COM6115 Text Processing (Autumn semester 2024-25)

Week 7 to week 11 recap

1. Why is Sentiment Analysis a relevant task in text processing?
2. Cite examples of “target of opinions”?
3. Give examples of factual (objective) and subjective sentences.
4. Map the following example to Bing Liu’s model: “I really hated the new Brooks shoes. The material seems cheap and the colour is very ugly. It is also extremely overpriced in comparison to previous editions” Jane Doe, 11th December 2023.
5. Explain the different granularity levels of sentiment analysis.
6. How does the binary lexicon-based approach to sentiment analysis work?
7. Which are the extra rules proposed for the gradable/graded lexicon-based approach?
8. Using the gradable/graded lexicon-based approach, classify the following sentences in positive, negative or neutral:
    a. The book is a very good read.
    b. The new iPhone is not pretty.
    c. I am feeling TERRIBLE today.

Lexicon	
good	2
pretty	2
terrible	-3
Intersifier	
very	1

9. Which are the five rules in VADER? Explain the rule for contrastive conjunction in detail.

10. Give the following training data, with the words highlighted used as features:

Doc	Words	Class
1	A **sensitive**, **moving**, **brilliant** work	**Positive**
2	An **edgy** thriller that delivers a **surprising** punch	**Positive**
3	A **sensitive**, **insightful**, **flamboyant** film	**Positive**
4	Neither **revelatory** nor truly **edgy** – merely crassly **flamboyant** and comedically **labored**	**Negative**
5	**Unlikable**, **uninteresting**, **unfunny**, and completely, utterly **inept**	**Negative**
6	A sometimes **incisive** and **sensitive** portrait, targeting **sensitive** topics, that is undercut by its **awkward** structure	**Negative**
7	It's a sometimes **interesting** remake that doesn't compare to the **brilliant** original	**Negative**

Calculate the the sentiment of "A not **surprising** thriller, with so many **unfunny** and **awkward** moments" using:

a. Multinomial Naive Bayes (without Laplace smoothing)

b. Multinomial Naive Bayes with Laplace smoothing

c. Binary Naives Bayes

11. Cite ways to improve the Naive Bayes approach?
12. Explain how we can evaluate a Naive Bayes classifier?
13. Explain relation extraction. Which are the three types of information that can be extracted?
14. Give examples of classes for Named Entity Recognition (NER).
15. Given the sentence "Google's founder, Sergey Brin is also an employee at Alphabet", which relations can be extracted?
16. Which are the four stages of Wakao et al. NER system?
17. Explain the BIO annotation. How would the following sentence be annotated?

    "The Rock visited London for his new movie and was seen near the Tower of London and the Shard"
18. Cite some features that can be used for BIO classification.
19. Explain the steps of the DIPRE algorithm.
20. Based on the diagram illustrating the word2vec model presented in class, which component is the word embeddings matrix?
21. Why is negative sampling applied in the context of the word2vec model?