

Quiz zur Klausurvorbereitung

VL Embeddings

Uni Heidelberg

SS 2019

Quiz

- Die folgenden Quizfragen unterscheiden sich von den Klausurfragen
- Sie dienen nur zur Klausurvorbereitung
- In der Klausur werden **keine** Multiple-Choice-Fragen kommen

SkipGram mit negativem Sampling

- A minimiert die Wahrscheinlichkeit von ungesehenen Wortpaaren.
- B maximiert die Wahrscheinlichkeit von ungesehenen Wortpaaren.
- C maximiert die Wahrscheinlichkeit von gesehenen Wortpaaren.

Subsampling von frequenten Worten wird benutzt, um

- A den Lernprozess zu verlangsamen und so lokale Minima zu vermeiden.
- B bessere Repräsentationen für seltene Worte zu lernen.
- C bessere Repräsentationen für unbekannte Worte zu lernen.

GloVe

Welche der folgenden Aussagen sind korrekt?

- A GloVe ist effizienter als eine SVD.
- B GloVe minimiert die Distanz zwischen dem Skalarprodukt zweier Worte und dem log der der beiden Worte im Korpus.
- C Glove gibt der Rekonstruktion von seltenen Worten im Korpus mehr Gewicht.

SkipGram, CBOW, GloVe

Welches Embedding-Modell trainiert Wortvektoren für Zielwort und Kontextworte und behält am Ende nur die Repräsentationen für die Zielworte?

- A SkipGram
- B CBOW
- C GloVe

SGNS

Mit der Anzahl der negativen Samples

- A steigt die Akkuratheit der Ergebnisse
- B steigt der Zeitbedarf beim Training
- C steigt das Risiko des Overfittings

Word2Vec (SkipGram, CBOW)

Wird im Word2Vec-Modell die Position eines Kontextwortes im Kontextfenster berücksichtigt, oder ignoriert das Modell diese Information?

- A Die Position wird berücksichtigt: Worte, die näher am Zentrumswort sind, werden stärker berücksichtigt.
- B Die Position wird berücksichtigt: Worte, die näher am Zentrumswort sind, werden weniger stark berücksichtigt.
- C Die Position eines Wortes im Kontextfenster spielt keine Rolle.

Dependenz-basierte Embeddings

Was ist der Unterschied zwischen dependenz-basierten Wortembeddings und SkipGram?

- A Dependenz-basierte Embeddings erfassen thematische Ähnlichkeit stärker als SkipGram.
- B SkipGram erfasst thematische Ähnlichkeit stärker als dependenz-basierte Embeddings.
- C Dependenz-basierte Embeddings haben höhere Akkuratheit in Analogie-Task als SkipGram.

Trainingsdaten

Was hat Einfluss darauf, welche Arten von Ähnlichkeiten in Embeddings kodiert werden?

- A Die Zusammensetzung des Korpus entscheidet, welche Informationen gelernt werden
- B Der Embedding-Algorithmus entscheidet, welche Informationen gelernt werden.
- C Die Größe des Korpus entscheidet, welche Informationen gelernt werden.

FastText I

- A FastText ist eine Erweiterung von CBOW.
- B FastText ist eine Erweiterung von SkipGram.
- C FastText ist eine Erweiterung von GloVe.

FastText II

Wodurch lernt FastText bessere Embeddings für seltene Worte?

- A Die Repräsentation eines Wortes basiert auf Buchstaben-Ngrammen, die zwischen Wörtern im gleichen Dokument geteilt werden.
- B FastText sagt die Wahrscheinlichkeit eines Wortes vorher, basierend auf der Summe der Worte im Kontext.
- C FastText lernt verlässlichere Repräsentationen mit Hilfe der gewichteten Kleinstquadratmethode (weighted least-squares objective).

Embeddings für Phrasen

Embeddings für Phrasen kann man erzeugen, indem man

- A die Summe der Embeddings für die einzelnen Worte bildet.
- B die Worte merkt und darauf Embeddings trainiert.
- C den Durchschnitt der Embeddings für die einzelnen Worte berechnet.

Stochastic Gradient Descent

- A Der Zeitbedarf für ein Parameterupdate steigt linear mit der Anzahl der Trainingsinstanzen.
- B Der Zeitbedarf für ein Parameterupdate steigt nicht mit der Anzahl der Trainingsinstanzen.
- C Der Zeitbedarf für ein Parameterupdate steigt exponentiell an.

Stochastic Gradient Descent

- A SGD wird oft zusammen mit Backpropagation benutzt.
- B SGD wird oft zusammen mit Adam und BFGS benutzt.
- C SGD ist eine Implementation der Kettenregel (chain rule).

Backpropagation

- A ist eine Prozedur zur Berechnung des Biasterms der Kostenfunktion.
- B ist eine Prozedur zur Berechnung des Gradienten der Kostenfunktion.
- C ist eine Prozedur zur Berechnung der Gewichte im Netzwerk.

Huffman-Bäume

- A werden benutzt, um die Anzahl an Trainingssamples zu minimieren.
- B werden benutzt, um die Anzahl an Parameterupdates beim Training zu minimieren.
- C werden benutzt, um Daten zu komprimieren.

Entropie I

- A Entropie ist die erwartete Anzahl an Bits, die man benötigt, um eine Zufallsvariable mit einem optimalen Kodierungsschema zu kodieren.
- B Entropie ist die erwartete Anzahl an Bits, die man benötigt, um eine Zufallsvariable mit einem suboptimalen Kodierungsschema zu kodieren.
- C Entropie ist die erwartete Anzahl an Bits, die man benötigt, um die Wahrscheinlichkeit einer Zufallsvariable zu berechnen.

Entropie II

- A Negative Log-Likelihood ist das Gleiche wie Entropy
- B Negative Log-Likelihood ist das Gleiche wie Surprisal
- C Negative Log-Likelihood ist das Gleiche wie Cross Entropy

Entropie III

- A Entropie ist immer größer als Cross-Entropie.
- B Cross-Entropie ist immer größer als Entropie.
- C Cross-Entropie ist entweder gleich groß oder größer als Entropie.

The End

Lösungen

Frage	Lösung	Frage	Lösung
1	A, C	10	–
2	B	11	A, B, C
3	A, B	12	B
4	A, B	13	A
5	B	14	B
6	C	15	B, C
7	B	16	A
8	A, B, C	17	C
9	B	18	C