Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents

Kentaro Inui, Shuya Abe, Kazuo Hara, Hiraku Morita, Chitose Sao,
Megumi Eguchi, Asuka Sumida, Koji Murakami, and Suguru Matsuyoshi
Nara Institute of Science and Technology
8916-5 Takayama Ikoma Nara 630-0192 Japan
{inui,shuya-a,kazuo-h,hiraku-m,chotose,megumi-e,asuka-s,kmurakami,matuyosi}@is.naist.jp

Abstract

This paper proposes a new UGC-oriented language technology application, which we call experience mining. Experience mining aims at automatically collecting instances of personal experiences as well as opinions from an explosive number of user generated contents (UGCs) such as weblog and forum posts and storing them in an experience database with semantically rich indices. After arguing the technical issues of this new task, we focus on the central problem, factuality analysis, among others and propose a machine learning-based solution as well as the task definition itself. Our empirical evaluation indicates that our factuality analysis task is sufficiently well-defined to achieve a high inter-annotator agreement and our Factorial CRF-based model considerably outperforms the baseline. We also present an application system, which currently stores over 50M experience instances extracted from 150M Japanese blog posts with semantic indices and is scheduled to start serving as an experience search engine for unrestricted users in October.

1 Introduction

The explosive spread of communication media on the Web, such as message forums and weblogs, allows Web users access to a rapidly increasing and massive amount personal experiences and opinions — a potential treasury of wisdom useful for making decisions, resolving troubles and avoiding problems, if only it were all indexed by well-organized user-friendly indices enabling users to find easily what they seek.

This potential is rapidly increasing interest in technologies to extract and analyze automatically personal opinions from such user generated contents (UGCs) as customer reviews and weblog posts. Hence, a new field of natural language processing called sentiment analysis or opinion mining is appearing [2, 12, 20, 5, 6, 7, 13]. As symbolized by the term *sentiment*, this trend of research has been focused on subjective statements such as *I like* and *is fabulous*.

Subjective information in sentiment analysis, however, is only half of the possible harvest from UGCs. UGCs contain not only subjective material but also a vast range of factual, objective statements describing such personal experiences as in (1).

(1) *On my way home, I (in a wheelchair) could not find my way out of Totsuka Station because all the elevators in the station building stop running at 11pm.*

Such information can indicate the concrete and objective reasons for sentiments or opinions, which are often crucial for decision making and problem solving.

In light of these newly emerging insights, we have been developing a language processing technology for fully automatic extraction of personal experiences as well as opinions from weblog and message forum posts, indexing them with semantically organized indices. In this paper, we use the term *experience* in a very broad sense that includes holding an opinion as well as hearing of an experience of others. So, to restate, our goal is to:

(a) collect personal experiences relevant to a broad range of topics including consumer products (automobiles, cellular phones, etc.), public places

(tourist sites, hospitals, etc.), social systems (administrative services, welfare systems, etc.), and

(b) store them all together in a huge database, called an *experience database*, where each experience is represented as a piece of structured information comprising such slots as *topic*, *experiencer*, *event type*, *factuality* and *source pointer* as in (2) below.

(2) a. **Topic object**: What the experience is about (e.g. *Totsuka Station* in the case of example (1) above)

b. **Experiencer**: Who experiences (the *author* of the text)

c. **Event expression**: What event is experienced (*could find my way out*)

d. **Event type**: The semantic type (and sentiment orientation if applicable) of the experienced event (*could find my way out* is a positive/desirable happening)

e. **Factuality**: Whether the event indeed took place or not i.e. the temporal and modal status of the event (*I couldn't find my way out* is an *affirmatively negated past event*)

f. **Source pointer**: A pointer to the source text

The key idea is to index experiences not just by topic keywords and authorship but by a combination of semantic indices such as *event types* and *factuality*. The event types categorize the main predicate of an experience into semantic categories such as *buying*, *using* and *positive/negative happening*. The factuality slot specifies the temporal and modal status of the event referred to by the main predicate of an experience, which indicates, for example, whether the event did indeed take place in the past or is just a hypothetical situation. In the above example, the occurrence of a positive/desirable event is affirmatively negated, from which we can recognize this experience as something undesirable, i.e., a *trouble*.

Once available, a DB of this type offers a wide range of applications. Semantic indices such as event types and temporal and modal attributes allow retrieval of, for example, *troubles experienced using a particular consumer product* or *complaints and requests regarding a particular local welfare system*. Furthermore, experiences collected from weblog posts, where authorship is identifiable, can also be used to profile an author (the blogger) and enable retrieval of *authors* by

such complex queries as *those who have not bought a particular product model while expressing interest in it* or *those who had been using a particular service regularly but have recently stopped using it*. Such retrieval possibilities turn the vast amount of UGCs into a valuable resource useful in evaluating public services and social systems as well as for corporate marketing and risk management.

2 Technical challenges

Our task can be decomposed into the following sequence of subtasks:

1. **Event mention extraction**: Given an input text, we first identify event mentions which may constitute an experience by simple dictionary lookup. For this purpose, we build a typologized lexicon of expressions of experiences as we briefly describe below. When a text is given, (1) for example, we identify *can find (my way out)* as a positive/desirable state and *(elevators) stop running* as a negative/undesirable happening.

2. **Entity-event relation extraction**: For each identified event mention, we next seek from the local context an entity about which the event is an experience (i.e., can be interpreted as an experience). *could not find* in (1), for example, can be considered to be an experience about *Totsuka Station* but not *home*.

3. **Factuality analysis**: If any appropriate entity mention is found, we then carry out factuality analysis to identify the factuality status of the event. By doing this, we can distinguish, for example, between events which actually took place and those merely surmised or desired by the author.

4. **Experiencer identification**: Finally, we identify the experiencer of each experience.

Each of the steps represents an interesting technical challenge. Entity-event relation extraction and experiencer identification have already been addressed in the context of opinion mining. Entity-event relation extraction is the task of identifying the relation instances between an evaluative expression and its subject in opinion mining whereas experiencer identification can be taken as an extension of the task of identifying opinion holders [8, etc.]. The other two steps, on the other hand, involve new challenges so far given paid little attention in opinion mining.

One major issue in event mention extraction is how to create a lexicon of event expressions with a sufficiently broad coverage. For the event typology, we currently assume that the following distinctions are useful for characterizing experiences:

(3) a. **Sentiment**: Predicative expressions of an emotion or subjective evaluation. Each has an sentiment orientation (i.e. positive or negative).
  - **Emotion**: *enjoy*, *disappointed*
  - **Evaluation**: *tasty*, *inconvenient*
  - **Reputation**: *popular*, *criticised*

 b. **Happening**: Predicative expressions referring to a non-volitional event or state which is related to the use of a topic object and has a sentiment orientation
  - **General**: *pass (an exam), get slim, do (something) on time*, *cheated*, *broken*, *run out*
  - **Availability**: *released*, *(system) go into effect*
  - **Usability**: *get used to*, *prohibited*

 c. **Action**: Predicative expressions referring to experiencers' volitional actions related to the use of a topic object. Sentiment orientations are not necessarily involved.
  - **Buying/Selecting**: *buy*, *get*, *apply to (a social system)*, *choose*
  - **Using**: *use*, *drive (a car)*
  - **Stopping**: *cancel*

Expressions of Emotion, Evaluation and Reputation can be largely imported from existing sentiment lexicons such as SentiWordNet for English and Kobayashi's sentiment lexicon [8] for Japanese. For Action expressions, on the other hand, our preliminary exploration into weblog posts reveals that most expressions can be covered by a relatively small list of predicates. To obtain those predicates, WordNet-like general purpose thesauri can be employed. In contrast to the above two classes, collecting Happening expressions with sentiment orientation is a new challenge given their wide variety. While we are currently employing lexicons such as those automatically acquired by Takamura et al.[18], we find much room for exploration in this field.

The last, but very important, subtask of experience mining is factuality analysis. We believe this task could serve as an important semantic component across a wide range of language technology applications. However, it has so far attracted surprisingly little attention in the literature. One major technical contribution of our present work is that we designed the task and gave a machine learning-based solution to it as we describe in the next section.

3 Factuality Analysis

3.1 Aims and background

For each event mention, we want to identify the temporal and modal status of the event entity referred to by the event mention. Namely, we want to know, for example:

- whether the event indeed took place, is intended to take place, or just hypothetical,

- whether the happening of the event is desired by the author or not, and

- whether the event is a single event, a series of repeated events, or a state.

To this end, one might consider adopting a highly formal representation like temporal logic. However, introducing such a logic-based representation would require extremely sophisticated language understating and the state-of-the-art technology has not reached that level.

Previous work on time and modality analysis can be seen in the context of event extraction. For example, in the ACE (Automatic Content Extraction) research program [1], which has been driving the major research trend in event extraction/reasoning, each event mention is supposed to be annotated with temporal and modal markers as in (4).

(4) a. **TENSE**: Past, Present, Future, Unspecified
  b. **POLARITY**: Positive, Negative
  c. **MODALITY**: Asserted, Other

This markup scheme, however, is too simple for our purpose. For example, ACE has only two labels for modality, Asserted and Others, while we need more than five or six distinct labels, as described below.

Another effort we should refer to is TimeML[14], a specification language for events and temporal expressions, which annotates event mentions with tense, aspect, polarity and modality information as in (5).

---

[1]http://projects.ldc.upenn.edu/ace/

(5) a. **TENSE**: Past, Present, Future, None, Infinitive, Present-Perfect, Past-Perfect

    b. **ASPECT**: Progressive, Perfective, Perfective-Progressive, None, Initiation, Culmination, Termination, Continuation, Reinitiation

    c. **POLARITY**: Positive, Negative

    d. **MODALITY**: must, may, should, would, could

    e. **S-LINK**: Modal, Factive, Counter-factive, Evidential, Negative-evidential, Conditional

While the labels are more fine-grained than those of ACE, the markup scheme of TimeML is, however, highly dependent on the syntax of the target language (currently only English and Chinese) and, more importantly, is too shallow to capture such factuality information as we require. In fact, researchers engaged in the TimeML project are currently developing a more semantic-oriented level of representation of factuality for the purpose of reasoning textual entailment[15].

3.2 Factuality markup scheme

Given this context, we have created a new markup scheme for annotating event mentions with factuality information. We annotate each event mention in a given text with a triplet ⟨*Event-time, Modality, Modality-time*⟩.

The *Event-time* slot represents the tense, aspect and polarity status of the event in question, consisting of three sub-slots *Past-Present-Future*. Each sub-slot is to be filled with one of the following ASPECT-POLARITY labels, denoting the aspect and polarity (negation) information:

(6) **ASPECT-POLARITY**: *Punctual* (*Pnc*), *State-Continuation* (*StC*), *Repetition* (*Rpt*), *Initiation* (*Int*), *Termination* (*Trm*), *Negation* (*Ngt*), *Uncommitted* (*Unc*)

where all but *Negation* and *Uncommitted* implicitly denote *Positive* in terms of polarity. *Uncommitted* denotes that the author does not say anything about whether the event takes place in the corresponding slot of time. An example is given in (7), where the *Event-time* of the event mention *using* is annotated with *Int-Rpt-Unc*.

(7) a. *I started using FireFox recently.*

    b. ⟨*Int-Rpt-Unc, Affirm, Unc-Pnc-Unc*⟩

In experience mining, it is often meaningful to distinguish between repeatedly happening events and single punctual events. For example, corporate marketers may seek customers who use their product repeatedly; and troubles which recur may well be more serious than single occurrence. It is also important to capture the initiation and termination of a repetitive or continuous event, for this Will enable a search, for example, *those who recently stopped using a particular social welfare system.*

The *Modality* slot specifies the author's mental or communicative attitude toward the event in question. As a set of possible values of this slot, we have so far identified the following classes based on several reference books on Japanese modality [10, etc.]:

(8) **MODALITY**: *Affirm, Infer, Doubt, Hear, Intend, Ask, Recommend, Hypothesize, Other*

For example, while the *Modality* of the event *Using* in (7a) is *Affirm*, the *Modality* of the event *possess* in the next sentence (9a) is interpreted as *Hear*.

(9) a. *I watched a TV program reporting isoflavone-rich foods possessed activity against cancer.*

    b. ⟨*Unc-StC-Unc, Hear, Pnc-Unc-Unc*⟩

An important point to note here is that unlike the modality labels defined in TimeML (see 3.1 above), our modality labels are defined at the semantic level. More specifically, in TimeML, each modality label simply corresponds to an auxiliary verb and each S-LINK label is also strictly associated with a small set of modality verbs; for example, Factive is associated with verbs such as *forget* and *regret*. However, on the other hand, what we want to do in factuality analysis is to identity the temporal and modal status of each event mention at a semantic level. For example, in Japanese, a modality value *Doubt* may be linguistically realized by such a verb as *utagau* (*doubt*) or an interrogative particle *ka*. There is also a range of adverbs and adverbial functional expressions that can be used to express a doubt. Some of them are highly context-dependent and are thus apparently ambiguous. To make a factuality analysis component applicable to experience mining, we need to handle these phenomena.

3.3 Training factuality analysis models

To automate the above factuality analysis task, we created a manually annotated corpus and trained a statistical machine learning-based model.

To create an annotated corpus, we first randomly sampled from our weblog corpus (see 4.1) sentences including any one of the three chosen topic keywords (beverage name, automobile name, shampoo name).

Table 1: The results of the experiments (label accuracy)

Model	Domain	Past	Pres	Fut	Mod
Baseline	all	.61	.61	.76	.66
SVM	beverage	.49	.52	.72	.82
SVM	automobile	.38	.48	.74	.84
SVM	shampoo	.53	.63	.80	.84
Fact. CRF	beverage	.66	.61	.90	.83
Fact. CRF	automobile	.75	.59	.88	.85
Fact. CRF	shampoo	.68	.58	.90	.85

We then asked two annotators to annotate with factuality tuples all the event mentions included in the sampled sentences. After rehearsing several times, the annotators came to exhibit a remarkable agreement on unseen data — the κ statistics citekappa was 0.68, where they were considered to agree for an event mention only if *all its slots agreed*. This figure indicates that our annotation scheme is reliable enough. We then re-sampled sentences for the same topic keywords, obtaining 2,646 sentences in total, and asked one of the above two annotators to annotate all 4,417 event mentions included in the obtained sentences.

As easily imagined, the distribution of the value of each slot is highly skewed. Therefore, a simple baseline is given by choosing the most common values for each slot (*Unc* for all the three sub-slots of *Event-time* and *Assert* for *Modality*). The results are shown in the baseline of Table 1. The *Modality-time* slot was neglected in the experiment because its value was *Unc-Pnc-Unc* (i.e. the present tense) over 95 percent of the time.

Our task is now restated as one of determining the values of the four slots $\langle Et_1\text{-}Et_2\text{-}Et_3, Mdl \rangle$. We have so far examined two machine learning models.

First, the three *Event-time* sub-slots Et_1, -Et_2 and -Et_3 may well be highly dependent on their neighbors. We therefore employed the SVM-HMM algorithm [19] to train an *Event-time* model so that it could optimize the labels of those three slots simultaneously and we used the SVM-Multiclass package [19] to train a *Modality* model, which took care of the *Modality* slot independently of the *Event-time* slots.

The second model we examined is more sophisticated. Besides the inter-dependency between the *Event-time* slots, each slot may well be dependent also on the *Modality* slot. Furthermore, the factuality of an
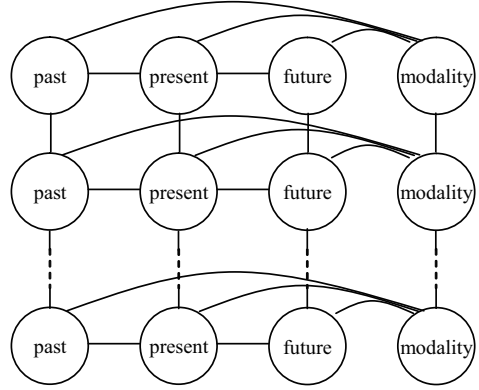

Figure 1: A graphical model representing the interdependencies between the factuality labels of neighboring event mentions

event mention is also likely to interact with that of any neighboring event mentions appearing in the same sentence. Such interdependencies led us to consider the graphical model illustrated in Figure 1. To train this mesh model, we employed the state-of-the-art GRMM toolkit, designed for the paradigm of conditional likelihood maximization [17]. This toolkit can deal with graph structures which include loops as in Figure 1. Training this type of mesh model on the basis of conditional likelihood maximization is also called *Factorial CRFs (Conditional Random Fields)*.

The feature set we used for both models included bag-of-words features with part-of-speech tags and lexemes extracted from neighboring base-NP/VP phrases and from the head phrase of the sentence.

3.4 Empirical evaluation

Finally, we conducted a three-fold cross-validation using our annotated corpus, where, for each fold, a model was trained on the data of two of the three domains (beverage, automobile, shampoo) and tested on the third domain. The results are shown in Table 1.

The tendency we observe from these figures is clear. First, the SVM-based model did not particularly outperform the baseline. This indicates the difficulty of the task, which is partly due to the skewness of the labels (i.e., the baseline is already quite high). Second, on the other hand, the Factorial CRF-based model substantially improved the accuracy for all the slots, which shows the importance of considering the interdependency between neighboring labels in this task.

Our error analysis revealed considerable room for improvement. Concretely, feature engineering is expected to be of great help — the present bag-of-words-based features set is doubtlessly too simple to rep-

resent complex combinations of Japanese auxiliary verbs and particles. While our factorial CRF-based model worked well across domains, for practical use, it would also be effective to extend the training data to a wide variety of other domains. We are planning to employ an active learning schema for efficient collection of informative training data.

4 Two application systems

Employing these components just described, we have developed two application systems each designed for distinct purposes. We briefly present each here.

4.1 Experience search for users/consumers

One system is designed for users of topic objects (consumer products, public places, social systems, etc.). Given one or more topic objects specified by a user of the system, the system provides the user with facilities for browsing bloggers' experiences related to those topic objects. Each experience instance is automatically classified into about ten *experience classes*. Each experience class is defined in terms of event types and factuality labels. For example, the experience class *Experienced troubles* is defined as a small number of combinations of event types and factuality labels including:

(10) a. *negative happening* and ⟨*Pnc-Unc-Unc, Affirm, Unc-Pnc-Unc*⟩ and
  b. *positive happening* and ⟨*Trm-Ngt-Unc, Affirm, Unc-Pnc-Unc*⟩.

By this classification, a user of the system can restrict a search to, for example, only *troubles experienced by the users of a specific topic object in question*.

To build the system, we first collected recent one year worth of Japanese weblog posts, which amounted to about 150M posts. We next collected a set of potential topic objects from Wikipedia[2]. From the categories under the *technology*, *culture* and *society* super categories in Wikipedia, we obtained about 15K keywords (i.e. topic objects) each corresponding to a Wikipedia article. To build a lexicon of event expressions, we (patially) manually checked semi-automatically collected candidate expressions, obtaining about 60K expressions. A majority of the lexicon comprised verbal phrases such as *ijou-o kanjiru* (*feel disorder*; negative happening) and *nedan-ga agaru* (*prices rise*; negative happening).

Figure 2: A snapshot of the summary view of the experience search engine

We next automatically extracted sentence-chunked texts from the weblog post set, and conducted tokenization and POS tagging with ChaSen[3] and dependency parsing with CaboCha[4]. We then carried out experience mining on the parsed texts and finally obtained over 50M experience instances related to one of our keywords and stored all of them in a relational database.

Figure 4 shows a snapshot of the system's view, where a summary of the search results for a query *Dogo Onsen* (hot spring), *Ikaho Onsen* and *Kurokawa Onsen*, is presented. For each given topic object, the system presents the number of bloggers who have described one or more experiences related to that topic object, where the bloggers are counted separately for each experience class. Furthermore, for each experience class, several major experience expressions are presented. Given this view, the user can overview the reported experiences for each topic object and compare them across different topic objects.

By clicking one of the experience classes (e.g. *trouble*), the user is led to another view, as shown in Figure 3, and can browse there all the mentions about troubles experienced during the use of the topic object. Characteristic of our system is that it presents experience-mentions blogger by blogger and ranks bloggers according to the number of their experience mentions about the queried topic object. We are assuming that the more experienced a person is with a given topic, the more he/she knows about it and the

Figure 3: A snapshot of the experience view showing troubles experienced while using *iPod*

Figure 4: A snapshot of the system view customized for corporate marketing research

more important his/her mentions about it are. Based on this assumption, the system also allows a user to browse a blogger's experiences with a topic object in chronological order, possibly a clue regarding the blogger's background (expert, confederate, etc.).

The system is scheduled to begin serving unrestricted users as of October this year at our Web site[5].

4.2 Blogger search for marketing research

We also extended the above system by enhancing the user interface as shown in Figure 4, which demonstrates how a user can specify complex queries comprising event type and factuality configurations. Collaborating with a major Internet service provider (the leading UGC-based marketing research business in Japan), we designed the user interface and defined the default set of experience classes based on a marketing theory [16, 4]: *Attention*, *Interest*, *Desire*, *Experience*, *Enthusiasm*, and *Share*. Those classes were straightforwardly defined in terms of our event types and factuality labels. Using those notions, a user of the system can seek, for example, *those who have not bought a particular product model while expressing interest in it* or *those who had been using a particular service regularly but recently stopped using it*.

5 Conclusion

In this paper we have proposed a new UGC-oriented language technology application called experience mining. Experience mining aims at automatically collecting instances of personal experiences as well as

---

opinions from an explosive number of UGCs such as weblog and forum posts and storing them in an experience database with semantically rich indices. Experience mining can be regarded as a substantial extension of opinion mining. Opinion mining has so far tended to aim at extracting sentiment information mainly from explicit evaluative or emotional expressions such as *useful* (positive) or *disturbing* (negative) [1, 3, 8, etc.]. On the other hand, experience mining covers all the descriptions of events that are related to any use of a wide variety of topic objects including so-called *implicit evaluative descriptions*.

We have also argued the technical issues of this new task. Focusing on factuality analysis, we have designed the task anew and given a machine learning-based solution to it. Our empirical evaluation indicates that the task is sufficiently well-defined to achieve a high inter-annotator agreement, and our factorial CRF-based model considerably outperforms the baseline. Furthermore, our technology will also benefit other types of applications. In the biomedical domain, for example, recognizing the factuality of each event mentioned in research papers is crucial, though very few researchers have addressed this issue [21, 9, 11].

We have also presented an application system, which currently stores over 50M experience instances with semantic indices — scheduled to start this October as an experience search engine for unrestricted users. Although we empirically evaluated the factuality analysis component, the experience search system as a whole is still to be evaluated from various angles, such as accuracy, utility and usability. An extrinsic

evaluation of the whole system is included in our future work.

Acknowledgements

References

[1] Y. Choi, E. Breck, and C. Cardie. Joint extraction of entities and relations for opinion recognition. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 431–439, Sydney, Australia, July 2006. Association for Computational Linguistics.

[2] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proc. of the 12th International World Wide Web Conference*, pages 519–528, 2003.

[3] A. Esuli and F. Sebastiani. Determining term subjectivity and term orientation for opinion mining. In *Proceedings of the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, 2006.

[4] H. M. Goldman. *How to Win Customers*. Pan Books, London, 1958.

[5] M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proc. of the Tenth International Conference on Knowledge Discovery and Data Mining*, pages 168–177, 2004.

[6] H. Kanayama and T. Nasukawa. Deeper sentiment analysis using machine translation technology. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 494–500, 2004.

[7] N. Kobayashi, R. Iida, K. Inui, and Y. Matsumoto. Opinion extraction using a learning-based anaphora resolution technique. In *The Second IJCNLP, Companion Volume to the Proceeding of Conference including Posters/Demos and Tutorial Abstracts*, pages 175–180, 2005.

[8] N. Kobayashi, K. Inui, and Y. Matsumoto. Opinion mining from web documents: Extraction and structurization. *Journal of the Japanese Society for Artificial Intelligence*, 22(2):227–238, 2007.

[9] M. Light, X. Y. Qiu, and P. Srinivasan. The language of bioscience: Facts, speculations, and statements in between. In L. Hirschman and J. Pustejovsky, editors, *Proceedings of the HLT-NAACL 2004 Workshop: BioLINK 2004, Linking Biological Literature, Ontologies and Databases*, pages 17–24, Boston, Massachusetts, USA, May 6 2004. Association for Computational Linguistics.

[10] T. Masuoka and Y. Takubo. *Fundamental Japanese Grammar*. Kuroshio, 1992. (in Japanese).

[11] B. Medlock and T. Briscoe. Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 992–999, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

[12] B. Pang and L. Lee. A sentiment education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics*, pages 271–278, 2004.

[13] A.-M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pages 339–346, 2005.

[14] J. Pustejovsky, J. M. Castano, R. Ingria, R. Sauri, R. J. Gaizauskas, A. Setzer, G. Katz, and D. R. Radev. Timeml: Robust specification of event and temporal expressions in text. In M. T. Maybury, editor, *New Directions in Question Answering*, pages 28–34. AAAI Press, 2003.

[15] R. Sauri and J. Pustejovsky. Determining modality and factuality for text entailment. In *Proceedings of the International Conference on Semantic Computing*, pages 509–516, Los Alamitos, CA, USA, 2007. IEEE Computer Society.

[16] E. Strong. Theories of selling. *Journal of Applied Psychology*, 9:75–86, 1925.

[17] C. Sutton. Grmm: A graphical models toolkit. http://mallet.cs.umass.edu, 2006.

[18] H. Takamura, T. Inui, and M. Okumura. Extracting semantic orientations of words using spin model. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL)*, pages 133–140, 2005.

[19] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research (JMLR)*, 6:1453 – 1484, September 2005.

[20] P. D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proc. of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, 2002.

[21] L. Zhou, G. B. Melton, S. Parsons, and G. Hripcsak. A temporal constraint structure for extracting temporal information from clinical narrative. *J. of Biomedical Informatics*, 39(4):424–439, 2006.