethyst,opal
(elegant)	0.477	M emerald,sapphire,ruby,pearl,amethyst,opal
alcohol	0.595	F champagne,chuhai,wine,highball,beer,whisky,shochu
(delicious)	0.243	M champagne,wine,beer,highball,chuhai,whisky,shochu
sports	0.501	F football,volleyball,tennis,baseball,table tennis,basketball,sumo,professional wrestling
(entertaining)	0.464	M baseball,football,tennis,volleyball,basketball,table tennis,professional wrestling,sumo
mammal	0.605	F dog,elephant,cat,horse,whale,bear,mouse,cow
(clever)	0.607	M dog,whale,elephant,cat,horse,bear,cow,mouse
conveyance	0.555	F Shinkansen,airplane,taxi,train,bicycle,bus,boat
(comfortable)	0.580	M Shinkansen,airplane,taxi,train,bus,bicycle,boat
food	0.351	F steak,pasta,pizza,curry,ramen,hamburger,fried rice
(yummy)	0.345	M steak,ramen,curry,pizza,pasta,fried rice,hamburger
instrument	0.626	F piano,flute,cello,guitar,harmonica,trumpet,drum
(soothing)	0.561	M piano,flute,guitar,cello,harmonica,trumpet,drum
bird	0.687	F dog,squirrel,penguin,rabbit,panda,lion,horse,lizard
(lovely)	0.714	M squirrel,rabbit,penguin,dog,panda,horse,lion,lizard
vegetable	0.267	F pumpkin,onion,spinach,napa cabbage,eggplant,cucumber,broccoli,sprout
(tasty)	0.307	M pumpkin,onion,eggplant,cucumber,spinach,napa cabbage,sprout,broccoli
fruit	0.542	F peach,melon,strawberry,grape,apple,persimmon,tangerine,cherry
(sweet)	0.555	M peach,melon,strawberry,grape,apple,cherry,tangerine,persimmon
appliance	0.815	F smartphone,PC,car navigation system,digital camera,printer,speaker
(useful)	0.795	M PC,smartphone,car navigation system,digital camera,printer,speaker
flesh	0.545	F chicken,pork,beef,lamb
(preferable)	0.607	M beef,chicken,pork,lamb
bird	0.679	F penguin,owl,sparrow,swan,quail,eagle,pheasant,chicken
(cute)	0.721	M penguin,sparrow,quail,owl,swan,pheasant,chicken,eagle
weather	0.780	F fine,snow,mist,rain,frost,gale,thunder,yellow sand
(likable)	0.724	M fine,snow,mist,frost,rain,gale,thunder,yellow sand
country	0.575	F UK,Spain,France,Thailand,America,India,Russia,China
(safe)	0.667	M UK,France,Spain,America,Thailand,India,Russia,China
conveyance	0.914	F airplane,Shinkansen,train,taxi,boat,bus,bicycle
(fast)	0.928	M airplane,Shinkansen,train,taxi,boat,bus,bicycle
café	0.509	F Starbucks,Doutor,Tully's,Saint Marc
(likable)	0.430	M Starbucks,Doutor,Tully's,Saint Marc
corner store	0.701	F 7-Eleven,Lawson,FamilyMart,Ministop
(useful)	0.699	M 7-Eleven,Lawson,FamilyMart,Ministop
automaker	0.681	F Toyota,Honda,Daihatsu,Mazda,Yamaha
(healthy)	0.633	M Toyota,Honda,Daihatsu,Mazda,Yamaha
fast-food chain	0.604	F MOS Burger,KFC,Mister Donut,McDonald's
(tasty)	0.560	M MOS Burger,KFC,Mister Donut,McDonald's
manufacturer	0.656	F Sony,Panasonic,Toshiba,Canon,Hitachi,Fujitsu,Seiko Epson
(famous)	0.683	M Sony,Panasonic,Toshiba,Canon,Hitachi,Fujitsu,Seiko Epson
temple	0.851	F Kinkaku-ji,Ginkaku-ji,Horyu-ji,Yakushi-ji,Zenko-ji,Chuson-ji,To-ji,Zojo-ji
(famous)	0.815	M Kinkaku-ji,Horyu-ji,Ginkaku-ji,Zenko-ji,Yakushi-ji,Chuson-ji,To-ji,Zojo-ji
fruit	0.885	F melon,apple,peach,persimmon,tangerine,grape,strawberry,cherry
(large)	0.873	M melon,apple,peach,persimmon,grape,tangerine,strawberry,cherry
country	0.876	F Thailand,India,Spain,America,China,France,UK,Russia
(warm)	0.878	M India,Thailand,Spain,China,America,France,UK,Russia
coffee shop	0.790	F 7-Eleven,FamilyMart,Lawson,Ministop
(numerous)	0.872	M 7-Eleven,Lawson,FamilyMart,Ministop
actress	0.547	F Ki Kitano,Yuka,Kazue Fukishi,Akina Minami,Yumiko Shaku,Rinka,Tomochika
(cute)	0.654	M Ki Kitano,Yuka,Kazue Fukishi,Akina Minami,Yumiko Shaku,Rinka,Tomochika

## Appendix A Domain-Specific Gold-Standard Orderings and System-Generated Orderings

TABLE A.2: The baseline system-generated ordering of all the female (F, FEMALE) / male workers (M, MALE).

Query	Ave. $\rho$	System-generated ordering
animal	-0.095	F whale,elephant,bear,mouse,cat,dog,cow,horse
(large)	0.143	M elephant,horse,whale,dog,bear,cat,mouse,cow
flower	0.810	F sakura,rose,camellia,sunflower,daisy,lavender,platycodon,lily
(beautiful)	0.714	M sakura,rose,camellia,sunflower,daisy,lavender,lily,platycodon
jewel	0.371	F pearl,emerald,amethyst,sapphire,ruby,opal
(elegant)	-0.714	M pearl,opal,amethyst,sapphire,emerald,ruby
alcohol	-0.107	F beer,champagne,whisky,shochu,highball,wine,chuha
(delicious)	-0.179	M beer,highball,champagne,shochu,wine,whisky,chuha
sports	-0.143	F volleyball,professional wrestling,tennis,table tennis,basketball,sumo,football,baseball
(entertaining)	-0.310	M table tennis,tennis,volleyball,basketball,professional wrestling,sumo,football,baseball
mammal	0.548	F dog,mouse,elephant,cat,whale,horse,bear,cow
(clever)	0.190	M mouse,dog,cat,whale,elephant,horse,bear,cow
conveyance	0.464	F Shinkansen,airplane,bicycle,bus,taxi,boat,train
(comfortable)	0.179	M Shinkansen,bicycle,bus,boat,airplane,taxi,train
food	0.464	F steak,pizza,fried rice,curry,ramen,pasta,hamburger
(yummy)	0.286	M pizza,steak,curry,fried rice,pasta,ramen,hamburger
instrument	-0.071	F guitar,drum,cello,flute,piano,trumpet,harmonica
(soothing)	-0.179	M guitar,drum,trumpet,cello,flute,piano,harmonica
bird	0.048	F penguin,lion,lizard,dog,squirrel,rabbit,panda,horse
(lovely)	0.024	M lion,penguin,squirrel,dog,lizard,rabbit,horse,panda
vegetable	0.619	F napa cabbage,onion,spinach,eggplant,pumpkin,sprout,cucumber,broccoli
(tasty)	0.500	M onion,napa cabbage,pumpkin,spinach,eggplant,sprout,cucumber,broccoli
fruit	-0.667	F tangerine,cherry,peach,apple,persimmon,strawberry,melon,grape
(sweet)	-0.143	M apple,tangerine,peach,cherry,melon,grape,strawberry,persimmon
appliance	-0.543	F smartphone,car navigation system,digital camera,speaker,printer,PC
(useful)	0.143	M digital camera,smartphone,car navigation system,speaker,PC,printer
flesh	0.000	F pork,lamb,chicken,beef
(preferable)	0.200	M pork,chicken,lamb,beef
bird	0.810	F penguin,owl,sparrow,quail,swan,eagle,chicken,pheasant
(cute)	0.690	M penguin,owl,sparrow,quail,chicken,pheasant,eagle,swan
weather	0.810	F mist,frost,fine,snow,rain,yellow sand,gale,thunder
(likable)	0.690	M mist,fine,frost,snow,thunder,rain,gale,yellow sand
country	0.690	F Thailand,Spain,France,UK,India,America,Russia,China
(safe)	-0.333	M Thailand,India,Russia,America,China,Spain,France,UK
conveyance	0.464	F Shinkansen,bicycle,boat,airplane,taxi,train,bus
(fast)	0.393	M Shinkansen,bicycle,boat,airplane,taxi,bus,train
café	-0.400	F Tully's,Saint Marc,Doutor,Starbucks
(likable)	0.200	M Tully's,Doutor,Starbucks,Saint Marc
corner store	0.800	F 7-Eleven,Ministop,Lawson,FamilyMart
(useful)	0.800	M 7-Eleven,Ministop,Lawson,FamilyMart
automaker	-0.800	F Honda,Toyota,Mazda,Daihatsu,Yamaha
(healthy)	-0.100	M Mazda,Toyota,Honda,Yamaha,Daihatsu
fast-food chain	0.400	F Mister Donut,MOS Burger,KFC,McDonald's
(tasty)	0.400	M Mister Donut,MOS Burger,KFC,McDonald's
manufacturer	0.071	F Sony,Panasonic,Fujitsu,Toshiba,Canon,Seiko Epson,Hitachi
(famous)	-0.036	M Sony,Toshiba,Fujitsu,Panasonic,Canon,Seiko Epson,Hitachi
temple	0.405	F Zenko-ji,To-ji,Yakushi-ji,Horyu-ji,Kinkaku-ji,Zojo-ji,Ginkaku-ji,Chuson-ji
(famous)	0.119	M Horyu-ji,Kinkaku-ji,Ginkaku-ji,Zojo-ji,Yakushi-ji,Zenko-ji,Chuson-ji,To-ji
fruit	-0.333	F melon,persimmon,apple,peach,cherry,tangerine,grape,strawberry
(large)	-0.238	M persimmon,melon,strawberry,peach,apple,cherry,grape,tangerine
country	-0.833	F Russia,UK,Spain,America,India,Thailand,China,France
(warm)	-0.833	M Russia,America,Spain,France,UK,Thailand,India,China
coffee shop	0.000	F FamilyMart-Eleven,Lawson,Ministop
(numerous)	-0.600	M Ministop,FamilyMart-Eleven,Lawson
actress	0.214	F Yuka,Rinka,Ki Kitano,Yumiko Shaku,Tomochika,Kazue Fukishi,Akina Minami
(cute)	0.536	M Yuka,Ki Kitano,Rinka,Yumiko Shaku,Tomochika,Kazue Fukishi,Akina Minami

## Appendix A Domain-Specific Gold-Standard Orderings and System-Generated Orderings

TABLE A.3: The domain-unaware system-generated ordering of all the female (F, FEMALE) / male workers (M, MALE).

Query	Ave. $\rho$	System-generated ordering
animal	0.214	F cat, whale, elephant, bear, mouse, dog, cow, horse
(large)	0.619	M elephant, whale, horse, dog, bear, cat, mouse, cow
flower	0.571	F sakura, sunflower, camellia, rose, daisy, lily, lavender, platycodon
(beautiful)	0.571	M sakura, rose, camellia, sunflower, daisy, lavender, lily, platycodon
jewel	0.886	F pearl, emerald, sapphire, ruby, opal, amethyst
(elegant)	-0.657	M pearl, opal, amethyst, sapphire, emerald, ruby
alcohol	-0.107	F champagne, beer, whisky, shochu, highball, wine, chuhai
(delicious)	0.321	M champagne, beer, shochu, highball, whisky, wine, chuhai
sports	-0.143	F volleyball, professional wrestling, table tennis, tennis, basketball, sumo, football, baseball
(entertaining)	-0.548	M professional wrestling, table tennis, tennis, volleyball, basketball, sumo, football, baseball
mammal	0.619	F dog, mouse, elephant, cat, whale, horse, bear, cow
(clever)	0.238	M mouse, dog, cat, elephant, whale, horse, bear, cow
conveyance	0.571	F Shinkansen, airplane, bus, bicycle, taxi, boat, train
(comfortable)	0.179	M Shinkansen, bicycle, bus, airplane, boat, taxi, train
food	0.250	F steak, pizza, fried rice, curry, ramen, hamburger, pasta
(yummy)	0.714	M ramen, steak, fried rice, pizza, curry, pasta, hamburger
instrument	-0.036	F guitar, drum, cello, flute, piano, trumpet, harmonica
(soothing)	-0.357	M guitar, drum, trumpet, cello, flute, piano, harmonica
bird	0.310	F squirrel, panda, penguin, lion, lizard, dog, rabbit, horse
(lovely)	-0.048	M lion, panda, penguin, squirrel, dog, lizard, rabbit, horse
vegetable	0.619	F napa cabbage, onion, spinach, eggplant, pumpkin, sprout, cucumber, broccoli
(tasty)	0.500	M napa cabbage, pumpkin, onion, spinach, eggplant, sprout, cucumber, broccoli
fruit	0.167	F melon, apple, strawberry, tangerine, cherry, peach, persimmon, grape
(sweet)	0.405	M apple, melon, strawberry, tangerine, peach, cherry, grape, persimmon
appliance	0.371	F smartphone, car navigation system, digital camera, speaker, printer, PC
(useful)	0.143	M digital camera, smartphone, car navigation system, speaker, PC, printer
flesh	0.400	F pork, beef, chicken, lamb
(preferable)	0.800	M beef, pork, chicken, lamb
bird	0.881	F sparrow, penguin, owl, quail, swan, pheasant, eagle, chicken
(cute)	0.786	M penguin, owl, sparrow, quail, pheasant, chicken, eagle, swan
weather	0.619	F mist, frost, fine, rain, snow, yellow sand, gale, thunder
(likable)	0.714	M mist, fine, frost, rain, gale, snow, thunder, yellow sand
country	0.643	F Thailand, Spain, France, India, UK, America, Russia, China
(safe)	-0.667	M Thailand, India, Russia, America, China, Spain, France, UK
conveyance	-0.107	F Shinkansen, bicycle, bus, train, airplane, boat, taxi
(fast)	0.214	M Shinkansen, bicycle, train, airplane, bus, taxi, boat
café	-0.800	F Tully's, Saint Marc, Doutor, Starbucks
(likable)	0.200	M Tully's, Doutor, Starbucks, Saint Marc
corner store	0.400	F 7-Eleven, Ministop, Lawson, FamilyMart
(useful)	0.400	M 7-Eleven, Ministop, Lawson, FamilyMart
automaker	0.500	F Honda, Mazda, Toyota, Daihatsu, Yamaha
(healthy)	0.200	M Mazda, Toyota, Honda, Yamaha, Daihatsu
fast-food chain	0.400	F Mister Donut, MOS Burger, KFC, McDonald's
(tasty)	0.400	M Mister Donut, MOS Burger, KFC, McDonald's
manufacturer	0.679	F Sony, Panasonic, Fujitsu, Canon, Toshiba, Seiko Epson, Hitachi
(famous)	0.643	M Sony, Toshiba, Fujitsu, Panasonic, Canon, Seiko Epson, Hitachi
temple	-0.048	F Zenko-ji, To-ji, Yakushi-ji, Kinkaku-ji, Horyu-ji, Zojo-ji, Ginkaku-ji, Chuson-ji
(famous)	0.714	M Horyu-ji, Kinkaku-ji, Ginkaku-ji, Zojo-ji, Yakushi-ji, Zenko-ji, Chuson-ji, To-ji
fruit	0.643	F melon, apple, persimmon, cherry, tangerine, peach, strawberry, grape
(large)	0.214	M melon, persimmon, strawberry, cherry, peach, apple, tangerine, grape
country	-0.381	F Russia, UK, Spain, America, India, Thailand, China, France
(warm)	-0.595	M Russia, America, Spain, France, UK, Thailand, India, China
coffee shop	0.800	F FamilyMart-Eleven, Lawson, Ministop
(numerous)	0.400	M FamilyMart-Eleven, Lawson, Ministop
actress	0.214	F Yuka, Rinka, Ki Kitano, Yumiko Shaku, Tomochika, Kazue Fukishi, Akina Minami
(cute)	0.393	M Yuka, Ki Kitano, Rinka, Yumiko Shaku, Tomochika, Kazue Fukishi, Akina Minami

## Appendix A Domain-Specific Gold-Standard Orderings and System-Generated Orderings

TABLE A.4: The domain-aware system-generated ordering of all the female (F, FEMALE) / male workers (M, MALE).

Query	Ave. $\rho$	System-generated ordering
animal	0.214	F cat, whale, elephant, bear, mouse, dog, cow, horse
(large)	0.619	M elephant, whale, horse, dog, bear, cat, mouse, cow
flower	0.619	F sakura, sunflower, rose, camellia, daisy, lily, lavender, platycodon
(beautiful)	0.571	M sakura, rose, camellia, sunflower, daisy, lavender, lily, platycodon
jewel	0.714	F pearl, emerald, amethyst, sapphire, ruby, opal
(elegant)	-0.657	M pearl, opal, amethyst, sapphire, emerald, ruby
alcohol	-0.143	F champagne, beer, shochu, whisky, highball, wine, chuhai
(delicious)	0.571	M champagne, beer, highball, shochu, wine, whisky, chuhai
sports	-0.095	F volleyball, professional wrestling, tennis, table tennis, basketball, sumo, football, baseball
(entertaining)	-0.405	M professional wrestling, table tennis, tennis, volleyball, basketball, football, sumo, baseball
mammal	0.619	F dog, mouse, elephant, cat, whale, horse, bear, cow
(clever)	0.262	M mouse, dog, cat, whale, elephant, horse, bear, cow
conveyance	0.571	F Shinkansen, airplane, bus, bicycle, taxi, boat, train
(comfortable)	0.179	M Shinkansen, bus, bicycle, airplane, boat, train, taxi
food	0.250	F steak, pizza, fried rice, curry, ramen, hamburger, pasta
(yummy)	0.714	M ramen, steak, fried rice, pizza, curry, pasta, hamburger
instrument	-0.036	F guitar, drum, cello, flute, piano, trumpet, harmonica
(soothing)	-0.357	M guitar, drum, trumpet, cello, flute, piano, harmonica
bird	0.333	F squirrel, penguin, lion, panda, lizard, dog, rabbit, horse
(lovely)	0.000	M lion, penguin, panda, squirrel, dog, lizard, rabbit, horse
vegetable	0.571	F napa cabbage, onion, eggplant, spinach, pumpkin, sprout, cucumber, broccoli
(tasty)	0.619	M onion, napa cabbage, pumpkin, eggplant, spinach, sprout, cucumber, broccoli
fruit	0.167	F melon, apple, strawberry, tangerine, cherry, peach, persimmon, grape
(sweet)	0.405	M apple, melon, strawberry, tangerine, peach, cherry, grape, persimmon
appliance	0.371	F smartphone, car navigation system, digital camera, speaker, printer, PC
(useful)	0.143	M digital camera, smartphone, car navigation system, speaker, PC, printer
flesh	0.600	F pork, chicken, lamb, beef
(preferable)	0.200	M pork, chicken, beef, lamb
bird	0.833	F sparrow, penguin, owl, quail, swan, chicken, eagle, pheasant
(cute)	0.667	M penguin, owl, sparrow, chicken, quail, pheasant, eagle, swan
weather	0.667	F mist, frost, fine, snow, rain, yellow sand, gale, thunder
(likable)	0.881	M mist, fine, frost, snow, rain, gale, thunder, yellow sand
country	0.524	F Thailand, France, Spain, India, America, UK, Russia, China
(safe)	-0.643	M Thailand, India, Russia, America, China, France, Spain, UK
conveyance	0.071	F Shinkansen, bicycle, airplane, bus, boat, train, taxi
(fast)	0.286	M Shinkansen, bicycle, airplane, train, bus, taxi, boat
café	-0.800	F Tully's, Saint Marc, Doutor, Starbucks
(likable)	0.200	M Tully's, Doutor, Starbucks, Saint Marc
corner store	0.400	F 7-Eleven, Ministop, Lawson, FamilyMart
(useful)	0.400	M 7-Eleven, Ministop, Lawson, FamilyMart
automaker	0.500	F Honda, Mazda, Toyota, Daihatsu, Yamaha
(healthy)	0.200	M Mazda, Toyota, Honda, Yamaha, Daihatsu
fast-food chain	0.400	F Mister Donut, MOS Burger, KFC, McDonald's
(tasty)	0.400	M Mister Donut, MOS Burger, KFC, McDonald's
manufacturer	0.679	F Sony, Panasonic, Fujitsu, Canon, Toshiba, Seiko Epson, Hitachi
(famous)	0.643	M Sony, Toshiba, Fujitsu, Panasonic, Canon, Seiko Epson, Hitachi
temple	0.048	F Zenko-ji, To-ji, Yakushi-ji, Horyu-ji, Kinkaku-ji, Ginkaku-ji, Zojo-ji, Chuson-ji
(famous)	0.714	M Horyu-ji, Kinkaku-ji, Ginkaku-ji, Zojo-ji, Yakushi-ji, Zenko-ji, Chuson-ji, To-ji
fruit	0.762	F melon, persimmon, apple, peach, cherry, tangerine, strawberry, grape
(large)	0.333	M melon, persimmon, strawberry, peach, cherry, apple, tangerine, grape
country	-0.381	F Russia, UK, Spain, America, India, Thailand, China, France
(warm)	-0.595	M Russia, America, Spain, France, UK, Thailand, India, China
coffee shop	0.800	F FamilyMart-Eleven, Lawson, Ministop
(numerous)	0.400	M FamilyMart-Eleven, Lawson, Ministop
actress	0.357	F Yuka, Rinka, Ki Kitano, Yumiko Shaku, Kazue Fukishi, Tomochika, Akina Minami
(cute)	0.393	M Yuka, Ki Kitano, Rinka, Yumiko Shaku, Tomochika, Kazue Fukishi, Akina Minami

# Appendix B

## Annotation Manual for Crowdsourcing (in Japanese)

### 概要

- 様々な観点に基づくモノの順序付け（例: 「動物（モノ）」を「かわいい（観点）」順に並べる）をしていただきます
  - － 順序はワーカーさん自身が思う順序に並べてください
  - － 分からないモノは Google 画像検索で外見を確認して構いません。ただし普通のインターネット検索で調べてはいけません。
  - － 順序付けの基準が末尾にもあるので、よくお読みください
- お渡しするもの
  - － モノの集合（例: ゴールデンレトリバー, トイプードル, チワワ）
  - － 様々な観点（例: かわいい（< – > かわいくない））
  - － モノの集合と観点の組を 30 組お渡しします
- 作成してもらうもの

## Appendix B Annotation Manual for Crowdsourcing (in Japanese)

- － モノの順序
- － 30 組分入力してください

### モノの順序の付け方

#### 差の大きさによる順序の表し方

2つのモノの間の順序は以下のように差の程度によって，書き分けてください．

程度	書き方	例（かわいい順）
(わざわざ言及する必要がないほど) 大きな差がある	1つ空行を入れて縦に並べる	ひよこ  ゴリラ
差がある	空行を入れずに縦に並べる	ひよこ つばめ
差が感じられない	スペースで区切って同じ行にならべる	ひよこ ハムスター

#### 作業手順（例）：

##### 問題例：

(かわいい < - > かわいくない)  
トイプードル ゴールデンレトリバー チワワ

が与えられた時，チワワがトイプードルより（わざわざ言及する必要が無いほど）かわいく，トイプードルがゴールデンレトリバーよりかわいいと（ワーカーさんご自身が）思われたら，以下のように記述してください．観点を一番上と一番下に（）（半角括弧）で記述することを忘れないようにしてください．



回答例:

(かわいい)  
チワワ  
  
トイプードル  
ゴールデンレトリバー  
(かわいくない)

### 順序関係が明確で無い場合

モノの集合の中にどちらが上位かどうしても判断が付かないものがある場合は、それらを同順位として記述して構いません。

(便利だ < - > 不便だ)  
冷蔵庫 洗濯機 掃除機 電子レンジ

が与えられた場合で、例えば、冷蔵庫と洗濯機の便利さに差が感じられない場合は、以下のように同順位のことをスペースで区切って横に並べるように記述してください。

(便利だ)  
冷蔵庫 洗濯機  
電子レンジ  
掃除機  
(不便だ)

冷蔵庫と洗濯機はどちらが先でも構いません。どうしても差が感じられないモノが3つ以上ある場合も同様にして順序付けをお願いします。

## 知らないモノがある場合

知らないモノがある場合は、Google 画像検索でのみ調べることが可能です。画像検索で調べても判断が付かないものがある場合は、自分が思うように書いてから、判断に自信がないことを示すために、名前のあとに？を付けて、以下のように記述してください。

(高い) エベレスト  富士山 高尾山? 筑波山 (低い)
---

## 順序付けの基準

- 順序付けは、自分自身が思うままの順序に並び替えてください。答えていただきたいのは、一般の人（多くの人）がどう思うかではなく、あなたがどう思うかです。
- 並び替えるときは「エベレストは富士山よりも高い」というように「A は B よりも C だ」といえると自分と思えれば「エベレスト > 富士山」と記述してください。
- モノの集合の中に、知らないモノや判断が付かないモノがある場合は、Google 画像検索を用いて、各対象を調べていただいて構いません（Google 画像検索以外の方法では調べないでください）。特に、知らないモノについて、順位をつけにくい観点もあるかもしれませんが、画像だけを見て判断をしてください。

## Appendix B Annotation Manual for Crowdsourcing (in Japanese)

- 与えられた観点が，モノのどの属性について言及しているかわからない場合（山について「大きい」は高さなのか裾野の広さなのかなど）は，自分が最もらしいと思う属性を選んで順序付けを行ってください。

# Bibliography

- [1] 仁科俊晴, 鍛冶伸裕, 吉永直樹, and 豊田正史. 対義形容詞対との相互情報量を利用した概念語の順序付け. *IPSJ SIG NL*, 2013(8):1–7, nov 2013.
- [2] Charles Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, 15(1):72–101, January 1904.
- [3] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 133–142, 2002.
- [4] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alexander J. Smola, and Vladimir N. Vapnik. Support vector regression machines. In *Advances in Neural Information Processing Systems 9, NIPS 1996*, pages 155–161. MIT Press, 1997.
- [5] Peter Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 417–424, 2002.
- [6] A Sumida, N Yoshinaga, and K Torisawa. Boosting precision and recall of hyponymy relation acquisition from hierarchical layouts in wikipedia. In *LREC-08*, 2008.
- [7] John Prager. *Open-Domain Question Answering*. Now Publishers Inc., 2007.

- [8] Hsin-Hsi Chen, Shih-Chung Tsai, and Jin-He Tsai. Mining tables from large scale HTML texts. In *Proceedings of the 18th conference on Computational linguistics (COLING)*, pages 166–172, 2000.
- [9] Minoru Yoshida, Kentaro Torisawa, and Jun’ichi Tsujii. Extracting attributes and their values from Web pages. In *Web Document Analysis: Challenges and Opportunities*, Series in Machine Perception and Artificial Intelligence, pages 179–200. World Scientific Publishing, Inc., 2003.
- [10] Sören Auer and Jens Lehmann. What have Innsbruck and Leipzig in common? Extracting semantics from wiki content. In *Proceedings of the 4th European Conference on The Semantic Web (ESWC)*, pages 503–517, 2007.
- [11] Fei Wu and Daniel S. Weld. Autonomously semantifying Wikipedia. In *Proceedings of the 16th ACM Conference on Information and Knowledge Management (CIKM)*, pages 41–50, 2007.
- [12] Naoki Yoshinaga and Kentaro Torisawa. Attribute-value acquisition from semi-structured texts. In *Proceedings of the Workshop of OntoLex07 - From Text to Knowledge: The Lexicon/Ontology Interface held at the sixth International Semantic Web Conference*, pages 55–66, 2007.
- [13] Hiroya Takamura and Jun’ichi Tsujii. Estimating numerical attributes by bringing together fragmentary clues. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 1305–1310, 2015.
- [14] Bo Pang and Lillian Lee. *Opinion Mining and Sentiment Analysis*. Now Publishers Inc., 2008.
- [15] Nitin Jindal and Bing Liu. Mining comparative sentences and relations. In *Proceedings of the 21st national conference on Artificial intelligence (AAAI)*, pages 1331–1336, 2006.

- [16] Takeshi Kurashima, Katsuji Bessho, Hiroyuki Toda, Toshio Uchiyama, and Ryoji Kataoka. Ranking entities using comparative relations. In *Proceedings of the 19th Conference on Database and Expert Systems Applications (DEXA)*, pages 124–133, 2008.
- [17] Ralph A. Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, December 1952.
- [18] Maksims N. Volkovs and Richard S. Zemel. A flexible generative model for preference aggregation. In *Proceedings of the 21st International Conference on World Wide Web (WWW)*, pages 479–488, 2012.
- [19] Shuzi Niu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. Stochastic rank aggregation. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 478–487, 2013.
- [20] Xi Chen, Paul N. Bennett, Kevyn Collins-Thompson, and Eric Horvitz. Pairwise ranking aggregation in a crowdsourced setting. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM)*, pages 193–202, 2013.
- [21] Karthik Raman and Thorsten Joachims. Methods for ordinal peer grading. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1037–1046, 2014.
- [22] Luana Făgărășan, Eva Maria Vecchi, and Stephen Clark. From distributional semantics to feature norms: grounding semantic models in human perceptual data. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 52–57, 2015.
- [23] Ken McRae, George S. Cree, Mark S. Seidenberg, and Chris McNorgan. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4):547–559, 2005.

- [24] Naoki Yoshinaga and Masaru Kitsuregawa. Polynomial to linear: Efficient classification with conjunctive features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1542–1551, 2009.
- [25] Naoki Yoshinaga and Masaru Kitsuregawa. Kernel slicing: Scalable online training with conjunctive features. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, pages 1245–1253, 2010.
- [26] Naoki Yoshinaga and Masaru Kitsuregawa. A self-adaptive classifier for efficient text-stream processing. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 1091–1102, 2014.
- [27] Daisuke Ikeda, Tomoyuki Nanno, and Manabu Okumura. Gender classification of blog authors. In *Proceedings of the 12th Annual Meeting for Natural Language Processing*, pages 356–359, 2006.
- [28] Alex Schulz, Aristotelis Hadjakos, Heiko Paulheim, Johannes Nachtwey, and Max Mühlhäuser. A Multi-Indicator Approach for Geolocalization of Tweets. In *Proceedings of The 7th International AAAI Conference on Weblogs and Social Media*, pages 573–582, 2013.
- [29] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.
- [30] Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, and Jenifer C. Lai. Class-based  $n$ -gram models of natural language. *Computational Linguistics*, 18(4):467–479, December 1992.

# Publications

## International Conferences (w/ Peer Review)

### Oral

- T. Iwanari, N. Yoshinaga, N. Kaji, T. Nishina, M. Toyoda, and M. Kitsuregawa, Ordering concepts based on common attribute intensity. In the 25th International Joint Conference on Artificial Intelligence (IJCAI-16), New York, 2016.

### System Demonstration

- T. Iwanari, K. Ohara, N. Yoshinaga, N. Kaji, M. Toyoda, and M. Kitsuregawa, Kotonush: Understanding Concepts Based on Values behind Social Media. In the 26th International Conference on Computational Linguistics (COLING-16), Osaka, 2016.



## Domestic Conferences (w/o Peer Review)

### Oral

- 岩成達哉, 吉永直樹, 豊田正史, 喜連川優, ランキング学習を用いた形容詞に基づく概念語の順序付け. 言語処理学会第 22 回年次大会 (NLP2016), 宮城, 2016.

### Poster

- 岩成達哉, 吉永直樹, 豊田正史, 喜連川優, 多様な手がかりを用いた形容詞に基づく概念語の順序付け. 第 8 回データ工学と情報マネジメントに関するフォーラム (第 14 回日本データベース学会年次大会) (DEIM2016), 福岡, 2016.

### System Demonstration

- 岩成達哉, 大原康平, 吉永直樹, 鍛冶伸裕, 豊田正史, 喜連川優, COMAVa: Concept Ordering Method for Acquiring Values. NLP 若手の会第 11 回シンポジウム (YANS2016), 和歌山, 2016.

## Scheduled Domestic Conferences (w/o Peer Review)

### Oral

- T. Iwanari, N. Yoshinaga, M. Toyoda, and M. Kitsuregawa, Inducing Writers' Values on Concept Ordering from Microblog. 第 9 回データ工学と情報マネジメントに関するフォーラム (第 15 回日本データベース学会年次大会) (DEIM2017), 岐阜, 2017.