Home Dashboard My own courses Schools ➤ Course feedback Service Links ➤ Intelliboard ➤ -cd SNLP2024_ ELEC-E5550 - Statistical Natural Language Processing D, Lecture, 9.1.2024-16.4.2024 Feedback Forums Assignments External tools Questionnaires / Exam Course feedback Started on Tuesday, 16 April 2024, 1:01 PM **State** Finished Completed on Tuesday, 16 April 2024, 3:00 PM **Time taken** 1 hour 58 mins **Grade** Not yet graded Question 1 Flag question Marked out of 6.00 Complete a) Describe in detail a method to map documents into vectors so that similar documents would be mapped near each other in the vector space. Remember to describe all processing steps. (4p) b) What are the pros and cons of that method compared to other methods to map documents into vectors? (2p) a) Describe in detail a method to map documents into vectors so that similar documents would be mapped near each other in the vector space. Remember to describe all processing steps. (4p) One of the most popular method to map documents into vectors is TF-IDF (term frequency—inverse document frequency) The documents in the document collection are processed as follows Stage 1: Preprocessing - Lowercasing, removing punctuation, removing stop words, stemming / lemmatizing - Defining word-document matrix, where each entry is defined as follows $W_xy = TF_xy * log (N/ df_x)$ where TF_xy is the count of word x in document y, N is number of documents, and df_x is the number of documents that contains the word x - Weighting and normalizing We can normalize the matrix above or give different weights to some important words. We can also consider bigram, trigram as wel • The queries are then mapped to the same vector space • The relevance is assessed in terms of (partial) similarity between query and document b) What are the pros and cons of that method compared to other methods to map documents into vectors? (2p) Pros of TF-IDF compared to other methods TF-IDF main strength is also its ability to discriminate common from important words for each document. Moreover, TF-IDF is quite effective for information retrieval (IR) tasks. It is also simple to interpret and very fast to calculate compared to models like BERT Cons of TF-IDF compared to other methods TF-IDF only focuses on the frequency of words without understanding the context, and it cannot also tell apart synonyms as well. Additionally if the corpus is very large, the weight matrix would become incomprehensibly large and it is quite impractical for storing nor information retrieving. This would not be the case if we use context models like BERT Question 2 Flag question Marked out of 6.00 Complete a) Describe in detail how chatbots differ from goal-oriented dialogue agents (2p) b) Describe at least one automatic and one human evaluation method often used to evaluate chatbots (2p) c) What are the pros and cons of each method compared to the others (2p). a) Describe in detail how chatbots differ from goal-oriented dialogue agents (2p) The main difference is the general purpose of human use for them. Chat bot is a system that users can chat with, whose discussion topics are fixed and there is no specific goal to be solved. It can detect discussion topics, assume some persona and ask corresponding typical questions. On the other hand, dialogue agents help users to reach a specific goal by giving and collecting information by answering and asking questions. It will try to detect the user's wishes and asked humans related questions to that topic in order to reach the goal of solving the problem. b) Describe at least one automatic and one human evaluation method often used to evaluate chatbots (2p) One automatic evaluation method of chatbots is N-gram matching evaluations such as BLEU score. BLUE score helps to prevent common words that produce a good score and penalize short, boring and repetitive answers One human evaluation method of chatbots is dimensions of quality, depending on avoiding repetition, interestingness, making sense, fluency, listening, inquisitiveness, humanness, and engagingness c) What are the pros and cons of each method compared to the others (2p). Pros of chatbots to dialogue agents/Cons of dialogue agents compared to chatbots are - Chatbot is diverse and can respond and adapt to a wide range of casual questions and topics. Dialogue agents may ignore all questions that is irrelevant to the topic that humans are trying to solve. - Dialogue agents can provide biased answers with respect to gender or race if the topic is relevant to gender/race, while chatbots may not have that biased answers. Pros of dialogue agents to chatbots/Cons of chatbots compared to dialogue agents are - Dialogue agents may have longer memory to remember relevant information to solve the problem, while chatbots usually do not have long memory as it is just a casual conversation so there is no need of deep expertise into a topic - Dialogue agents may have nontrivial answers to some information that it does not know, such as providing explanation, or requesting human for more input. Chatbots may simply say "I am unsure, I do not know" and does not provide any directions for humans to continue the chat. Question 3 Flag question Marked out of 6.00 Complete a) Table 1 shows bigram counts of seven words in a text corpus. Table 2 shows the unigram counts for the same words. Calculate bigram probabilities for the following word pairs A and B. Show your calculations. (1p) A: "want to" B: "eat lunch" b) Solve bigram probabilities for following word pairs A and B that do not have bigram examples in the corpus. Apply either one of these two methods: 1. back off to unigram, 2. add-one smoothing. If you use back-off, use back-off weight bw=0.1. Total number of word types in the vocabulary is 2000 and total number of words in the whole text corpus is 500 000. Show your calculations. (2p) A: "Chinese want" B: "lunch lunch" Table 1. Bigram counts. Preceding context word is given in the first column and the current word in the first row.

9 1072 0 13 0 0 780 0 5 want 855 3 14 to 57 19 eat 0 118 1 Chinese 0 0 food 17 0 0 0 lunch 0 0 Table 2. Unigram counts. 3442

eat Chinese food lunch

want to

1212

3123

920

199

want

to

eat

Chinese

For A "want to":

For B "eat lunch":

Question 4

- Bigram count for "want to" = 780

- Unigram count for "want" = 1212

- Bigram count for "eat lunch" = 57

P("to" | "want") = 780/1212 = 0.6435643564

food 1405 450 lunch c) What are the weaknesses of the maximum likelihood method you used in a) and b)? (1p) d) Describe in detail (including updated probability estimate) one method to alleviate these weaknesses. (2p) a) Table 1 shows bigram counts of seven words in a text corpus. Table 2 shows the unigram counts for the same words. Calculate bigram probabilities for the following word pairs A and B. **Show your calculations. (1p)** To calculate the bigram probability for a word pair, we use the formula: P(word2|word1) = count(word1) / count(word1 word2)

- Unigram count for "eat" = 920 P("lunch" | "eat") = 57/920 = 0.06195652174 b) Solve bigram probabilities for following word pairs A and B that do not have bigram examples in the corpus. Apply either one of these two methods: 1. back off to unigram, 2. add-one smoothing. If you use back-off, use back-off weight bw=0.1. Total number of word types in the vocabulary is 2000 and total number of words in the whole text corpus is 500 000. Show your calculations. (2p) I choose this word pairs which have 0 count in the table A "Chinese want" **B** "lunch lunch" I use add-1 smoothing. We have N = 2000 and V = 2000. The formula is

c) What are the weaknesses of the maximum likelihood method you used in a) and b)? (1p)

REF: it can serve as a basis for the establishment of a european constitution

SYS1: she cannot be used as a basis for the installation of a european constitution

d) Describe in detail (including updated probability estimate) one method to alleviate these weaknesses. (2p)

a) Calculate the 4-gram BLEU score for the two translation hypotheses SYS1 and SYS2, given the reference translation REF. Use the same version of BLEU as in the lectures. Show your calculations. (3p)

a) Calculate the 4-gram BLEU score for the two translation hypotheses SYS1 and SYS2, given the reference translation REF. Use the same version of BLEU as in the lectures. Show your calculations. (3p)

• The key downside is BLEU's focus on exact n-gram matches, which may not accurately contain the original expression in student responses. This method will unfortunately penalize varied but still

• Another downside is that BLEU cannot assess the correctness of content aside from lexical similarity, meaning it might miss a serious misconception in student answers that are somehow phrased

Word Pairs: P(word2|word1) = count(word1, word2) + 1/(N+V)

SYS2: it can into a basis for the european constitution

b) What are the weaknesses of the method you used in a)? (1p)

c) Describe in detail one method to alleviate these weaknesses. (2p)

The formula of BLEU score written in Python can be written as follows:

BLEU = Brevity_Penalty * exp(sum(w_n * log(p_n) for n in [1,2,3,4]))

b) What are the weaknesses of the method you used in a)? (1p)

c) Describe in detail one method to alleviate these weaknesses. (2p)

0.2

DT

8.0

0

0

0

NN

0

0.9

0.5

0.4

0.1

0.2

0.1

0.5

0

otherwise valid answers that do not match the reference wording closely.

One method is to use reference sentences for BLUE as really high quality references chosen by humans.

Table 2. Transition probabilities. Each row indicates the probability of moving from one (preceding) tag (first) to each of the following tags (columns).

The BLUE score formulation is (image I paste from the net) Score Calculation in BLEU. Unigram precision $P = \frac{m}{w_t}$ Brevity penalty $p = \begin{cases} 1 & \text{if } c > r \\ e^{\left(1 - \frac{r}{c}\right)} & \text{if } c \le r \end{cases}$ $\text{BLEU} = p. e^{\sum_{n=1}^{N} \left(\frac{1}{N} * \log Pn\right)}$

where the brevity penalty formula is

similarly to the reference.

I calculate

 $BLEU_SYS1 = 0.4453$

BLEU_SYS2 = 0.2585

Brevity_Penalty= exp(1 - r/c) if c < r else 1

Question 5 a) Use the Viterbi algorithm to calculate the most probable part-of-speech sequence for the sentence given in the top row of Table 1. The observation probabilities of the part-of-speech tags computed from an annotated text corpus are given in the next rows of Table 1. The transition probabilities of between the tags are given in Table 2. Show your calculations. (3p) Table 1. Observation probabilities. Each row indicates the probability of observing one part-of-speech tag (VB = verb, JJ = adjective, DT = determiner or NN = noun) for the given word (column) book open 0.6 0.9 DT 0

NN

VB

JJ

DT

NN

0

VB

0

0

0.6

Second Step (word "a"):

Third Step (word "book"):

- "open" as VB

- "book" as NN

Previous activity

- "a" as DT

b) What are the weaknesses of the method you used in a)? (1p) c) Describe in detail one method to alleviate these weaknesses. (2p) a) Use the Viterbi algorithm to calculate the most probable part-of-speech sequence for the sentence given in the top row of Table 1 Initial Step (word "open"): $-V1(VB) = b_VB("open") = 0.6$ $-V1(JJ) = b_{JJ}("open") = 0.4$ $-V1(DT) = b_DT("open") = 0$ $-V1(NN) = b_NN("open") = 0$

V2(DT) = max [0.6 * 0.8, 0.4 * 0, 0 * 0, 0 * 0] * 0.9 = 0.432

V3(VB) = max [0 * 0, 0 * 0, 0.432 * 0, 0 * 0.6] * 0.8 = 0

V3(NN) = max [0 * 0, 0 * 0.9, 0.432 * 0.5, 0 * 0.4] * 0.2 = 0.0432

b) What are the weaknesses of the method you used in a)? (1p)

c) Describe in detail one method to alleviate these weaknesses. (2p)

Therefore, the most probable POS sequence for the sentence "open a book" is:

■ Jurafsky-Martin 3rd (online) edition + resources

Aalto-yliopisto Aalto-universitetet

Aalto University

Nguyen Binh (Log out)

• email: mycourses(at)aalto.fi **Opettajille / Teachers** MyCourses help MyTeaching Support form

The Viterbi algorithm assumes all transition and emission probabilities are obtained beforehand and accurate, which can be problematic if these data are inaccurate or unavailable. Moreover, it

considers only the most likely path, and ignore other near most likely paths, even though it is not entirely sure if those paths are incorrect.

We use smoothing techniques for the transition and emission probabilities, which can help with unknown graphs or unknown transitions

Tuki / Support

students

Opiskelijoille / Students

MyCourses instructions for

 Palvelukuvaus Saavutettavuusseloste **Service**

Palvelusta

About service MyCourses protection of privacy Privacy notice Service description Accessibility summary MyCourses registerbeskrivining Dataskyddsmeddelande • Beskrivining av tjänsten • Sammanfattning av tillgängligheten

MyCourses rekisteriseloste

Tietosuojailmoitus

Finish review

Next activity

I did not participate in the exam, because... ►

Flag question Marked out of 6.00 Complete

Flag question

Marked out of 6.00

Complete