

Рецензия на доклад Олега Найдина по теме «Обучение векторных моделей без учителя в задачах информационного поиска»

Рецензент: Анна Потапенко

В данном докладе рассматривается задача обучения векторных моделей языка. Эта область активно развивается, каждый месяц появляются новые статьи с теоретическими результатами и успешными прикладными исследованиями. В связи с большим потоком работ, в доклад было сложно включить полный обзор существующих методов с необходимыми для понимания деталями. Докладчик постарался охватить большое число работ: DSSM, N-gram Language Model, Word2Vec, Paragraph Vector, RNNLM, Skip-thought Vectors, и другие. При этом по большинству моделей была сказана лишь основная идея или показана архитектура сети, что было явно недостаточно для аудитории. Не хватило общих постановок задач, логических связей между рассматриваемыми моделями, понимания, почему именно они включены в обзор, а другие нет.

Во второй части доклада были показаны эксперименты на данных одного из соревнований Kaggle с наиболее популярными моделями в данной области: CBOW, Skip-gram и GloVe. Было очень любопытно увидеть и проанализировать эти результаты, однако докладчик слишком быстро проговорил постановку экспериментов (описание данных, меры качества), также этой информации нет на слайдах. В будущем хотелось бы, чтобы цели и постановка экспериментов были четко зафиксированы, тогда результаты будут лучше восприниматься и, что важно, будут более воспроизводимы другими исследователями.

В целом хочется отметить, что докладчик очень хорошо ориентируется в данной области и имеет большой потенциал для проведения интересного исследования. На данном этапе не была озвучена даже предполагаемая тема диссертационной работы или предполагаемый план исследования на ближайшее время. Возможно, именно это сыграло ключевую роль в восприятии доклада и (лично у меня) создало впечатление случайного набора моделей и экспериментов. Хочется пожелать докладчику поскорее определиться с планом исследования и поблагодарить за подборку полезных ссылок, которыми несколько слушателей семинара уже воспользовались!