# Investigation of sentence structure in domain adaptation for sentiment classification

Anthony Gentile

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2013

Reading Committee:

Gina-Anne Levow, Chair

Michael Tjalve

Program Authorized to Offer Degree:
Computational Linguistics

University of Washington

**Abstract**


Investigation of sentence structure in domain adaptation for sentiment classification

Anthony Gentile

Chair of the Supervisory Committee:
Gina-Anne Levow
Linguistics

A popular use case of computational linguistics is the identification of sentiment in text. Many current methods for sentiment classification focus on word features within sentences of a text. These methods employ different mathematical and computational techniques to achieve increasing accuracies. Additionally, these techniques are being applied to domain adaptation for sentiment classification which allow sentiment classifiers to be even more flexible. This thesis intends to show the relevance of sentence structure in combination with word features for determining sentiment and the benefits to be seen in domain adaptation contexts. By using part-of-speech (POS) representations for sentences in the Amazon product reviews dataset we find that there is useful sentiment information to be gleaned from sentence structures. This information can be subsequently used by classifiers to improve sentiment classification accuracies.

# TABLE OF CONTENTS

# LIST OF EQUATIONS

## LIST OF TABLES

# GLOSSARY

CLASSIFIER: A computational tool that inputs data and assigns it to a set of categories.

CORPORA: Large, structured sets of textual data, usually in an electronic format.

DOMAIN ADAPTATION: The process of applying a classifier trained on one distribution of data to another. Most work in this field aims to keep cross-domain accuracy high.

HYPERPLANE: A geometrical plane that cuts a N-dimensional space into two halves.

N-GRAM: A sequence of the textual items in a text with a length of *n*. Given the example sentence "The cat died", unigrams for this text would be "The" "cat" and "died". Bigrams for this text would be "The cat" and "cat died".

PART-OF-SPEECH: A syntactic class of words such as *adjective* or *noun* which is defined by its syntactic functionality.

POLARITY: The positive or negative character of a word, sentence, text or idea expressed in a language.

SENTIMENT: An opinion regarding a particular situation, event or thing. Often labeled as positive, negative or neutral when discussing sentiment classification.

## ACKNOWLEDGMENTS

I am incredibly grateful to my family, fiancé, and the professors in the Computational Linguistics program at the University of Washington. Without their time and support, I would not have been able to complete this work. I would especially like to thank my adviser, Dr. Gina-Anne Levow, for being a wonderful instructor and always being in bright spirits.

## DEDICATION

To my family and fiancé, Breanne, who puts up with me when I am being a "cranky-pants".

# Chapter 1

# INTRODUCTION

As the information stored on computers throughout the world grows, many want to derive usefulness from this mass amount of data for different applications. One such use case where computational linguistics comes into play is in determining whether the attitude of what people are talking about online is positive or negative. The outcome of this sentiment detection can be particularly interesting to companies that wish to know what people are saying about a particular product that they make. Another example beneficiary would be election candidates wanting to know how a particular campaign is being received by the public. This thesis will delve into the inner workings of sentiment classification, highlight some current approaches to the problem, how it may be improved by analyzing sentence structure, and how the information gleaned from sentence structures is beneficial in a domain adaptation scenario.

## *1.1 Background for Determining Sentiment*

How does one go about determining whether or not a sentence, paragraph or an entire text is negative or positive? One might start by looking at the text and checking if there are words with *positive* or *negative* connotation associated with them. This connotation is determined by a particular society within a language group. In English, some example positive and negative words are *fantastic*, *awesome*, *awful*, and *terrible*. However, the word *brilliant* may have a different connotation if the speaker if from the United Sates versus England. In the United States, an English speaker would often use the word brilliant to describe something or someone of great intelligence, whereas in England, brilliant is common slang for something that is outstanding. Once it has been determined what the *positive* and *negative* words should be for the text in question one could then count up these words into separate negative and positive tallies. If the tally for negative is greater, the sentence could be classified as negative, likewise for positive, and if they are the same, we could consider it to be of neutral polarity.

This would be a good start and sentiment classifiers have been built on exactly this premise, employing various machine learning methods, such as Naive Bayes, maximum entropy, and Support Vector Machines [8]. The accuracy of these classifiers do a good job, but allow for several false positive cases. Take for example the following movie review: "If they would've focused more on the plot, that would have been a great movie!". This example, while having the word *great* in it, we know to be of negative sentiment. There are many other phrases one could utter that use positive or negative words in abundance but actually convey the opposite sentiment of these words. Sometimes this can be detected by looking for negation words such as *not* and analyzing sentence trees, but overall, it can be a difficult task.

The problem becomes even more complicated when looking beyond the sentence level into paragraphs. Some sentences might convey positive sentiment while others convey negative sentiment. How is the sentiment of the paragraph as a whole determined in this case? Do particular sentences and their classified sentiment outweigh other sentences in the same paragraph? Do speakers tend to convey overall sentiment near the end of a text? Is this language specific? For example, take another example movie review: "I don't know why the theater was out of popcorn, but my friend was really upset about it. This put us in a bad mood and then the people in front of us kept talking. Despite all that, the crude humor and being slow at times, I loved it!". We can see by this example that if we are just counting positive and negative words on a sentence level, this text would likely be classified as negative, but in reality the last part of the last sentence, "I loved it !", indicates that the movie itself received positive sentiment.

This also brings into question the target of sentiment in a text. In the previous example, positive sentiment was expressed to a particular movie, while negative sentiment was expressed about the theater experience. Being the director of the movie in question or the manager of the theater would determine which target of sentiment you might care about. Another difficulty with identifying sentiment is determining the role of parts of speech and how they affect the sentiment in a sentence. For example, there are different uses for the word hate in "This movie is about a hate crime" versus "I hate this movie". The first does not convey negative sentiment, but a part-of-speech (POS) tagger and sentiment classifier may have trouble correctly tagging and classifying it. To solve this problem, one could use a method called *word sense disambiguation* where POS tags are added to word features during training and testing of a classifier. This method was used by Pang et al. (2002) while Chesley et al. (2006) and Nicholls and Song (2009) leveraged other techniques with POS tags to improve sentiment classification.

We see that language allows us to be very expressive and flexible in the way we convey ideas. The movie review example helps show how it can be difficult and increasingly complex to implement machine learning techniques to classify sentiment with accuracy comparable to that of a native speaker of the language.

## 1.2 Domain Specific Knowledge

In addition to specific word features, we may also look for key phrases in the context which we are talking about, or the domain. If we are talking about the movie reviews domain then phrases such as "two thumbs up" and "left the theater" give further indication as to the sentiment of the text. If, however, we saw these phrases in a different domain, such as pet supplies, they may not provide the same boost in accuracy when classifying the sentiment. Researchers have delved into helping keep the accuracy strong when applying a sentiment classifier trained on one domain to another. This is known as domain adaptation and some interesting techniques have been researched.

## 1.3 The Purpose of This Thesis

While it has been discussed that determining sentiment can be difficult and encapsulates many different ideas and techniques for obtaining high accuracies, there is one technique that we have yet to speak about and which doesn't come up in the literature as often. This has to do with the structure of the sentences in the text. What might they tell us about sentiment? Do people tend to say positive things in certain ways and likewise negative things in other ways? If so, then this information would be quite useful alongside the other techniques in determining sentiment. Perhaps sentence structure is helpful in different ways depending on the domain. These are the questions that this thesis intends to answer and to show how some shallow processing on POS representations of sentences can provide insight into the benefit of using sentence structure to improve single domain and multiple domain sentiment classification.

# Chapter 2

# LITERATURE REVIEW

In this chapter, I will review pertinent publications regarding sentiment classification and domain adaptation as it pertains to the topic of this thesis and the techniques described in the introduction.

## *2.1 Initial Techniques for Sentiment Classification*

Pang et al. (2002)[21] describes the problem of sentiment detection very well and how different machine learning methods including Naive Bayes (NB)[23][22], maximum entropy (MaxEnt)[3], and support vector machines (SVMs)[8] compare to one another. Pang et al. (2002) showed SVMs winning out compared to the other two methods. It was also shown how the use of unigrams versus bigrams in sentiment classification can produce differing accuracies and in the case of movie reviews, unigrams obtained higher accuracies. We see in further publications the favored use of Support Vector Machines for the sentiment classification problem instead of other methods. For my experiments, I will also test these three machine learning methods to see if SVMs are truly the best way to go. This publication also forms my reasoning for providing multiple distributions of n-grams in my experiments. By experimenting with all these combinations, I hope to show how sentiment classification can be quite domain specific concerning classifier accuracies.

Nicholls and Song (2009)[19] optimize weights for POS categories to improve the accuracy of sentiment classifiers. This effectively would boost a word feature if the corresponding POS category was adjectives and likewise for other categories. This is an interesting approach as it hints at the importance of looking more into the syntax instead of just word features. This publication helped steer me into the idea of looking at sentence structures in regard to sentiment analysis.

Derczynski et al. (2013)[9] delve into the POS tagging side of sentiment classification as many sentiment analysis approaches factor in parts of speech in some manner. They focus on the unique dataset that is Twitter. The text found in Twitter is quite different from other internet sources as it is restricted to 140 characters and often contains references to other Twitter users or hash tagged topics. Derczynski et al. (2013) propose a POS tagger model that caters to Twitter and helps improve accuracy on the POS tagging side. This particularly helps stress the importance of the data you are trying to classify and how improvement to POS tagging can subsequently result in improved accuracies for the sentiment classifiers that leverage POS tagging in some manner. This publication helped me review the model I was using for the Stanford Part-Of-Speech Tagger[26] and sparked my interest for future experiments using different POS tagging models.

## 2.2 Classification Methods

Naive Bayes is a probabilistic classification method that leverages Bayes' theorem as shown in Equation 2.1 which computes the probability of $A$ given $B$.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Equation 2.1: Bayes' Theorem

The main distinguishing feature of this classification method is that it considers features independently from other features in the same class. This method has been seen with criticism as well as support in the machine learning community because the independence of all features in a class is often not an accurate assumption and yet it still tends to perform well. It definitely has a place as a contender for the sentiment classification problem. To compute the probability of a class using Naive Bayes, the formula shown in Equation 2.2 can be used which expresses the probability of class $C$ given a set of features $(F_1, \ldots, F_n)$ where $Z$ is a scaling constant.

$$p(C|F_1, \dots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^{n} p(F_i|C)$$

Equation 2.2: Probability using Naive Bayes

Maximum entropy classification, which is commonly viewed as a form of Occam's razor (when given multiple hypotheses, choose the one with the least amount of assumptions), differs from Naive Bayes in that it allows for a large amount of features but does not assume the features are independent. From the training data, the distribution with the maximum entropy is selected from the features provided. This distribution will have the least amount of assumptions. Maximum entropy is determined using the formula shown in Equation 2.3 where the probability of class $y$ given instance $x$ where $\lambda_i$ are weights for corresponding features and $f_i$ are binary feature functions indicating the occurrence of a feature in instance $x$ with label $y$ and $Z$ is a normalization term.

$$P(y|x) = \frac{e^{\lambda_0(y) + \Sigma_j \lambda_j f_j(x,y)}}{Z}$$

Equation 2.3: Maximum entropy

Support vector machines[11] are models for detecting patterns in which probabilistic measures are not used, but rather instances of different classes are separated along a geometric hyperplane in an optimal way which produces the largest margin achievable on either side of the hyperplane. The larger the margin between the two classes the better the expected generalization of the classifier on new data. SVMs leverage kernels which are mathematical functions that allow data that is not linearly separable to be placed unto a higher dimensional space, where it can be separated linearly. SVMs are often implemented with a soft margin, allowing some instances to fall on the wrong side of the separating hyperplane. This is a means of handling outliers in the data so that a workable hyperplane can exist.

### 2.3 Domain Adaptation

In 2006, Blitzer et al. (2006) showcased structural correspondence learning (SCL) where pivot features were identified and leveraged on MEDLINE corpora sentences to provide for better accuracies in domain adaptation for sentiment analysis. This led to their work, Blitzer et al. (2007) which introduced the Amazon product reviews dataset. They combined SCL with mutual information to highlight improvement for domain adaptation for sentiment analysis. These works are particularly important as the Amazon product reviews dataset has gone on to become a heavily used dataset for many publications working on domain adaptation for sentiment classification. Such publications include Li and Zong (2008) who created a framework to combine labeled data from various source domains with labeled data from the target domain to improve classifier accuracy. Pan et al. (2010) also used the dataset to improve cross domain accuracy with Spectral Feature Alignment which takes domain independent features and domain dependent features and aligns them to better discover the sentiment of unseen words in the target domain. I will also use this dataset to show how sentence structure may improve sentiment classification in a domain adaptation scenario.

In 2006, Hinton et al. (2006) made progress on deep belief nets which helped progress deep learning architectures forward and spurred publications of deep learning approaches to sentiment analysis. Bengio (2009) reintroduces deep architectures as a promising approach for machine learning which subsequently helps lead approaches to sentiment classification using deep learning techniques. Glorot et al. (2011) leverages these techniques, specifically, Stacked Denoising Autoencoders, using the Amazon product reviews dataset, as his predecessors did, to show promising improvement in a domain adaptation scenario. Glorot et al. (2011) compared the accuracy of their work to the work of Blitzer et al. (2007), Li and Zong (2008), and Pan et al. (2010) to display higher accuracy with their approach on the same dataset. While deep learning approaches appear very promising, they can be complicated to understand. Socher et al. (2013), whose work is most similar to mine, leverages neural nets, another deep learning approach, to improve sentiment classification.

### 2.4 Sentence Structure

When the idea of analyzing sentence structures for sentiment classification came to my mind, I found that Awais (2011) had a similar concept and used sentence substructures to apply sentiment classification to citations. I found this approach interesting, but not quite on par with what I wanted to discover about sentence structures. It wasn't until after I carried out my experiments that Socher et al. (2013) published their work, which is most similar to mine. It shows how finer grained sentence structure representations, leveraging neural nets, can be used to improve sentiment analysis. These sentence structures and substructures can be used to recognize the sentiment of a sentence as a whole, instead of relying on the words occurring in particular clauses. This is particularly useful for the example edge cases I described in the introduction. While our works are similar, mine is a much more shallow approach and serves to show the importance of sentence structure in sentiment analysis as a whole, whereas Socher et al. (2013) goes deeper into solving how sentence structures can be manipulated to detect sentiment accurately.

# Chapter 3
# METHODOLOGY

In this chapter, I will describe steps taken to produce part-of-speech (POS) sentence representations from the Amazon product reviews dataset that will form the basis of the experiments.

## ***3.1 The Dataset***

I chose to use the Amazon product reviews dataset which contains 38,548 reviews for products spanning across 25 different domains, including books, DVDs, electronics, and video games. This data was introduced by Blitzer et al. (2007) and subsequently used by Li and Zong (2008), Pan et al. (2010) and Glorot et al. (2011) in their work with domain adaptation for sentiment classification. By using this data set, I was able to leverage a large amount of labeled text categorized into separate domains.

## ***3.2 Parts of Speech***

To determine whether sentence structures can be beneficial to sentiment classification we must first obtain these structures. While one might go the route of working with tree structures that contain parent-child nodes as produced by tools like the Stanford Parser[15], I wanted to look at more shallow representations that would be computationally cheaper to produce and work with. I decided to use the Stanford Part-Of-Speech Tagger[26] to obtain POS representations for sentences. By applying the Stanford POS Tagger to the following sentence "This movie is great!" it would produce the corresponding POS tag pairs: "This/DT movie/NN is/VBZ great/JJ". These tags identify the following parts of speech: DT - Determiner, NN - Noun, VBZ - Verb, 3rd person singular present, and JJ - Adjective. Table 3.1 describes all the POS tags that are a part of the Penn Treebank Project[17]. Once the POS tags are obtained for each sentence in the reviews, we can then form a simple sentence representation by concatenating these tags together with underscores. Using the previous "This/DT movie/NN is/VBZ great/JJ" example, the sentence representation would be DT_NN_VBZ_JJ. These representations could be very long depending on the how long the sentence is. For the experiments in this thesis, the representations are not truncated in any way except for the symbolic POS tags that are removed as described in section 4.1.

| Tag | Description |
|---|---|
| CC | Coordinating conjunction |
| CD | Cardinal number |
| DT | Determiner |
| EX | Existential there |
| FW | Foreign word |
| IN | Preposition or subordinating conjunction |
| JJ | Adjective |
| JJR | Adjective, comparative |
| JJS | Adjective, superlative |
| LS | List item marker |
| MD | Modal |
| NN | Noun, singular or mass |
| NNS | Noun, plural |
| NNP | Proper noun, singular |
| NNPS | Proper noun, plural |
| PDT | Predeterminer |
| POS | Possessive ending |
| PRP | Personal pronoun |
| PRP$ | Possessive pronoun |
| RB | Adverb |
| RBR | Adverb, comparative |
| RBS | Adverb, superlative |
| RP | Particle |
| SYM | Symbol |
| TO | to |
| UH | Interjection |
| VB | Verb, base form |
| VBD | Verb, past tense |
| VBG | Verb, gerund or present participle |
| VBN | Verb, past participle |
| VBP | Verb, non-3rd person singular present |
| VBZ | Verb, 3rd person singular present |
| WDT | Wh-determiner |
| WP | Wh-pronoun |
| WP$ | Possessive wh-pronoun |
| WRB | Wh-adverb |

Table 3.1: Penn Treebank Project - Part-Of-Speech Tags

### *3.3 Representation Statistics*

From the 38,548 reviews contained in the dataset, 233,411 sentence representations were obtained, 198,115 of them being unique. Of the 233,411 structures, 103,351 of them came from positive labeled reviews (44.28%) while 130,060 of them came from negative labeled reviews (55.72%)

### *3.4 Structure Counts*

Once the sentence structures for all the sentences in all the reviews have been obtained it becomes trivial to count them up. In doing so, some interesting information presents itself. The top twenty representations in respect to polarity are shown in Table 3.2 and Table 3.3. The bold faced rows indicate representations that have a 2:1 ratio or better for their given polarities. What this tells us is that there are particular sentence structures that tend to happen more frequently in positive reviews than in negative reviews and vice versa. In some cases the ratio is very high, such as PRP_RB_VB_DT_NN, which is derived from sentences like "I highly recommend this product". It occurred 83 times in positive reviews and only once in a negative review. This is important information when we want to determine sentiment. If we run across this particular sentence structure we may want to consider it more likely that the review in question is positive. It is the job of the classifier to perform these probabilistic calculations. The mutual information (MI) is also shown for the representations in Table 3.2 and Table 3.3 and is a measure of mutual dependence for these representations and the sentences in the respective positive or negative labeled reviews. The formula for mutual information is shown in Equation 3.1 where information $I$ that $X$ and $Y$ share is computed.

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right),$$

Equation 3.1: Mutual Information

| **POS Rep.** | **# Positive** | **# Negative** | **MI** | **Positive Review Example** | **Negative Reviews Example** |
|---|---|---|---|---|---|
| **JJ NN** | 414 | 169 | 0.0005273877 | Great product. | Big mistake. |
| **DT NN VBZ JJ** | 258 | 99 | 0.0003547609 | This movie is great. | This game is horrible. |
| **PRP VBP DT NN** | 224 | 38 | 0.0005954746 | I love this product! | I hate this product. |
| **RB VBN** | 176 | 50 | 0.0003290579 | Highly recommended! | Very disappointed. |
| **DT VBZ DT JJ NN** | 165 | 42 | 0.0003375984 | This is an excellent magazine. | This is a terrible product. |
| RB JJ | 157 | 198 | 0.0000000013 | Very satisfied. | Very disappointing. |
| **PRP VBP PRP** | 156 | 11 | 0.0005750067 | I love it! | I hate it. |
| **PRP VBZ JJ** | 123 | 45 | 0.0001787133 | It is amazing. | It's pathetic. |
| DT NN VBZ RB JJ | 118 | 83 | 0.0000520275 | The quality is very good. | This game is really bad. |
| **DT JJ NN** | 104 | 36 | 0.0001606259 | An irresistible combination. | A true nightmare. |
| **VB PRP** | 102 | 27 | 0.0002022557 | Thank you. | Forget it! |
| PRP VBZ RB JJ | 91 | 56 | 0.0000568904 | It smells so good. | It seems somewhat deceptive. |
| **PRP VBZ DT JJ NN** | 86 | 21 | 0.0001815437 | It is a definite winner. | It's no big deal. |
| **PRP RB VB DT NN** | 83 | 1 | 0.0003881358 | I highly recommend this product | They totally misrepresent this product. |
| **PRP VBP RB JJ IN DT NN** | 80 | 14 | 0.0002089947 | I am very pleased with this product. | I am very disappointed in this camera |
| NNP NNP | 80 | 75 | 0.0000103712 | Fiestaware ROCKS! | BIG Dissapointment. |
| **NNP NNP NNP** | 78 | 21 | 0.0001525247 | The Rebecca Review | Amazon Reviewer AKNapa |
| RB JJ NN | 73 | 40 | 0.0000582958 | Very light weight. | Very poor service. |
| **PRP RB VB PRP** | 70 | 0 | 0.0003525604 | I highly recommend it. | |
| **DT NNS VBP JJ** | 67 | 22 | 0.0001093028 | These burgers are unbelievable. | These knives are cheap. |

Table 3.2: Top twenty occurring POS representations in positive reviews

If one looks at the higher ranking MI POS representations for positive polarity such as PRP_VBP_DT_NN, PRP_VBP_PRP, and JJ_NN one sees that in the corresponding example sentences "I love this product", "I love it", and "Excellent product" that one would expect that a VBP or JJ of negative polarity could be interchanged to produce likely sentences of "I hate this product", "I hate it", and "Horrible product". While this does occur, it is interesting that the occurrences happening for positive polarity outweigh the negative occurrences substantially.

| **POS Rep.** | **# Negative** | **# Positive** | **MI** | **Negative Review Example** | **Positive Review Example** |
|---|---|---|---|---|---|
| RB JJ | 198 | 157 | 0.0000583426 | Very disappointing. | Very comfortable. |
| JJ NN | 169 | 414 | 0.0001784178 | Big mistake! | Excellent product. |
| DT NN VBZ JJ | 99 | 258 | 0.0001285082 | This product is worthless. | This product is great! |
| **WP DT NN** | 92 | 27 | 0.0001670656 | What a disappointment. | What a timesaver. |
| **PRP VBD RB JJ** | 86 | 30 | 0.0001317124 | I was very disappointed. | I was very impressed! |
| DT NN VBZ RB JJ | 83 | 118 | 0.0000022571 | The viewfinder is very dim. | The case is very stylish. |
| **VBP RB VB PRP NN** | 82 | 0 | 0.0004130157 | Don't waste your money | |
| NNP NNP | 75 | 80 | 0.0000032625 | BIG Dissapointment. | Fiestaware ROCKS! |
| DT NN | 74 | 56 | 0.0000258241 | No response. | No brainer. |
| **VB PRP NN** | 59 | 13 | 0.0001339620 | Save your money. | Do your research. |
| **VBP RB VB DT NN** | 59 | 3 | 0.0002337426 | Do not buy this product. | Do not miss this book! |
| PRP VBZ RB JJ | 56 | 91 | 0.0000071316 | It is not beautiful. | It is very sturdy. |
| NN NN | 56 | 53 | 0.0000068296 | Caveat emptor | -Michael Cran |
| NNP NN | 52 | 48 | 0.0000074127 | Buyer beware | Grade: A- |
| PRP MD RB VB DT NN | 51 | 44 | 0.0000104422 | I would not recommend this product. | I would highly recommend this product. |
| RB VBN | 50 | 176 | 0.0001498650 | Very disappointed. | Highly recommended. |
| **PRP VBD RB JJ IN DT NN** | 47 | 8 | 0.0001246168 | I was very unhappy with this product. | I was very pleased with this purchase. |
| PRP VBD JJ | 47 | 38 | 0.0000128111 | It was awful! | I was impressed. |
| PRP VBP RB JJ | 46 | 66 | 0.0000014519 | I am so disappointed. | I am very pleased. |
| PRP VBZ JJ | 45 | 123 | 0.0000678121 | It is defective. | It is great! |

Table 3.3: Top twenty occurring POS representations in negative reviews

In the higher ranking MI POS representations for negative polarity such as VBP_RB_VB_PRP_NN and VBP_RB_VB_DT_NN the corresponding example sentences "Don't waste your money" and "Do not buy this product" are imperative sentences (give a request or command). This seemingly indicates that it is likely to see imperative sentences in negative context much more so than imperative sentences in positive context, an example being VB_DT_NN, "Buy this product!".

### *3.5 POS Tagging Discrepancies*

It can be seen that there are cases where the Stanford POS tagger doesn't accurately tag particular words or reveals interesting scenarios. For example, with the sentence structure NNP NNP NNP we likely have a name or signature of the reviewer. It is interesting to see that for this structure it is more likely that the reviews will be positive. Does this indicate that those who leave their name are more likely to give positive reviews? Could it suggest people are being paid or otherwise influenced to produce more positive reviews than negative? With NNP NNP we see that the POS tagger mistakenly tags ROCKS and BIG as proper nouns. We must not forget that we will also find a positive example for a representation in a negative review and vice versa. This is seen with the structure RB VBN for the sentence "Very disappointed" occurring in a positive review. This structure, while having a negative polarity, is outweighed by the positive polarity of the entire text.

### *3.6 Classification Methods*

The classifiers that will be used for the experiments in this thesis use Naive Bayes, maximum entropy, and support vector machines. These classifiers will be used to perform baseline classifications of the reviews using unigrams, bigrams, and unigrams+bigrams together. I will then add in the sentence structures to see how they might affect the accuracies of the classifiers. Once this is achieved, we will target particular sentence structures identified by the bold faced rows in Table 3.2 and Table 3.3. By running the classifiers with a selected set of sentence representations that are more likely to occur and have a stronger leaning to a given polarity we will see if there is perhaps a set of sentence structures that could be applied to domains in a more generic manner to improve accuracy overall.

# Chapter 4

# IMPLEMENTATION

In this chapter, I will describe in more detail the technical process by which the steps to obtaining the results were performed.

## *4.1 Building POS Representations*

When using the Stanford POS tagger, the *english-left3words-distsim.tagger* model was used for the sake of improved speed on larger datasets. In processing these sentences the following POS symbols were discarded: :, ’,’, ., “, $, ”, #, ), (, –, and SYM. I decided that for these experiments punctuation, symbolic, and numerical tags would not be incorporated. This means that the following review sentence: "shipping, to quick to be true." has the corresponding POS tags NN , TO JJ TO VB JJ. The ’,’ POS tag will be discarded to leave NN TO JJ TO VB JJ. By doing this, I hope to obtain a smaller subset of possible POS sentence representations at this preprocessing stage instead of trying to group NN TO JJ TO VB JJ and NN , TO JJ TO VB JJ together in later stages. Once POS symbols are discarded, I simply join the tags with underscores to allow for an easier feature representation for the classifier to consume. The resulting feature for the POS sentence representation being NN_TO_JJ_TO_VB_JJ.

## *4.2 Domain Statistics*

In Table 4.1 one can see the unique unigram and bigram counts for each domain as well as the total unique unigram and bigram counts across all domains. This means that if the unigram *faucet* came up more than once in the *Kitchen & Housewares* it is only counted once in the table. Table 4.1 does not refer to the unique n-grams seen only in a specific domain (if the unigram *faucet* were to occur *only* in *Kitchen & Housewares*).

| **Domain** | **# of unique unigram features** | **# of unique bigram features** |
|---|---|---|
| Apparel | 12,392 | 46,851 |
| Automotive | 7,911 | 26,685 |
| Baby | 15,266 | 69,890 |
| Beauty | 12,877 | 48,193 |
| Books | 36,485 | 137,051 |
| Camera & Photo | 18,017 | 78,936 |
| Cell Phones & Service | 10,447 | 39,602 |
| Computer & Video Games | 21,685 | 86,468 |
| DVD | 38,470 | 143,395 |
| Electronics | 19,418 | 80,590 |
| Gourmet Food | 11,283 | 37,654 |
| Grocery | 10,912 | 39,309 |
| Health & Personal Care | 16,282 | 64,601 |
| Jewelry & Watches | 8,992 | 32,618 |
| Kitchen & Housewares | 16,600 | 69,022 |
| Magazines | 18,599 | 67,084 |
| Music | 31,697 | 110,724 |
| Musical Instruments | 4,925 | 14,530 |
| Office Products | 5,503 | 16,138 |
| Outdoor Living | 11,641 | 44,381 |
| Software | 18,771 | 80,641 |
| Sports & Outdoors | 18,047 | 72,804 |
| Tools & Hardware | 1,498 | 3,359 |
| Toys & Games | 17,595 | 69,756 |
| Video | 32,114 | 114,466 |
| | | |
| **All Domains** | 175,642 | 1,005,396 |

Table 4.1: Unique unigram and bigram counts within each domain

## *4.3 MALLET*

MALLET: MAchine Learning for LanguagE Toolkit[18] is a useful document classifier written in Java leveraging various algorithms such as Naive Bayes, Maximum Entropy and more. It was used for classification in my experiments using Naive Bayes and Maximum Entropy. For the experiments in this paper, MaxEnt classification is performed using a Gaussian prior variance of 1.0.

### 4.4 LIBSVM

LIBSVM[6], was used for the classification using support vector machines performed in the experiments. Similar publications using the Amazon product reviews dataset for sentiment analysis also leverage SVM classifiers. Classification was run with a linear kernel type. LIBSVM uses a sparse format so that zero values do not need to be captured for training files. This can cause training time to be longer, but keeps LIBSVM flexible for sparse cases.

### 4.5 Classification Details

Word features were binarized as opposed to using occurrence counts. Subsequent vector files were made in a binarized fashion, where positive polarity was indicated by 1 and negative polarity indicated by 0. We add unigrams, bigrams, or unigrams+bigrams as features, not capping by any particular frequency. Then additional classifications are performed with the addition of the POS sentence representation features and selected set. Classification in cross-domain scenarios is achieved by training the classifiers, in their different variations, against a particular source domain and then applying those trained classifiers to the test set of a different target domain.

# Chapter 5

# EXPERIMENTS

In my experiments, I wanted to test how consideration of sentence structure may improve a sentiment classifier. To do this I wanted to ensure that no particular classifier or n-gram distribution was favored. The following experiments use three separate classification methods (Naive Bayes, MaxEnt, and SVMs), and three different n-gram distributions (unigrams, bigrams, and unigrams+bigrams). Trigrams were not included as they suffered considerably lower accuracies than unigram, bigrams, or unigrams+bigrams combination.

## *5.1 Baseline*

To produce a baseline, the data was classified using three different methods and with three different n-gram distributions before any addition of POS sentence representations are added. The training data is 80% of the reviews for the respective domains and their positive and negative distributions. The remaining 20% is the test data. The classifiers are trained on the information contained in the training set and then accuracies are gleaned by applying the classifiers to the test set. As an example, the *Books* domain contains 1,000 positive reviews and 1,000 negative reviews. 80% from each set is taken to be used as training data, the remaining 20% for each set will be used as test data. This results in 1,600 reviews for training and 400 reviews for testing.

To provide a means of comparison for in-domain classification accuracies, one can refer to Table 5.1 which shows the in-domain gold standard provided by Li and Zong (2008) whose approach largely mimicked Blitzer et al. (2007). This is a linear classifier using unigrams and bigrams, trained to minimize Huber loss with stochastic gradient descent (Zhang 2004). In comparison, I show the results of the baseline Naive Bayes and SVM linear classifiers using unigrams and bigrams, trained on the same domains but without stochastic gradient descent. I chose not to minimize Huber loss with stochastic gradient descent during training, because, while applicable to linear classifiers such as Naive Bayes and SVMs, it would not be applicable to MaxEnt. I, therefore, decided to keep the baseline consistent across classifiers, even if this meant the SVM domain adaptation baseline would not be as easily comparable to other publications on the same dataset which used the aforementioned approach. Considering the results for Li and Zong (2008) and the Naive Bayes basline classifier are fairly competitive, it would suggest that minimizing Huber loss is not the solitary difference in the approaches and that the baseline for Li and Zong (2008) was constructed with some additional calculations that I did not perform.

| Domain | Li and Zong (2008) with SGD | Baseline Naive Bayes | Baseline SVM |
|---|---|---|---|
| Books | 81.0% | **82.8%** | 72.5% |
| DVDs | **83.7%** | 77.8% | 76.8% |
| Electronics | **84.2%** | 82.3% | 77.8% |
| Kitchen | 84.0% | **85.3%** | 82.5% |

Table 5.1: In-domain classification baseline comparison versus Li and Zong (2008)

In Tables 5.3, 5.4, and 5.5 one can see the resulting accuracies of the classifiers against the dataset. The bold faced cells indicate the best accuracy for the given domain. One will notice that for particular domains, unigrams, bigrams, or a combination of unigrams+bigrams fare better than in other domains.

### *5.2 Addition of POS representations*

Tables 5.3, 5.4, and 5.5 also show the addition of the POS sentence representations. Also of note are the particular improvements when POS representations are added, disregarding the highest accuracy for the domain. For example, in the SVM classification the *Books* domain unigrams were improved by 1% when POS representations were added, even though the highest accuracy was achieved with unigrams+bigrams + POS representations.

### 5.3 Using a Selected Set

After the previous experiment, I asked myself what if I only add sentence structure information to the classifier when it is within the selected set of sentence structures that have at least a 2:1 ratio from the top 20 most occurring representations for their respective polarities, as shown in Table 3.2 and Table 3.3. Table 5.2 reiterates this selected set of 21 representations. The last columns of Tables 5.3, 5.4, and 5.5 show the accuracies when the selected set is incorporated.

| **POS Representation** | **Polarity** |
|---|---|
| JJ_NN | Positive |
| DT_NN_VBZ_JJ | Positive |
| PRP_VBP_DT_NN | Positive |
| RB_VBN | Positive |
| DT_VBZ_DT_JJ_NN | Positive |
| PRP_VBP_PRP | Positive |
| PRP_VBZ_JJ | Positive |
| DT_JJ_NN | Positive |
| VB_PRP | Positive |
| PRP_VBZ_DT_JJ_NN | Positive |
| PRP_RB_VB_DT_NN | Positive |
| PRP_VBP_RB_JJ_IN_DT_NN | Positive |
| NNP_NNP_NNP | Positive |
| PRP_RB_VB_PRP | Positive |
| DT_NNS_VBP_JJ | Positive |
| WP_DT_NN | Negative |
| PRP_VBD_RB_JJ | Negative |
| VBP_RB_VB_PRP_NN | Negative |
| VB_PRP_NN | Negative |
| VBP_RB_VB_DT_NN | Negative |
| PRP_VBD_RB_JJ_IN_DT_NN | Negative |

Table 5.2: Selected set of high occurring 2:1 or greater ratio representations

| **Domain** | **Train** | **Test** | **Unigrams** | **Bigrams** | **Uni+Bi** | **Uni+POS** | **Bi+POS** | **Uni+Bi+POS** | **Uni+SS** | **Bi+SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Apparel | 1600 | 400 | 80.25% | 81.75% | **83.25%** | 79.75% | 81.00% | 83.00% | 80.25% | 81.75% | **83.25%** |
| Automotive | 589 | 147 | 80.27% | **81.63%** | 79.59% | 79.59% | **81.63%** | 79.59% | 80.27% | **81.63%** | 79.59% |
| Baby | 1520 | 380 | 76.58% | 82.37% | 81.05% | 77.37% | **83.42%** | 81.32% | 76.58% | 82.37% | 81.05% |
| Beauty | 1194 | 299 | 74.92% | 78.60% | 73.91% | 72.24% | **78.93%** | 73.24% | 74.92% | 78.60% | 73.91% |
| Books | 1600 | 400 | 79.75% | 79.75% | **82.75%** | 80.00% | 80.25% | **82.75%** | 79.75% | 79.75% | **82.75%** |
| Camera & Photo | 1599 | 400 | 84.75% | **86.50%** | 86.00% | 85.00% | 86.00% | 86.00% | 84.75% | **86.50%** | 86.00% |
| Cell Phones & Service | 818 | 205 | **80.49%** | 80.00% | 77.07% | 76.59% | 80.00% | 76.10% | **80.49%** | 80.00% | 77.07% |
| Computer & Video Games | 1166 | 292 | 94.12% | **96.92%** | **96.92%** | 95.55% | **96.92%** | **96.92%** | 94.18% | **96.92%** | **96.92%** |
| DVD | 1600 | 400 | 74.25% | **78.75%** | 77.75% | 74.75% | 78.25% | 78.50% | 74.25% | **78.75%** | 77.75% |
| Electronics | 1600 | 400 | 80.75% | 79.75% | **82.25%** | 81.25% | 79.75% | **82.25%** | 80.75% | 79.75% | **82.25%** |
| Gourmet Food | 966 | 242 | 83.06% | 85.95% | 85.12% | 83.88% | **86.36%** | 85.12% | 83.06% | 85.95% | 85.12% |
| Grocery | 1082 | 270 | 79.63% | **80.00%** | 78.52% | 77.78% | 78.52% | 78.89% | 79.63% | **80.00%** | 78.52% |
| Health & Personal Care | 1600 | 400 | 84.00% | 83.50% | **85.25%** | 84.25% | 83.00% | **85.25%** | 84.00% | 83.50% | **85.25%** |
| Jewelry & Watches | 1034 | 258 | 78.68% | **80.62%** | 78.29% | 77.13% | 80.23% | 77.91% | 78.68% | **80.62%** | 78.29% |
| Kitchen & Housewares | 1600 | 400 | 83.75% | 84.25% | **85.25%** | 84.00% | 84.50% | 85.00% | 83.75% | 84.25% | **85.25%** |
| Magazines | 1576 | 394 | 81.22% | 79.95% | 81.22% | **81.47%** | 80.46% | **81.47%** | 81.22% | 79.95% | 81.22% |
| Music | 1600 | 400 | 79.00% | 78.25% | 80.00% | **80.75%** | 78.25% | 79.75% | 79.00% | 78.75% | 80.00% |
| Musical Instruments | 265 | 67 | **85.07%** | **85.07%** | **85.07%** | **85.07%** | **85.07%** | **85.07%** | **85.07%** | **85.07%** | **85.07%** |
| Office Products | 345 | 86 | 82.56% | **84.88%** | **84.88%** | 83.72% | **84.88%** | **84.88%** | 82.56% | **84.88%** | **84.88%** |
| Outdoor Living | 1062 | 265 | 76.98% | 76.98% | 76.60% | 76.60% | **77.36%** | 76.60% | 76.98% | 76.98% | 76.60% |
| Software | 1532 | 383 | 83.03% | 82.25% | 83.81% | 82.25% | 82.25% | **84.33%** | 83.03% | 82.25% | 83.81% |
| Sports & Outdoors | 1600 | 400 | 80.50% | 80.25% | **83.25%** | 82.00% | 81.25% | 82.50% | 80.50% | 80.25% | **83.25%** |
| Tools & Hardware | 89 | 23 | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** |
| Toys & Games | 1600 | 400 | 80.75% | 84.75% | 83.00% | 80.75% | **85.00%** | 82.50% | 80.75% | 84.75% | 83.00% |
| Video | 1600 | 400 | **81.75%** | 79.75% | 81.25% | 81.00% | 79.50% | 81.25% | **81.75%** | 79.50% | 81.25% |
| **Overall Accuracy** | | | 80.96% | 82.00% | **82.27%** | 80.92% | 82.02% | 82.20% | 80.96% | 82.01% | **82.27%** |

Table 5.3: Naive Bayes Classifier Accuracies with POS sentence representations and selected set added

For Naive Bayes classification, the unigrams distribution sees very little top accuracies, however, the best increase resulting from the addition of POS representations or selected set is in the *Music* domain with 1.75% when POS representations are added to unigrams. Most improvements and top accuracies are seen in the bigrams and unigrams+bigrams distributions varying when POS representations and selected set is added.

| Domain | Train | Test | Unigrams | Bigrams | Uni+Bi | Uni+POS | Bi+POS | Uni+Bi+POS | Uni+SS | Bi+SS | Uni+Bi+SS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Apparel | 1600 | 400 | 79.50% | 80.25% | 82.00% | 79.50% | 80.00% | **82.50%** | 79.50% | 80.25% | 82.00% |
| Automotive | 589 | 147 | 83.67% | 80.95% | **86.39%** | 84.35% | 80.95% | **86.39%** | 83.67% | 80.95% | **86.39%** |
| Baby | 1520 | 380 | 76.05% | 77.63% | 78.42% | 76.32% | **78.95%** | 78.16% | 76.05% | 77.63% | 78.42% |
| Beauty | 1194 | 299 | 79.26% | 78.60% | 80.60% | **80.94%** | 78.26% | **80.94%** | 79.26% | 78.60% | 80.60% |
| Books | 1600 | 400 | **76.50%** | 73.00% | 76.00% | 76.00% | 73.00% | 75.75% | **76.50%** | 73.00% | 76.00% |
| Camera & Photo | 1599 | 400 | 83.50% | 81.25% | **84.00%** | 83.75% | 81.25% | 83.75% | 83.50% | 81.25% | **84.00%** |
| Cell Phones & Service | 818 | 205 | 79.02% | 75.12% | **83.41%** | 78.54% | 75.12% | 82.93% | 79.02% | 75.12% | **83.41%** |
| Computer & Video Games | 1166 | 292 | 97.26% | 97.26% | **97.60%** | 97.26% | 97.26% | **97.60%** | 97.26% | 97.26% | **97.60%** |
| DVD | 1600 | 400 | **79.25%** | 74.50% | 78.00% | 79.00% | 75.00% | 78.25% | **79.25%** | 74.50% | 78.00% |
| Electronics | 1600 | 400 | **79.75%** | 74.75% | 78.75% | 79.50% | 75.00% | 78.75% | **79.75%** | 74.75% | 78.75% |
| Gourmet Food | 966 | 242 | **87.19%** | 85.12% | 86.78% | **87.19%** | 85.12% | 86.78% | **87.19%** | 85.12% | 86.78% |
| Grocery | 1082 | 270 | **85.93%** | 79.63% | 84.44% | 85.56% | 80.00% | 84.07% | **85.93%** | 79.63% | 84.44% |
| Health & Personal Care | 1600 | 400 | 82.00% | 80.50% | **84.50%** | 81.50% | 81.25% | **84.50%** | 82.00% | 80.50% | **84.50%** |
| Jewelry & Watches | 1034 | 258 | 86.43% | 84.50% | **87.60%** | 85.66% | 84.11% | 87.21% | 86.43% | 84.50% | **87.60%** |
| Kitchen & Housewares | 1600 | 400 | 83.25% | 81.00% | **85.50%** | 82.75% | 81.00% | 84.75% | 83.25% | 81.00% | **85.50%** |
| Magazines | 1576 | 394 | 80.20% | 81.22% | **82.49%** | 80.20% | 80.71% | 82.23% | 80.20% | 81.29% | **82.49%** |
| Music | 1600 | 400 | **80.25%** | 74.75% | 79.75% | 80.00% | 74.00% | 79.25% | **80.25%** | 74.75% | 79.50% |
| Musical Instruments | 265 | 67 | **86.57%** | 85.07% | 85.07% | **86.57%** | 85.07% | 85.08% | **86.57%** | 85.07% | 85.07% |
| Office Products | 345 | 86 | **86.05%** | 84.88% | 84.88% | **86.05%** | 84.88% | 84.88% | **86.05%** | 84.88% | 84.88% |
| Outdoor Living | 1062 | 265 | **81.89%** | 78.49% | 81.51% | **81.89%** | 77.74% | 81.51% | **81.89%** | 78.49% | 81.51% |
| Software | 1532 | 383 | 82.51% | 78.33% | **83.03%** | 81.98% | 77.55% | 82.77% | 82.51% | 78.33% | **83.03%** |
| Sports & Outdoors | 1600 | 400 | 78.50% | 79.00% | **81.75%** | 78.00% | 80.00% | 81.50% | 78.50% | 79.00% | **81.75%** |
| Tools & Hardware | 89 | 23 | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** |
| Toys & Games | 1600 | 400 | 79.25% | 80.25% | **82.50%** | 80.25% | 80.25% | **82.50%** | 79.25% | 80.25% | **82.50%** |
| Video | 1600 | 400 | 81.50% | 80.50% | 83.00% | 81.25% | 80.25% | **83.25%** | 81.50% | 80.50% | 83.00% |
| **Overall Accuracy** | | | 81.67% | 79.7% | **82.75%** | 81.6% | 79.73% | 82.63% | 81.67% | 79.71% | 82.74% |

Table 5.4: Maximum Entropy Classifier Accuracies with POS sentence representations and selected set added

With MaxEnt we see the opposite of Naive Bayes where there is a strong showing in the unigrams and unigrams+bigrams, but no top accuracies in bigrams (except in the outlying *Tools & Hardware* domain, c.f. section 6.3). The best increase is 1.68% in the *Beauty* domain when POS representations are added to unigrams. This is the top accuracy for the domain with MaxEnt.

| **Domain** | **Train** | **Test** | **Unigrams** | **Bigrams** | **Uni+Bi** | **Uni+POS** | **Bi+POS** | **Uni+Bi+POS** | **Uni+SS** | **Bi+SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Apparel | 1600 | 400 | **79.75%** | 78.25% | 78.50% | **79.75%** | 78.50% | 78.50% | **79.75%** | 78.25% | 78.50% |
| Automotive | 589 | 147 | 80.95% | 80.27% | 80.95% | 80.95% | 80.95% | **81.63%** | 80.95% | 80.27% | 80.95% |
| Baby | 1520 | 380 | 76.84% | 76.84% | 74.74% | **77.37%** | 76.58% | 73.95% | 76.84% | 76.84% | 73.95% |
| Beauty | 1194 | 299 | 78.26% | **80.60%** | 78.60% | 78.26% | 79.60% | 78.93% | 78.26% | **80.60%** | 78.60% |
| Books | 1600 | 400 | 74.25% | 72.50% | 76.00% | 75.25% | 73.00% | **76.25%** | 74.25% | 72.50% | 76.00% |
| Camera & Photo | 1599 | 400 | 80.75% | 80.75% | 81.75% | **82.00%** | 80.50% | 81.75% | 80.50% | 80.75% | 81.75% |
| Cell Phones & Service | 818 | 205 | 76.10% | 74.15% | 76.10% | **77.56%** | 73.17% | 76.10% | 76.10% | 74.15% | 76.10% |
| Comp. & Video Games | 1166 | 292 | 95.55% | 96.58% | 96.92% | 95.55% | 96.92% | **97.26%** | 95.55% | 96.58% | 96.92% |
| DVD | 1600 | 400 | **78.50%** | 72.25% | 76.75% | **78.50%** | 73.50% | 76.75% | **78.50%** | 72.25% | 77.00% |
| Electronics | 1600 | 400 | 77.00% | 74.50% | 77.75% | 76.50% | 74.25% | **78.00%** | 77.00% | 74.50% | **78.00%** |
| Gourmet Food | 966 | 242 | 86.78% | 84.30% | 85.54% | **87.19%** | 84.71% | 85.95% | 86.78% | 84.30% | 85.54% |
| Grocery | 1082 | 270 | 85.19% | 82.22% | 84.81% | **86.30%** | 81.11% | 84.81% | 85.19% | 82.22% | 84.81% |
| Health & Personal Care | 1600 | 400 | 79.85% | 80.25% | 81.75% | 80.00% | 80.5% | **82.25%** | 79.50% | 80.25% | 82.00% |
| Jewelry & Watches | 1034 | 258 | 84.88% | 84.50% | **87.21%** | 84.88% | 84.50% | 86.43% | 84.88% | 84.50% | **87.21%** |
| Kitchen & Housewares | 1600 | 400 | 82.25% | 81.00% | 82.50% | 82.25% | 81.25% | 82.50% | 82.25% | 81.00% | **82.75%** |
| Magazines | 1576 | 394 | **81.22** | 80.20% | 79.70% | 80.96% | 80.96% | 79.70% | **81.22%** | 80.20% | 79.70% |
| Music | 1600 | 400 | 77.50% | 73.00% | 77.50% | 77.00% | 73.50% | 77.25% | 77.50% | 73.00% | **77.75%** |
| Musical Instruments | 265 | 67 | **89.55%** | 85.07% | 85.07% | 88.06% | 85.07% | 85.07% | **89.55%** | 85.07% | 85.07% |
| Office Products | 345 | 86 | **86.05%** | **86.05%** | **86.05%** | **86.05%** | **86.05%** | **86.05%** | **86.05%** | **86.05%** | **86.05%** |
| Outdoor Living | 1062 | 265 | 81.13% | 79.62% | **81.89%** | 81.51% | 78.87% | **81.89%** | 81.13% | 79.62% | **81.89%** |
| Software | 1532 | 383 | **81.46%** | 74.67% | 79.63% | **81.46%** | 74.93% | 79.37% | **81.46%** | 74.67% | 79.63% |
| Sports & Outdoors | 1600 | 400 | 77.75% | 78.75% | 77.50% | 77.25% | **80.75%** | 77.25% | 77.75% | 78.75% | 77.50% |
| Tools & Hardware | 89 | 23 | **100%** | **100%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** | **100.00%** |
| Toys & Games | 1600 | 400 | 78.00% | **82.50%** | 80.25% | 79.25% | 82.25% | 80.50% | 78.00% | **82.50%** | 80.50% |
| Video | 1600 | 400 | 79.75% | 81.50% | **82.00%** | 79.75% | 81.75% | 81.75% | 79.75% | 81.50% | **82.00%** |
| **Overall Accuracy** | | | 80.46% | 79.33% | 80.62% | **80.67%** | 79.51% | 80.62% | 80.43% | 79.33% | 80.66% |

Table 5.5: SVM Classifier Accuracies with POS sentence representations and selected set added

For SVMs, top accuracies for domains mostly occur in unigrams or unigram+bigram distributions, bigram alone achieves only a couple top accuracies. The best increase is 2.00% in the *Sports & Outdoors* domain when POS representations are added to bigrams. This is the top accuracy for the domain with SVM.

### *5.4 Percentage Changes*

Tables 5.6, 5.7, and 5.8 gives the percentage change in accuracy when POS representations and selected set are added to unigram, bigram, and unigram+bigram distributions. Bold face cells indicate non-zero percent changes. This reiterates the information found in Tables 5.3, 5.4, and 5.5 and is done simply to present the changes in accuracy due to addition of POS representations and selected set in a clearer manner.

| **Domain** | **Unigrams+POS** | **Bigrams+POS** | **Uni+Bi+POS** | **Unigram+SS** | **Bigram+SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|
| Apparel | **-0.50%** | **-0.75%** | **-0.25%** | 0.00% | 0.00% | 0.00% |
| Automotive | **-0.68%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Baby | **0.79%** | **1.05%** | **0.27%** | 0.00% | 0.00% | 0.00% |
| Beauty | **-2.68%** | **0.33%** | **-0.67%** | 0.00% | 0.00% | 0.00% |
| Books | **0.25%** | **0.50%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Camera & Photo | **0.25%** | **-0.50%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Cell Phones & Service | **-3.90%** | 0.00% | **-0.97%** | 0.00% | 0.00% | 0.00% |
| Computer & Video Games | **1.43%** | 0.00% | 0.00% | **0.06%** | 0.00% | 0.00% |
| DVD | **0.50%** | **-0.50%** | **0.75%** | 0.00% | 0.00% | 0.00% |
| Electronics | **0.50%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Gourmet Food | **0.82%** | **0.41%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Grocery | **-1.85%** | **-1.48%** | **0.37%** | 0.00% | 0.00% | 0.00% |
| Health & Personal Care | **0.25%** | **-0.50%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Jewelry & Watches | **-1.55%** | **-0.39%** | **-0.38%** | 0.00% | 0.00% | 0.00% |
| Kitchen & Housewares | **0.25%** | **0.25%** | **-0.25%** | 0.00% | 0.00% | 0.00% |
| Magazines | **0.25%** | **0.51%** | **0.25%** | 0.00% | 0.00% | 0.00% |
| Music | **1.75%** | 0.00% | **-0.25%** | 0.00% | **0.50%** | 0.00% |
| Musical Instruments | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Office Products | **1.16%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Outdoor Living | **-0.38%** | **0.38%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Software | **-0.78%** | 0.00% | **0.52%** | 0.00% | 0.00% | 0.00% |
| Sports & Outdoors | **1.50%** | **1.00%** | **-0.75%** | 0.00% | 0.00% | 0.00% |
| Tools & Hardware | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Toys Games | 0.00% | **0.25%** | **-0.50%** | 0.00% | 0.00% | 0.00% |
| Video | **-0.75%** | **-0.25%** | 0.00% | 0.00% | **-0.25%** | 0.00% |

Table 5.6: Naive Bayes Classifier Accuracy Changes

| **Domain** | **Unigrams+POS** | **Bigrams+POS** | **Uni+Bi+POS** | **Unigram+SS** | **Bigram+SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|
| Apparel | 0.00% | **-0.25%** | **0.50%** | 0.00% | 0.00% | 0.00% |
| Automotive | **0.68%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Baby | **0.27%** | **1.32%** | **-0.26%** | 0.00% | 0.00% | 0.00% |
| Beauty | **1.68%** | **-0.34%** | **0.34%** | 0.00% | 0.00% | 0.00% |
| Books | **-0.50%** | 0.00% | **-0.25%** | 0.00% | 0.00% | 0.00% |
| Camera & Photo | **0.25%** | 0.00% | **-0.25%** | 0.00% | 0.00% | 0.00% |
| Cell Phones & Service | **-0.48%** | 0.00% | **-0.48%** | 0.00% | 0.00% | 0.00% |
| Computer & Video Games | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| DVD | **-0.25%** | **0.50%** | **0.25%** | 0.00% | 0.00% | 0.00% |
| Electronics | **-0.25%** | **0.25%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Gourmet Food | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Grocery | **-0.37%** | **0.37%** | **-0.37%** | 0.00% | 0.00% | 0.00% |
| Health & Personal Care | **-0.50%** | **0.75%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Jewelry & Watches | **-0.77%** | **-0.39%** | **-0.39%** | 0.00% | 0.00% | 0.00% |
| Kitchen & Housewares | **-0.50%** | 0.00% | **-0.75%** | 0.00% | 0.00% | 0.00% |
| Magazines | 0.00% | **-0.51%** | **-0.26%** | 0.00% | **0.07%** | 0.00% |
| Music | **-0.25%** | **-0.75%** | **-0.50%** | 0.00% | 0.00% | **-0.25%** |
| Musical Instruments | 0.00% | 0.00% | **0.01%** | 0.00% | 0.00% | 0.00% |
| Office Products | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Outdoor Living | 0.00% | **-0.75%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Software | **-0.53%** | **-0.78%** | **-0.26%** | 0.00% | 0.00% | 0.00% |
| Sports & Outdoors | **-0.50%** | **1.00%** | **-0.25%** | 0.00% | 0.00% | 0.00% |
| Tools & Hardware | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Toys & Games | **1.00%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Video | **-0.25%** | **-0.25%** | **0.25%** | 0.00% | 0.00% | 0.00% |

Table 5.7: Maximum Entropy Classifier Accuracy Changes

| **Domain** | **Unigrams+POS** | **Bigrams+POS** | **Uni+Bi+POS** | **Unigram+SS** | **Bigram+SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|
| Apparel | 0.00% | **0.25%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Automotive | 0.00% | **0.68%** | **0.68%** | 0.00% | 0.00% | 0.00% |
| Baby | **0.53%** | **-0.26%** | **-0.79%** | 0.00% | 0.00% | **-0.79%** |
| Beauty | 0.00% | **-1.00%** | **0.33%** | 0.00% | 0.00% | 0.00% |
| Books | **1.00%** | **0.50%** | **0.25%** | 0.00% | 0.00% | 0.00% |
| Camera & Photo | **1.25%** | **-0.25%** | 0.00% | **-0.25%** | 0.00% | 0.00% |
| Cell Phones & Service | **1.46%** | **-0.98%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Comp. & Video Games | 0.00% | **0.34%** | **0.34%** | 0.00% | 0.00% | 0.00% |
| DVD | 0.00% | **1.25%** | 0.00% | 0.00% | 0.00% | **0.25%** |
| Electronics | **-0.50%** | **-0.25%** | **0.25%** | 0.00% | 0.00% | **0.25%** |
| Gourmet Food | **0.41%** | **0.41%** | **0.41%** | 0.00% | 0.00% | 0.00% |
| Grocery | **1.11%** | **-1.11%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Health & Personal Care | **0.15%** | **0.25%** | **0.50%** | **-0.35%** | 0.00% | **0.25%** |
| Jewelry & Watches | 0.00% | 0.00% | **-0.78%** | 0.00% | 0.00% | 0.00% |
| Kitchen & Housewares | 0.00% | **0.25%** | 0.00% | 0.00% | 0.00% | **0.25%** |
| Magazines | **-0.26%** | **0.76%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Music | **-0.50%** | **0.50%** | **-0.25%** | 0.00% | 0.00% | **0.25%** |
| Musical Instruments | **-1.49%** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Office Products | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Outdoor Living | **0.38%** | **-0.75%** | 0.00% | 0.00% | 0.00% | 0.00% |
| Software | 0.00% | **0.26%** | **-0.26%** | 0.00% | 0.00% | 0.00% |
| Sports & Outdoors | **-0.50%** | **2.00%** | **-0.25%** | 0.00% | 0.00% | 0.00% |
| Tools & Hardware | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Toys & Games | **1.25%** | **-0.25%** | **0.25%** | 0.00% | 0.00% | **0.25%** |
| Video | 0.00% | **0.25%** | **-0.25%** | 0.00% | 0.00% | 0.00% |

Table 5.8: SVM Classifier Accuracy Changes

## 5.5 Rankings

Table 5.9 serves to provide a reference of the best and worst performing classifier and n-gram distribution variations. It shows that there is no clearly dominant classifier and that it seems to be very domain specific. This is interesting considering Pang et al. (2002) showed SVMs winning out in their paper given the specific domain of movie reviews. The closest domain to movie reviews in the Amazon product reviews dataset, DVDs, shows MaxEnt to be the top classifier.

| Domain | Highest | Classifier Variant | Lowest | Classifier Variant |
|---|---|---|---|---|
| Apparel | 83.25% | NB (Uni+Bi, Uni+Bi+SS) | 78.25% | SVM (Bigrams, Bi+SS) |
| Automotive | 86.39% | MaxEnt (Uni+Bi, Uni+Bi+POS, Uni+Bi+SS) | 79.59% | NB (Uni+Bi, Uni+POS, Uni+Bi+POS, Uni+Bi+SS) |
| Baby | 83.42% | NB (Bi+POS) | 73.95% | SVM (Uni+Bi+POS, Uni+Bi+SS) |
| Beauty | 80.94% | MaxEnt (Uni+POS, Uni+Bi+POS) | 72.24% | NB (Uni+POS) |
| Books | 82.75% | NB (Uni+Bi, Uni+Bi+POS, Uni+Bi+SS) | 72.5% | SVM (Bigrams, Bi+SS) |
| Camera & Photo | 86.50% | NB (Bigrams, Bi+SS) | 80.5% | SVM (Bi+POS, Uni+SS) |
| Cell Phones & Service | 83.41% | MaxEnt (Uni+Bi, Uni+Bi+SS) | 73.17% | SVM (Bi+POS) |
| Computer & Video Games | 97.60% | MaxEnt (Uni+Bi, Uni+Bi+POS, Uni+Bi+SS) | 94.12% | NB (Uni) |
| DVD | 79.25% | MaxEnt (Uni, Uni+SS) | 72.25% | SVM (Bigrams, Bi+SS) |
| Electronics | 82.25% | NB (Uni+Bi, Uni+Bi+POS, Uni+Bi+SS) | 74.25% | SVM (Bi+POS) |
| Gourmet Food | 87.19% | SVM (Uni+POS), MaxEnt (Uni, Uni+POS, Uni+SS) | 83.06% | NB (Uni, Uni+SS) |
| Grocery | 86.30% | SVM (Uni+POS) | 77.78% | NB (Uni+POS) |
| Health & Personal Care | 85.25% | NB (Uni+Bi, Uni+Bi+POS, Uni+Bi+SS) | 79.5% | SVM (Uni+SS) |
| Jewelry & Watches | 87.60% | MaxEnt (Uni+Bi, Uni+Bi+SS) | 77.13% | NB (Uni+POS) |
| Kitchen & Housewares | 85.50% | MaxEnt (Uni+Bi, Uni+Bi+SS) | 81.00% | MaxEnt (Bigrams) |
| Magazines | 82.49% | MaxEnt (Uni+Bi, Uni+Bi+SS) | 79.7% | SVM (Uni+Bi, Uni+Bi+POS, Uni+Bi+SS) |
| Music | 80.75% | NB (Uni+POS) | 73.00% | SVM (Bigrams, Bi+SS) |
| Musical Instruments | 89.55% | SVM (Uni, Uni+SS) | 85.07% | NB, MaxEnt, SVM (Numerous Variants) |
| Office Products | 86.05% | SVM, MaxEnt (Numerous Variants) | 82.56% | NB (Uni, Uni+SS) |
| Outdoor Living | 81.89% | SVM (Uni+Bi, Uni+Bi+SS, Uni+Bi+POS), MaxEnt (Uni, Uni+POS, Uni+SS) | 76.6% | NB (Uni+POS, Uni+Bi, Uni+Bi+POS, Uni+Bi+SS) |
| Software | 84.33% | NB (Uni+Bi+POS) | 74.67% | SVM (Bigrams, Bi+SS) |
| Sports & Outdoors | 83.25% | NB (Uni+Bi, Uni+Bi+SS) | 77.25% | SVM (Uni+POS, Uni+Bi+POS) |
| Tools & Hardware | 100.00% | All Variants | 100.00% | All Variants |
| Toys & Games | 85.00% | NB (Bi+POS) | 78.00% | SVM (Uni, Uni+SS) |
| Video | 83.25% | MaxEnt (Uni+Bi+POS) | 79.5% | NB (Bi+POS, Bi+SS) |

Table 5.9: Best and worst classifier variant accuracy per domain

## 5.6 Domain Adaptation

After experimenting with the addition of POS representations and a selected set, I decided to see how the sentence representations would perform in a domain adaptation scenario, using the same domains in the dataset that were used in Blitzer et al. (2007), Li and Zong (2008), Pan et al. (2010) and Glorot et al. (2011). To better understand any gains we may see with the selected set, Table 5.10 and Table 5.11 display the occurrence count of the selected set POS representations in positive and negative reviews for each domain during training and testing.

| POS Representation | Polarity | | Apparel | Auto | Baby | Beauty | Books | Camera | Cell | Computer | DVD | Electronics | Gourmet | Grocery | Health |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| JJ_NN | Positive | Train -> | 24/8 | 11/0 | 14/8 | 15/3 | 15/5 | 18/15 | 15/5 | 19/5 | 10/6 | 28/9 | 17/4 | 15/2 | 18/10 |
| | | Test -> | 7/2 | 4/0 | 4/1 | 3/0 | 2/1 | 10/2 | 6/1 | 3/3 | 2/2 | 3/1 | 5/0 | 4/1 | 2/4 |
| DT_NN_VBZ_JJ | Positive | Train -> | 6/7 | 10/2 | 11/1 | 22/4 | 6/4 | 11/7 | 9/1 | 11/5 | 11/11 | 9/2 | 16/0 | 10/1 | 8/11 |
| | | Test -> | 2/0 | 0/0 | 2/0 | 2/0 | 2/3 | 2/1 | 5/1 | 2/1 | 2/1 | 1/0 | 3/1 | 4/0 | 2/1 |
| PRP_VBP_DT_NN | Positive | Train -> | 14/1 | 2/0 | 22/1 | 24/0 | 5/3 | 5/2 | 3/1 | 10/2 | 4/3 | 6/1 | 7/1 | 12/1 | 15/4 |
| | | Test -> | 1/1 | 1/0 | 2/0 | 8/0 | 0/0 | 4/1 | 0/0 | 0/0 | 1/0 | 4/1 | 3/0 | 2/0 | 4/0 |
| RB_VBN | Positive | Train -> | 6/2 | 4/1 | 9/2 | 9/0 | 9/1 | 9/2 | 1/0 | 5/0 | 7/2 | 10/0 | 9/5 | 8/1 | 5/1 |
| | | Test -> | 2/2 | 1/0 | 0/0 | 3/0 | 3/0 | 0/4 | 1/0 | 0/0 | 1/0 | 0/2 | 0/1 | 2/0 | 4/0 |
| DT_VBZ_DT_JJ_NN | Positive | Train -> | 5/2 | 5/1 | 5/1 | 3/2 | 5/2 | 10/3 | 7/3 | 2/1 | 7/2 | 6/4 | 5/0 | 7/0 | 6/3 |
| | | Test -> | 2/0 | 0/1 | 2/0 | 1/0 | 2/0 | 4/1 | 2/0 | 1/0 | 4/0 | 1/2 | 0/0 | 2/0 | 4/0 |
| PRP_VBP_PRP | Positive | Train -> | 8/0 | 5/0 | 12/2 | 15/1 | 2/0 | 3/0 | 4/0 | 2/0 | 1/0 | 6/0 | 6/0 | 4/0 | 5/0 |
| | | Test -> | 0/0 | 0/0 | 3/0 | 2/0 | 2/0 | 1/2 | 0/0 | 0/0 | 0/0 | 1/0 | 2/1 | 2/0 | 2/1 |
| PRP_VBZ_JJ | Positive | Train -> | 4/0 | 2/0 | 9/4 | 12/5 | 5/1 | 5/1 | 4/1 | 8/1 | 4/3 | 5/0 | 8/0 | 2/1 | 13/4 |
| | | Test -> | 0/1 | 0/0 | 2/0 | 1/1 | 0/0 | 1/0 | 0/0 | 2/1 | 0/1 | 1/1 | 0/0 | 3/0 | 2/1 |
| DT_JJ_NN | Positive | Train -> | 6/6 | 2/0 | 1/1 | 4/1 | 5/4 | 3/1 | 1/1 | 2/5 | 6/3 | 4/0 | 1/0 | 2/0 | 1/1 |
| | | Test -> | 1/0 | 0/0 | 1/0 | 4/0 | 0/0 | 2/0 | 2/1 | 0/3 | 2/1 | 3/0 | 0/0 | 0/0 | 0/0 |
| VB_PRP | Positive | Train -> | 6/1 | 4/0 | 3/0 | 3/1 | 3/2 | 3/1 | 2/2 | 3/0 | 5/8 | 2/2 | 5/0 | 5/0 | 5/0 |
| | | Test -> | 0/0 | 1/1 | 2/0 | 0/0 | 0/0 | 1/0 | 1/0 | 0/0 | 0/0 | 3/0 | 3/0 | 2/0 | 0/0 |
| PRP_VBZ_DT_JJ_NN | Positive | Train -> | 2/1 | 2/0 | 0/1 | 0/0 | 1/2 | 2/1 | 1/0 | 0/1 | 3/3 | 2/0 | 0/0 | 3/0 | 6/1 |
| | | Test -> | 0/0 | 2/0 | 1/0 | 2/0 | 0/0 | 0/1 | 0/0 | 0/0 | 0/1 | 1/0 | 0/0 | 0/0 | 0/0 |
| PRP_RB_VB_DT_NN | Positive | Train -> | 3/0 | 4/0 | 6/0 | 3/0 | 4/0 | 2/0 | 1/0 | 1/0 | 2/0 | 4/1 | 2/0 | 2/0 | 9/0 |
| | | Test -> | 1/0 | 0/0 | 1/0 | 1/0 | 1/0 | 2/0 | 1/0 | 0/0 | 2/0 | 2/0 | 1/0 | 1/0 | 2/0 |
| PRP_VBP_RB_JJ_IN_DT_NN | Positive | Train -> | 4/1 | 2/1 | 3/0 | 1/0 | 1/0 | 7/0 | 1/1 | 0/0 | 0/0 | 6/1 | 1/0 | 0/1 | 5/0 |
| | | Test -> | 0/0 | 0/0 | 3/0 | 0/0 | 0/1 | 0/1 | 1/0 | 0/0 | 0/0 | 3/0 | 1/0 | 0/0 | 0/0 |
| NNP_NNP_NNP | Positive | Train -> | 0/2 | 0/0 | 2/0 | 15/0 | 4/4 | 1/1 | 0/0 | 4/0 | 4/1 | 0/0 | 12/1 | 2/0 | 3/2 |
| | | Test -> | 0/0 | 0/0 | 0/0 | 2/0 | 1/1 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 2/0 | 0/1 | 1/3 |
| PRP_RB_VB_PRP | Positive | Train -> | 2/0 | 3/0 | 4/0 | 4/0 | 3/0 | 3/0 | 0/0 | 1/0 | 5/0 | 1/0 | 5/0 | 3/0 | 2/0 |
| | | Test -> | 2/0 | 0/0 | 1/0 | 3/0 | 0/0 | 0/0 | 1/0 | 0/0 | 1/0 | 0/0 | 0/0 | 1/0 | 2/0 |
| DT_NNS_VBP_JJ | Positive | Train -> | 7/1 | 3/0 | 1/2 | 3/0 | 2/1 | 2/1 | 1/0 | 5/0 | 1/1 | 3/1 | 4/0 | 5/0 | 2/0 |
| | | Test -> | 3/1 | 0/0 | 2/0 | 0/1 | 0/1 | 0/0 | 0/0 | 1/0 | 0/0 | 0/0 | 2/0 | 0/0 | 3/1 |
| WP_DT_NN | Negative | Train -> | 0/1 | 0/0 | 3/5 | 2/3 | 1/8 | 0/4 | 1/0 | 2/1 | 0/8 | 2/7 | 0/1 | 1/0 | 1/5 |
| | | Test -> | 0/1 | 0/0 | 0/0 | 0/0 | 1/0 | 0/1 | 1/0 | 0/1 | 0/1 | 0/0 | 1/1 | 1/0 | 1/0 |
| PRP_VBD_RB_JJ | Negative | Train -> | 1/5 | 0/0 | 1/13 | 1/2 | 2/6 | 2/0 | 0/2 | 0/4 | 2/4 | 0/1 | 1/2 | 3/1 | 2/4 |
| | | Test -> | 0/2 | 0/0 | 2/2 | 0/2 | 0/3 | 0/2 | 0/0 | 0/0 | 1/0 | 1/0 | 0/0 | 1/0 | 1/1 |
| VBP_RB_VB_PRP_NN | Negative | Train -> | 0/2 | 0/2 | 0/10 | 0/2 | 0/7 | 0/0 | 0/3 | 0/2 | 0/1 | 0/6 | 0/1 | 0/0 | 0/5 |
| | | Test -> | 0/1 | 0/0 | 0/1 | 0/0 | 0/1 | 0/0 | 0/0 | 0/0 | 0/2 | 0/1 | 0/0 | 0/0 | 0/2 |
| VB_PRP_NN | Negative | Train -> | 0/2 | 1/1 | 0/6 | 1/6 | 1/2 | 0/2 | 0/4 | 1/1 | 2/5 | 0/3 | 0/1 | 0/1 | 0/2 |
| | | Test -> | 0/0 | 0/0 | 0/0 | 1/0 | 1/0 | 0/0 | 0/0 | 1/1 | 0/2 | 0/0 | 0/0 | 0/0 | 0/0 |
| VBP_RB_VB_DT_NN | Negative | Train -> | 0/1 | 0/1 | 0/3 | 0/0 | 1/4 | 0/5 | 0/2 | 0/0 | 0/1 | 0/7 | 0/1 | 0/1 | 1/5 |
| | | Test -> | 0/1 | 0/0 | 0/4 | 0/0 | 1/0 | 0/4 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 |
| PRP_VBD_RB_JJ_IN_DT_NN | Negative | Train -> | 0/4 | 0/0 | 0/4 | 1/6 | 0/3 | 0/0 | 0/0 | 0/1 | 0/3 | 0/1 | 0/0 | 2/2 | 0/0 |
| | | Test -> | 0/0 | 1/0 | 0/3 | 0/2 | 0/0 | 1/1 | 0/0 | 0/1 | 0/0 | 0/0 | 0/0 | 0/0 | 0/1 |

Table 5.10: Occurrences of selected set POS representations by domain (seen in positive reviews/seen in negative reviews)

| POS Representation | Polarity | | Jewelry | Kitchen | Magazines | Music | Musical Inst. | Office | Outdoor | Software | Sports | Tools | Toys | Video |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| JJ_NN | Positive | Train -> | 20/2 | 11/3 | 7/2 | 10/13 | 1/0 | 5/2 | 16/6 | 18/13 | 25/7 | 1/0 | 16/9 | 2/7 |
| | | Test -> | 6/1 | 1/4 | 3/5 | 2/1 | 2/0 | 0/0 | 3/0 | 5/5 | 4/0 | 1/0 | 5/5 | 0/0 |
| DT_NN_VBZ_JJ | Positive | Train -> | 15/3 | 8/3 | 6/2 | 11/6 | 4/0 | 4/0 | 14/2 | 9/4 | 11/8 | 0/0 | 4/6 | 3/1 |
| | | Test -> | 4/0 | 4/2 | 2/0 | 1/0 | 0/0 | 0/0 | 4/1 | 1/1 | 4/0 | 0/0 | 2/0 | 0/0 |
| PRP_VBP_DT_NN | Positive | Train -> | 11/0 | 9/1 | 5/4 | 6/3 | 0/0 | 4/0 | 9/2 | 5/3 | 7/0 | 0/0 | 9/1 | 0/0 |
| | | Test -> | 3/0 | 3/0 | 3/0 | 3/0 | 0/0 | 0/0 | 0/0 | 0/1 | 3/0 | 0/0 | 1/0 | 0/0 |
| RB_VBN | Positive | Train -> | 7/1 | 8/2 | 7/2 | 14/4 | 3/0 | 1/0 | 3/3 | 5/1 | 8/4 | 0/0 | 8/4 | 0/0 |
| | | Test -> | 0/0 | 0/0 | 0/2 | 3/0 | 0/0 | 1/0 | 2/1 | 1/0 | 1/0 | 0/0 | 3/1 | 0/0 |
| DT_VBZ_DT_JJ_NN | Positive | Train -> | 3/0 | 11/2 | 5/0 | 9/2 | 2/0 | 0/0 | 9/0 | 5/4 | 5/0 | 0/0 | 4/7 | 0/0 |
| | | Test -> | 3/0 | 2/0 | 2/0 | 3/1 | 0/0 | 0/0 | 2/0 | 1/0 | 3/0 | 0/0 | 4/0 | 0/0 |
| PRP_VBP_PRP | Positive | Train -> | 11/1 | 9/2 | 8/0 | 4/1 | 0/0 | 2/0 | 4/1 | 11/0 | 10/0 | 0/0 | 5/0 | 0/0 |
| | | Test -> | 4/0 | 4/0 | 1/0 | 3/0 | 0/0 | 0/0 | 0/0 | 2/0 | 2/0 | 0/0 | 2/0 | 0/0 |
| PRP_VBZ_JJ | Positive | Train -> | 3/0 | 7/4 | 1/1 | 4/6 | 0/0 | 1/0 | 1/0 | 1/3 | 5/3 | 0/0 | 1/0 | 1/0 |
| | | Test -> | 3/1 | 0/3 | 3/1 | 0/0 | 0/0 | 0/0 | 3/0 | 0/0 | 3/0 | 0/0 | 0/0 | 0/0 |
| DT_JJ_NN | Positive | Train -> | 4/0 | 3/2 | 11/2 | 6/3 | 1/0 | 1/0 | 3/0 | 6/5 | 4/0 | 0/0 | 2/0 | 1/2 |
| | | Test -> | 1/0 | 1/0 | 1/0 | 4/0 | 0/0 | 0/0 | 0/0 | 1/0 | 0/0 | 0/0 | 2/0 | 0/0 |
| VB_PRP | Positive | Train -> | 6/0 | 6/2 | 4/0 | 6/2 | 2/0 | 0/0 | 0/0 | 5/1 | 1/3 | 1/0 | 4/1 | 2/0 |
| | | Test -> | 3/0 | 0/1 | 0/0 | 3/1 | 0/0 | 0/0 | 0/0 | 3/1 | 2/0 | 1/0 | 1/0 | 0/0 |
| PRP_VBZ_DT_JJ_NN | Positive | Train -> | 8/0 | 3/1 | 12/0 | 5/3 | 0/0 | 0/0 | 3/0 | 4/4 | 2/0 | 2/0 | 5/0 | 0/0 |
| | | Test -> | 4/0 | 4/0 | 1/0 | 0/0 | 0/0 | 0/0 | 1/0 | 1/1 | 3/0 | 0/0 | 1/1 | 0/0 |
| PRP_RB_VB_DT_NN | Positive | Train -> | 3/0 | 2/0 | 1/0 | 0/0 | 0/0 | 2/0 | 5/0 | 0/0 | 4/0 | 0/0 | 2/0 | 0/0 |
| | | Test -> | 0/0 | 2/0 | 1/0 | 0/0 | 0/0 | 0/0 | 1/0 | 0/0 | 2/0 | 0/0 | 1/0 | 0/0 |
| PRP_VBP_RB_JJ_IN_DT_NN | Positive | Train -> | 6/0 | 3/2 | 1/0 | 0/1 | 1/0 | 2/0 | 8/0 | 6/2 | 5/0 | 0/0 | 2/1 | 0/0 |
| | | Test -> | 2/1 | 3/0 | 0/0 | 0/0 | 0/0 | 1/0 | 0/0 | 0/0 | 4/0 | 0/0 | 0/0 | 0/0 |
| NNP_NNP_NNP | Positive | Train -> | 1/0 | 0/0 | 10/8 | 3/1 | 1/0 | 1/0 | 1/0 | 3/1 | 2/0 | 2/0 | 1/2 | 0/0 |
| | | Test -> | 0/0 | 0/0 | 0/0 | 1/2 | 0/0 | 0/0 | 0/0 | 3/0 | 0/0 | 0/0 | 0/0 | 0/0 |
| PRP_RB_VB_PRP | Positive | Train -> | 3/0 | 5/0 | 3/0 | 2/0 | 0/0 | 3/0 | 2/0 | 2/0 | 2/0 | 0/0 | 0/0 | 0/0 |
| | | Test -> | 0/0 | 1/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 1/0 | 0/0 | 0/0 | 1/0 | 0/0 |
| DT_NNS_VBP_JJ | Positive | Train -> | 5/1 | 1/3 | 0/0 | 2/1 | 3/0 | 2/0 | 3/0 | 0/2 | 0/0 | 0/0 | 1/2 | 1/0 |
| | | Test -> | 0/0 | 1/0 | 0/0 | 1/0 | 0/0 | 0/0 | 0/0 | 0/0 | 1/0 | 0/0 | 0/2 | 0/0 |
| WP_DT_NN | Positive | Train -> | 1/1 | 4/8 | 0/3 | 1/8 | 0/0 | 0/0 | 1/2 | 0/10 | 0/2 | 0/0 | 0/9 | 1/1 |
| | | Test -> | 0/1 | 0/2 | 0/0 | 0/2 | 0/0 | 1/0 | 0/0 | 1/1 | 0/0 | 0/0 | 1/1 | 0/0 |
| PRP_VBD_RB_JJ | Positive | Train -> | 1/0 | 1/7 | 2/2 | 0/5 | 0/0 | 1/0 | 0/0 | 0/4 | 1/3 | 0/0 | 2/9 | 0/2 |
| | | Test -> | 0/1 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 1/2 | 0/0 | 0/0 | 0/0 |
| VBP_RB_VB_PRP_NN | Positive | Train -> | 0/0 | 0/6 | 0/5 | 0/3 | 0/1 | 0/0 | 0/4 | 0/2 | 0/2 | 0/0 | 0/9 | 0/1 |
| | | Test -> | 0/0 | 0/0 | 0/0 | 0/0 | 0/1 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/2 | 0/0 |
| VB_PRP_NN | Positive | Train -> | 1/1 | 1/1 | 0/4 | 1/3 | 0/0 | 0/0 | 0/1 | 0/1 | 1/3 | 0/0 | 1/4 | 0/0 |
| | | Test -> | 0/0 | 0/0 | 1/0 | 0/0 | 0/0 | 1/0 | 0/0 | 0/3 | 0/1 | 0/0 | 0/0 | 0/0 |
| VBP_RB_VB_DT_NN | Positive | Train -> | 0/0 | 0/6 | 0/0 | 0/4 | 0/0 | 0/0 | 0/1 | 0/2 | 0/2 | 0/0 | 0/3 | 0/0 |
| | | Test -> | 0/1 | 0/1 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/2 | 0/0 | 0/0 | 0/0 | 0/0 |
| PRP_VBD_RB_JJ_IN_DT_NN | Positive | Train -> | 1/2 | 0/0 | 0/2 | 0/1 | 0/0 | 1/0 | 0/0 | 0/3 | 1/1 | 0/0 | 0/4 | 0/0 |
| | | Test -> | 0/0 | 0/0 | 0/0 | 0/1 | 0/1 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 | 0/0 |

Table 5.11: Occurrences of selected set POS representations by domain (seen in positive reviews/seen in negative reviews) - continued

Tables 5.12, 5.14, and 5.16 showcase the result of training the classifiers with the POS sentence representations and the selected set and then applying those trained classifiers to the test data of a different domain. We can see that across the board the accuracies have suffered because of the domain change.

| **Source -> Target** | **Unigrams** | **Bigrams** | **Uni+Bi** | **Uni+POS** | **Bi+POS** | **Uni+Bi+POS** | **Uni+SS** | **Bi+SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|---|---|---|
| Books -> DVDs | 51.25% | 49.25% | 51.25% | 52.00% | 48.75% | 54.25% | 50.25% | 48.00% | **55.00%** |
| Books -> Electronics | 52.00% | 50.25% | 49.00% | 52.75% | 51.75% | **53.25%** | 52.50% | 51.00% | 51.75% |
| Books -> Kitchen | **55.00%** | 46.00% | 51.50% | 48.25% | 48.50% | 50.25% | 49.00% | 48.75% | 49.50% |
| DVDs -> Books | 50.00% | 52.50% | 51.25% | 53.00% | **54.25%** | 52.00% | 53.25% | 53.50% | 49.75% |
| DVDs -> Electronics | 51.50% | 53.25% | 53.00% | 54.75% | 49.00% | 50.50% | **55.50%** | 47.50% | 50.25% |
| DVDs -> Kitchen | **51.75%** | 48.50% | 51.25% | 49.75% | 50.75% | 45.50% | 51.50% | 50.50% | 46.75% |
| Electronics -> Books | 50.00% | 50.75% | 48.50% | 53.00% | 52.50% | 56.50% | 54.75% | 50.50% | **57.00%** |
| Electronics -> DVDs | 51.75% | 48.25% | 54.50% | 50.50% | 49.00% | 54.50% | 49.00% | 47.25% | **55.00%** |
| Electronics -> Kitchen | 47.50% | 46.00% | 47.00% | **52.25%** | 47.75% | 46.25% | 52.00% | 48.50% | 48.00% |
| Kitchen -> Books | 52.50% | 50.50% | 53.75% | 51.25% | **54.50%** | 51.00% | 50.75% | 53.50% | 54.25% |
| Kitchen -> DVDs | 54.00% | **54.25%** | 52.50% | 50.75% | 48.25% | 51.50% | 51.25% | 49.00% | 51.50% |
| Kitchen -> Electronics | **53.75%** | 51.00% | 48.50% | 49.00% | 46.25% | 50.00% | 50.50% | 47.25% | 49.75% |

Table 5.12: Domain Adaptation with Naive Bayes Classifier

With Naive Bayes we see some considerable improvement, as much as 8.50% in *Electronics* to *Books*, when the selected set is added. POS representations and selected set contribute both generous increases as well as generous decreases in accuracy. Table 5.13 provides easy reference to the percentage change in accuracy for the domain adaptation experiment using Naive Bayes. Bold faced cells indicate the top positive percentage gain for the respective domain.

| **Source -> Target** | **Unigram POS** | **Bigram POS** | **Uni+Bi POS** | **Unigram SS** | **Bigram SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|
| Books -> DVDs | 0.75% | -0.50% | 3.00% | -1.00% | -1.25% | **3.75%** |
| Books -> Electronics | 0.75% | 1.50% | **4.25%** | 0.50% | 0.75% | 2.75% |
| Books -> Kitchen | -6.75% | 2.50% | -1.25% | -6.00% | **2.75%** | -2.00% |
| DVDs -> Books | 3.00% | 1.75% | 0.75% | **3.25%** | 1.00% | -1.50% |
| DVDs -> Electronics | 3.25% | -4.25% | -2.50% | **4.00%** | -5.75% | -2.75% |
| DVDs -> Kitchen | -2.00% | **2.25%** | -5.75% | -0.25% | 2.00% | -4.50% |
| Electronics -> Books | 3.00% | 1.75% | 8.00% | 4.75% | -0.25% | **8.50%** |
| Electronics -> DVDs | -1.25% | **0.75%** | 0.00% | -2.75% | -1.00% | 0.50% |
| Electronics -> Kitchen | **4.75%** | 1.75% | -0.75% | 4.50% | 2.50% | 1.00% |
| Kitchen -> Books | -1.25% | **4.00%** | -2.75% | -1.75% | 3.00% | 0.50% |
| Kitchen -> DVDs | -3.25% | -6.00% | -1.00% | -2.75% | -5.25% | -1.00% |
| Kitchen -> Electronics | -4.75% | -4.75% | **1.50%** | -3.25% | -3.75% | 1.25% |

Table 5.13: Accuracy changes for domain adaptation using Naive Bayes

| **Source -> Target** | **Unigrams** | **Bigrams** | **Uni+Bi** | **Uni+POS** | **Bi+POS** | **Uni+Bi+POS** | **Uni+SS** | **Bi+SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|---|---|---|
| Books -> DVDs | 56.00% | 50.00% | 54.50% | 52.50% | 49.50% | **56.25%** | 52.50% | 50.25% | 56.00% |
| Books -> Electronics | **53.75%** | 50.25% | 53.25% | 51.50% | 52.75% | 53.25% | 52.25% | 53.00% | 53.50% |
| Books -> Kitchen | 48.50% | 49.25% | **53.75%** | 48.25% | 50.50% | 50.75% | 49.25% | 50.50% | 50.50% |
| DVDs -> Books | 47.25% | 50.50% | 50.50% | 51.50% | 52.50% | 52.25% | 51.25% | **52.75%** | 52.25% |
| DVDs -> Electronics | 52.75% | 48.50% | **55.25%** | 50.75% | 46.00% | 54.75% | 52.00% | 45.25% | 54.50% |
| DVDs -> Kitchen | **56.00%** | 50.50% | 50.00% | 53.25% | 52.50% | 46.25% | 53.50% | 52.75% | 47.00% |
| Electronics -> Books | 52.25% | 48.50% | 52.50% | 50.75% | 49.75% | 53.75% | 52.00% | 51.00% | **54.50%** |
| Electronics -> DVDs | 53.25% | 53.00% | **57.00%** | 53.25% | 53.25% | 55.50% | 54.00% | 54.50% | 55.75% |
| Electronics -> Kitchen | 52.25% | 49.25% | 52.25% | 52.50% | 47.25% | **54.25%** | 51.75% | 47.00% | 53.75% |
| Kitchen -> Books | 54.00% | 49.50% | 55.50% | 52.00% | 50.50% | **55.75%** | 52.00% | 50.25% | 55.00% |
| Kitchen -> DVDs | **56.50%** | 54.50% | 45.75% | 50.75% | 50.75% | 47.00% | 51.50% | 49.00% | 47.00% |
| Kitchen -> Electronics | 55.75% | 49.00% | 52.50% | **57.75%** | 48.75% | 52.25% | 56.50% | 49.50% | 51.50% |

Table 5.14: Domain Adaptation with MaxEnt

MaxEnt shows similar swings in increases and decreases in accuracy depending on the variation for a given domain. *DVD* to *Books* shows the top increase in accuracy of 4.25% when POS representations are added to unigrams. Table 5.15 provides easy reference to the percentage change in accuracy for the domain adaptation experiment using Maximum Entropy. Bold faced cells indicate the top positive percentage gain for the respective domain.

| **Source -> Target** | **Unigram POS** | **Bigram POS** | **Uni+Bi POS** | **Unigram SS** | **Bigram SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|
| Books -> DVDs | -3.50% | -0.50% | **1.75%** | -3.50% | 0.25% | 1.50% |
| Books -> Electronics | -2.25% | 2.50% | 0.00% | -1.50% | **2.75%** | 0.25% |
| Books -> Kitchen | -0.25% | **1.25%** | -3.00% | 0.75% | **1.25%** | -3.25% |
| DVDs -> Books | **4.25%** | 2.00% | 1.75% | 4.00% | 2.25% | 1.75% |
| DVDs -> Electronics | -2.% | -2.50% | -0.50% | -0.75% | -3.25% | -0.75% |
| DVDs -> Kitchen | -2.75% | 2.00% | -3.75% | -2.50% | **2.25%** | -3.00% |
| Electronics -> Books | -1.50% | 1.25% | 1.25% | -0.25% | **2.50%** | 2.00% |
| Electronics -> DVDs | 0.00% | 0.25% | -1.50% | 0.75% | **1.50%** | -1.25% |
| Electronics -> Kitchen | 0.25% | -2.00% | **2.00%** | -0.50% | -2.25% | 1.50% |
| Kitchen -> Books | -2.00% | **1.00%** | 0.25% | -2.00% | 0.75% | -0.50% |
| Kitchen -> DVDs | -5.75% | -3.75% | **1.25%** | -5.00% | -5.50% | **1.25%** |
| Kitchen -> Electronics | **2.00%** | -0.25% | -0.25% | 0.75% | 0.50% | -1.00% |

Table 5.15: Accuracy changes for domain adaptation using Maximum Entropy

| **Source -> Target** | **Unigrams** | **Bigrams** | **Uni+Bi** | **Uni+POS** | **Bi+POS** | **Uni+Bi+POS** | **Uni+SS** | **Bi+SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|---|---|---|
| Books -> DVDs | **56.25%** | 48.25% | 51.50% | 53.50% | 48.75% | 52.50% | 51.75% | 50.00% | 52.00% |
| Books -> Electronics | 53.00% | 48.25% | 51.75% | 52.25% | 52.75% | 53.25% | 52.50% | 51.25% | **53.75%** |
| Books -> Kitchen | 47.50% | 49.25% | 49.00% | 49.50% | 49.50% | 51.50% | 52.50% | **52.75%** | 51.25% |
| DVDs -> Books | 48.75% | 49.50% | 50.25% | 49.25% | 51.75% | 52.25% | 48.00% | 51.75% | **52.50%** |
| DVDs -> Electronics | 51.25% | 48.25% | **53.75%** | 53.00% | 49.25% | **53.75%** | 52.00% | 52.75% | 53.25% |
| DVDs -> Kitchen | **56.75%** | 49.25% | 51.75% | 54.25% | 50.50% | 45.00% | 46.00% | 54.75% | 45.25% |
| Electronics -> Books | 55.25% | 50.50% | 52.75% | 49.50% | 49.75% | 55.00% | 52.25% | 54.50% | **55.50%** |
| Electronics -> DVDs | 53.50% | 52.00% | 54.50% | 55.75% | 54.25% | **56.25%** | 55.25% | 50.50% | 56.00% |
| Electronics -> Kitchen | 53.00% | 51.25% | 52.25% | 51.75% | 47.25% | 54.25% | **56.25%** | 49.50% | 53.50% |
| Kitchen -> Books | **54.25%** | 50.25% | 52.50% | 52.00% | 51.00% | 53.00% | 53.00% | 52.50% | 52.50% |
| Kitchen -> DVDs | **55.75%** | 53.50% | 50.75% | 51.75% | 50.00% | 48.75% | 49.50% | 50.50% | 48.50% |
| Kitchen -> Electronics | 54.50% | 50.75% | **56.75%** | **56.75%** | 49.25% | 52.00% | 54.50% | 53.50% | 52.00% |

Table 5.16: Domain Adaptation with SVM Classifier

Like Naive Bayes and MaxEnt, SVM shows accuracy increases, but it also shows the worst decrease in accuracy by -10.75% (*DVD* to *Kitchen*). Table 5.17 provides easy reference to the percentage change in accuracy for the domain adaptation experiment using SVMs. Bold faced cells indicate the top positive percentage gain for the respective domain.

| **Source -> Target** | **Unigram POS** | **Bigram POS** | **Uni+Bi POS** | **Unigram SS** | **Bigram SS** | **Uni+Bi+SS** |
|---|---|---|---|---|---|---|
| Books -> DVDs | -2.75% | 0.50% | 1.00% | -4.50% | **1.75%** | 0.50% |
| Books -> Electronics | -0.75% | **4.50%** | 1.50% | -0.50% | 3.00% | 2.00% |
| Books -> Kitchen | 2.00% | 0.25% | 2.50% | **5.00%** | 3.50% | 2.25% |
| DVDs -> Books | 0.50% | **2.25%** | 2.00% | -0.75% | **2.25%** | **2.25%** |
| DVDs -> Electronics | 1.75% | 1.00% | 0.00% | 0.75% | **4.50%** | -0.50% |
| DVDs -> Kitchen | -2.50% | 1.25% | -6.75% | -10.75% | **5.50%** | -6.50% |
| Electronics -> Books | -5.75% | -0.75% | 2.25% | -3.00% | **4.00%** | 2.75% |
| Electronics -> DVDs | **2.25%** | **2.25%** | 1.75% | 1.75% | -1.50% | 1.50% |
| Electronics -> Kitchen | -1.25% | -4.00% | 2.00% | **3.25%** | -1.75% | 1.25% |
| Kitchen -> Books | -2.25% | 0.75% | 0.50% | -1.25% | **2.25%** | 0.00% |
| Kitchen -> DVDs | -4.00% | -3.50% | -2.00% | -6.25% | -3.00% | -2.25% |
| Kitchen -> Electronics | 2.25% | -1.50% | -4.75% | 0.00% | **2.75%** | -4.75% |

Table 5.17: Accuracy changes for domain adaptation using SVMs

Overall, we see some nice jumps in accuracy for selected set in the domain adaptation scenario. One such example is seen going from the *Electronics* domain to the *Books* domain. To understand why this (and other accuracy changes) happened, one could look at Table 5.10 to derive the training and test numbers seen in Table 5.18. Bold faced rows in Table 5.18 indicate where there are non-zero values in both training and test sets.

| **POS Representation** | **Polarity** | **Electronics - Train (+/-)** | **Books - Test (+/-)** |
|---|---|---|---|
| **JJ_NN** | **Positive** | **28/9** | **2/1** |
| **DT_NN_VBZ_JJ** | **Positive** | **9/2** | **2/3** |
| PRP_VBP_DT_NN | Positive | 6/1 | 0/0 |
| **RB_VBN** | **Positive** | **10/0** | **3/0** |
| **DT_VBZ_DT_JJ_NN** | **Positive** | **6/4** | **2/0** |
| **PRP_VBP_PRP** | **Positive** | **6/0** | **2/0** |
| PRP_VBZ_JJ | Positive | 5/0 | 0/0 |
| DT_JJ_NN | Positive | 4/0 | 0/0 |
| VB_PRP | Positive | 2/2 | 0/0 |
| PRP_VBZ_DT_JJ_NN | Positive | 2/0 | 0/0 |
| **PRP_RB_VB_DT_NN** | **Positive** | **4/1** | **1/0** |
| **PRP_VBP_RB_JJ_IN_DT_NN** | **Positive** | **6/1** | **0/1** |
| NNP_NNP_NNP | Positive | 0/0 | 1/1 |
| PRP_RB_VB_PRP | Positive | 1/0 | 0/0 |
| **DT_NNS_VBP_JJ** | **Positive** | **3/1** | **0/1** |
| **WP_DT_NN** | **Negative** | **2/7** | **1/0** |
| **PRP_VBD_RB_JJ** | **Negative** | **0/1** | **0/3** |
| **VBP_RB_VB_PRP_NN** | **Negative** | **0/6** | **0/1** |
| **VB_PRP_NN** | **Negative** | **0/3** | **1/0** |
| **VBP_RB_VB_DT_NN** | **Negative** | **0/7** | **1/0** |
| PRP_VBD_RB_JJ_IN_DT_NN | Negative | 0/1 | 0/0 |

Table 5.18: Selected set occurrences in Electronics training data and Books test data

One will notice that there exists the possibility that we trained correctly for particular POS representations in the *Electronics* domain that were then seen in the test data of the *Books* domain. However, one will also notice where several false positives could occur. To find out what actually happened, further digging into the classification results is needed. Using the SVM results of the *Electronics* to *Books* domain adaptation results seen in

Table 5.16, one can see there was an accuracy change of 52.75% with unigrams+bigrams to 55.50% with unigrams+bigrams and selected set. Reviewing the classification differences between the two, there are 14 instances where the classification changed from positive to negative or negative to positive with the addition of the selected set where a POS representation in the selected set was present in the review. 9 of the 14 changes were correctly classified and 5 of the 14 changes were classified incorrectly. An example of the help given to the classifier by the selected set is the following review that was incorrectly classified as negative with unigrams+bigrams: "For someone like myself that is new to selling and licensing art,this book is so informative. I am so glad that I bought it. It is also a good reference book to keep on hand. I recommend it.". Looking at unigrams+bigrams alone, it was the decision of the SVM classifier that this review be classified as negative. Which we know to be wrong. With the addition of the selected set, the POS representation PRP_VBP_PRP seen in the sentence "I recommend it" contributes information to the classifier to correctly classify the review as positive. It should be clarified that with the addition of selected set features during training of the classifier, the decision making of the classifier will be altered for all reviews. It would be incorrect to say the presence of a selected set representation in a review is the sole determinant for the classifier's decision. The additional knowledge gained when selected set features were added during training contributes to the overall decision making of the classifier.

# Chapter 6
# DISCUSSION

## *6.1 Outcome of POS representations*

The results of the experiments reveal some interesting numbers. Across all classifier methods we see the addition of POS representations not only accounts for many top ranking accuracies for the domains, but will often increase the accuracy of a distribution, even if it is not the top accuracy for that domain. While the percentage point increases and decreases may seem subtle, in the field of computational linguistics, single percent points can be quite substantial. Given very large datasets, a 2.00% increase in accuracy could be the difference in a large amount of documents being correctly classified. Additionally, we see that the addition of the selected set results in only a couple of high ranking results where the same high ranking result was not already achieved by the corresponding n-gram distribution alone. We see this happen in the SVM classifier for the domains of *Kitchen & Housewares* and *Music*.

It is my belief that the domain and data will be strong indicators as to what classification method is best. It would seem, however, that the easiest method of determining which classifier method is appropriate for your domain and data, would be to run experiments as I have done in this paper.

## *6.2 Domain Adaptation*

It is interesting to see that the selected set, which performed badly in the domain classification experiments, does significantly better in the domain adaptation experiments. In some cases, it considerably increases the accuracy up to 8.50% (Naive Bayes, *Electronics* -> *Books*), while at its worst, it effects the accuracy by -10.75% (SVM, *DVD* -> *Kitchen*). It would appear that like the individual domain experiments, the best distribution is variable to the domains involved and that the top accuracy may occur with the addition of POS sentence representations, the selected set, or no addition whatsoever. It was shown how the selected set can explicitly change classification in domain adaptation for the better. Even so, the numbers indicate that POS sentence representations and the selected set as they are constructed in this thesis are not a blanket improvement and their use should be considered on a classifier and domain level.

### *6.3 Outlying Domains*

The *Tools & Hardware* domain shows an accuracy of 100% in every variant of the classifiers. This is due to the small sample size of the training and test data and that the training data sufficiently encompasses the n-grams that present themselves in the test data. It does not contribute to the findings, but is included to show the entirety of the Amazon product reviews dataset. The *Computer & Video Games* domain is also different in the fact that the accuracies are substantially higher than other domains, with the lowest being 94.12%. After manually reviewing many of the training and test data reviews, I encountered several reviews similar to the following: "Its patheticic. You cant do anything, graphics are horrble", "This is without a doubt the worste game I have ever purchased.", and "This game is a joke! It is immpossible for a child to get past level 3.". These examples show misspellings and unambiguous sentiment. While the misspellings may contribute to some classification errors, overall, I believe that the misspellings coupled with the amount of profanity seen in the reviews may indicate that the domain is comprised of reviewers that are in a younger age distribution. It is my theory that this contributes to the number of sentences that have numerous occurrences of one-sided polar adjectives and are less likely to be ambiguous regarding sentiment in the way the classifiers are looking to detect it. More review is needed to confirm this theory.

# Chapter 7

# CONCLUSION

I believe that the experiments done as a part of this thesis help show the importance of sentence structure as a component in a sentiment classifier. It was shown how shallow sentence representations created by POS tags improve accuracies in several domains across differing classifier methods and n-gram distributions. It was exciting to see the positive and negative counts in Table 3.2 and Table 3.3 as they identify particular sentence structures that strongly lean toward a particular polarity. This suggests that there are indeed cases where particular sentiment is conveyed in a specific grammatical way. It is also important to mention how much the domain comes into play in regard to the best n-gram set for a given classifier. With some domains, unigrams yielded the best accuracies, while in others it was the combination of unigrams+bigrams. Being aware of which n-gram distribution is best suited for your source and target domains can be of significant importance when performing domain adaptation. Likewise, we saw in Table 5.9 the importance of selecting a classifier method that is best suited to the domain in question. It can be surmised by the experiments that word features do indeed tell a lot in regard to sentiment analysis, but I believe that sentence structure is a significant piece of the puzzle for solving the remaining edge cases and the nuances we see in language. The better we solve these edge cases, the better applications we can produce to mimic human understanding of language.

## *7.1 Continued Work*

As shown by Socher et al. (2013), more fine-grained sentence structures help in better determining the role of sentence structure in sentiment analysis. I believe they are on a promising path by using recursive neural networks and that further improvement may be had by delving into deep learning approaches[2]. I, also, would like to obtain finer grained sentence structures in this manner, and see if calculations such as graph edit distance may glean interesting results. Secondly, I would like to see how the use of a different POS tagger model, such as the GATE Twitter part-of-speech tagger, may affect the experiments. Thirdly, I would like to combine the approaches, if feasible, of Pan et al. (2010), Glorot et al. (2011), and Socher et al. (2013) on the same Amazon product reviews dataset using a SVM classifier. In doing so, I would like to see if they work well together and increase the accuracy of domain adaptation more so than they do by themselves. Lastly, I would like to expand upon and open source the code I used in my experiments to provide a framework for others who are seeking to find the best combination of classifier method and n-gram distribution suited to their domain.

# BIBLIOGRAPHY

[1] Awais Athar. Sentiment analysis of citations using sentence structure-based features. In *ACL (Student Session)*, pages 81–87, 2011.

[2] Yoshua Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.

[3] Adam L Berger, Vincent J Della Pietra, and Stephen A Della Pietra. A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1):39–71, 1996.

[4] John Blitzer, Mark Dredze, and Fernando Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, volume 7, pages 440–447, 2007.

[5] John Blitzer, Ryan McDonald, and Fernando Pereira. Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 120–128. Association for Computational Linguistics, 2006.

[6] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:1–27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[7] Paula Chesley, Bruce Vincent, Li Xu, and Rohini K Srihari. Using verbs and adjectives to automatically classify blog sentiment. In *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pages 27–29, 2006.

[8] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[9] Leon Derczynski, Alan Ritter, Sam Clark, and Kalina Bontcheva. Twitter part-of-speech tagging for all: Overcoming sparse and noisy data. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, 2013.

[10] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 513–520, 2011.

[11] Marti A. Hearst, ST Dumais, E Osman, John Platt, and Bernhard Scholkopf. Support vector machines. *Intelligent Systems and their Applications, IEEE*, 13(4):18–28, 1998.

[12] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006.

[13] Hal Daume III and Daniel Marcu. Domain adaptation for statistical classifiers. *Journal of Artificial Intelligence Research*, 26(1):101–126, 2006.

[14] Jing Jiang and ChengXiang Zhai. Instance weighting for domain adaptation in nlp. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 265–271, 2007.

[15] Dan Klein and Christopher D. Manning. Accurate unlexicalized parsing. *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pages 423–430, 2003.

[16] Shoushan Li and Chengqing Zong. Multi-domain adaptation for sentiment classification: Using multiple classifier combining methods. In *Natural Language Processing and Knowledge Engineering, 2008. NLP-KE'08.*, pages 1–8. IEEE, 2008.

[17] Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330, 1993.

[18] Andrew Kachites McCallum. MALLET: A Machine Learning for Language Toolkit. `http://mallet.cs.umass.edu/`, 2002.

[19] Chris Nicholls and Fei Song. Improving sentiment analysis with part-of-speech weighting. In *Machine Learning and Cybernetics, 2009 International Conference on*, volume 3, pages 1592–1597. IEEE, 2009.

[20] Sinno Jialin Pan, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang, and Zheng Chen. Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th International Conference on World Wide Web*, pages 751–760. ACM, 2010.

[21] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics, 2002.

[22] Jason D Rennie, Lawrence Shih, Jaime Teevan, and David R Karger. Tackling the poor assumptions of naive bayes text classifiers. In *ICML*, volume 3, pages 616–623. Washington DC), 2003.

[23] Irina Rish. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46, 2001.

[24] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. *EMNLP*, 2013.

[25] Songbo Tan, Xueqi Cheng, Yuefen Wang, and Hongbo Xu. Adapting naive bayes to domain adaptation for sentiment analysis. In *Advances in Information Retrieval*, pages 337–349. Springer, 2009.

[26] Kristina Toutanova and Christopher D Manning. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13*, pages 63–70. Association for Computational Linguistics, 2000.

[27] Tong Zhang. Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the twenty-first international conference on Machine learning*, page 116. ACM, 2004.

Appendix A

## RESOURCES

- Amazon product reviews dataset

  `http://www.cs.jhu.edu/~mdredze/datasets/sentiment/`

- Stanford Parser

  `http://nlp.stanford.edu:8080/parser/`

- Stanford Part-Of-Speech Tagger

  `http://nlp.stanford.edu/downloads/tagger.shtml`

- LIBSVM: A library for Support Vector Machines

  `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

- MALLET: A Machine Learning for Language Toolkit

  `http://mallet.cs.umass.edu/`

- Penn Treebank Project

  `http://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html`

- GATE Twitter part-of-speech tagger

  `https://gate.ac.uk/wiki/twitter-postagger.html`

# VITA

Anthony Gentile is a consulting web architect and graduate student at the University of Washington. He welcomes your comments to `agentile@uw.edu`.