Tutorial 4: Lexical semantics and Text Classification

Status: Completed
Your identity is visible during marking.
Marks: 8 / 8

Submission deadline: 30 Aug 2019 23:55, 45 days left

Responsible for the exercises: Bill Noble (bill.noble@gu.se)

Part I: Lexical semantics

Q1: WordNet

Consider this sentence:

> *Swedes like to fish for sea bass.*

Q1a - Using WordNet, determine how many senses there are for each of the open-class words in this sentence. How many distinct combinations of senses are there for this sentence?

Q1b - Now tag each open-class word in the sentence with its correct tag (WordNet sense number). Was choosing the correct sense always a straightforward task? Report on any difficulties you encountered. **Note:** You will have to select *show sense numbers* in the display dropdown.

Q2: FrameNet

Consider the following sentence:

> *Eva baked some saffron buns for Tom using her new oven on Santa Lucia (13th December).*

Q2a - Search the Berkeley FrameNet for a suitable frame which you think fits this situation. Report the name of the frame you chose, listing all of the frame elements which you think apply to this sentence and the parts of the sentence they correspond to.

Q2a - Rewrite the sentence with the same frame and core elements, but three new non-core elements.

Part II: Text classification

Q3. Naïve Bayes WSD:

[See: lecture 9, p.23]

Note: You should be able to do this exercise by hand. If you use code, make sure it's clear how you got each answer.

Let's build a word sense disambiguation model for the word cricket.

We have the following two senses (the prior probability is given):

	sense	**prior probability**
v1	cricket#1	P(v1) = 0.4
v2	cricket#2	p(v2) = 0.6

We will use the following words as features in a bag-of-words representation of context.

	word
a1	score
a2	game
a3	bat
a4	insect

And we are given the following training data:

a1	a2	a3	a4	label
1	1	0	0	v2
0	0	1	0	v2
1	1	0	0	v2
1	0	1	1	v2
0	0	1	0	v1
0	0	0	1	v1
1	0	0	1	v1
0	0	0	1	v1

Q3a - For each ai and vj, compute the conditional probabilities P(ai=0 | vj) and P(ai=1 | vj):

attribute	value	v1	v2
a1	0		
	1		
a2	0		
	1		
a3	0		
	1		
a4	0		
	1		

Here is our testing data:

1. the main object of **cricket#2** is to score the most runs
2. the early form of **cricket#2** differed from the modern game in key aspects
3. **cricket#2** is a bat and ball game played between two teams of eleven players
4. in **cricket#2** you score runs by striking the ball bowled at the wicket with the bat
5. a bat is a flying mammal that eat insects such the **cricket#1**

6. in international **cricket**#2 the game is adjudicated by three umpires
7. **cricket#1**

Q3b - For each sentence in the testing data, use the Baysian model defined by the *prior probabilities* for vj (given) and the *conditional probabilites* for aj (part a).

Q3c - Complete the confusion matrix for **cricket#2** [**note**: In this case a *negative* means guessing cricket#1]**.** Compute precision, recall, and F-score.

Q4. If you have a hammer...

Choose two of the following tasks, and explain how you would model them as a classification task.

- Diacritic restoration (e.g. *cote* to *côté, alska* to *älska*)
- Language identification
- Lexical selection in machine translation
- Sentiment analysis
- Spam filtering

No need to write code, just explain on a high level: **(a)** what features are relevant, **(b)** if you think a certain classifier would be particularly suitable (or unsuitable) for the task, and **(c)** what kind of training data you would need.

Bill Noble , 8 Jan 2019 17:24

Status set to: Revision required

Hemanth Kumar Battula , 22 Jan 2019 17:31

File name: Tutorial4.pdf Listen (567,8 KB)

Status set to: Revision submitted

Bill Noble , 6 Feb 2019 13:09

Status set to: Completed

Mark set to: 8

Grade set to: G

Comment: Q1 - Good
Q2 - Good
Q3 - Good
Q4 - Very good!