

Рецензия

на доклад

"Возможности использования локального контекста в вероятностных тематических моделях"

Доклад Анны Потапенко состоял из двух частей.

Первая часть состояла из обзора методов векторного представления слов, учитывающих их контекст: методы Word2Vec (Skip-Gram) и GloVe. Данные методы позволяют не только представить слова в некотором пространстве малой размерности, но и выявить смысловые зависимости между словами по их контексту.

Для каждого из методов была объяснена его идея и разобрана математическая постановка с некоторыми выкладками.

Вторая часть посвящена дальнейшим направлениям исследований докладчика в области PLSA (Probabilistic Latent Semantic Analysis) и роли описанных выше методов в них.

Одной из идей докладчика является использование "псевдо-документов", построенных с помощью метода word2vec, при выявлении тем на корпусе документов. При использовании "псевдо-документов" вместе с обычными документами, данный подход аналогичен мультимодальной тематической модели.

Далее докладчик изложила мотивацию использования контекста в тематических моделях, связанную с тем, что в наиболее представленные темы часто попадают слова из общей лексики, которые зашумляют терминологию выявленной темы. Регуляризация по контексту может позволить избежать этот казус.

В завершение Анна коротко рассказала про другой способ комбинирования LDA и word2vec — Gaussian LDA.

Доклад был насыщен математическими постановками задач и примерами. Докладчик складно отвечал на частые и всесторонние вопросы.