

Detecting Experiences from Weblogs

Keun Chan Park, Yoonjae Jeong and Sung Hyon Myaeng

Department of Computer Science

Korea Advanced Institute of Science and Technology

{keunchan, hybris, myaeng}@kaist.ac.kr

Abstract

Weblogs are a source of human activity knowledge comprising valuable information such as facts, opinions and personal experiences. In this paper, we propose a method for mining personal experiences from a large set of weblogs. We define experience as knowledge embedded in a collection of activities or events which an individual or group has actually undergone. Based on an observation that experience-revealing sentences have a certain linguistic style, we formulate the problem of detecting experience as a classification task using various features including tense, mood, aspect, modality, experiencer, and verb classes. We also present an activity verb lexicon construction method based on theories of lexical semantics. Our results demonstrate that the activity verb lexicon plays a pivotal role among selected features in the classification performance and shows that our proposed method outperforms the baseline significantly.

1 Introduction

In traditional philosophy, human beings are known to acquire knowledge mainly by reasoning and experience. Reasoning allows us to draw a conclusion based on evidence, but people tend to believe it firmly when they experience or observe it in the physical world. Despite the fact that direct experiences play a crucial role in making a firm decision and solving a problem, people often resort to indirect experiences by reading written materials or asking around. Among many sources people resort to, the Web has become the largest one for human experiences, especially with the proliferation of weblogs.

While Web documents contain various types of information including facts, encyclopedic knowledge, opinions, and experiences in general,

personal experiences tend to be found in weblogs more often than other web documents like news articles, home pages, and scientific papers. As such, we have begun to see some research efforts in mining experience-related attributes such as time, location, topic, and experiencer, and their relations from weblogs (Inui et al., 2008; Kurashima et al., 2009).

Mined experiences can be of practical use in wide application areas. For example, a collection of experiences from the people who visited a resort area would help planning what to do and how to do things correctly without having to spend time sifting through a variety of resources or rely on commercially-oriented sources. Another example would be a public service department gleaned information about how a park is being used at a specific location and time.

Experiences can be recorded around a frame like “who did what, when, where, and why” although opinions and emotions can be also linked. Therefore attributes such as location, time, and activity and their relations must be extracted by devising a method for selecting experience-containing sentences based on verbs that have a particular linguistics case frame or belong to a “do” class (Kurashima et al., 2009). However, this kind of method may extract the following sentences as containing an experience:

- [1] If Jason arrives on time, I'll buy him a drink.
- [2] Probably, she will laugh and dance in his funeral.
- [3] Can anyone explain what is going on here?
- [4] Don't play soccer on the roads!

None of the sentences contain actual experiences because hypotheses, questions, and orders have not actually happened in the real world. For experience mining, it is important to ensure a sentence mentions an event or passes a factuality test to contain experience (Inui et al., 2008).

In this paper, we focus on the problem of detecting experiences from weblogs. We formulate

Class	Examples
State	like, know, believe
Activity	run, swim, walk
Achievement	recognize, realize
Accomplishment	paint (a picture), build (a house)

Table 1. Vendler class examples

the problem as a classification task using various linguistic features including tense, mood, aspect, modality, experiencer, and verb classes.

Based on our observation that experience-revealing sentences tend to have a certain linguistic style (Jijkoun et al., 2010), we investigate on the roles of various features. The ability to detect experience-revealing sentences should be a precursor for ensuring the quality of extracting various elements of actual experiences.

Another issue addressed in this paper is automatic construction of a lexicon for verbs related to activities and events. While there have been well-known studies about classifying verbs based on aspectual features (Vendler, 1967), thematic roles and selectional restrictions (Fillmore, 1968; Somers, 1987; Kipper et al., 2008), valence alternations and intuitions (Levin, 1993) and conceptual structures (Fillmore and Baker, 2001), we found that none of the existing lexical resources such as Framenet (Baker et al., 2003) and Verbnets (Kipper et al., 2008) are sufficient for identifying experience-revealing verbs. We introduce a method for constructing an activity/event verb lexicon based on Vendler's theory and statistics obtained by utilizing a web search engine.

We define *experience* as knowledge embedded in a collection of activities or events which an individual or group has actually undergone¹. It can be subjective as in opinions as well as objective, but our focus in this article lies in objective knowledge. The following sentences contain objective experiences:

- [5] I ran with my wife 3 times a week until we moved to Washington, D.C.
- [6] Jane and I hopped on a bus into the city centre.
- [7] We went to a restaurant near the central park.

Whereas sentences like the following contain subjective knowledge:

- [8] I like your new style. You're beautiful!
- [9] The food was great, the interior too.

Subject knowledge has been studied extensively for various functions such as identification, po-

larity detection, and holder extraction under the names of opinion mining and sentiment analysis (Pang and Lee, 2008).

In summary, our contribution lies in three aspects: 1) conception of experience detection, which is a precursor for experience mining, and specific related tasks that can be tackled with a high performance machine learning based solution; 2) examination and identification of salient linguistic features for experience detection; 3) a novel lexicon construction method with identification of key features to be used for verb classification.

The remainder of the paper is organized as follows. Section 2 presents our lexicon construction method with experiments. Section 3 describes the experience detection method, including experimental setup, evaluation, and results. In Section 4, we discuss related work, before we close with conclusion and future work in Section 5.

2 Lexicon Construction

Since our definition of experience is based on activities and events, it is critical to determine whether a sentence contains a predicate describing an activity or an event. To this end, it is quite conceivable that a lexicon containing activity / event verbs would play a key role. Given that our ultimate goal is to extract experiences from a large amount of weblogs, we opt for increased coverage by automatically constructing a lexicon rather than high precision obtainable by manually crafted lexicon.

Based on the theory of Vendler (1967), we classify a given verb or a verb phrase into one of the two categories: *activity* and *state*. We consider all the verbs and verb phrases in WordNet (Fellbaum, 1998) which is the largest electronic lexical database. In addition to the linguistic schemata features based on Vendler's theory, we used thematic role features and an external knowledge feature.

2.1 Background

Vendler (1967) proposes that verb meanings can be categorized into four basic classes, *states*, *activities*, *achievements*, and *accomplishments*, depending on interactions between the verbs and their aspectual and temporal modifiers. Table 1 shows some examples for the classes.

Vendler (1967) and Dowty (1979) introduce linguistic schemata that serve as evidence for the classes.

¹ [http://en.wikipedia.org/wiki/Experience_\(disambiguation\)](http://en.wikipedia.org/wiki/Experience_(disambiguation))

Linguistic Schemata	<i>bs</i>	<i>prs</i>	<i>prp</i>	<i>pts</i>	<i>ptp</i>
No schema	■	■	■	■	■
Progressive			■		
Force	■				
Persuade	■				
Stop			■		
For	■	■	■	■	■
Carefully	■	■	■	■	■

Table 2. Query matrix. The “■” indicates that the query is applied. *No Schema* indicates that no schema is applied when the word itself is a query. *bs*, *prs*, *prp*, *pts*, *ptp* correspond to base form, present simple (3rd person singular), present participle, past simple and past participle, respectively.

Below are the six schemata we chose because they can be tested automatically: *progressive*, *force*, *persuade*, *stop*, *for*, and *carefully* (An asterisk denotes that the statement is awkward).

- *States* cannot occur in *progressive* tense:
John is running.
John is liking.*
- *States* cannot occur as complements of *force* and *persuade*:
John forced harry to run.
John forced harry to know.*
John persuaded harry to know.*
- *Achievements* cannot occur as complements of *stop*:
John stopped running.
John stopped realizing.*
- *Achievements* cannot occur with time adverbial *for*:
John ran for an hour.
John realized for an hour.*
- *State* and *achievement* cannot occur with adverb *carefully*:
John runs carefully.
John knows carefully.*

The schemata are not perfect because verbs can shift classes due to various contextual factors such as arguments and senses. However, a verb certainly has its fundamental class that is its most *natural* category at least in its dominant use.

The four classes can further be grouped into two genuses: a genus of processes going on in time and the other that refers to non-processes. *Activity* and *accomplishment* belong to the former whereas *state* and *achievement* belong to the latter. As can be seen in table 1, *states* are rather immanent operations and *achievements* are those occur in a single moment or operations related to

perception level. On the other hand, *activity* and *accomplishment* are processes (transeunt operations) in traditional philosophy. We henceforth call the first genus *activity* and the latter *state*. Our aim is to classify verbs into the two genuses.

2.2 Features based on Linguistic Schemata

We developed a relatively simple computational testing method for the schemata. Assuming that an awkward expression like, “John is liking something” won’t occur frequently, for example, we generated a co-occurrence based test for the first linguistic schema using the Web as a corpus. By issuing a search query, *((be OR am OR is OR was OR were OR been) and ? ing)* where ‘?’ represents the verb at hand, to a search engine, we can get an estimate about how the verb is likely to belong to *state*. A test can be generated for each of the schemata in a similar way.

For completeness, we considered all the verb forms (i.e., 3rd person singular present, present participle, simple past, past participle) available. However, some of the patterns cannot be applied to some forms. For example, other forms except the base form cannot come as a complement of *force* (e.g., *force to runs.**). Therefore, we created a query matrix which represents all query patterns we have applied, in table 2.

Based on the query matrix in table 2, we issued queries for all the verbs and verb phrases from WordNet to a search engine. We used the Google news archive search for two reasons. First, since news articles are written rather formally compared to weblogs and other web pages, the statistics obtained for a test would be more reliable. Second, Google provides an advanced option to retrieve snippets containing the query word. Normally, a snippet is composed of 3~5 sentences.

The basic statistics we consider are *hit count*, *candidate sentence count* and *correct sentence count* which we use the notations $H_{ij}(w)$, $S_{ij}(w)$, and $C_{ij}(w)$, respectively, where w is a word, i the linguistic schema and j the verb form from the query matrix in table 2. $H_{ij}(w)$ was directly gathered from the Google search engine. $S_{ij}(w)$ is the number of sentences containing the word w in the search result snippets. $C_{ij}(w)$ is the number of correct sentences matching the query pattern among the candidate sentences. For example, the *progressive* schema for a verb “build” can retrieve the following sentences.

[10] ..., New-York, is building one of the largest ...

[11] Is building an artifact?

“Building” in the first example is a progressive verb, but the one in second is a noun, which does not satisfy the linguistic schema. For a POS and grammatical check of a candidate sentence, we used the Stanford POS tagger (Toutanova et al., 2003) and Stanford dependency parser (Klein and Manning, 2003).

For each linguistic schema, we derived three features: *Absolute hit ratio*, *Relative hit ratio* and *Valid ratio* for which we use the notations $A_i(w)$, $R_i(w)$ and $V_i(w)$, respectfully, where w is a word and i a linguistic schema. The index j for summations represents the j -th verb form. They are computed as follows.

$$\begin{aligned} A_i(w) &= \frac{\sum_j H_{ij}(w)}{H_i(*)} \\ R_i(w) &= \frac{\sum_j H_{ij}(w)}{\sum_j H_{NoScheme}(w)} \\ V_i(w) &= \frac{\sum_j C_{ij}(w)}{\sum_j S_{ij}(w)} \end{aligned} \quad (1)$$

Absolute hit ratio is computed the extent to which the target word w occurs with the i -th schema over all occurrences of the schema. The denominator is the *hit count* of wild card “*” matching any single word with the schema pattern from Google (e.g., $H_i(*)$, the progressive test *hit count* is 3.82×10^8). *Relative hit ratio* computes the extent to which the target word w occurs with the i -th schema over all occurrences of the word. The denominator is the sum of all verb forms. *Valid ratio* means the fraction of correct sentences among candidate sentences. The weight of a linguistic schema increases as the *valid ratio* gets high. With the three different ratios, $A_i(w)$, $R_i(w)$ and $V_i(w)$, for each test, we can generate a total of 18 features.

2.3 Features based on case frames

Since the hit count via Google API sometimes returns unreliable results (e.g., when the query becomes too long in case of long verb phrases), we also consider additional features. While our initial observation indicated that the existing lexical resources would not be sufficient for our goal, it occurred to us that the linguistic theory behind them would be worth exploring as generating additional features for categorizing verbs for the two classes. Consider the following examples:

[12] John(D) believed(V) the story(O).

[13] John(A) hit(V) him(O) with a bat(I).

The subject of a state verb is dative (D) as in [12] whereas the subject for an action verb takes the agent (A) role. In addition, a verb with the instrument (I) role tends to be an action verb. From these observations, we can use the distribution of cases (thematic roles) for a verb in a corpus. Activity verbs are expected to have high frequency of agent and instrument roles than state verbs. Although a verb may have more than one case frame, it is possible to determine which thematic roles used more dominantly.

We utilize two major resources of lexical semantics, Verbnet (Kipper et al., 2008) based on the theory of Levin (1993), and Framenet (Baker et al., 2003), which is based on Fillmore (1968). Levin (1993) demonstrated that syntactic alternations can be the basis for groupings of verbs semantically and accord reasonably well with linguistic intuitions. Verbnet provides 274 verb classes with 23 thematic roles covering 3,769 verbs based on their alternation behaviors with thematic roles annotated. Framenet defines 978 semantic frames with 7,124 unique semantic roles, covering 11,583 words including verbs, nouns, adverbs, etc.

Using Verbnet alone does not suit our needs because it has a relatively small number of example sentences. Framenet contains a much larger number of examples but the vast number of semantic roles presents a problem. In order to get meaningful distributions for a manageable number of thematic roles, we used Semlink (Loper et al., 2007) that provides a mapping between Framenet and Verbnet and uses a total of 23 thematic roles of Verbnet for the annotated corpora of the two resources. By the mapping, we obtained distributions of the thematic roles for 2,868 unique verbs that exist in both of the resources. For example, the verb “construct” has high frequencies with *agent*, *material* and *product* roles.

2.4 Features based on how-to instructions

Ryu et al. (2010) presented a method for extracting action steps for how-to goals from eHow² a website containing a large number of how-to instructions. The authors attempted to extract actions comprising a verb and some ingredients like an object entity from the documents based on syntactic patterns and a CRF based model.

Since each extracted action has its probability, we can use the value as a feature for state / activity verb classification. However, a verb may appear in different contexts and can have multiple

² <http://www.ehow.com>

Feature	ME		SVM	
	<i>Prec.</i>	<i>Recall</i>	<i>Prec.</i>	<i>Recall</i>
All 43	68%	50%	83%	75%
Top 30	72%	52%	83%	75%
Top 20	83%	76%	85%	77%
Top 10	89%	88%	91%	78%

Table 3. Classification Performance

Class	Examples
Activity	act, battle, build, carry, chase, drive, hike, jump, kick, sky dive, tap dance, walk, ...
State	admire, believe, know, like, love, ...

Table 4. Classified Examples

probability values. To generate a single value for a verb, we combine multiple probability values using the following sigmoid function:

$$E(w) = \frac{1}{1 + e^{-t}} \quad (2)$$

$$t = \sum_{d \in D_w} P_d(w)$$

Evidence of a word w being an action in eHow is denoted as $E(w)$ where variable t is the sum of individual action probability values in D_w , the set of documents from which the word w has been extracted as an action. The higher probability a word gets and the more frequent the word has been extracted as an action, the more evidence we get.

2.5 Classification

For training, we selected 80 seed verbs from Dowty’s list (1979) which are representative verbs for each Vendler (1967) class. The selection was based on the lack of word sense ambiguity.

One of our classifiers is based on Maximum Entropy (ME) models that implement the intuition that the best model will be the one that is consistent with the set of constraints imposed by the evidence, but otherwise is as uniform as possible (Berger et al., 1996). ME models are widely used in natural language processing tasks for its flexibility to incorporate a diverse range of features. The other one is based on Support Vector Machine (Chang and Lin, 2001) which is the state-of-the-art algorithm for many classification tasks. We used RBF kernel with the default settings (Hsu et al., 2009) because it is been known to show moderate performance using multiple feature compositions.

The features we considered are a total of 42 real values: 18 from linguistic schemata, 23 the-

matic role distributions, and one from eHow. In order to examine which features are discriminative for the classification, we used two well known feature selection methods, Chi-square and information gain.

2.6 Results

Table 3 shows the classification performance values for different feature selection methods. The evaluation was done on the training data with 10-fold cross validation.

Note that the precision and recall are macro-averaged values across the two classes, *activity* and *state*. The most discriminative features were *absolute ratio* and *relative ratio* in conjunction with the *force*, *stop*, *progressive*, and *persuade* schemata, the role distribution of *experiencer*, and the eHow evidence.

It is noteworthy that eHow evidence and the distribution of *experiencer* got into the top 10. Other thematic roles did not perform well because of the data sparseness. Only a few roles (e.g., experience, agent, topic, location) among the 23 had frequency values other than 0 for many verbs. Data sparseness affected the linguistic schemata as well. Many of the verbs had zero hit counts for the *for* and *carefully* schemata. It is also interesting that the validity ratio $V_i(w)$ was not shown to be a good feature-generating statistic.

We finally trained our model with the top 10 features and classified all WordNet verbs and verb phrases. For actual construction of the lexicon, 11,416 verbs and verb phrases were classified into the two classes roughly equally. We randomly sampled 200 items and examined how accurately the classification was done. A total of 164 items were correctly classified, resulting in 82% accuracy. Some examples from the classification are shown in table 4.

A further analysis of the results show that most of the errors occurred with domain-specific verbs (e.g., *ablactate*, *alkalify*, and *transaminate* in chemistry) and multi-word verb phrases (e.g., *turn a nice dime*; *keep one’s shoulder to the wheel*). Since many features are computed based on Web resources, rare verbs cannot be classified correctly when their hit ratios are very low. The domain-specific words rarely appear in Framenet or e-how, either.

3 Experience Detection

As mentioned earlier, experience-revealing sentences tend to have a certain linguistic style.

Having converted the problem of experience detection for sentences to a classification task, we focus on the extent to which various linguistic features contribute to the performance of the binary classifier for sentences. We also explain the experimental setting for evaluation, including the classifier and the test corpus.

3.1 Linguistic features

In addition to the verb class feature available in the verb lexicon constructed automatically, we used tense, mood, aspect, modality, and experiencer features.

Verb class: The feature comes directly from the lexicon since a verb has been classified into a state or activity verb. The predicate part of the sentence to be classified for experience is looked up in the lexicon without sense disambiguation.

Tense: The tense of a sentence is important since an experience-revealing sentence tends to use past and present tense. Future tenses are not experiences in most cases. We use POS tagging (Toutanova et al., 2003) for tense determination, but since the Penn tagset provides no future tenses, they are determined by exploiting modal verbs such as “will” and future expressions such “going to”.

Mood: It is one of distinctive forms that are used to signal the modal status of a sentence. We consider three mood categories: indicative, imperative and subjunctive. We determine the mood of a sentence by a small set of heuristic rules using the order of POS occurrences and punctuation marks.

Aspect: It defines the temporal flow of a verb in the activity or state. Two categories are used: *progressive* and *perfective*. This feature is determined by the POS of the predicate in a sentence.

Modality: In linguistics, modals are expressions broadly associated with notions of possibility. While modality can be classified at a fine level (e.g., epistemic and deontic), we simply determine whether or not a sentence includes a modal marker that is involved in the main predicate of the sentence. In other words, this binary feature is determined based on the existence of a modal verb like “can”, “shall”, “must”, and “may” or a phrase like “have to” or “need to”. The dependency parser is used to ensure a modal marker is indeed associated with the main predicate.

Experiencer: A sentence can or cannot be treated as containing an experience depending on the subject or experiencer of the verb (note that this is different from the experiencer role in a case frame). Consider the following sentences:

[14] The stranger messed up the entire garden.

[15] His presence messed up the whole situation.

The first sentence is considered an experience since the subject is a person. However, the second sentence with the same verb is not, because the subject is a non-animate abstract concept. That is, a non-animate noun can hardly constitute an experience. In order to make a distinction, we use the dependency parser and a named-entity recognizer (Finkel et al., 2005) that can recognize person pronouns and person names.

3.2 Classification

To train our classifier, we first crawled weblogs from Wordpress³, one of the most popular blog sites in use today. Wordpress provides an interface to search blog posts with queries. In selecting experience-containing blog posts, we used location names such as Central Park, SOHO, Seoul and general place names such as airport, subway station, and restaurant because blog posts with some places are expected to describe experiences rather than facts or thoughts.

We crawled 6,000 blog posts. After deleting non-English and multi-media blog posts for which we could not obtain any meaningful text data, the number became 5,326. We randomly sampled 1,000 sentences⁴ and asked three annotators to judge whether or not individual sentences are considered containing an experience based on our definition. For maximum accuracy, we decided to use only those sentences all the three annotators agreed, resulting in a total of 568 sentences.

While we tested several classifiers, we chose to use two different classifiers based on SVM and Logistic Regression for the final experimental results because they showed the best performance.

3.3 Results

For comparison purposes, we take the method of Kurashima et al. (2005) as our baseline because the method was used in subsequent studies (Kurashima et al., 2006; Kurashima et al., 2009) where experience attributes are extracted. We briefly describe the method and present how we implemented it.

The method first extracts all verbs and their dependent phrasal unit from candidate sentences.

³ <http://wordpress.com>

⁴ It was due to the limited human resources, but when we increased the number at a later stage, the performance increase was almost negligible.

Feature	Logistic Regression		SVM	
	<i>Prec.</i>	<i>Recall</i>	<i>Prec.</i>	<i>Recall</i>
Baseline	32.0%	55.1%	25.3%	44.4%
Lexicon	77.5%	76.0%	77.5%	76.0%
Tense	75.1%	75.1%	75.1%	75.1%
Mood	75.8%	60.3%	75.8%	60.3%
Aspect	26.7%	51.7%	26.7%	51.7%
Modality	79.8%	70.5%	79.8%	70.5%
Experiencer	54.3%	53.5%	54.3%	53.5%
All included	91.9%	91.7%	91.7%	91.4%

Table 5. Experience Detection Performance

The candidate goes through three filters before it is treated as experience-containing sentence. First, the candidates that do not have an objective case (Fillmore, 1968) are eliminated because their definition of experience as “action + object”. This was done by identifying the object-indicating particle (case marker) in Japanese. Next, the candidates belonging to “become” and “be” statements based on Japanese verb types are filtered out. Finally, the candidate sentences including a verb that indicates a movement are eliminated because the main interest was to identify an activity in a place.

Although their definition of experience is somewhat different from ours (i.e., “action + object”), they used the method to generate candidate sentences from which various experience attributes are extracted. From this perspective, the method functioned like our experience detection. Put differently, the definition and the method by which it is determined were much cruder than the one we are using, which seems close to our general understanding.⁵

The three filtering steps were implemented as follows. We used the dependency parser for extracting objective cases using the *direct object* relation. The second step, however, could not be applied because there is no grammatical distinction among “do, be, become” statements in English. We had to alter this step by adopting the approach of Inui et al. (2008). The authors propose a lexicon of experience expression by collecting hyponyms from a hierarchically structured dictionary. We collected all hyponyms of words “do” and “act”, from WordNet (Fellbaum, 1998). Lastly, we removed all the verbs that are under the hierarchy of “move” from WordNet.

We not only compared our results with the baseline in terms of precision and recall but also

⁵ This is based on our observation that the three annotators found their task of identifying experience sentences not difficult, resulting in a high degree of agreements.

Feature	Logistic Regression		SVM	
	<i>Prec.</i>	<i>Recall</i>	<i>Prec.</i>	<i>Recall</i>
Baseline	32.0%	55.1%	25.3%	44.4%
-Lexicon	84.6%	84.6%	83.1%	81.2%
-Tense	87.3%	87.1%	86.8%	86.5%
-Mood	89.5%	89.5%	89.3%	89.2%
-Aspect	90.8%	90.5%	89.0%	88.6%
-Modality	89.5%	89.5%	82.8%	82.8%
-Experiencer	91.5%	91.4%	91.1%	90.8%
All included	91.9%	91.7%	91.7%	91.4%

Table 6. Experience Detection Performance without Individual Features

evaluated individual features for their importance in experience detection (classification). The evaluation was conducted with 10-fold cross validation. The results are shown in table 5.

The performance, especially precision, of the baseline is much lower than those of the others. The method devised for Japanese doesn’t seem suitable for English. It seems that the linguistic styles shown in experience expressions are different from each other. In addition, the lexicon we constructed for the baseline (i.e., using the WordNet) contains more errors than our activity lexicon for activity verbs. Some hyponyms of an activity verb may not be activity verbs. (e.g., “appear” is a hyponym of “do”).

There is almost no difference between the Logistic Regression and SVM classifiers for our methods although SVM was inferior for the baseline. The performance for the best case with all the features included is very promising, closed to 92% precision and recall. Among the features, the lexicon, i.e., verb classes, gave the best result when each is used alone, followed by *modality*, *tense*, and *mood*. *Aspect* was the worst but close to the baseline. This result is very encouraging for the automatic lexicon construction work because the lexicon plays a pivotal role in the overall performance.

In order to see the effect of including individual features in the feature set, precision and recall were measured after eliminating a particular feature from the full set. The results are shown in table 6. Although the absence of the lexicon feature hurt the performance most badly, still the performance was reasonably high (roughly 84 % in precision and recall for the Logistic Regression case). Similar to table 5, the aspect and experience features were the least contributors as the performance drops are almost negligible.

4 Related Work

Experience mining in its entirety is a relatively new area where various natural language processing and text mining techniques can play a significant role. While opinion mining or sentiment analysis, which can be considered an important part of experience mining, has been studied quite extensively (see Pang and Lee's excellent survey (2008)), another sub-area, factuality analysis, begins to gain some popularity (Inui et al., 2008; Sauri, 2008). Very few studies have focused explicitly on extracting various entities that constitute experiences (Kurashima et al., 2009) or detecting experience-containing parts of text although many NLP research areas such as named entity recognition and verb classification are strongly related. The previous work on experience detection relies on a handcrafted lexicon.

There have been a number of studies for verb classification (Fillmore, 1968; Vendler, 1967; Somers, 1982; Levin, 1993; Fillmore and Baker, 2001; Kipper et al., 2008) that are essential for construction of an activity verb lexicon, which in turn is important for experience detection. Most similar to our work was done by Siegel and McKeown (2000), who attempted to categorize verbs into state or event classes based on 14 tests similar to those of Vendler's. They attempted to compute co-occurrence statistics from a corpus. The *event* class, however, includes *activity*, *accomplishment*, and *achievement*. Similarly, Zaccrone and Lenci (2008) attempted to categorize verbs in Italian into the four Vendler classes using the Vendler tests by using a tagged corpus. They focused on existence of arguments such as subject and object that should co-occur with the linguistic features in the tests.

The main difference between the previous work and ours lies in the goal and scope of the work. Since our work is specifically geared toward domain-independent experience detection, we attempted to maximize the coverage by using all the verbs in WordNet, as opposed to the verbs appearing in a particular domain-specific corpus (e.g., medicine domain) as done in the previous work. Another difference is that while we are not limited to a particular domain, we did not use extensive human-annotated corpus other than using the 80 seed verbs and existing lexical resources.

5 Conclusion and Future Work

We defined experience detection as an essential task for experience mining, which is restated as

determining whether individual sentences contain experience or not. Viewing the task as a classification problem, we focused on identification and examination of various linguistic features such as verb class, tense, aspect, mood, modality, and experience, all of which were computed automatically. For verb classes, in particular, we devised a method for classifying all the verbs and verb phrases in WordNet into the *activity* and *state* classes. The experimental results show that verb and verb phrase classification method is reasonably accurate with 91% precision and 78% recall with manually constructed gold standard consisting of 80 verbs and 82% accuracy for a random sample of all the WordNet entries. For experience detection, the performance was very promising, closed to 92% in precision and recall when all the features were used. Among the features, the verb classes, or the lexicon we constructed, contributed the most.

In order to increase the coverage even further and reduce the errors in lexicon construction, i.e., verb classification, caused by data sparseness, we need to devise a different method, perhaps using domain specific resources.

Given that experience mining is a relatively new research area, there are many areas to explore. In addition to refinements of our work, our next step is to develop a method for representing and extracting actual experiences from experience-revealing sentences. Furthermore, considering that only 13% of the blog data we processed contain experiences, an interesting extension is to apply the methodology to extract other types of knowledge such as facts, which are not necessarily experiences.

Acknowledgments

This research was supported by the IT R&D program of MKE/KEIT under grant KI001877 [Locational/Societal Relation-Aware Social Media Service Technology], and by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) [NIPA-2010-C1090-1011-0008].

Reference

Eiji Aramaki, Yasuhide Miura, Masatsugu Tonoike, Tomoko Ohkuma, Hiroshi Mashiuchi, and Kazuhiko Ohe. 2009. TEXT2TABLE: Medical Text Summarization System based on Named Entity

- Recognition and Modality Identification. In *Proceedings of the Workshop on BioNLP*.
- Collin F. Baker, Charles J. Fillmore, and Beau Cronin. 2003. The Structure of the Framenet Database. *International Journal of Lexicography*.
- Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra. 1996. A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics*.
- Chih-Chung Chang and Chih-Jen Lin. 2001. LIBSVM : a Library for Support Vector Machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- David R. Dowty. 1979. *Word meaning and Montague Grammar*. Reidel, Dordrecht.
- Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.
- Charles J. Fillmore. 1968. The Case for Case. In Bach and Harms (Ed.): *Universals in Linguistic Theory*.
- Charles J. Fillmore and Collin F. Baker. 2001. Frame Semantics for Text Understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*.
- Jenny R. Finkel, Trond Grenager, and Christopher D. Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of ACL*.
- Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. 2009. A Practical Guide to Support Vector Classification. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Kentaro Inui, Shuya Abe, Kazuo Hara, Hiraku Morita, Chitose Sao, Megumi Eguchi, Asuka Sumida, Koji Murakami, and Suguru Matsuyoshi. 2008. Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents. In *Proceedings of the International Conference on Web Intelligence*.
- Valentin Jijkoun, Maarten de Rijke, Wouter Weerkamp, Paul Ackermans and Gijs Geleijnse. 2010. Mining User Experiences from Online Forums: An Exploration. In *Proceedings of NAACL HLT Workshop on Computational Linguistics in a World of Social Media*.
- Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A Large-scale Classification of English Verbs. *Language Resources and Evaluation Journal*.
- Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. In *Proceedings of ACL*.
- Takeshi Kurashima, Ko Fujimura, and Hidenori Okuda. 2009. Discovering Association Rules on Experiences from Large-Scale Blog Entries. In *Proceedings of ECIR*.
- Takeshi Kurashima, Taro Tezuka, and Katsumi Tanaka. 2005. Blog Map of Experiences: Extracting and Geographically Mapping Visitor Experiences from Urban Blogs. In *Proceedings of WISE*.
- Takeshi Kurashima, Taro Tezuka, and Katsumi Tanaka. 2006. Mining and Visualizing Local Experiences from Blog Entries. In *Proceedings of DEXA*.
- John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of ICML*.
- Beth Levin. 1993. *English verb classes and alternations: A Preliminary investigation*. University of Chicago press.
- Edward Loper, Szu-ting Yi, and Martha Palmer. 2007. Combining Lexical Resources: Mapping Between PropBank and Verbnet. In *Proceedings of the International Workshop on Computational Linguistics*.
- Bo Pang and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis, *Foundations and Trends in Information Retrieval*.
- Jihee Ryu, Yuchul Jung, Kyung-min Kim and Sung H. Myaeng. 2010. Automatic Extraction of Human Activity Knowledge from Method-Describing Web Articles. In *Proceedings of the 1st Workshop on Automated Knowledge Base Construction*.
- Roser Saurí. 2008. A Factuality Profiler for Eventualities in Text. PhD thesis, Brandeis University.
- Eric V. Siegel and Kathleen R. McKeown. 2000. Learning Methods to Combine Linguistic Indicators: Improving Aspectual Classification and Revealing Linguistic Insights. In *Computational Linguistics*.
- Harold L. Somers. 1987. *Valency and Case in Computational Linguistics*. Edinburgh University Press.
- Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL*.
- Zeno Vendler. 1967. *Linguistics in Philosophy*. Cornell University Press.
- Alessandra Zarcone and Alessandro Lenci. 2008. Computational Models of Event Type Classification in Context. In *Proceedings of LREC*.