**Section 1: Semantics**

**Question 1A (10 Marks)**  
Consider the following sentences:

The **black** horse jumped over the **white** fence.  
The **black** fence was too high for the **black** horse.  
The **white** horse could not have jumped over the **black** fence.

The **black** horse jumped successfully over the **black** fence.

The **white** fence was also too high for the **white** horse.

Using cosine similarity, compute the distributional semantic distance between the vectors for ‘black’ and ‘white’ constructed on the basis of the sentences above with a context window of two tokens. Give the vectors as well as the similarity score.

**Question 1B (5 Marks)**

How can **dependency parsing** be used in a vector space model?

**Question 1C (5 Marks)**

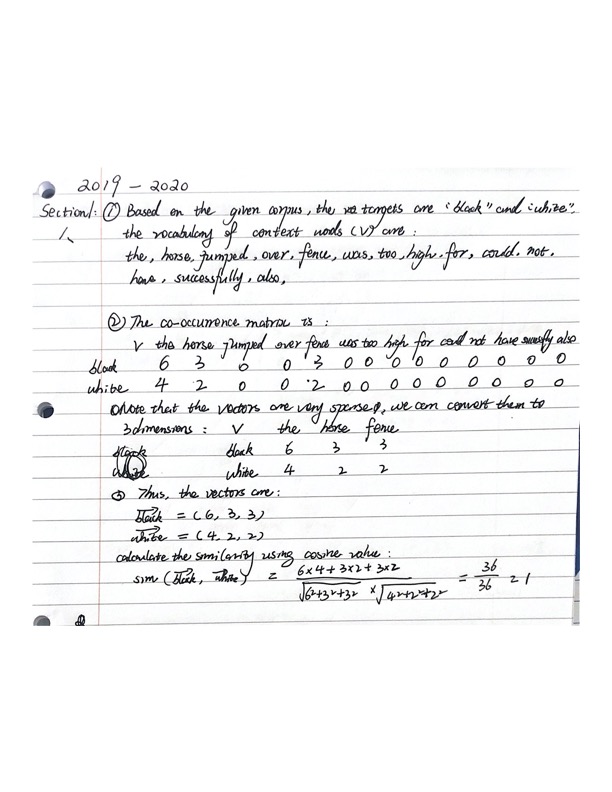
What is the opposite of a hyponym?

**Question 1D (5 Marks)**

Give 2 examples of types of semantic roles and construct a sentence in which both of these can be used. Annotate the relevant words in the sentence for each of the 2 roles.

**Answer:**

1. **Question 1A: 这个题答案错了，应该是前后两个词**

****

1. **Question 1B:**

**Answer:** To get the vector representation of a word in a vector space model, we can define the context around the word and this is where part-of-speech tagging and dependency parsing can be useful. They can be used to restrict what content is being covered in a fixed size context window.

Besides, we can use dependency parsing to define a structured vector space model as dependency parsing can help in identifying useful clusters of words. For example:

eat-V mushrooms-N-DOBJ

eat-V onions-N-DOBJ

eat-V potatoes-N-DOBJ

While defining the context for verbs, if we do dependency parsing and consider the associated N-DOBJ as the context of that verb, we can get an interpretation where words that have similar meanings like mushroom, banana, onion, etc. can end up in a cluster together, while other words like ship, boat, truck, car, etc. can end up in a different cluster, where these words share some characteristics among themselves in the same cluster but don't share similar properties across different clusters.

1. **Question 1C**

**Answer:** Hypernyms are taxonomic relations between word senses in WordNet. They are the opposite of hyponyms which indicate a word with a broad meaning that more specific words fall under. For example, “degree” is a hypernym of “Master’s degree”.

1. **Question 1D**

**Answer:**

AGENT, EXPERIENCER, INSTRUMENT, etc.

For example: [Kristina **AGT**] hit [Scott **EXP**] [with a baseball **INS**].

There are several actors/items (“entities”) with different “roles” in the “hitting” event.

In this example, Kristina is the volitional causer of this event. Scott is the experiencer of this event, and baseball is the instrument used in this event.

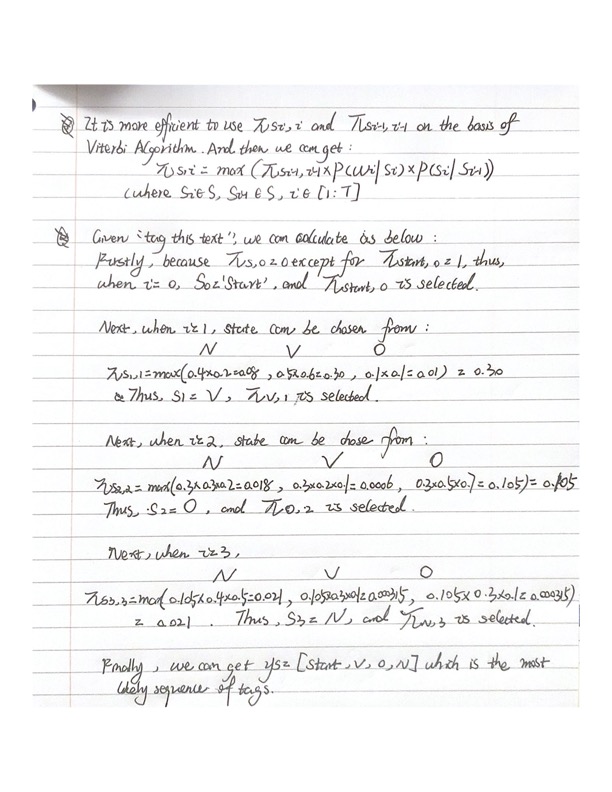
**Section 2: Part-of-speech tagging**

**Question 2A (10 Marks)**

Consider a Hidden Markov Model with the following probabilities (Start designates the

start state):

By using the Viterbi algorithm or otherwise, what is the most likely sequence of tags for the text “tag this text”?



**Question 2B (5 Marks)**

What is the advantage of the Viterbi algorithm over an exhaustive search of all possible

part-of-speech tag combinations?

**Answer:**

* Viterbi algorithm is an efficient method of optimum sequence estimation of a finite-state process. It can avoid exponential complexity and is very efficient. Complexity of an exhaustive search is O(TNT); while complexity of Viterbi algorithm is O(TN2).
* The intuition behind the Viterbi algorithm is to use dynamic programming to reduce the number of computations by storing the calculations that are repeated.
* It enables polynomial time calculation of:

Most likely sequence (and its probability)

Total probability of a sequence (language model)

**Question 2C (10 Marks)**

Give one advantage and disadvantage of unsupervised learning of a Hidden Markov Model by means of the Baum-Welch algorithm as opposed to supervised learning of the model. Why might you wish to combine supervised and unsupervised learning for Hidden Markov Models?

**Answer:**

**Advantage of unsupervised learning:**

* For using Baum Welch (EM-based) HMM to learn the matrices (emission and transition), we do not need manually annotate the dataset which is rather expensive and time-consuming. By contrast, for supervised learning, labels are hard to find and thus the amount of training data is limited.
* Baum-Welch algorithm is an expectation-maximisation algorithm. It generates the most likely hidden transition probabilities as well as the most likely set of emission probabilities.

**Disadvantage of unsupervised learning:**

* The BWA cannot be guaranteed to converge to the global maximum likelihood since it is only proved to converge to a local optimum.
* The computational complexity makes unsupervised learning a very expensive method compared with supervised learning.

**Why to combine supervised and unsupervised learning for HMM?**

A semi-supervised learning method is most widely used for HMM in which some tags are labelled, and some are not. This method is less expensive and easier to obtain a high performance. We can achieve the result based on the labelled tags (simply counting and basic probabilistic calculation) which can increase the training speed and accuracy, and meanwhile, for the left part (the un-labelled ones) the method of unsupervised learning is adopted to explore the data and deliver a high performance.

For example, in speech recognition, we can use semi-supervised learning to train the system on known data and also make it adapt to an individual speaker.

**Section 3: Sentiment Analysis**

**Question 3A (10 Marks)**

Consider the following sentences from product reviews with sentiment scores, using a range of -1, 0, +1:

-1 The food was **great**, but the service in this restaurant was **unfriendly**.

+1 The steak was cooked to **perfection** and the service was **great**.  
0 The **best** part of the meal was the beer.

**Note**: the sentiment lexicon provides a list of positive and negative words. Calculate percentage of positive/negative words in text to be classified. Highest percentage determines sentiment.

**Answer:**

Here, the sentiment scores give an overall sentiment.

We know that the sentiment score of sentence 1 is “-1” and the sentiment score of sentence 2 is “+1”. Also, they both have the word “great”. Thus, “great” is neutral, “unfriendly” is negative, and “perfection” is positive. Besides, the sentiment score of sentence 3 is 0. Thus, “best”.

Overall, the sentiment lexicon which contains a list of positive and negative words can be shown as below:

**Pos [perfection] NEG [unfriendly]**

**Question 3B (10 Marks)**

Recall that a **count vector** represents the proportions of negative and positive sentiment in a review text. Using the sentiment lexicon you created in question 3A, calculate the count vector for each review sentence given above.

**Answer:**

Sentence1: POS:0/12=0, NEG=1/12=0.083

The count vector for sentence one is [POS: 0, NEG, 0.083]

#f(0, 0.083)=NEG

Sentence2: POS:1/11=0.091, NEG=0

The count vector for sentence one is [POS: 0.091, NEG, 0]

3f(0.091,0)=POS

Sentence3: POS:0, NEG=0

The count vector for sentence one is [POS: 0, NEG, 0]

#f(0,0)=Neutral

**Question 3C (5 Marks)**

Aspect-based sentiment analysis:

|  |  |  |
| --- | --- | --- |
|  | aspect | sentiment |
| Sentence1 | food | Neutral |
| Sentence1 | service | NEG |
| Sentence2 | steak | POS |
| Sentence2 | service | Neutral |
| Sentence3 | beer | Neutral |

**Section 4: Information Extraction & Knowledge Graphs**

Consider the following sentence:

“David Robert Joseph Beckham is a former professional footballer who played for Manchester United and the England national team.”

**Question 4A (10 Marks)**

Annotate the sentence above for Named Entity types ‘person’ (PER), ‘organization’ (ORG) and ‘location’ (LOC) by the use of the IOB tagging scheme. Explain the reasoning behind your annotations.

**Answer:**

David (B-PER) Robert (I-PER)Joseph(I-PER) Beckham(I-PER) is (O) a(O) former(O) professional (O) footballer(O) who(O) played(O) for(O) Manchester (B-ORG) United(I-ORG) and(O) the(O) England(B-ORG) national(I-ORG) team(I-ORG).

The entities identified are: David Robert Joseph Beckham-PER, Manchester United-ORG, England national team-ORG. According to the IOB tagging scheme, the beginning of entity is annotated with “B”, the words inside of entity is annotated with “I”, and the words outside of an entity is annotated with “O”.

**Question 4B (10 Marks)**

The sentence above provides a positive instance for extracting the ‘played for’ relation between a ‘person’ (footballer) and an ‘organization’ (football team). Give a negative instance for this relation.

原来是: Positive instance for PER/ORG/POS>POS(PER,ORG)

所以：Negative instance for PER/ORG/POS>POS(PER,ORG)

[PER David Robert Joseph Beckham] said that his friend is a former professional footballer who ***played for*** [ORG the ARSENAL team].

**Question 4C (5 Marks)**

How can clustering be used in taxonomy extraction? What are potential problems with this approach?

Use vector space model

* Construct vectors for all terms (use pre-trained models)
* Compute cosine similarity between vectors
* Build clusters over similar vectors

Challenge: Labelling of clusters unclear and therefore difficult to evaluate.