



ممیزی (momayezi@ut.ac.ir)

به نام پروردگار
پردازش زبان طبیعی تمرین سوم
موعد تحویل: ۲۷ اسفند ۹۸



سؤال اول

از بین گزاره های زیر، کدام یک در رابطه با Skip-gram صحیح است؟ به طور مختصر توضیح دهید.

- ☐ از روی کلماتی که در اطراف یک کلمه می آیند، آن کلمه را پیش بینی می کند.
- ☐ بردار نهایی برای یک کلمه از میانگین و یا مجموع بردار ورودی v و بردار خروجی u متناظر با آن کلمه به دست می آید.
- ☐ برای یک مجموعه اسناد کوچک، عملکرد بهتری نسبت به GloVe دارد.
- ☐ از مقدار احتمال وقوع هم زمان کلمات در حالت کلی بهره می برد.

سؤال دوم

ماتریس word-context زیر با سه کلمه chocolate، data و book و سه context شامل cake، the و study را در نظر بگیرید و با توجه به آن به سوالات زیر پاسخ دهید.

	cake	the	study
chocolate	10	20	0
data	8	20	0
book	1	20	10

الف) با استفاده از مقادیر احتمال $P(w)$ ، $P(c)$ و $P(w, c)$ ، به ازای ترکیب هر کلمه و هر context مقدار Maximum likelihood estimation متناظر با آن کلمه و آن context را محاسبه کنید.

ب) با استفاده از قسمت قبل، مقدار PPMI متناظر با ترکیب هر کلمه و هر context را محاسبه کنید.

ج) مقدار شباهت کسینوسی بردارهای PPMI برای (chocolate, data) و (chocolate, book) را محاسبه کنید.

سؤال سوم

بردارهای one-hot و dense را نظر معایب و مزایا با هم مقایسه کنید. همچنین برای هر یک از معیارهای ارزیابی داخلی و خارجی بردارها، یک مثال بزنید. آیا می توانید شرایطی را بیان کنید که در آن استفاده از بردارهای one-hot نسبت به بردارهای dense برتری داشته باشد؟

سؤال چهارم

Word2Vec مجموعه ای از الگوریتم های embedding را در بر می گیرد که کاربرد های فراوانی در خارج از حوزه متن نیز دارد. فرض کنید شما مسئول یک سیستم پیشنهاد دهنده¹ یک فروشگاه آنلاین هستید. این فروشگاه مجموعه داده ای از آیتم هایی که معمولاً به طور همزمان داخل سبد مشتریان قرار می گیرد را در اختیار شما قرار می دهد. چگونه از این مجموعه داده و روشی مشابه Word2Vec برای پیشنهاد آیتم های جدید به مشتریان استفاده می کنید؟

¹ Recommender System

سؤال پنجم

یکی از مسائلی که در زبان و به طور مشخص در پردازش زبان طبیعی وجود دارد، تشخیص جملات طعنه آمیز² در زبان است. چالش اصلی این مسئله در این است که در این گونه جملات، معنای تلویحی جملات مد نظر است که دقیقاً خلاف معنای ظاهری آن هاست. به طور مثال یک کاربر در یک فروشگاه اینترنتی می تواند نظر خود درباره یک کالا را این گونه بیان کند: «این محصول این قدر کیفیت بالایی داشت که همان روز اول خراب شد.» همان طور که مشخص است، در این جا منظور از داشتن کیفیت بالا دقیقاً برعکس معنای ظاهری و کیفیت پایین محصول مد نظر کاربر است. چگونه می توان از Word Embedding برای حل این مسئله استفاده کرد؟

راهنمایی: می توانید از روشی که در مقاله زیر ذکر شده استفاده کنید.

[Sarcastic or Not: Word Embeddings to Predict the Literal or Sarcastic Meaning of Words](#)

لطفاً به قواعد حل تمرین که در CECM قرار داده شده است توجه کنید.

² Sarcastic