Learning Extraction Patterns for Subjective Expressions*

Ellen Riloff
School of Computing
University of Utah
Salt Lake City, UT 84112
riloff@cs.utah.edu

Janyce Wiebe
Department of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260
wiebe@cs.pitt.edu

Abstract

This paper presents a bootstrapping process that learns linguistically rich extraction patterns for subjective (opinionated) expressions. High-precision classifiers label unannotated data to automatically create a large training set, which is then given to an extraction pattern learning algorithm. The learned patterns are then used to identify more subjective sentences. The bootstrapping process learns many subjective patterns and increases recall while maintaining high precision.

1 Introduction

Many natural language processing applications could benefit from being able to distinguish between factual and subjective information. Subjective remarks come in a variety of forms, including opinions, rants, allegations, accusations, suspicions, and speculations. Ideally, information extraction systems should be able to distinguish between factual information (which should be extracted) and non-factual information (which should be discarded or labeled as uncertain). Question answering systems should distinguish between factual and speculative answers. Multi-perspective question answering aims to present multiple answers to the user based upon speculation or opinions derived from different sources. Multi-document summarization systems need to summarize different opinions and perspectives. Spam filtering systems must recognize rants and emotional tirades, among other things. In general, nearly any system that seeks to identify information could benefit from being able to separate factual and subjective information.

Some existing resources contain lists of subjective words (e.g., Levin's *desire* verbs (1993)), and some empirical methods in NLP have automatically identified adjectives, verbs, and N-grams that are statistically associated with subjective language (e.g., (Turney, 2002; Hatzivassiloglou and McKeown, 1997; Wiebe, 2000; Wiebe et al., 2001)). However, subjective language can be exhibited by a staggering variety of words and phrases. In addition, many subjective terms occur infrequently, such as strongly subjective adjectives (e.g., *preposterous*, *unseemly*) and metaphorical or idiomatic phrases (e.g., *dealt a blow*, *swept off one's feet*). Consequently, we believe that subjectivity learning systems must be trained on extremely large text collections before they will acquire a subjective vocabulary that is truly broad and comprehensive in scope.

To address this issue, we have been exploring the use of bootstrapping methods to allow subjectivity classifiers to learn from a collection of unannotated texts. Our research uses high-precision subjectivity classifiers to automatically identify subjective and objective sentences in unannotated texts. This process allows us to generate a large set of labeled sentences automatically. The second emphasis of our research is using extraction patterns to represent subjective expressions. These patterns are linguistically richer and more flexible than single words or N-grams. Using the (automatically) labeled sentences as training data, we apply an extraction pattern learning algorithm to automatically generate patterns representing subjective expressions. The learned patterns can be used to automatically identify more subjective sentences, which grows the training set, and the entire process can then be bootstrapped. Our experimental results show that this bootstrapping process increases the recall of the high-

*This work was supported by the National Science Foundation under grants IIS-0208798, IIS-0208985, and IRI-9704240. The data preparation was performed in support of the Northeast Regional Research Center (NRRC) which is sponsored by the Advanced Research and Development Activity (ARDA), a U.S. Government entity which sponsors and promotes research of import to the Intelligence Community which includes but is not limited to the CIA, DIA, NSA, NIMA, and NRO.

precision subjective sentence classifier with little loss in precision. We also find that the learned extraction patterns capture subtle connotations that are more expressive than the individual words by themselves.

This paper is organized as follows. Section 2 discusses previous work on subjectivity analysis and extraction pattern learning. Section 3 overviews our general approach, describes the high-precision subjectivity classifiers, and explains the algorithm for learning extraction patterns associated with subjectivity. Section 4 describes the data that we use, presents our experimental results, and shows examples of patterns that are learned. Finally, Section 5 summarizes our findings and conclusions.

2 Background

2.1 Subjectivity Analysis

Much previous work on subjectivity recognition has focused on document-level classification. For example, (Spertus, 1997) developed a system to identify inflammatory texts and (Turney, 2002; Pang et al., 2002) developed methods for classifying reviews as positive or negative. Some research in genre classification has included the recognition of subjective genres such as editorials (e.g., (Karlgren and Cutting, 1994; Kessler et al., 1997; Wiebe et al., 2001)).

In contrast, the goal of our work is to classify individual sentences as subjective or objective. Document-level classification can distinguish between "subjective texts", such as editorials and reviews, and "objective texts," such as newspaper articles. But in reality, most documents contain a mix of both subjective and objective sentences. Subjective texts often include some factual information. For example, editorial articles frequently contain factual information to back up the arguments being made, and movie reviews often mention the actors and plot of a movie as well as the theatres where it's currently playing. Even if one is willing to discard subjective texts in their entirety, the objective texts usually contain a great deal of subjective information in addition to facts. For example, newspaper articles are generally considered to be relatively objective documents, but in a recent study (Wiebe et al., 2001) 44% of sentences in a news collection were found to be subjective (after editorial and review articles were removed).

One of the main obstacles to producing a sentence-level subjectivity classifier is a lack of training data. To train a document-level classifier, one can easily find collections of subjective texts, such as editorials and reviews. For example, (Pang et al., 2002) collected reviews from a movie database and rated them as positive, negative, or neutral based on the rating (e.g., number of stars) given by the reviewer. It is much harder to obtain collections of individual sentences that can be easily identified as subjective or objective. Previous work on sentence-level subjectivity classification (Wiebe et al., 1999) used training corpora that had been manually annotated for subjectivity. Manually producing annotations is time consuming, so the amount of available annotated sentence data is relatively small.

The goal of our research is to use high-precision subjectivity classifiers to automatically identify subjective and objective sentences in unannotated text corpora. The high-precision classifiers label a sentence as subjective or objective when they are confident about the classification, and they leave a sentence unlabeled otherwise. Unannotated texts are easy to come by, so even if the classifiers can label only 30% of the sentences as subjective or objective, they will still produce a large collection of labeled sentences. Most importantly, the high-precision classifiers can generate a much larger set of labeled sentences than are currently available in manually created data sets.

2.2 Extraction Patterns

Information extraction (IE) systems typically use lexico-syntactic patterns to identify relevant information. The specific representation of these patterns varies across systems, but most patterns represent role relationships surrounding noun and verb phrases. For example, an IE system designed to extract information about hijackings might use the pattern *hijacking of* <*x*>, which looks for the noun *hijacking* and extracts the object of the preposition *of* as the hijacked vehicle. The pattern <*x*> *was hijacked* would extract the hijacked vehicle when it finds the verb *hijacked* in the passive voice, and the pattern <*x*> *hijacked* would extract the hijacker when it finds the verb *hijacked* in the active voice.

One of our hypotheses was that extraction patterns would be able to represent subjective expressions that have noncompositional meanings. For example, consider the common expression *drives (someone) up the wall*, which expresses the feeling of being annoyed with something. The meaning of this expression is quite different from the meanings of its individual words (*drives*, *up*, *wall*). Furthermore, this expression is not a fixed word sequence that could easily be captured by N-grams. It is a relatively flexible construction that may be more generally represented as <*x*> *drives* <*y*> *up the wall*, where *x* and *y* may be arbitrary noun phrases. This pattern would match many different sentences, such as "George drives me up the wall," "She drives the mayor up the wall," or "The nosy old man drives his quiet neighbors up the wall."

We also wondered whether the extraction pattern representation might reveal slight variations of the same verb or noun phrase that have different connotations. For example, you can say that *a comedian bombed last night*, which is a subjective statement, but you can't express this sentiment with the passive voice of *bombed*. In Section 3.2, we will show examples of extraction patterns representing subjective expressions which do in fact exhibit both of these phenomena.

A variety of algorithms have been developed to automatically learn extraction patterns. Most of these algorithms require special training resources, such as texts annotated with domain-specific tags (e.g., AutoSlog (Riloff, 1993), CRYSTAL (Soderland et al., 1995), RAPIER (Califf, 1998), SRV (Freitag, 1998), WHISK (Soderland, 1999)) or manually defined keywords, frames, or object recognizers (e.g., PALKA (Kim and Moldovan, 1993) and LIEP (Huffman, 1996)). AutoSlog-TS (Riloff, 1996) takes a different approach, requiring only a corpus of unannotated texts that have been separated into those that are related to the target domain (the "relevant" texts) and those that are not (the "irrelevant" texts). Most recently, two bootstrapping algorithms have been used to learn extraction patterns. Meta-bootstrapping (Riloff and Jones, 1999) learns both extraction patterns and a semantic lexicon using unannotated texts and seed words as input. ExDisco (Yangarber et al., 2000) uses a bootstrapping mechanism to find new extraction patterns using unannotated texts and some seed patterns as the initial input.

For our research, we adopted a learning process very similar to that used by AutoSlog-TS, which requires only relevant texts and irrelevant texts as its input. We describe this learning process in more detail in the next section.

3 Learning and Bootstrapping Extraction Patterns for Subjectivity

We have developed a bootstrapping process for subjectivity classification that explores three ideas: (1) high-precision classifiers can be used to automatically identify subjective and objective sentences from unannotated texts, (2) this data can be used as a training set to automatically learn extraction patterns associated with subjectivity, and (3) the learned patterns can be used to grow the training set, allowing this entire process to be bootstrapped.

Figure 1 shows the components and layout of the bootstrapping process. The process begins with a large collection of unannotated text and two high precision subjectivity classifiers. One classifier searches the unannotated corpus for sentences that can be labeled as subjective with high confidence, and the other classifier searches for sentences that can be labeled as objective with high confidence. All other sentences in the corpus are left unlabeled. The labeled sentences are then fed to an extraction pattern learner, which produces a set of extraction patterns that are statistically correlated with the subjective sentences (we will call these the *subjective patterns*). These patterns are then used to identify more sentences within the unannotated texts that can be classified as subjective. The extraction pattern learner can then retrain using the larger training set and the process repeats. The subjective patterns can also be added to the high-precision subjective sentence classifier as new features to improve its performance. The dashed lines in Figure 1 represent the parts of the process that are bootstrapped.

In this section, we will describe the high-precision sentence classifiers, the extraction pattern learning process, and the details of the bootstrapping process.

3.1 High-Precision Subjectivity Classifiers

The high-precision classifiers (HP-Subj and HP-Obj) use lists of lexical items that have been shown in previous work to be good subjectivity clues. Most of the items are single words, some are N-grams, but none involve syntactic generalizations as in the extraction patterns. Any data used to develop this vocabulary does not overlap with the test sets or the unannotated data used in this paper.

Many of the subjective clues are from manually developed resources, including entries from (Levin, 1993; Ballmer and Brennenstuhl, 1981), Framenet lemmas with frame element *experiencer* (Baker et al., 1998), adjectives manually annotated for polarity (Hatzivassiloglou and McKeown, 1997), and subjectivity clues listed in (Wiebe, 1990). Others were derived from corpora, including subjective nouns learned from unannotated data using bootstrapping (Riloff et al., 2003).

The subjectivity clues are divided into those that are strongly subjective and those that are weakly subjective, using a combination of manual review and empirical results on a small training set of manually annotated data. As the terms are used here, a strongly subjective clue is one that is seldom used without a subjective meaning, whereas a weakly subjective clue is one that commonly has both subjective and objective uses.

The high-precision subjective classifier classifies a sentence as subjective if it contains two or more of the strongly subjective clues. On a manually annotated test set, this classifier achieves 91.5% precision and 31.9% recall (that is, 91.5% of the sentences that it selected are subjective, and it found 31.9% of the subjective sentences in the test set). This test set consists of 2197 sentences, 59% of which are subjective.

The high-precision objective classifier takes a different approach. Rather than looking for the presence of lexical items, it looks for their absence. It classifies a sentence as objective if there are no strongly subjective clues and at most one weakly subjective clue in the current, previous, and next sentence combined. Why doesn't the objective classifier mirror the subjective classifier, and consult its own list of strongly objective clues? There are certainly lexical items that are statistically correlated with the *ob-*

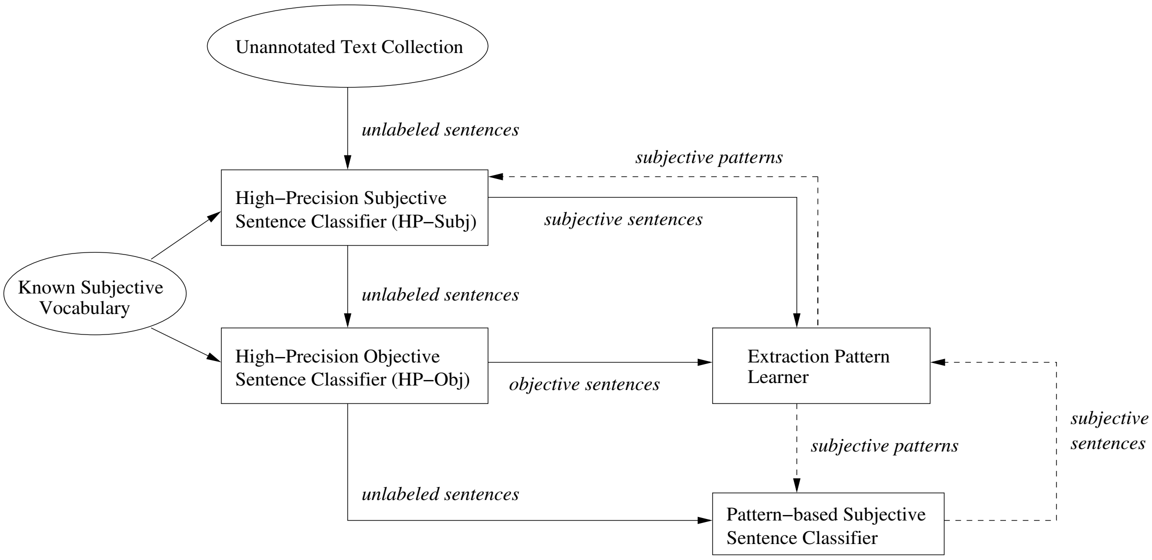

Figure 1: Bootstrapping Process

jective class (examples are cardinal numbers (Wiebe et al., 1999), and words such as *per*, *case*, *market*, and *total*), but the presence of such clues does not readily lead to high precision objective classification. Add sarcasm or a negative evaluation to a sentence about a dry topic such as stock prices, and the sentence becomes subjective. Conversely, add objective topics to a sentence containing two strongly subjective words such as *odious* and *scumbag*, and the sentence remains subjective.

The performance of the high-precision objective classifier is a bit lower than the subjective classifier: 82.6% precision and 16.4% recall on the test set mentioned above (that is, 82.6% of the sentences selected by the objective classifier are objective, and the objective classifier found 16.4% of the objective sentences in the test set). Although there is room for improvement, the performance proved to be good enough for our purposes.

3.2 Learning Subjective Extraction Patterns

To automatically learn extraction patterns that are associated with subjectivity, we use a learning algorithm similar to AutoSlog-TS (Riloff, 1996). For training, AutoSlog-TS uses a text corpus consisting of two distinct sets of texts: "relevant" texts (in our case, subjective sentences) and "irrelevant" texts (in our case, objective sentences). A set of syntactic templates represents the space of possible extraction patterns.

The learning process has two steps. First, the syntactic templates are applied to the training corpus in an exhaustive fashion, so that extraction patterns are generated for (literally) every possible instantiation of the templates that appears in the corpus. The left column of Figure 2 shows the syntactic templates used by AutoSlog-TS. The right column shows a specific extraction pattern that was learned during our subjectivity experiments as an instantiation of the syntactic form on the left. For example, the pattern *<subj> was satisfied*[1] will match any sentence where the verb *satisfied* appears in the passive voice. The pattern *<subj> dealt blow* represents a more complex expression that will match any sentence that contains a verb phrase with head=*dealt* followed by a direct object with head=*blow*. This would match sentences such as "The experience dealt a stiff blow to his pride." It is important to recognize that these patterns look for specific syntactic constructions produced by a (shallow) parser, rather than exact word sequences.

SYNTACTIC FORM	EXAMPLE PATTERN
<subj> passive-verb	<subj> was satisfied
<subj> active-verb	<subj> complained
<subj> active-verb dobj	<subj> dealt blow
<subj> verb infinitive	<subj> appear to be
<subj> aux noun	<subj> has position
active-verb <dobj>	endorsed <dobj>
infinitive <dobj>	to condemn <dobj>
verb infinitive <dobj>	get to know <dobj>
noun aux <dobj>	fact is <dobj>
noun prep <np>	opinion on <np>
active-verb prep <np>	agrees with <np>
passive-verb prep <np>	was worried about <np>
infinitive prep <np>	to resort to <np>

Figure 2: Syntactic Templates and Examples of Patterns that were Learned

[1]This is a shorthand notation for the internal representation.

PATTERN	FREQ	%SUBJ
<subj> was asked	11	100%
<subj> asked	128	63%
<subj> is talk	5	100%
talk of <np>	10	90%
<subj> will talk	28	71%
<subj> put an end	10	90%
<subj> put	187	67%
<subj> is going to be	11	82%
<subj> is going	182	67%
was expected from <np>	5	100%
<subj> was expected	45	42%
<subj> is fact	38	100%
fact is <dobj>	12	100%

Figure 3: Patterns with Interesting Behavior

The second step of AutoSlog-TS's learning process applies all of the learned extraction patterns to the training corpus and gathers statistics for how often each pattern occurs in subjective versus objective sentences. AutoSlog-TS then ranks the extraction patterns using a metric called RlogF (Riloff, 1996) and asks a human to review the ranked list and make the final decision about which patterns to keep.

In contrast, for this work we wanted a fully automatic process that does not depend on a human reviewer, and we were most interested in finding patterns that can identify subjective expressions with high precision. So we ranked the extraction patterns using a conditional probability measure: the probability that a sentence is subjective given that a specific extraction pattern appears in it. The exact formula is:

$$Pr(subjective \mid pattern_i) = \frac{subjfreq(pattern_i)}{freq(pattern_i)}$$

where $subjfreq(pattern_i)$ is the frequency of $pattern_i$ in subjective training sentences, and $freq(pattern_i)$ is the frequency of $pattern_i$ in all training sentences. (This may also be viewed as the precision of the pattern on the training data.) Finally, we use two thresholds to select extraction patterns that are strongly associated with subjectivity in the training data. We choose extraction patterns for which $freq(pattern_i) \geq \theta_1$ and $Pr(subjective \mid pattern_i) \geq \theta_2$.

Figure 3 shows some patterns learned by our system, the frequency with which they occur in the training data (**FREQ**) and the percentage of times they occur in subjective sentences (**%SUBJ**). For example, the first two rows show the behavior of two similar expressions using the verb *asked*. 100% of the sentences that contain *asked* in the passive voice are subjective, but only 63% of the sentences that contain *asked* in the active voice are subjective. A human would probably not expect the active and passive voices to behave so differently. To understand why this is so, we looked in the training data and found that the passive voice is often used to query someone about a specific opinion. For example, here is one such sentence from our training set: "Ernest Bai Koroma of RITCORP was asked to address his supporters on his views relating to 'full blooded Temne to head APC'." In contrast, many of the sentences containing *asked* in the active voice are more general in nature, such as "The mayor asked a newly formed JR about his petition."

Figure 3 also shows that expressions using *talk* as a noun (e.g., "Fred is the talk of the town") are highly correlated with subjective sentences, while *talk* as a verb (e.g., "The mayor will talk about...") are found in a mix of subjective and objective sentences. Not surprisingly, longer expressions tend to be more idiomatic (and subjective) than shorter expressions (e.g., *put an end (to)* vs. *put*; *is going to be* vs. *is going*; *was expected from* vs. *was expected*). Finally, the last two rows of Figure 3 show that expressions involving the noun *fact* are highly correlated with subjective expressions! These patterns match sentences such as *The fact is...* and *... is a fact*, which apparently are often used in subjective contexts. This example illustrates that the corpus-based learning method can find phrases that might not seem subjective to a person intuitively, but that are reliable indicators of subjectivity.

4 Experimental Results

4.1 Subjectivity Data

The text collection that we used consists of English-language versions of foreign news documents from FBIS, the U.S. Foreign Broadcast Information Service. The data is from a variety of countries. Our system takes unannotated data as input, but we needed annotated data to evaluate its performance. We briefly describe the manual annotation scheme used to create the gold-standard, and give interannotator agreement results.

In 2002, a detailed annotation scheme (Wilson and Wiebe, 2003) was developed for a government-sponsored project. We only mention aspects of the annotation scheme relevant to this paper. The scheme was inspired by work in linguistics and literary theory on *subjectivity*, which focuses on how opinions, emotions, etc. are expressed linguistically in context (Banfield, 1982). The goal is to identify and characterize expressions of *private states* in a sentence. *Private state* is a general covering term for opinions, evaluations, emotions, and speculations (Quirk et al., 1985). For example, in sentence (1) the writer is expressing a negative evaluation.

(1) "The time has come, gentlemen, for Sharon, the assassin, to realize that injustice cannot last long."

Sentence (2) reflects the private state of Western countries. Mugabe's use of *overwhelmingly* also reflects a private state, his positive reaction to and characterization of his victory.

(2) "Western countries were left frustrated and impotent after Robert Mugabe formally declared that he had overwhelmingly won Zimbabwe's presidential election."

Annotators are also asked to judge the strength of each private state. A private state may have *low*, *medium*, *high* or *extreme* strength.

To allow us to measure interannotator agreement, three annotators (who are not authors of this paper) independently annotated the same 13 documents with a total of 210 sentences. We begin with a strict measure of agreement at the sentence level by first considering whether the annotator marked any private-state expression, of any strength, anywhere in the sentence. If so, the sentence is subjective. Otherwise, it is objective. The average pairwise percentage agreement is 90% and the average pairwise κ value is 0.77.

One would expect that there are clear cases of objective sentences, clear cases of subjective sentences, and borderline sentences in between. The agreement study supports this. In terms of our annotations, we define a sentence as borderline if it has at least one private-state expression identified by at least one annotator, and all strength ratings of private-state expressions are *low*. On average, 11% of the corpus is borderline under this definition. When those sentences are removed, the average pairwise percentage agreement increases to 95% and the average pairwise κ value increases to 0.89.

As expected, the majority of disagreement cases involve low-strength subjectivity. The annotators consistently agree about which are the clear cases of subjective sentences. This leads us to define the gold-standard that we use when evaluating our results. A sentence is *subjective* if it contains at least one private-state expression of medium or higher strength. The second class, which we call *objective*, consists of everything else.

4.2 Evaluation of the Learned Patterns

Our pool of unannotated texts consists of 302,163 individual sentences. The HP-Subj classifier initially labeled roughly 44,300 of these sentences as subjective, and the HP-Obj classifier initially labeled roughly 17,000 sentences as objective. In order to keep the training set relatively balanced, we used all 17,000 objective sentences and 17,000 of the subjective sentences as training data for the extraction pattern learner.

17,073 extraction patterns were learned that have $frequency \geq 2$ and $Pr(subjective \mid pattern_i) \geq .60$ on the training data. We then wanted to determine whether the extraction patterns are, in fact, good indicators of subjectivity. To evaluate the patterns, we applied different subsets of them to a test set to see if they consistently occur in subjective sentences. This test set consists of 3947 sentences, 54% of which are subjective.

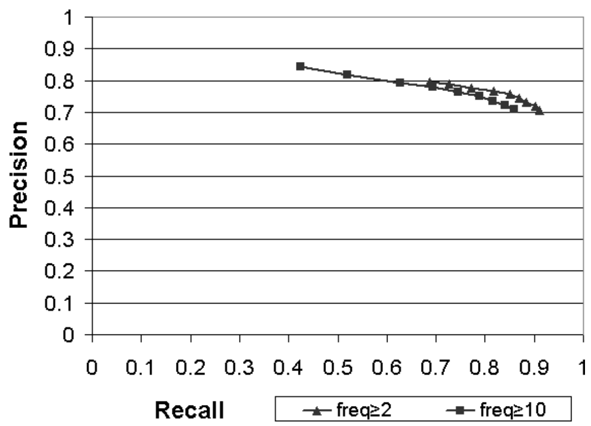

Figure 4: Evaluating the Learned Patterns on Test Data

Figure 4 shows sentence recall and pattern (instance-level) precision for the learned extraction patterns on the test set. In this figure, precision is the proportion of pattern instances found in the test set that are in subjective sentences, and recall is the proportion of subjective sentences that contain at least one pattern instance.

We evaluated 18 different subsets of the patterns, by selecting the patterns that pass certain thresholds in the training data. We tried all combinations of $\theta_1 = \{2,10\}$ and $\theta_2 = \{.60,.65,.70,.75,.80,.85,.90,.95,1.0\}$. The data points corresponding to θ_1=2 are shown on the upper line in Figure 4, and those corresponding to θ_1=10 are shown on the lower line. For example, the data point corresponding to θ_1=10 and θ_2=.90 evaluates only the extraction patterns that occur at least 10 times in the training data and with a probability $\geq$ .90 (i.e., at least 90% of its occurrences are in subjective training sentences).

Overall, the extraction patterns perform quite well. The precision ranges from 71% to 85%, with the expected tradeoff between precision and recall. This experiment confirms that the extraction patterns are effective at recognizing subjective expressions.

4.3 Evaluation of the Bootstrapping Process

In our second experiment, we used the learned extraction patterns to classify previously unlabeled sentences from the unannotated text collection. The new subjective sentences were then fed back into the Extraction Pattern Learner to complete the bootstrapping cycle depicted by the rightmost dashed line in Figure 1. The Pattern-based Subjective Sentence Classifier classifies a sentence as subjective if it contains at least one extraction pattern with $\theta_1 \geq 5$ and $\theta_2 \geq 1.0$ on the training data. This process produced approximately 9,500 new subjective sentences that were previously unlabeled.

Since our bootstrapping process does not learn new objective sentences, we did not want to simply add the new subjective sentences to the training set, or it would become increasingly skewed toward subjective sentences. Since HP-Obj had produced roughly 17,000 objective sentences used for training, we used the 9,500 new subjective sentences along with 7,500 of the previously identified subjective sentences as our new training set. In other words, the training set that we used during the second bootstrapping cycle contained exactly the same objective sentences as the first cycle, half of the same subjective sentences as the first cycle, and 9,500 brand new subjective sentences.

On this second cycle of bootstrapping, the extraction pattern learner generated many new patterns that were not discovered during the first cycle. 4,248 new patterns were found that have $\theta_1 \geq 2$ and $\theta_2 \geq .60$. If we consider only the strongest (most subjective) extraction patterns, 308 new patterns were found that had $\theta_1 \geq 10$ and $\theta_2 \geq 1.0$. This is a substantial set of new extraction patterns that seem to be very highly correlated with subjectivity.

An open question was whether the new patterns provide additional coverage. To assess this, we did a simple test: we added the 4,248 new patterns to the original set of patterns learned during the first bootstrapping cycle. Then we repeated the same analysis that we depict in Figure 4. In general, the recall numbers increased by about 2-4% while the precision numbers decreased by less, from 0.5-2%.

In our third experiment, we evaluated whether the learned patterns can improve the coverage of the high-precision subjectivity classifier (HP-Subj), to complete the bootstrapping loop depicted in the top-most dashed line of Figure 1. Our hope was that the patterns would allow more sentences from the unannotated text collection to be labeled as subjective, without a substantial drop in precision. For this experiment, we selected the learned extraction patterns that had $\theta_1 \geq 10$ and $\theta_2 \geq 1.0$ on the training set, since these seemed likely to be the most reliable (high precision) indicators of subjectivity.

We modified the HP-Subj classifier to use extraction patterns as follows. All sentences labeled as subjective by the original HP-Subj classifier are also labeled as subjective by the new version. For previously unlabeled sentences, the new version classifies a sentence as subjective if (1) it contains two or more of the learned patterns, or (2) it contains one of the clues used by the original HP-Subj classifier and at least one learned pattern. Table 1 shows the performance results on the test set mentioned in Section 3.1 (2197 sentences) for both the original HP-Subj classifier and the new version that uses the learned extraction patterns. The extraction patterns produce a 7.2 percentage point gain in coverage, and only a 1.1 percentage point drop in precision. This result shows that the learned extraction patterns do improve the performance of the high-precision subjective sentence classifier, allowing it to classify more sentences as subjective with nearly the same high reliability.

HP-Subj		**HP-Subj w/Patterns**	
Recall	*Precision*	*Recall*	*Precision*
32.9	91.3	40.1	90.2

Table 1: Bootstrapping the Learned Patterns into the High-Precision Sentence Classifier

Table 2 gives examples of patterns used to augment the HP-Subj classifier which do not overlap in non-function words with any of the clues already known by the original system. For each pattern, we show an example sentence from our corpus that matches the pattern.

5 Conclusions

This research explored several avenues for improving the state-of-the-art in subjectivity analysis. First, we demonstrated that high-precision subjectivity classification can be used to generate a large amount of labeled training data for subsequent learning algorithms to exploit. Second, we showed that an extraction pattern learning technique can learn subjective expressions that are linguistically richer than individual words or fixed phrases. We found that similar expressions may behave very differently, so that one expression may be strongly indicative of subjectivity but the other may not. Third, we augmented our original high-precision subjective classifier with these newly learned extraction patterns. This bootstrapping process resulted in substantially higher recall with a minimal loss in precision. In future work, we plan to experiment with different configurations of these classifiers, add new subjective language learners in the bootstrapping process, and address the problem of how to identify new objective sentences during bootstrapping.

6 Acknowledgments

We are very grateful to Theresa Wilson for her invaluable programming support and help with data preparation.

References

C. Baker, C. Fillmore, and J. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the COLING-ACL-98*.

T. Ballmer and W. Brennenstuhl. 1981. *Speech Act Classification: A Study in the Lexical Analysis of English Speech Activity Verbs*. Springer-Verlag.

A. Banfield. 1982. *Unspeakable Sentences*. Routledge and Kegan Paul, Boston.

seems to be <dobj>	I am pleased that there now **seems to be** broad political consensus ...
underlined <dobj>	Jiang's subdued tone . . . **underlined** his desire to avoid disputes ...
pretext of <np>	On the **pretext of** the US opposition . . .
atmosphere of <np>	Terrorism thrives in an **atmosphere of** hate . . .
<subj> reflect	These are fine words, but they do not **reflect** the reality . . .
to satisfy <dobj>	The pictures resemble an attempt **to satisfy** a primitive thirst for revenge . . .
way with <np>	... to ever let China use force to have its **way with** ...
bring about <np>	"Everything must be done by everyone to **bring about** de-escalation," Mr Chirac added.
expense of <np>	at the **expense of** the world's security and stability
voiced <dobj>	Khatami ... **voiced** Iran's displeasure.
turn into <np>	... the surging epidemic could **turn into** "a national security threat," he said.

Table 2: Examples of Learned Patterns Used by HP-Subj and Sample Matching Sentences

M. E. Califf. 1998. *Relational Learning Techniques for Natural Language Information Extraction*. Ph.D. thesis, Tech. Rept. AI98-276, Artificial Intelligence Laboratory, The University of Texas at Austin.

Dayne Freitag. 1998. Toward General-Purpose Learning for Information Extraction. In *Proceedings of the ACL-98*.

V. Hatzivassiloglou and K. McKeown. 1997. Predicting the Semantic Orientation of Adjectives. In *Proceedings of the ACL-EACL-97*.

S. Huffman. 1996. Learning information extraction patterns from examples. In Stefan Wermter, Ellen Riloff, and Gabriele Scheler, editors, *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, pages 246–260. Springer-Verlag, Berlin.

J. Karlgren and D. Cutting. 1994. Recognizing Text Genres with Simple Metrics Using Discriminant Analysis. In *Proceedings of the COLING-94*.

B. Kessler, G. Nunberg, and H. Schütze. 1997. Automatic Detection of Text Genre. In *Proceedings of the ACL-EACL-97*.

J. Kim and D. Moldovan. 1993. Acquisition of Semantic Patterns for Information Extraction from Corpora. In *Proceedings of the Ninth IEEE Conference on Artificial Intelligence for Applications*.

Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.

B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proceedings of the EMNLP-02*.

R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman, New York.

E. Riloff and R. Jones. 1999. Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping. In *Proceedings of the AAAI-99*.

E. Riloff, J. Wiebe, and T. Wilson. 2003. Learning Subjective Nouns using Extraction Pattern Bootstrapping. In *Proceedings of the Seventh Conference on Computational Natural Language Learning (CoNLL-03)*.

E. Riloff. 1993. Automatically Constructing a Dictionary for Information Extraction Tasks. In *Proceedings of the AAAI-93*.

E. Riloff. 1996. Automatically Generating Extraction Patterns from Untagged Text. In *Proceedings of the AAAI-96*.

S. Soderland, D. Fisher, J. Aseltine, and W. Lehnert. 1995. CRYSTAL: Inducing a Conceptual Dictionary. In *Proceedings of the IJCAI-95*.

S. Soderland. 1999. Learning Information Extraction Rules for Semi-Structured and Free Text. *Machine Learning*, 34(1-3):233–272.

E. Spertus. 1997. Smokey: Automatic Recognition of Hostile Messages. In *Proceedings of the IAAI-97*.

P. Turney. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the ACL-02*.

J. Wiebe, R. Bruce, and T. O'Hara. 1999. Development and Use of a Gold Standard Data Set for Subjectivity Classifications. In *Proceedings of the ACL-99*.

J. Wiebe, T. Wilson, and M. Bell. 2001. Identifying Collocations for Recognizing Opinions. In *Proceedings of the ACL-01 Workshop on Collocation: Computational Extraction, Analysis, and Exploitation*.

J. Wiebe. 1990. *Recognizing Subjective Sentences: A Computational Investigation of Narrative Text*. Ph.D. thesis, State University of New York at Buffalo.

J. Wiebe. 2000. Learning Subjective Adjectives from Corpora. In *Proceedings of the AAAI-00*.

T. Wilson and J. Wiebe. 2003. Annotating Opinions in the World Press. In *Proceedings of the ACL SIGDIAL-03*.

R. Yangarber, R. Grishman, P. Tapanainen, and S. Huttunen. 2000. Automatic Acquisiton of Domain Knowledge for Information Extraction. In *Proceedings of COLING 2000*.