

**Your Name:**

**Your Andrew ID:**

## **Homework 3**

### **Collaboration and Originality**

1. Did you receive help of any kind from anyone in obtaining your data for this assignment (Yes or No)?  
It is not necessary to describe discussions with the instructor or TA.

If you answered Yes, provide the name(s) of anyone who provided help, and describe the type of help that you received.

2. Did you give help of any kind to anyone in obtaining their data for this assignment (Yes or No)?

If you answered Yes, provide the name(s) of anyone that you helped, and describe the type of help that you provided.

3. Are you the author of every word of your report (Yes or No)?

If you answered No:

- a. identify the text that you did not write,
- b. explain where it came from, and
- c. explain why you used it.

Your Name

Your Andrew ID

## Homework 3

### 1 Experiment #1: Baselines

#### 1.1 Results

Movie							
	Category	NB			SVM		
		P	R	F1	P	R	F1
Baseline 1		0.001	0.001	0.001	0.001	0.001	0.001
	Macro Average						
Baseline 2							
	Macro Average						
Custom							
	Macro Average						

Epinions3							
	Category	NB			SVM		
		P	R	F1	P	R	F1
Baseline 1		0.001	0.001	0.001	0.001	0.001	0.001
	Macro Average						
Baseline 2							
	Macro Average						
Custom							
	Macro Average						

Apple2							
	Category	NB			SVM		
		P	R	F1	P	R	F1
Baseline 1		0.001	0.001	0.001	0.001	0.001	0.001
	Macro Average						
Baseline 2							
	Macro Average						

<b>Custom</b>							
	<b>Macro Average</b>						

## 1.2 Custom Representations

Describe the custom representations that you used for each dataset. Explain why you made these choices.

## 1.3 Analysis

Discuss the differences among the three baselines and the two different algorithms. Pay particular attention to differences in accuracy, and what may have caused them.

## 2 Experiment #2: Bigram vs. POS Features

### 2.1 Results

<b>Movie</b>							
	<b>Category</b>	<b>NB</b>			<b>SVM</b>		
		<b>P</b>	<b>R</b>	<b>F1</b>	<b>P</b>	<b>R</b>	<b>F1</b>
<b>Custom Bigram</b>		0.001	0.001	0.001	0.001	0.001	0.001
	<b>Macro Average</b>						
<b>Custom POS</b>							
	<b>Macro Average</b>						

<b>Epinions3</b>							
	<b>Category</b>	<b>NB</b>			<b>SVM</b>		
		<b>P</b>	<b>R</b>	<b>F1</b>	<b>P</b>	<b>R</b>	<b>F1</b>
<b>Custom Bigram</b>		0.001	0.001	0.001	0.001	0.001	0.001
	<b>Macro Average</b>						
<b>Custom POS</b>							
	<b>Macro Average</b>						

<b>Apple2</b>							
	<b>Category</b>	<b>NB</b>			<b>SVM</b>		
		<b>P</b>	<b>R</b>	<b>F1</b>	<b>P</b>	<b>R</b>	<b>F1</b>
<b>Custom Bigram</b>		0.001	0.001	0.001	0.001	0.001	0.001

	<b>Macro Average</b>						
<b>Custom POS</b>							
	<b>Macro Average</b>						

## 2.2 Analysis

Discuss your results. Do phrase features help? Is there any difference in accuracy and/or the number of features generated? Does each dataset and/or learning algorithm behave the same way?

## 3 Experiment #3: Two Classes vs. Three Classes

### 3.1 Results

<b>Apple2 vs. Apple3</b>							
	<b>Category</b>	<b>NB</b>			<b>SVM</b>		
		<b>P</b>	<b>R</b>	<b>F1</b>	<b>P</b>	<b>R</b>	<b>F1</b>
<b>Apple2</b>		0.001	0.001	0.001	0.001	0.001	0.001
	<b>Macro Average</b>						
<b>Apple3</b>							
	<b>Macro Average</b>						

### 3.2 Analysis

Discuss your results. How does the addition of an extra class affect the learning algorithms? Which pairs of classes are most often confused? Why might that be?

## 4 Analysis #1: Sentiment Analysis Pipeline

As discussed in class, sentiment analysis systems often contain a sequence of classifiers. The three epinions datasets used in HW2 and HW3 are examples of data encountered in different parts of a sentiment analysis system for Ford automobiles. epinions1 is a first stage that finds posts **about automobiles**. epinions2 is a second stage that finds posts **about Ford automobiles**. epinions3 is a third stage that finds posts that **recommend Ford automobiles**.

Discuss the differences that you observed between classification accuracy on the **three** epinions datasets used in HW2 and HW3. What trends did you observe across the various experiments? Which datasets are easier or harder, and why? Are there interesting or meaningful differences in the effectiveness of different text representations or learning algorithms on these three datasets?

You may find it convenient to include a table or two that summarize(s) results from HW2 and compare(s) them to results from HW3 to support your arguments.

## **5 Effectiveness of the Different Representational Choices**

Do you see any trends? Do some techniques consistently work better than others? Do any of the choices affect the stability of results? Do any of the differences seem significant?

There is some overlap between this section and observations that you may make in Experiments 1-3. However, in this section you have an opportunity comment on trends that span different experiments, which you can't really do in the earlier sections.

## **6 Effectiveness of the Different Learning Algorithms**

Do you see any trends? Does one algorithm consistently work better than another? Is there a difference in the stability of the algorithms? Do any of the differences seem significant?

There is some overlap between this section and observations that you may make in Experiments 1-3. However, in this section you have an opportunity comment on trends that span different experiments, which you can't really do in the earlier sections.

## **7 Sentiment Analysis as a Text Categorization Problem**

Discuss how the results in HW3 relate to the results in HW2. Was sentiment detection in HW3 easier or more difficult than the categorization done in HW2? Why, or why not? What could be done to improve the accuracy of sentiment analysis of these datasets?

## **8 Text Analysis Problem**

Suppose that you work for Virgin Mobile Australia. The company is able to monitor the web browsing behavior of its mobile phone customers. Your job is to use this information to build a profile of each customer's interests across about 100 broad topic categories that Virgin Mobile can use for marketing and advertising purposes. You decide that your topic categories will be the top two levels of the Open Directory Project (DMOZ). Describe how you would accomplish this task. Be clear about what data is used, how you obtain the data, how the data is used to create a profile of each customer, and what a customer profile looks like. Your solution must be scalable to a large population of customers.