

Product Name Classification for Product Instance Distinction

Hye-Jin Min

CS Department, KAIST
Daejeon, Republic of Korea
hjmin@nlp.kaist.ac.kr

Jong C. Park

CS Department, KAIST
Daejeon, Republic of Korea
park@cs.kaist.ac.kr

Abstract

Product names with a temporal cue in a product review often refer to several product instances purchased at different times. Previous approaches to product entity recognition and temporal information analysis do not take into account such temporal cues and thus fail to distinguish different product instances. We propose to formulate the resolution of such product names as a classification problem by utilizing time expressions, event features and other temporal cues for a classifier in two stages, detecting the existence of such temporal cues and identifying the purchase time. The empirical results show that term-based features and existing event-based features together enhance the performance of product instance distinction.

1 Introduction

Traditional work on product entity recognition has been conducted on competing products for comparative opinion mining from forum data (Jindal and Liu, 2006; Ding et al., 2009; Li et al., 2010), but not on the same type of products purchased at different times, thus failing to distinguish products at the instance level. The use of temporal information would help to make such distinction, but previous studies of temporal information have been made only for the detection and determination of temporal relations between time expressions and events, through the relevant shared tasks, or TempEval-1 and TempEval-2 tasks (Pustejovsky et al., 2003; Verhagen et al., 2009; Verhagen et al., 2010), but not for the distinction of products.

There is evidence that temporal relations between product instances of the same type are found quite often in product reviews, to give rise to important differences in the respective opinion of the reviewer. Consider Examples (1) and (2)¹ below, with two product names *other Levis 501s* and *these new ones* that refer to the product instances that the customer bought. While the former refers to the past purchase, the latter refers to the recent purchase.

- (1) My husband has [*other Levis 501s*]_a and [*these new ones*]_b are different in the weight of fabric (light), (...) We are not happy with *these jeans*.
- (2) a. I don't wear boots and I wear *these jeans* when I ride my bike.
b. *Jeans* were exactly like one purchased from Khol's or Sears.
c. I'm done with buying *jeans* online.

Resolving such different product instances properly is found crucial to identifying long-term customers, among others, whose opinions count at least as important as those of human annotators for influential reviews (Min and Park, 2012). Moreover, it is also crucial to identifying such long-term customer's sentiment change over several purchases of the same product (cf. Min and Park, 2012).

First, we note that sortal anaphoric expressions such as *these jeans* in (1) indicate the presence of a temporal cue but may also refer to the whole product as shown in (2a). We also note that the product name without a demonstrative or definite article as shown in (2b) may refer to the purchased

¹ The examples are taken from the reviews at Amazon.com.

product instance, unlike the one without temporal cue in (2c) that refers to a generic object.

We thus argue that, for the proper resolution of such product names, it is important to see if the given product name bears temporal information and to identify the temporal order among the product instances. We propose to formulate the resolution of such product names as a classification problem, by utilizing time expressions, event features and other temporal cues as relevant features for a classifier. We construct the classifier in two stages, first detecting the existence of such temporal cues and then detecting the recency of the purchase time. The proposed features are utilized in conjunction with the event-based temporal features in the TempEval task and the experience mining task.

We employ a support vector machine (SVM) classifier with cost-sensitive learning by taking minor classes into consideration. The empirical results show that the term-based features and existing event-based features can be made to work in different combinations to enhance the overall performance for product instance distinction.

We also apply our results of product name classification to two applications. One is to classify product reviews with respect to a customers' sentiment change (cf. Min and Park, 2012). The other is to automatically rate product reviews based on the detected sentiment in the given review. Our results show that the results of the product name classification are important for distinguishing the sentiment towards the recent purchase from the sentiments towards other purchases in the past.

The rest of the paper is organized as follows. Section 2 reviews the related work. Section 3 examines product names with temporal information. Section 4 compares event and time expression features with term-based temporal features. Section 5 shows the classification results and Section 6 discusses classification errors. Section 7 shows the applications and Section 8 concludes the paper with further work.

2 Related Work

Product entity recognition has been conducted to identify comparative opinions between competing products. Jindal and Liu (2006) proposed label sequential rules to detect comparable entities based

on association rule mining. Ding and colleagues (2009) added the process of filtering with pruning patterns about brand and model names for product names of comparable entities. Li and colleagues (2010) used weakly-supervised bootstrapping to detect comparable entities with sequential seeds derived from typical comparative questions. Our work is slightly different from the previous work in that each instance purchased at a time is regarded as an independent 'entity'. Thus, the temporal information about each candidate target works as an essential clue as compared to the previous work. We will discuss why we did not apply the approach in the previous work in Section 5.

Systems were developed for TempEval tasks by utilizing features from time expressions and events. For tasks C and E, TRIOS (Uzzaman and Allen, 2010) achieved the best performance on English news data. The system employed a Markov Logic Network-based classifier with feature-based first-order logic formulae. The utilized features are event-related, timex-related, and TLINK event-time signal. The system JU_CSE_TEMP that showed the second best performance (Kolaya et al., 2010) also utilized a similar time expression and event features for a CRF-based classifier. As events in the news reports are mostly in the form of a verb phrase, they focused on verb-related clues for event features.

Researches to mining experiences in user-generated web documents addressed the issue of distinguishing 'experience sentences' from others. Park and colleagues (2010) proposed a discrimination method based on the linguistic properties of the mentioned events in such sentences, including verb class, tense, aspect, mood, modality and experiencer. Since we focus on classification at the term level, we utilize not only time expression and event-based features but also term-based temporal features.

3 Product Names

3.1 Temporal Class vs. Atemporal Class

We use temporal class to include product names whose instances are purchased and used by the customer of a given review (Ex. (3) ~ (8)). We also use atemporal class to include product names that refer to generic objects or those with unknown purchase time (Ex. (9) ~ (13)). One might think the

product name in (3) should be classified as the atemporal class because it refers to the brand and the model name of the given product. However, we perceive that it also implies that the customer has been wearing the product for such a long time if we look at the time expression ‘for 22 years’ and ‘since I was 16’. Since the purpose of the classification in this paper is to identify long-term customers who have purchased the product several times, we classify such name as the temporal class. On the contrary, the product name with ‘no article’ such as (10) or (12) should be classified as the atemporal class as it is just used in order to describe a generic object not an product instance.

Temporal Class

- (3) I have been wearing *Levis 501* since I was 16 - which means that for 22 years I keep returning to the classic, and it always fits just right.
- (4) In the past (at least up until 2 or 3 years ago) my husband would have about *4 pairs of jeans*.
- (5) This is [*the second pair of levis*]_a I ordered through amazon. It was identical to [*the first pair*]_b except the first was made in Lesotho the second in Egypt.
- (6) Now, I’m doomed for eternity to constantly checking the back of *my other two pairs of jeans* every time I wear them to make sure there isn’t another big rip in them.
- (7) Currently I have over *10 pairs* that have never been worn because there is no quality control at Levi Strauss.
- (8) *My oldest pair*, which is holding up well, is over 10 years old.

Atemporal Class

- (9) No matter how tall or short you are, these are *the best jeans* you can buy.
- (10) *Real 501s* are made of 14 oz canvas-like material.
- (11) Of course this is simply a matter of personal preference on how you want *your jeans* to fit.
- (12) I will not buy *501s* again unless they are from trusted chain retail store.
- (13) [*These good old 501s*]_a with their slightly-inconvenient button-fly are [*the best-looking jeans*]_b on a middle-aged body I’ve found.

3.2 Recent Purchase vs. Past Purchase

Unlike relatively formal documents such as news reports, most of the temporal information mentioned in a product review is somewhat vague,

as in ‘a few months ago’. It thus makes sense to consider simply two sub-temporal classes, or ‘recent purchase’ and ‘past purchase’ subclasses, in order to tell apart whether or not the customer has experiences in the product over an extended period of time. Some names refer to both the recent purchases and past purchases, requiring compound class as well.

Recent Purchase (P_r): includes product names that refer to the product instances that are most recently purchased (Ex. (1b), (5a)).

Past Purchase (P_p): includes product names that refer to the product instances purchased prior to the most recent purchase (Ex. (1a), (4), (5b), (6), (8)).

Recent&Past Purchase (P_{r&p}): includes (3), (7).

3.3 Annotation

In order to establish the proof-of-concept, we crawled 382 product reviews of men’s jeans from Amazon.com. Note that there is no annotated corpus yet for product names that bear temporal information. Two annotators performed the task of annotating the text span for product names and the class for each name. The inter-annotator agreement score (kappa statistic) for the classification is substantially high (0.72). The disagreements resulted mostly from names with no articles, which made it ambiguous to determine whether they have temporal information or not. After setting-up more fine-grained guidelines for such cases, the annotators adjusted them together. Table 1 shows the distribution of the annotated product names.²

| Temporal | | | Atemporal |
|----------------|----------------|----------------------|-----------|
| P _r | P _p | P _{r&p} | |
| 326 | 63 | 51 | 390 |
| 440 | | | |
| 830 | | | |

Table 1: Distribution of product names

4 Linguistic Features for Classification

4.1 Temporal Features in the Literature

Time Expressions:

past duration/simple past/present

Following the previous work on the TempEval-2 task C (UzZaman and Allen, 2010; Kolaya et al., 2010), we utilized types and values of time

² The annotated data is available at <http://nlp.kaist.ac.kr/resources/>.

expression as the basic features. For example, the ‘duration’ type of expression such as ‘for 22 years’ as in (3) can work as an important clue for the ‘recent & past purchase’ subclass. For the present purpose, we only need to know each expression’s type and its symbolic value (e.g., past duration instead of ‘DURATION’ and ‘P22Y’ for ‘for 22 years’). Hence, we used the ‘duration’ relation for product use and the ‘before’ relation between several purchases among the relations defined in the Timebank Corpus (Pustejovsky et al., 2003). We re-classified them into three types as shown in Table 2, based on TYPE/VALUE information in the <TIMEX> tag, and set each type as the feature with the value either true or false.

| TIMEX2/TIMEX3 | | POS | Type | Exam ples |
|---------------|----------------------|--------------|------------------|-----------------------|
| TYPE | VALUE | IN/RB | | |
| DURAT ION | PNY | for until | Past duration | for 22 years |
| DATE | PRESE NT_RE F; | - | Present | Curre ntly, now |
| | PND, PNW | in/ago | | A few days ago |
| | PAST_ REF; | - | Simple past | before |
| | PNM, PNY | in/ago | | 2 years ago |

Table 2: Types of time expression

Event Features: We adopted the event features for the TempEval-2 C and such as event class, tense, aspect, and polarity. We also adopted several verb-related features including verb class, tense, aspect, mood, modality and experiencer in order to determine whether the given sentence is experience-revealing or not (Park et al., 2010), because the first stage of the classification into temporal and atemporal classes in the present work is similar to their work in that the temporal aspect of events is considered as important for detecting mentions about the actual experiences.

In addition to the features adopted from the previous work, we considered syntactic and semantic types of verbs, because the whole phrase including the verb and the name can be a generic object as shown in (2c), or because copular verbs are frequently used to express opinions, which characterize the products as shown in (9). Since we focus on classification at the term level, unlike at the sentence level in their work, we utilize three

types of verbs: 1) the verb type taking the given name as an argument (v1, ‘have’ in (7)); 2) the verb type which functions in a relative clause modifying the name (v2, ‘worn’ in (7)); and 3) the matrix verb to which the name belongs (v3, ‘have’ in (7)).

1) Tense & Aspect:

present/past/future/present perfect/past perfect/
present progressive/past progressive/ present
perfect progressive/past perfect progressive

Tense and aspect information of a verb can also be an important clue because instances already purchased would be mentioned in the sentence with the past tense and the perfect aspect. For example, the tense and aspect of the v1 type verbs as shown in (3) and (4) and those of the v2 type verb as shown in (5) indicate that the product names are classified into the temporal class. By contrast, future and present tenses are more likely to suggest that the relevant names refer to generic objects, as shown in (10) and (12).

2) Syntactic type: gerund/to-inf/general verb

We set the indication of whether the v1 type verb is the main verb or not as another feature because the whole phrase including the verb and the term can be a generic object as shown in (2c).

3) Semantic type:

purchase/possess/purchasing process/
copular verbs/emotional expressions

While verbs such as ‘purchase’, ‘possess’ or emotional expressions are most likely used in the sentence that contains product names in the temporal class as shown in (4), copular verbs are frequently used to express opinions, which describe the product instances as shown in (9).

4) Event class:

OCCURRENCE/I_ACTION/I_STATE/STATE/
REPORTING/PERCEPTION/ASPECTUAL

The event class of a verb may also affect the classification of the given product name into the temporal or atemporal class. For example, events such as ‘know’ or ‘want’ (I_STATE) as shown in (11) are related to the atemporal class. On the other hand, events such as ‘buy’, ‘have’, or ‘wear’ (OCCURRENCE) as shown in (3) and (4) are related to the temporal class. We semi-automatically annotated the event class of each verb in our corpus by following the event annotation guidelines of the TempEval-2 task.³

³ <http://www.timeml.org/tempeval2/tempeval2-trial/guidelines/EventGuidelines-050409.pdf>

5) Event polarity: positive/negative

Negation often reveals the product name to be classified into the atemporal class such as a generic object *original Levis* in the sentence “all they sell are Levis Signature jeans, and not *original Levis*.” We set either the positive or negative value depending on the presence of negation about the verb.

Experiencer:

I or we/you/3rd person/product/ETC

We adopted the experiencer feature from Park and others (2010) but we further distinguished 1st person subject from 2nd person subject and 3rd person subject with the intuition that the sentence with 2nd person or 3rd person is likely to contain a suggestion for potential customers as shown in (11).

If clause: true/false

The clause with the *if* or *unless* marker mostly expresses condition or supposition, and hence, the product name in such a clause may refer to the atemporal type as shown in (12). We set either true or false depending on the existence of the *if* or *unless* clause in the sentence.

4.2 Term-based Temporal Features

We also consider term-based features. Customers may use specific cue words (e.g., *these new ones*) and/or ordinal words (e.g., *the second pair I bought*) in order to distinguish a product instance from others that are purchased at different times. We also add clues for identifying the coreference relations between the product names (Soon et al., 2001).

Temporal cue words: recent purchase-related cue /past purchase-related cue

Adjectives within a product name or adverbs modifying the verb that takes the name as an argument furnish the given product name’s temporal information (e.g., *new* for *these new ones* as in (1b), *currently* for *10 pairs* as in (7), and *oldest* for *my oldest pair* in (8)). In order to avoid counting the word with a different sense in the given context such as *these good old 501s* as in (13a), we filter out such cases from a bi-gram model for our corpus. For each type of cue, its value is set to either true or false.

Quantity-based cue words:

(one, two, several)/(first, second, nth)

Cardinal or ordinal numbers in the given product name (e.g., (4) and (5), respectively) may also

work as good clues for the membership in the temporal class. In particular, ordinal numbers may also suggest temporal orders among several product names as shown in Example (5). The name *the second pair* refers to a more recently purchased instance as compared to the name *the first pair*, which obviously refers to the past purchase. After detecting cardinal or ordinal number words in the given product name, we set the categorical value of the cardinal and ordinal number features.

Determiner/JJ: this/the/a(n)/(any)other/another

The product names containing ‘this’/‘these’ may refer to the recent purchase(s) (e.g., (1b)), and the product names containing an indefinite article ‘a’/‘an’ or the definite article ‘the’ may also be used to indicate an instance of a given product (e.g., (5)). The determiner ‘another’ or the adjective ‘other’ is often used for separating one instance from others, and hence, the instances mentioned in the same review with different temporal information can be disambiguated (e.g., *other Levis 501s* in (1a) and (6)). For each determiner or adjective, its value is set to either true or false.

Possessive pronouns: my/your/his/her/their

The possessive pronoun ‘my’ indicates that the given product name is more likely to refer to the instance that the customer possesses currently (e.g., *my other two pairs of jeans* in (6) and *my oldest pair* in (8)). On the other hand, the possessive pronoun ‘your’ indicates that it is more likely to refer either to a generic object or to the instance with an unknown purchase time (e.g., *your jeans* in (11)). For each pronoun type, its value is set to either true or false.

Keywords for an instance or an entire product: instance/class/brand/model

While keywords such as ‘pair’ or ‘item’ (e.g., (4)) are somewhat likely to indicate temporal information of the given product name, keywords such as ‘product’ or the product category ‘jeans’ or ‘pants’ are less likely to indicate such temporal information (e.g., (11)). In addition, the brand name and the model name, for example, ‘Levis’ and ‘501’, respectively, in the name *Levis 501*, can be utilized as keywords for an entire product or an instance. We set four feature values of instance, class (an entire product), brand name and model name to either true or false.

Argument type:

(object, subject, complement, object of preposition)

Our intuition is that opinion sentences are mostly expressed with either adjectives or descriptive product names so that copular verbs are frequently used, whereas experience sentences are mostly expressed with general verbs. This suggests that some types of argument of verbs such as complement or subject serve to describe a particular instance or generic object (e.g., (9), (10)). On the other hand, the object term of a verb (e.g., ‘wear’, ‘purchase’ or ‘possess’) denotes or refers to the instance with temporal information (e.g., (3), (7)). We use one of the categories as the feature value for the argument type.

Table 3 summarizes the term-based temporal features discussed in this section.

| Type | Sub types: values | Examples |
|----------------------------------|---|-------------------------------------|
| Temporal cue words (cue) | recent purchase-cue/past purchase cue: {true, false} | new (1b), currently (7), oldest (8) |
| Quantity-based cue words (quant) | cardinal: {one, two, several}/ordinal: {first, second, nth} | 4 (4), second (5a) |
| DET/JJ (co-refer) | this/the/a(n)/other/a nother: {true, false} | other (1a), these (1b) |
| PRP\$ (co-refer) | my/your/his/her/the ir: {true, false} | my (8); your (11) |
| Instance/Product (co-refer) | instance/class/brand/model: {true, false} | pair, (4), jeans (11) |
| Argument type (arg) | argument type: {object, subject, complement, object of preposition} | compl (9), subj (10), obj (4) |

Table 3: Product name-based Temporal Features

To detect the syntactic features discussed above, such as tense, aspect, polarity, argument types of product names and syntactic types of verb, we utilized the dependency parse tree from the Stanford parser (Klein and Manning, 2003). For time expressions, we employed the rule-based time expression tagger⁴ which covers time expressions according to the TIMEX2 2001 guidelines. For the semantic types of verb, we manually collected frequent verbs related to purchase and emotional expressions.

⁴ http://fofoca.mitre.org/taggers/timex2_taggers.html

5 Experiment

5.1 Experimental Setup

We used annotated product names for the classification experiment since our main focus is on classifying the product names into suitable temporal classes. For comparison, we conducted an experiment on product name extraction based on a parser with a regular grammar (RegexpParser in the NLTK; Bird and Loper, 2004). For the experiment, we also utilized predefined product name patterns for pruning unrelated candidates from the NP chunks. We achieved the F1 score of 88.1%. We believe that the performance can be improved further by bootstrapping the patterns, but this process is left as future work.

We employed the LIBSVM toolkit with RBF kernel for both stages of product name classification (Chang and Lin, 2011). We used annotated product names for the experiment. As Table 1 shows, the distribution of the product names in the temporal class is skewed. To improve the performance with such a skewed class distribution, we incorporated cost-sensitive learning for the second stage of the classification (McCarthy et al., 2005). We empirically varied the penalty cost factors for minor classes (P_p and $P_{r\&p}$).

We used the prediction accuracy, precision, recall and F1 scores for each class by 10-fold cross validation for the first stage. For the second stage, we used the geometric mean (G-mean), which is known as a good indicator for the performance as it is independent of the data distribution between classes (Kubat and Matwin, 1997). The G-mean score can be calculated as follows.

| Actual \ Predicted | Positive | Negative |
|--------------------|----------|----------|
| Positive | TP | FN |
| Negative | FP | TN |

$$ACC^+ = \frac{TP}{TP + FN} \quad ACC^- = \frac{TN}{FP + TN}$$

$$G\text{-mean} = \sqrt{(ACC^+ \cdot ACC^-)}$$

For the comparison with the previous work on product entity recognition, we carefully implemented a class sequential pattern mining (CSR) method (Ding et al., 2009). We considered the adjectives as cue words such as ‘new’, ‘previous’ and the nouns for the product (e.g., pants) and the instance (e.g., pair). The length of

the sequence is 11. The mined patterns cut by the threshold 0.01, 0.005, 0.001, 0.0005, 0.0001 are 5, 10, 35, 52, and 23 respectively from the overall 895 patterns. The example pattern are ‘DT ENT/NNS’ (threshold: 0.01; e.g., the jeans) and ‘DT ENT/NNS IN NN NNS’ (threshold: 0.001; e.g., these jeans in size 34x30). However, such patterns are not so promising for our purpose because of the following reasons. First, the high-score patterns are very short (the length is less than 3) and simple, so newly discovered names must be short as well. In fact, the pattern ‘DT ENT/NNS IN NN NNS’ is more helpful than ‘DT ENT/NNS’ in spite of its lower support score. One of the reasons why shorter patterns get high score is the CSR depends on frequency when generating a new pattern from the current pattern. Second, for our purpose we split the keyword sets into two sub sets for each subclass (P_r and P_p). However, due to the small amount of sequence data sets for the P_p , the minded patterns are not effective. Thus, we argue that the mined patterns by CSR are not that effective for product name distinction.

Instead, we employed the CRF++ toolkit⁵ for a CRF-based classifier that has been popular for the named entity recognition (NER) task. The CRF-based classifier was also compared with Ding and colleagues’ work (2009). We regarded the co-refer features in Table 3 as the term-based baseline feature sets.

5.2 Classification Results

Table 5 shows the first stage of the classification result. The system achieved the best averaged accuracy 79.0% (ANOVA, $F(6,63) = 6.03$; $p = .000$). The best F1 scores for the temporal class and the atemporal class are 80.2% (ANOVA, $F(6,63) = 7.14$; $p = .000$) and 77.4% (ANOVA, $F(6,63) = 5.43$; $p = .000$), respectively. As for the second stage of the classification, the best G-mean scores for the P_r and P_p subclasses, and the P_r and $P_{r\&p}$ subclasses are 0.74 (ANOVA, $F(7,72) = 4.17$; $p = .001$) and 0.82 (ANOVA, $F(6,63) = 2.43$; $p = .035$) when the costs are set to 4 and 5, respectively, as shown in Figure 1. The best combination of features for classifying into the P_p subclass is different from that for classifying into the $P_{r\&p}$ subclass. This suggests that while the contribution of time expressions and event features

is critical to distinguishing product names in the P_r subclass from those in the $P_{r\&p}$ subclass, time expressions and term-based features are critical to distinguishing product names in the P_r subclass from those in the P_p subclass. Overall, combining time expressions, event-based features and term-based features is found to enhance the performance of temporal cue identification and temporal instance distinction.

Table 4 shows the classification results by the CRF-based classifier. The scores for the P_p subclass or $P_{r\&p}$ class are quite low. These results imply that the temporal information-based features are crucial to such subclasses. This also happens to the SVM-based classifier with the same feature sets (Term-based Base) as shown in Table 5 and Figure 1.

| Class\Score | P | R | F1 |
|-------------|------|------|------|
| P_r | 60.9 | 59.4 | 60.1 |
| P_p | 27.8 | 9.6 | 14.0 |
| $P_{r\&p}$ | 35.7 | 13.4 | 19.4 |
| Atemporal | 61.8 | 48.8 | 54.4 |
| Overall | 59.9 | 48.2 | 53.4 |

Table 4: The classification results by the CRF-based classifier.

6 Error Analysis and Discussion

We analyzed errors from each stage of classification and listed major errors related to temporal features as follows.

Misclassified by dominant temporal features:

The dominant temporal features may trigger misclassification into atemporal class as shown in (14) and (15).

- (14) Levis have been around forever and will continue to be because it is a great product.
- (15) Starting almost a year ago it is an absolute fact that *these jeans* no longer hold up for years the way they used to.

Although present perfect tense and simple past time expression were detected, the names do not refer to the particular product instance that the customer of the review purchased. In this case, either the argument or the syntactic type of a verb should be taken into account with more weight.

⁵ <http://crfpp.googlecode.com/>

| Feature set | Temporal class | | | Atemporal class | | | Acc. |
|---|----------------|-------------|-------------|-----------------|-------------|-------------|-------------|
| | P | R | F1 | P | R | F1 | |
| Base (timex + event features: v1, v2) | 72.3 | 65.6 | 68.6 | 65.1 | 71.5 | 68.0 | 68.4 |
| Base + cue | 74.9 | 66.8 | 70.4 | 66.6 | 74.3 | 70.1 | 70.3 |
| Base + cue + quant | 74.6 | 67.9 | 70.9 | 67.2 | 73.6 | 70.0 | 70.6 |
| Base + cue + quant + arg | 71.1 | 75.4 | 73.0 | 70.5 | 65.1 | 67.4 | 70.6 |
| Base + cue + quant + arg + co-refer | 80.5 | 80.6 | 80.2 | 78.3 | 77.1 | 77.4 | 79.0 |
| Term-based base (co-refer) | 68.3 | 76.1 | 71.8 | 69.4 | 59.7 | 63.7 | 68.4 |
| Timex + term-based (co-refer + cue + quant) | 74.3 | 76.3 | 75.2 | 72.2 | 69.7 | 70.8 | 73.2 |

Table 5: The classification results for temporal cue identification

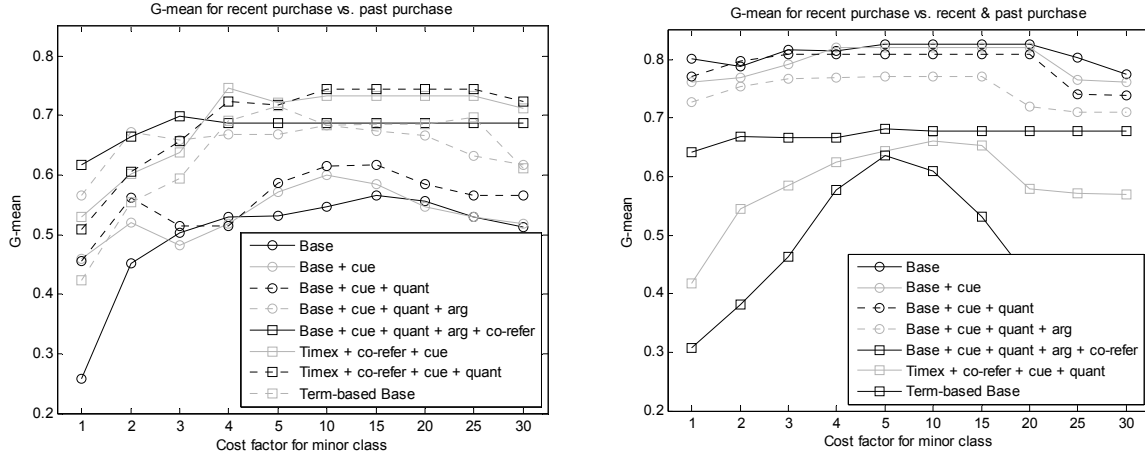


Figure 1: The classification results for temporal instance distinction (G-mean)

Contrastive relation between past purchase and recent purchase in the same sentence: The present work used the feature values extracted for each product name only for the given product name. However, the following case of errors as shown in (16) would be handled properly if the feature values for one product name are shared with its adjacent product name. Contrastive words (e.g., ‘new’ vs. ‘worn out’) and the syntactic structure of the sentence (e.g., ‘infinitive’) indicate that two given names may refer to different instances.

(16) To sum it up, these were *some new pants* to replace *some worn out ones*.

Chained coreference relation: We did not consider the coreference relation between adjacent product names due to the rarity of such cases. However, if the first mentioned product name has a coreference relation with all the following product names as shown in (17), the classification for each name may not be meaningful.

(17) Until this week, I had been wearing *levis* all my life and in recent years was only wearing *501s* for all occasions. Currently I have over *10 pairs* that have never been worn (...) The labels all match in size however *some of the*

jeans are at least a full size smaller at the waist and *some of the pairs* have the correct waist but very narrow legs.

In order to deal with contrastive relations and coreference relations properly, we may have to incorporate the pair or cluster-based term classification model that shares features among mentioned expressions, determining the degree of which is left as future work.

7 Applications

7.1 Sentiment Change over Time

An old customer with long-term experiences sometimes expresses her sentiment change over product reviews as shown in Example (18). While customer A in (18a) expresses her sentiment change on the given product, customer B in (18b) reports that his sentiment on the product has been positive.

(18) a. My husband has *other Levis 501s* and *these new ones* are different in the weight of fabric (light), fit (tighter in the leg and crotch) and color of stitching (white). (...) They are made in Mexico, *his older ones* are made in

Colombia (...) We are not happy with *these pants* but have already washed and used them.

b. I have been wearing *501s* since I can remember. These are just as good as *my original ones*. (...) hey they are *501s* hard to go wrong with these.

In (18a), we see that the sentiment towards the past purchase (e.g., *other Levis 501s, his older ones*) is positive but that the one towards the recent purchase (e.g., *these new ones, these pants*) is negative. Based on this difference in sentiment with respect to product instance, we can identify such sentiment change expressed in a product review by simple heuristic rules. For example, if the major polarity towards P_r is ‘Negative’ and the major polarity towards P_p is ‘Positive’ in a given review we classify it as the review with a sentiment change ‘positive to negative’.

In order to apply some heuristic rules to sentiment change identification, we performed product-wise sentiment detection (cf. Min and Park, 2012). As for target detection, we utilized our results of product name classification. As for polarity classification, we utilized the ‘compositionality-based polarity propagation’ method (Moilanen and Pulman, 2007; Min and Park, 2011). As for target-sentiment association, in order to determine whether each candidate target is associated with the detected sentiment, we applied the association rules in order to prevent a generic object from being associated with the sentiment (cf. Min and Park, 2012).

Table 6 shows the classification results in the same data sets as used in product name classification. We believe that these results can be utilized to cluster customer reviews in a novel way to help customers make their decision more wisely on re-purchase as well as on their first purchase.

| Sentiment Change | P | R | F |
|------------------|------|------|------|
| PtoN | 0.69 | 0.48 | 0.56 |
| PtoP | 0.37 | 0.91 | 0.53 |
| No change | 0.96 | 0.89 | 0.92 |

Table 6: The classification results of sentiment change⁶

⁶ We annotated a sentiment change with the code ‘ $S(P_p)$ to $S(P_r)$ ’, where $S(P_p)$ is the sentiment toward the past purchase, $S(P_r)$ the sentiment towards the recent purchase, and P denotes positive, and N negative. The inter-annotator agreement is 0.734 (Kappa; $p < .001$). The reason why ‘NtoP’

After analyzing the major errors, we observed that sentiment detection regarding target product names at the instance level is crucial to the class PtoN. However, we also realized that detecting the existence of the product names in the $P_{r\&p}$ subclass is also significant for the class of PtoP, (i.e., 28% of the instances require the detection of such clues for correct classification).

7.2 Review Rating with Enhanced Credibility

Based on the results of identifying the sentiment change, we implemented an automatic review rating system. The system assigns +1 to the clause/sentence in the ‘Positive’ class and -1 to the clause/sentence in the ‘Negative’ class with respect to a product instance. It then calculates the total score for the rating by applying positive weights to the ‘Positive’ class if the sentiment is maintained (i.e., PtoP) and negative weights to the ‘Negative’ class if the sentiment is changed (i.e., PtoN). We compared the performance of our system with that of the baseline system. The baseline system calculates the rating based on the detected sentiment in each clause without utilizing the results of the product instance distinction or the sentiment change identification. We showed two ratings calculated by both systems to 6 evaluators and asked them to choose the more credible rating. We tested 140 reviews of 7 products (20 reviews of each product; we randomly selected 20 reviews from the reviews of each product except the review cases where the two ratings are even.). The ratings calculated by our system are chosen more often than the ratings by the baseline system. Overall, the ratings by our system were preferred (statistically significant at the level 0.001). We believe that our rating system is considered more credible because of product instance distinction and sentiment change identification.

8 Conclusion

In this paper, we proposed linguistically meaningful novel features for classifying product names at the instance level in customer reviews about a particular product with respect to the

and ‘NtoN’ categories are missing is that we found only one or two examples classified as in these categories from the sample reviews. We think that the customers with such experiences tend to express their opinions in the forum sites rather than in online shopping web sites. We leave this issue as future work.

instance's temporal information. We formulated the problem as a classification problem with respect to the existence of temporal cues and the recency of the purchase time. The results show that combining time expressions, event-based features and term-based features does enhance the performance of product name classification with a statistical significance. Two applications, 'sentiment change identification' and 'automatic review rating', also show that the results of product instance classification are useful. For future work, we will impose further constraints against penalizing minority. We will also look into effective clues for handling more complex cases such as contrastive and coreference relations.

Acknowledgements

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No.2011-0018262), and in part by the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Knowledge Economy of Korea. We thank the three anonymous reviewers for helpful comments and insightful suggestions.

References

Steven Bird, and Edward Loper. 2004. NLTK: The Natural Language Toolkit, In *Proceedings of the ACL demonstration session*, pages 214-217.

Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3): 1-27, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

Xiaowen Ding, Bing Liu, and Lei Zang. 2009. Entity discovery and assignment for opinion mining applications, In *Proceeding of KDD'09*, Paris, France.

Nithin Jindal and Bing Liu. 2006. Mining Comparative Sentences and Relations, In *Proceedings of AAAI*, pages 1331-1336.

Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. In *Proceedings of the 41th ACL*, pages 423-430.

Anup Kumar Kolya, Asif Ekbal, and Sivaji Bandyopadhyay. 2010. JU_CSE_TEMP: A First Step towards Evaluating Events, Time Expressions and Temporal Relations, In *Proceedings of SemEval-2010*, pages 345-350, Uppsala, Sweden.

Miroslav Kubat and Stan Matwin. 1997. Addressing the curse of imbalanced data sets: One-sided sampling. In *Proceedings of the 14th ICML*, pages 179-186.

Shasha Li, Chin-Yew Lin, Young-In Song, and Zhoujun Li. 2010. Comparable Entity Mining from Comparative Questions, In *Proceedings of the 48th ACL*, pages 650-658, Uppsala, Sweden.

Kate McCarthy, Bibi Zabar, and Gary Weiss. 2005. Does Cost-Sensitive Learning Beat Sampling for Classifying Rare Classes?, In *Proceedings of UDBM'05*, pages 69-77, Chicago, Illinois, USA.

Hye-Jin Min and Jong C. Park. 2011. Detecting and Blocking False Sentiment Propagation, In *Proceedings of IJCNLP*, pages 354-362.

Hye-Jin Min and Jong C. Park. 2012. Identifying Helpful Reviews Based on Customer's Mentions about Experiences, *Expert Systems With Applications*, 39(15): 11830-11838, Elsevier.

Hye-Jin Min and Jong C. Park. 2012. Product-wise Sentiment Detection for Sentiment Change Identification (draft).

Karo Moilanen and Stephen Pulman. 2007. Sentiment Composition, In *Proceedings of RANLP*, pages 378-382, Borovets, Bulgaria.

Keun Chan Park, Yoonjae Jeong, and Sung Hyoon Myaeng. 2010. Detecting Experience from Weblogs, In *Proceedings of the 48th ACL*, pages 1464-1472.

James Pustejovsky, Jose Castano, Robert Ingria, Roser Sauri, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text, *the 5th International Workshop on Computational Semantics*.

Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4): 521-544.

Naushad UzZaman and James F. Allen. 2010. TRIPS and TRIOS System for TempEval-2: Extracting Temporal Information from Text, In *Proceedings of SemEval-2010*, pages 276-283, Uppsala, Sweden.

Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Jessica Moszkowicz, and James Pustejovsky. 2009. The tempeval challenge: identifying temporal relations in text. *Language Resources and Evaluation*, 43(2): 161-179.

Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. SemEval-2010 Task 13: TempEval-2, In *Proceedings of the 5th SemEval-2010*, pages 57-52, Uppsala, Sweden.