Automatic Classification of Semantic Relations
between Facts and Opinions

Koji Murakami†   Eric Nichols‡   Junta Mizuno†‡   Yotaro Watanabe‡
Hayato Goto†   Megumi Ohki†   Suguru Matsuyoshi†   Kentaro Inui‡   Yuji Matsumoto†
†Nara Institute of Science and Technology
‡Tohoku University
{kmurakami,matuyosi,hayato-g,megumi-o,matsu}@is.naist.jp
{eric-n,junta-m,inui}@ecei.tohoku.ac.jp

Abstract

Classifying and identifying semantic relations between facts and opinions on the Web is of utmost importance for organizing information on the Web, however, this requires consideration of a broader set of semantic relations than are typically handled in Recognizing Textual Entailment (RTE), Cross-document Structure Theory (CST), and similar tasks. In this paper, we describe the construction and evaluation of a system that identifies and classifies semantic relations in Internet data. Our system targets a set of semantic relations that have been inspired by CST but that have been generalized and broadened to facilitate application to mixed fact and opinion data from the Internet. Our system identifies these semantic relations in Japanese Web texts using a combination of lexical, syntactic, and semantic information and evaluate our system against gold standard data that was manually constructed for this task. We will release all gold standard data used in training and evaluation of our system this summer.

1   Introduction

The task of organizing the information on the Internet to help users find facts and opinions on their topics of interest is increasingly important as more people turn to the Web as a source of important information. The vast amounts of research conducted in NLP on automatic summarization, opinion mining, and question answering are illustrative of the great interest in making relevant information easier to find. Providing Internet users with thorough information re-

quires recognizing semantic relations between both facts and opinions, however the assumptions made by current approaches are often incompatible with this goal. For example, the existing semantic relations considered in Recognizing Textual Entailment (RTE) (Dagan et al., 2005) are often too narrow in scope to be directly applicable to text on the Internet, and theories like Cross-document Structure Theory (CST) (Radev, 2000) are only applicable to facts or second-hand reporting of opinions rather than relations between both.

As part of the STATEMENT MAP project we proposed the development of a system to support information credibility analysis on the Web (Murakami et al., 2009b) by automatically summarizing facts and opinions on topics of interest to users and showing them the evidence and conflicts for each viewpoint. To facilitate the detection of semantic relations in Internet data, we defined a sentence-like unit of information called the *statement* that encompasses both facts and opinions, started compiling a corpus of *statements* annotated with semantic relations (Murakami et al., 2009a), and begin constructing a system to automatically identify semantic relations between statements.

In this paper, we describe the construction and evaluation of a prototype semantic relation identification system. We build on the semantic relations proposed in RTE and CST and in our previous work, refining them into a streamlined set of semantic relations that apply across facts and opinions, but that are simple enough to make automatic recognition of semantic relations between statements in Internet text possible.Our semantic relations are [AGREEMENT], [CONFLICT], [CONFINEMENT], and [EVIDENCE]. [AGREEMENT] and [CONFLICT] are expansions of the [EQUIVALENCE] and [CONTRADICTION]

relations used in RTE. [CONFINEMENT] and [EVIDENCE] are new relations between facts and opinions that are essential for understanding how statements on a topic are inter-related.

Our task differs from opinion mining and sentiment analysis which largely focus on identifying the polarity of an opinion for defined parameters rather than identify how facts and opinions relate to each other, and it differs from web document summarization tasks which focus on extracting information from web page structure and contextual information from hyperlinks rather than analyzing the semantics of the language on the webpage itself.

We present a system that automatically identifies semantic relations between *statements* in Japanese Internet texts. Our system uses *structural alignment* to identify *statement* pairs that are likely to be related, then classifies semantic relations using a combination of lexical, syntactic, and semantic information. We evaluate cross-statement semantic relation classification on sentence pairs that were taken from Japanese Internet texts on several topics and manually annotated with a semantic relation where one is present. In our evaluation, we look closely at the impact that each of the resources has on semantic relation classification quality.

The rest of this paper is organized as follows. In Section 2, we discuss related work in summarization, semantic relation classification, opinion mining, and sentiment analysis, showing how existing classification schemes are insufficient for our task. In Section 3, we introduce a set of cross-sentential semantic relations for use in the opinion classification needed to support information credibility analysis on the Web. In Section 4, we present our cross-sentential semantic relation recognition system, and discuss the algorithms and resources that are employed. In Section 5, we evaluate our system in a semantic relation classification task. In Section 6, we discuss our findings and conduct error analysis. Finally, we conclude the paper in Section 7.

2 Related Work

2.1 Recognizing Textual Entailment

Identifying logical relations between texts is the focus of Recognizing Textual Entailment, the task of deciding whether the meaning of one text is entailed from another text. A major task in the RTE Challenge (Recognizing Textual Entailment Challenge) is classifying the semantic relation between a Text (T) and a Hypothesis (H) into [ENTAILMENT], [CONTRADICTION], or [UNKNOWN]. Over the last several years, several corpora annotated with thousands of (T,H) pairs have been constructed for this task. In these corpora, each pair was tagged indicating its related task (e.g. Information Extraction, Question Answering, Information Retrieval or Summarization).

The RTE Challenge has successfully employed a variety of techniques in order to recognize instances of textual entailment, including methods based on: measuring the degree of lexical overlap between bag of words (Glickman et al., 2005; Jijkoun and de Rijke, 2005), the alignment of graphs created from syntactic or semantic dependencies (Marsi and Krahmer, 2005; MacCartney et al., 2006), statistical classifiers which leverage a wide range of features (Hickl et al., 2005), or reference rule generation (Szpektor et al., 2007). These approaches have shown great promise in RTE for entailment pairs in the corpus, but more robust models of recognizing logical relations are still desirable.

The definition of contradiction in RTE is that T contradicts H if it is very unlikely that both T and H can be true at the same time. However, in real documents on the Web, there are many pairs of examples which are contradictory in part, or where one statement confines the applicability of another, as shown in the examples in Table 1.

2.2 Cross-document Structure Theory

Cross-document Structure Theory (CST), developed by Radev (2000), is another task of recognizing semantic relations between sentences. CST is an expanded rhetorical structure analysis based on Rhetorical Structure Theory (RST: (William and Thompson, 1988)), and attempts to describe the semantic relations that exist between two or more sentences from different source documents that are related to the same topic, as well as those that come from a single source document. A corpus of cross-document sentences annotated with CST relations has also been constructed (The CSTBank Corpus: (Radev et al., 2003)). CSTBank is organized into clusters of topically-related articles. There are 18 kinds of semantic relations in this corpus, not limited to [EQUIVALENCE] or [CONTRADICTION], but also including [JUDGEMENT], [ELABORATION], and [RE-

Query	Matching sentences	Output
キシリトールは虫歯予防に効果がある Xylitol is effective at preventing cavities.	キシリトールの含まれている量が多いほどむし歯予防の効果は高いようです The cavity-prevention effects are greater the more Xylitol is included. キシリトールがお口の健康維持や虫歯予防にも効果を発揮します Xylitol shows effectiveness at maintaining good oral hygiene and preventing cavities. キシリトールの虫歯抑制効果についてはいろいろな意見がありますが実際は効果があるわけではありません There are many opinions about the cavity-prevention effectiveness of Xylitol, but it is not really effective.	同意 [AGREEMENT]. 同意 [AGREEMENT] 対立 [CONFLICT]
還元水は健康に良い Reduced water is good for the health.	弱アルカリ性のアルカリイオン還元水があなたと家族の健康を支えます Reduced water, which has weak alkaline ions, supports the health of you and your family. 還元水は活性酸素を除去すると言われ健康を維持してくれる働きをもたらす Reduced water is said to remove active oxygen from the body, making it effective at promoting good health. 美味しくても酸化させる水は健康には役立ちません Even if oxidized water tastes good, it does not help one's health.	同意 [AGREEMENT] 同意 [AGREEMENT] 対立 [CONFLICT]
イソフラボンは健康維持に効果がある Isoflavone is effective at maintaining good health.	大豆イソフラボンをサプリメントで過剰摂取すると健康維持には負の影響を与える結果となります Taking too much soy isoflavone as a supplement will have a negative effect on one's health	限定 [CONFINEMENT]

Table 1: Example semantic relation classification.

FINEMENT]. Etoh *et al.* (Etoh and Okumura, 2005) constructed a Japanese Cross-document Relation Corpus, and they redefined 14 kinds of semantic relations in their corpus.

CST was designed for objective expressions because its target data is newspaper articles related to the same topic. Facts, which can be extracted from newspaper articles, have been used in conventional NLP research, such as Information Extraction or Factoid Question Answering. However, there are a lot of opinions on the Web, and it is important to survey opinions in addition to facts to give Internet users a comprehensive view of the discussions on topics of interest.

2.3 Cross-document Summarization Based on CST Relations between Sentences

Zhang and Radev (2004) attempted to classify CST relations between sentence pairs extracted from topically related documents. However, they used a vector space model and tried multi-class classification. The results were not satisfactory. This observation may indicate that the recognition methods for each relation should be developed separately. Miyabe *et al.* (2008) attempted to recognize relations that were defined in a Japanese cross-document relation corpus (Etoh and Okumura, 2005). However, their target relations were limited to [EQUIVALENCE] and [TRANSITION]; other relations were not targeted. Recognizing [EVIDENCE] is indispensable for organizing information on the Internet. We need to develop satisfactory methods of [EVIDENCE] recognition.

2.4 Opinion Mining and Sentiment Analysis

Subjective statements, such as opinions, have recently been the focus of much NLP research including review analysis, opinion extraction, opinion question answering, and sentiment analysis. In the corpus constructed in the Multi-Perspective Question Answering (MPQA) Project (Wiebe et al., 2005), individual expressions are tagged that correspond to explicit mentions of private states, speech event, and expressive subjective elements.

The goal of opinion mining to extract expressions with polarity from texts, not to recognize semantic relations between sentences. Sentiment analysis also focus classifying subjective expressions in texts into positive/negative classes. In comparison, although we deal with sentiment information in text, our objective is to recognize semantic relations between sentences. If a user's query requires positive/negative information, we will also need to extract sentences including sentiment expression like in opinion mining, however, our semantic relation, [CONFINEMENT], is more precise because it identifies the condition or scope of the polarity. Queries do not necessarily include sentiment information; we also accept queries that are intended to be a statement of fact. For example, for the query "Xylitol is effective at preventing cavities." in Table 1, we extract a variety of sentences from the Web and recognize semantic relations between the query and many kinds of sentences.

3 Semantic Relations between Statements

In this section, we define the semantic relations that we will classify in Japanese Internet texts as well as their corresponding relations in RTE and CST. Our goal is to define semantic relations that are applicable over both fact and opinions, making them more appropriate for handling Internet texts. See Table 1 for real examples.

3.1 [AGREEMENT]

A bi-directional relation where statements have equivalent semantic content on a shared topic. Here we use *topic* in a narrow sense to mean that the semantic contents of both statements are relevant to each other.

The following is an example of [AGREEMENT] on the topic of *bio-ethanol environmental impact*.

(1) a. Bio-ethanol is good for the environment.
   b. Bio-ethanol is a high-quality fuel, and it has the power to deal with the environment problems that we are facing.

Once relevance has been established, [AGREEMENT] can range from strict logical entailment or identical polarity of opinions.

Here is an example of two statements that share a broad topic, but that are not classified as [AGREEMENT] because *preventing cavities* and *tooth calcification* are not intuitively relevant.

(2) a. Xylitol is effective at preventing cavities.
   b. Xylitol advances tooth calcification.

3.2 [CONFLICT]

A bi-directional relation where statements have negative or contradicting semantic content on a shared topic. This can range from strict logical contradiction to opposite polarity of opinions. The next pair is a [CONFLICT] example.

(3) a. Bio-ethanol is good for our earth.
   b. There is a fact that bio-ethanol further the destruction of the environment.

3.3 [EVIDENCE]

A uni-directional relation where one statement provides justification or supporting evidence for the other. Both statements can be either facts or opinions. The following is a typical example:

(4) a. I believe that applying the technology of cloning must be controlled by law.

   b. There is a need to regulate cloning, because it can be open to abuse.

The *statement* containing the evidence consists of two parts: one part has a [AGREEMENT] or [CONFLICT] with the other *statement*, the other part provides support or justification for it.

3.4 [CONFINEMENT]

A uni-directional relation where one statement provides more specific information about the other or quantifies the situations in which it applies. The pair below is an example, in which one *statement* gives a condition under which the other can be true.

(5) a. Steroids have side-effects.
   b. There is almost no need to worry about side-effects when steroids are used for local treatment.

4 Recognizing Semantic Relations

In order to organize the information on the Internet, we need to identify the [AGREEMENT], [CONFLICT], [CONFINEMENT], and [EVIDENCE] semantic relations. Because identification of [AGREEMENT] and [CONFLICT] is a problem of measuring semantic similarity between two *statements*, it can be cast as a sentence alignment problem and solved using an RTE framework. The two sentences do not need to be from the same source.

However, the identification of [CONFINEMENT] and [EVIDENCE] relations depend on contextual information in the sentence. For example, conditional statements or specific discourse markers like "because" act as important cues for their identification. Thus, to identify these two relations across documents, we must first identify [AGREEMENT] or [CONFLICT] between sentences in different documents and then determine if there is a [CONFINEMENT] or [EVIDENCE] relation in one of the documents.

Furthermore, the surrounding text often contains contextual information that is important for identifying these two relations. Proper handling of surrounding context requires discourse analysis and is an area of future work, but our basic detection strategy is as follows:

1. Identify a [AGREEMENT] or [CONFLICT] relation between the Query and Text
2. Search the Text sentence for cues that identify [CONFINEMENT] or [EVIDENCE]

3. Infer the applicability of the [CONFINE-MENT] or [EVIDENCE] relations in the Text to the Query

4.1 System Overview

We have finished constructing a prototype system that detects semantic relation between *statements*. It has a three-stage architecture similar to the RTE system of MacCartney *et al.* (2006):

1. Linguistic analysis
2. Structural alignment
3. Feature extraction for detecting [EVIDENCE] and [CONFINEMENT]
4. Semantic relation classification

However, we differ in the following respects.

First, our relation classification is broader than RTE's simple distinction between [ENTAILMENT], [CONTRADICTION], and [UNKNOWN]; in place of [ENTAILMENT] and [CONTRADICTION, we use broader [AGREEMENT] and [CONFLICT] relations. We also consider cover gradations of applicability of statements with the [CONFINEMENT] relation.

Second, we conduct structural alignment with the goal of aligning semantic structures. We do this by directly incorporating dependency alignments and predicate-argument structure information for both the user query and the Web text into the alignment scoring process. This allows us to effectively capture many long-distance alignments that cannot be represented as lexical alignments. This contrasts with MacCartney *et al.* (2006), who uses dependency structures for the Hypothesis to reduce the lexical alignment search space but do not produce structural alignments and do not use the dependencies in detecting entailment.

Finally, we apply several rich semantic resources in alignment and classification: extended modality information that helps align and classify structures that are semantically similar but divergent in tense or polarity; and lexical similarity through ontologies like WordNet.

4.2 Linguistic Analysis

In order to identify semantic relations between the user *query* (Q) and the sentence extracted from Web *text* (T), we first conduct syntactic and semantic linguistic analysis to provide a basis for alignment and relation classification.

For syntactic analysis, we use the Japanese dependency parser CaboCha (Kudo and Mat-
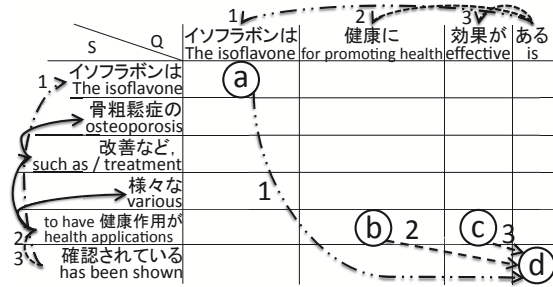

Figure 1: An example of structural alignment

sumoto, 2002) and the predicate-argument structure analyzer ChaPAS (Watanabe et al., 2010). CaboCha splits the Japanese text into phrase-like chunks and represents syntactic dependencies between the chunks as edges in a graph. ChaPAS identifies predicate-argument structures in the dependency graph produced by CaboCha.

We also conduct extended modality analysis using the resources provided by Matsuyoshi *et al.* (2010), focusing on tense, modality, and polarity, because such information provides important clues for the recognition of semantic relations between *statements*.

4.3 Structural Alignment

In this section, we describe our approach to structural alignment. The structural alignment process is shown in Figure 1. It consists of the following two phases:

1. lexical alignment
2. structural alignment

We developed a heuristic-based algorithm to align chunk based on lexical similarity information. We incorporate the following information into an alignment confidence score that has a range of 0.0-1.0 and align chunk whose scores cross an empirically-determined threshold.

- surface level similarity: identical content words or cosine similarity of chunk contents
- semantic similarity of predicate-argument structures

    predicates we check for matches in predicate entailment databases (Hashimoto et al., 2009; Matsuyoshi et al., 2008) considering the default case frames reported by ChaPAS

    arguments we check for synonym or hypernym matches in the Japanese WordNet (2008) or the Japanese hypernym collection of Sumida *et al.* (2008)

T: 畑 で　　　　　　　農薬 を　　使用する
(field) (in)　(agricultural chemicals) (ACC)　(use)

H: 畑 に　　　　　　　農薬 を　　散布する
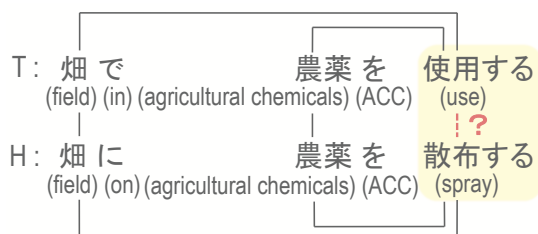(field) (on)　(agricultural chemicals) (ACC)　(spray)

Figure 2: Determining the compatibility of semantic structures

We compare the predicate-argument structure of the query to that of the text and determine if the argument structures are compatible. This process is illustrated in Figure 2 where the T(ext) "Agricultural chemicals are used in the field." is aligned with the H(ypothesis) "Over the field, agricultural chemicals are sprayed." Although the verbs *used* and *sprayed* are not directly semantically related, they are aligned because they share the same argument structures. This lets up align predicates for which we lack semantic resources. We use the following information to determine predicate-argument alignment:

- the number of aligned children
- the number of aligned case frame arguments
- the number of possible alignments in a window of n chunk
- predicates indicating existence or quantity. E.g. *many*, *few*, *to exist*, etc.
- polarity of both parent and child chunks using the resources in (Higashiyama et al., 2008; Kobayashi et al., 2005)

We treat structural alignment as a machine learning problem and train a Support Vector Machine (SVM) model to decide if lexically aligned chunks are semantically aligned.

We train on gold-standard labeled alignment of 370 sentence pairs. This data set is described in more detail in Section 5.1. As features for our SVM model, we use the following information:

- the distance in edges in the dependency graph between parent and child for both sentences
- the distance in chunks between parent and child in both sentences
- binary features indicating whether each chunk is a predicate or argument according to ChaPAS
- the parts-of-speech of first and last word in each chunk

- when the chunk ends with a case marker, the case of the chunk , otherwise *none*
- the lexical alignment score of each chunk pair

4.4 Feature Extraction for Detecting Evidence and Confinement

Once the structural alignment system has identified potential [AGREEMENT] or [CONFLICT] relations, we need to extract contextual cues in the Text as features for detecting [CONFINEMENT] and [EVIDENCE] relations. Conditional statements, degree adverbs, and partial negation, which play a role in limiting the scope or degree of a *query*'s contents in the *statement*, can be important cues for detecting the these two semantic relations. We currently use a set of heuristics to extract a set of expressions to use as features for classifying these relations using SVM models.

4.5 Relation Classification

Once the structural alignment is successfully identified, the task of semantic relation classification is straightforward. We also solve this problem with machine learning by training an SVM classifier. As features, we draw on a combination of lexical, syntactic, and semantic information including the syntactic alignments from the previous section. The feature set is:

alignments We define two binary function, $ALIGN_{word}(q_i, t_m)$ for the lexical alignment and $ALIGN_{struct}((q_i, q_j), (t_m, t_k))$ for the structural alignment to be true if and only if the node $q_i, q_j \in Q$ has been semantically and structurally aligned to the node $t_m, t_k \in T$. Q and T are the (Q)uery and the (T)ext, respectively. We also use a separate feature for a score representing the likelihood of the alignment.

modality We have a feature that encodes all of the possible polarities of a predicate node from modality analysis, which indicates the utterance type, and can be *assertive*, *volitional*, *wish*, *imperative*, *permissive*, or *interrogative*. Modalities that do not represent opinions (i.e. *imperative*, *permissive* and *interrogative*) often indicate [OTHER] relations.

antonym We define a binary function $ANTONYM(q_i, t_m)$ that indicates if the pair is identified as an antonym. This information helps identify [CONFLICT].

Relation	Measure	3-class		Cascaded 3-class		Δ
[AGREEMENT]	precision	0.79	(128 / 162)	0.80	(126 / 157)	+0.01
[AGREEMENT]	recall	0.86	(128 / 149)	0.85	(126 / 149)	-0.01
[AGREEMENT]	f-score	0.82		0.82		-
[CONFLICT]	precision	0	(0 / 5)	0.36	(5 / 14)	+0.36
[CONFLICT]	recall	0	(0 / 12)	0.42	(5 / 12)	+0.42
[CONFLICT]	f-score	0		0.38		+0.38
[CONFINEMENT]	precision	0.4	(4 / 10)	0.8	(4 / 5)	+0.4
[CONFINEMENT]	recall	0.17	(4 / 23)	0.17	(4 / 23)	-
[CONFINEMENT]	f-score	0.24		0.29		+0.05

Table 2: Semantic relation classification results comparing 3-class and cascaded 3-class approaches

negation To identify negations, we primarily rely on a predicate's *Actuality* value, which represents epistemic modality and existential negation. If a predicate pair $ALIGN_{word}(q_i, t_m)$ has mismatching actuality labels, the pair is likely a [OTHER].

contextual cues This set of features is used to mark the presence of any contextual cues that identify of [CONFINEMENT] or [EVIDENCE] relations in a chunk . For example, "ので (because)" or "ため (due to)" are typical contextual cues for [EVIDENCE], and "とき (when)" or "ならば (if)" are typical for [CONFINEMENT].

5 Evaluation

5.1 Data Preparation

In order to evaluate our semantic relation classification system on realistic Web data, we constructed a corpus of sentence pairs gathered from a vast collection of webpages (2009a). Our basic approach is as follows:

1. Retrieve documents related to a set number of topics using the Tsubaki[1] search engine
2. Extract real sentences that include major subtopic words which are detected based on TF/IDF in the document set
3. Reduce noise in data by using heuristics to eliminate advertisements and comment spam
4. Reduce the search space for identifying sentence pairs and prepare pairs, which look feasible to annotate
5. Annotate corresponding sentences with [AGREEMENT], [CONFLICT], [CONFINEMENT], or [OTHER]

Although our target semantic relations include [EVIDENCE], they difficult annotate consistently, so we do not annotate them at this time. Expanding our corpus and semantic relation classifier to handle [EVIDENCE] remains and area of future work.

The data that composes our corpus comes from a diverse number of sources. A hand survey of a random sample of the types of domains of 100 document URLs is given below. Half of the URL domains were not readily identifiable, but the known URL domains included governmental, corporate, and personal webpages. We believe this distribution is representative of information sources on the Internet.

type	count
academic	2
blogs	23
corporate	10
governmental	4
news	5
press releases	4
q&a site	1
reference	1
other	50

We have made a partial release of our corpus of sentence pairs manually annotated with the correct semantic relations[2]. We will fully release all the data annotated semantic relations and with gold standard alignments at a future date.

5.2 Experiment Settings

In this section, we present results of empirical evaluation of our proposed semantic relation classification system on the dataset we constructed in the previous section. For this experiment, we use SVMs as described in Section 4.5

---

[1] http://tsubaki.ixnlp.nii.ac.jp/

[2] http://stmap.naist.jp/corpus/ja/index.html (in Japanese)

to classify semantic relations into one of the four classes: [AGREEMENT], [CONFLICT], [CONFINEMENT], or [OTHER] in the case of no relation. As data we use 370 sentence pairs that have been manually annotated both with the correct semantic relation and with gold standard alignments. Annotations are checked by two native speakers of Japanese, and any sentence pair where annotation agreement is not reached is discarded. Because we have limited data that is annotated with correct alignments and semantic relations, we perform five-fold cross validation, training both the structural aligner and semantic relation classifier on 296 sentence pairs and evaluating on the held out 74 sentence pairs. The figures presented in the next section are for the combined results on all 370 sentence pairs.

5.3 Results

We compare two different approaches to classification using SVMs:

3-class semantic relations are directly classified into one of [AGREEMENT], [CONFLICT], and [CONFINEMENT] with all features described in 4.5

cascaded 3-class semantic relations are first classified into one of [AGREEMENT], [CONFLICT] without contextual cue features. Then an additional judgement with all features determines if [AGREEMENT] and [CONFLICT] should be reclassified as [CONFINEMENT]

Initial results using the **3-class** classification model produced high f-scores for [AGREEMENT] but unfavorable results for [CONFLICT] and [CONFINEMENT]. We significantly improved classification of [CONFLICT] and [CONFINEMENT] by adopting the **cascaded 3-class** model. We present these results in Table 2 and successfully recognized examples in Table 1.

6 Discussion and Error Analysis

We constructed a prototype semantic relation classification system by combining the components described in the previous section. While the system developed is not domain-specific and capable of accepting queries on any topic, we evaluate its semantic relation classification on several queries that are representative of our training data.

Figure 3 shows a snapshot of the semantic relation classification system and the various semantic relations it recognized for the query.

	Baseline	Structural Alignment	Upper-bound
Precision	0.44 (56/126)	0.52 (96/186)	0.74 (135/183)
Recall	0.30 (56/184)	0.52 (96/184)	0.73 (135/184)
F1-score	0.36	0.52	0.74

Table 3: Comparison of lexical, structural, and upper-bound alignments on semantic relation classification

In the example (6), recognized as [CONFINEMENT] in Figure 3, our system correctly identified negation and analyzed the description "Xylitol alone can not completely" as playing a role of requirement.

(6) a. キシリトールは虫歯予防に効果がある
(Xylitol is effective at preventing cavities.)
b. キシリトールだけでは完全な予防は出来ません
(Xylitol alone can not completely prevent cavities.)

Our system correctly identifies [AGREEMENT] relations in other examples about reduced water from Table 1 by structurally aligning phrases like "promoting good health" and "supports the health" to "good for the health."

These examples show how resources like (Matsuyoshi et al., 2010) and WordNet (Bond et al., 2008) have contributed to the relation classification improvement of structural alignment over them baseline in Table 3. Focusing on similarity of syntactic and semantic structures gives our alignment method greater flexibility.

However, there are still various examples which the system cannot recognized correctly. In examples on cavity prevention, the phrase "effective at preventing cavities" could not be aligned with "can prevent cavities" or "good for cavity prevention," nor can "cavity prevention" and "cavity-causing bacteria control."

The above examples illustrate the importance of the role played by the alignment phase in the whole system's performance.

Table 3 compares the semantic relation classification performance of using lexical alignment only (as the baseline), lexical alignment and structural alignment, and, to calculate the maximum possible precision, classification using correct alignment data (the upper-bound). We can
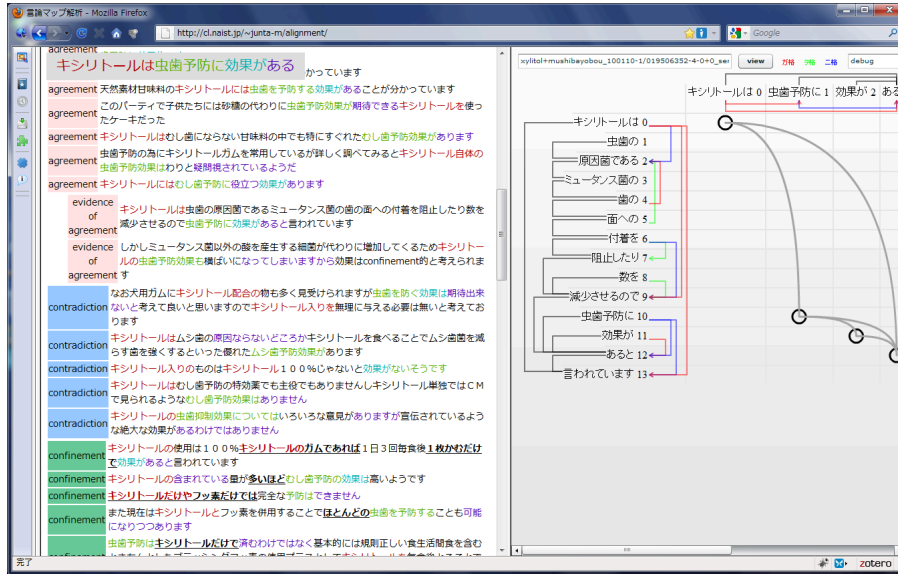
Figure 3: Alignment and classification example for the query *"Xylitol is effective at preventing cavities."*

see that structural alignment makes it possible to align more words than lexical alignment alone, leading to an improvement in semantic relation classification. However, there is still a large gap between the performance of structural alignment and the maximum possible precision. Error analysis shows that a big cause of incorrect classification is incorrect lexical alignment. Improving lexical alignment is a serious problem that must be addressed. This entails expanding our current lexical resources and finding more effective methods of apply them in alignment.

The most serious problem we currently face is the feature engineering necessary to find the optimal way of applying structural alignments or other semantic information to semantic relation classification. We need to conduct a quantitative evaluation of our current classification models and find ways to improve them.

7 Conclusion

Classifying and identifying semantic relations between facts and opinions on the Web is of utmost importance to organizing information on the Web, however, this requires consideration of a broader set of semantic relations than are typically handled in RTE, CST, and similar tasks. In this paper, we introduced a set of cross-sentential semantic relations specifically designed for this task that apply over both facts and opinions. We

presented a system that identifies these semantic relations in Japanese Web texts using a combination of lexical, syntactic, and semantic information and evaluated our system against data that was manually constructed for this task. Preliminary evaluation showed that we are able to detect [AGREEMENT] with high levels of confidence. Our method also shows promise in [CONFLICT] and [CONFINEMENT] detection. We also discussed some of the technical issues that need to be solved in order to identify [CONFLICT] and [CONFINEMENT].

Acknowledgments

References

Bond, Francis, Hitoshi Isahara, Kyoko Kanzaki, and Kiyotaka Uchimoto. 2008. Boot-strapping a wordnet using multiple existing wordnets. In *Proc. of the 6th International Language Resources and Evaluation (LREC'08)*.

Dagan, Ido, Oren Glickman, and Bernardo Magnini. 2005. The pascal recognising textual entailment challenge. In *Proc. of the PASCAL Challenges Workshop on Recognising Textual Entailment*.

Etoh, Junji and Manabu Okumura. 2005. Cross-document relationship between sentences corpus.

In *Proc. of the 14th Annual Meeting of the Association for Natural Language Processing*, pages 482–485. (in Japanese).

Glickman, Oren, Ido Dagan, and Moshe Koppel. 2005. Web based textual entailment. In *Proc. of the First PASCAL Recognizing Textual Entailment Workshop*.

Hashimoto, Chikara, Kentaro Torisawa, Kow Kuroda, Masaki Murata, and Jun'ichi Kazama. 2009. Large-scale verb entailment acquisition from the web. In *Conference on Empirical Methods in Natural Language Processing (EMNLP2009)*, pages 1172–1181.

Hickl, Andrew, John Williams, Jeremy Bensley, Kirk Roberts, Bryan Rink, and Ying Shi. 2005. Recognizing textual entailment with lcc's groundhog system. In *Proc. of the Second PASCAL Challenges Workshop*.

Higashiyama, Masahiko, Kentaro Inui, and Yuji Matsumoto. 2008. Acquiring noun polarity knowledge using selectional preferences. In *Proc. of the 14th Annual Meeting of the Association for Natural Language Processing*.

Jijkoun, Valentin and Maarten de Rijke. 2005. Recognizing textual entailment using lexical similarity. In *Proc. of the First PASCAL Challenges Workshop*.

Kobayashi, Nozomi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi, and Toshikazu Fukushima. 2005. Collecting evaluative expressions for opinion extraction. *Journal of natural language processing*, 12(3):203–222.

Kudo, Taku and Yuji Matsumoto. 2002. Japanese dependency analysis using cascaded chunking. In *Proc of CoNLL 2002*, pages 63–69.

MacCartney, Bill, Trond Grenager, Marie-Catherine de Marneffe, Daniel Cer, and Christopher D. Manning. 2006. Learning to recognize features of valid textual entailments. In *Proc. of HLT/NAACL 2006*.

Marsi, Erwin and Emiel Krahmer. 2005. Classification of semantic relations by humans and machines. In *Proc. of ACL-05 Workshop on Empirical Modeling of Semantic Equivalence and Entailment*, pages 1–6.

Matsuyoshi, Suguru, Koji Murakami, Yuji Matsumoto, and Kentaro Inui. 2008. A database of relations between predicate argument structures for recognizing textual entailment and contradiction. In *Proc. of the Second International Symposium on Universal Communication*, pages 366–373, December.

Matsuyoshi, Suguru, Megumi Eguchi, Chitose Sao, Koji Murakami, Kentaro Inui, and Yuji Matsumoto. 2010. Annotating event mentions in text with modality, focus, and source information. In *Proc. of the 7th International Language Resources and Evaluation (LREC'10)*, pages 1456–1463.

Miyabe, Yasunari, Hiroya Takamura, and Manabu Okumura. 2008. Identifying cross-document relations between sentences. In *Proc. of the 3rd International Joint Conference on Natural Language Processing (IJCNLP-08)*, pages 141–148.

Murakami, Koji, Shouko Masuda, Suguru Matsuyoshi, Eric Nichols, Kentaro Inui, and Yuji Matsumoto. 2009a. Annotating semantic relations combining facts and opinions. In *Proceedings of the Third Linguistic Annotation Workshop*, pages 150–153, Suntec, Singapore, August. Association for Computational Linguistics.

Murakami, Koji, Eric Nichols, Suguru Matsuyoshi, Asuka Sumida, Shouko Masuda, Kentaro Inui, and Yuji Matsumoto. 2009b. Statement map: Assisting information credibility analysis by visualizing arguments. In *Proc. of the 3rd ACM Workshop on Information Credibility on the Web (WICOW 2009)*, pages 43–50.

Radev, Dragomir, Jahna Otterbacher, and Zhu Zhang. 2003. CSTBank: Cross-document Structure Theory Bank. http://tangra.si.umich.edu/clair/CSTBank.

Radev, Dragomir R. 2000. Common theory of information fusion from multiple text sources step one: Cross-document structure. In *Proc. of the 1st SIGdial workshop on Discourse and dialogue*, pages 74–83.

Sumida, Asuka, Naoki Yoshinaga, and Kentaro Torisawa. 2008. Boosting precision and recall of hyponymy relation acquisition from hierarchical layouts in wikipedia. In *Proc. of the 6th International Language Resources and Evaluation (LREC'08)*.

Szpektor, Idan, Eyal Shnarch, and Ido Dagan. 2007. Instance-based evaluation of entailment rule acquisition. In *Proc. of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 456–463.

Watanabe, Yotaro, Masayuki Asahara, and Yuji Matsumoto. 2010. A structured model for joint learning of argument roles and predicate senses. In *Proceedings of the 48th Annual Meeting of the Association of Computational Linguistics (to appear)*.

Wiebe, Janyce, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210.

William, Mann and Sandra Thompson. 1988. Rhetorical structure theory: towards a functional theory of text organization. *Text*, 8(3):243–281.

Zhang, Zhu and Dragomir Radev. 2004. Combining labeled and unlabeled data for learning cross-document structural relationships. In *Proc. of the Proceedings of IJC-NLP*.