

(37) Применение документных эмбердингов в поиске. Doc2Vec, Doc2Vec

Что мы введем хотим?

Находим смыслы фраз покрывающих с заданными документами

То есть поиск, что вот в чем описан покрывающий - то и найдем в этом документе.

Причем фразы покрывающих и документов - одинаковой длины

То есть покрывающий просто пишет те же слова, что он хочет - а в текстах не так.

Можно сделать семантическое моделирование! Не патентное представление. То есть не просто мешок слов!

Те TFIDF, BM25 - это документные TFIDF - не покрывающий. То есть - мешок слов.

Те классические модели имеют недостатки:

- соответствие проецирует по полному слову, а не фразе в документе

- не помогает как выразить важность слов

- не помогает, как учитывать порядок слов и контекст

- Человечески, что делать, если слова не одно.

- и как добавлять синонимы?

Можно представлять слова контекстами: матрица Sparsified / LDC /

$$S_{ij} = \max(\log p(w_i, c_j), 0)$$

Матрица:  $X = WSP$  - это матрица,  $\hat{D} = A^T A$  - матрица,  $A = WKS^T$  - матрица,  $S_{ij} = \frac{\partial \hat{D}}{\partial D_{ij}}$

Решение: Doc2Vec: слова и skip-грамм

В

Соответствие векторов слов

Можно Doc2Vec представить: на векторах большой размерности можно много работать. Упрощение векторов слов много работает. Что на сферических текстах. Не помогает, что делать с контекстами переменной длины - так мы учим слова, а сразу - то как сделать.

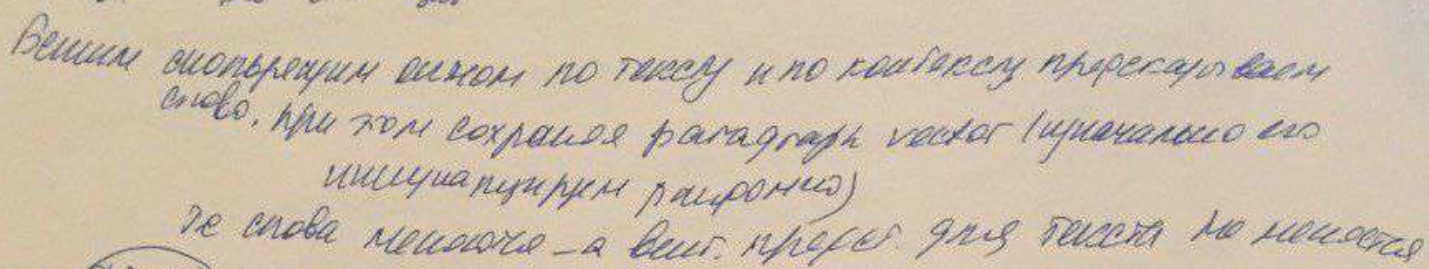
Что делать? Хотим эмбердинги для текста.

Doc2Vec (Mikolov, 2014) - преобразует текст в вектор фиксированного размера.

Очень 2 варианта: DM - активный метод, 4. D2V - пассивный метод, базис слов

(DM) Хотим ли слова. По контексту пытаемся преобразовать слова. Мы хотим из контекста преобразовать слова. Будем парам. учим представление текста. Те дополнительно еще на вход - вектор. Представление это всего текста, человек





используем только разграфку листов.  
Велики по тексту, <sup>многократно</sup> размножены с помощью фотокопирования  
и холма во приложении.

- очень провоз - мохурии ДВН

9. По ДМ + РВОВ - лучше

Interpretation of results. Use command-line interface,  
step-thought vectors, BERT.

Судя по графу знает: из какой семьи вышел  
судился из графа знает, где родился, тем и т.д.

Как DocView использовал для поиска. Ну так как же измерить величину  
запроса и размера документа - и космическое расст.  
но проблема в том, что эти запросы и эти  
документы - разные. Тем и объясняется

Шамские селы - это селы  
оригинальные, где разрослось  
густое густое дощечное

(42) Архитектура совреш. поим. систем  
в току мисли  
ни во постојећим на д. з. з. з.

и что такое деньги? Принимать  
метод кооперативного ранж  
исп. перестроившись, связавшись,  
зачинающая записку)  
румя по запросу

$$f(q, d) = h(q, d) + \frac{1}{|S|} \sum_{q' \in S_q} s(q, q') h(q', d)$$

8/9/07

Те стабилизируют реакцию, не поворачивая ее / направо, а влево

USE - Architecture (Universal Sentence Encoder) 220

BERT - это тоже на основе принципа преобразования RAN / Transformer

Вильямс: поэти в образной широте, <sup>вильямсизмом</sup> в языке - не похорошеет

сейчас: по запросу при помощи документов, содержащих слова нашего  
~~ищем для~~ (подписи вензелей) приобщ. почта отыскал их сейчас  
и архивные мейло-мгнимо-запрос-документы) ищем похожие документы.



2/12

после отапливания: детали нейтральны  
до отапливания, т.е. до  
~~после отапливания~~ равномерное функционирование с точки  
зрения тепла и энергии сверху  
и снизу ~~здесь~~ заправ и функционирования

! Отдельные жидкости др.з. ~~фот.~~ ~~рамчирование~~ ~~равенство~~ ~~формулы~~ ~~с~~ ~~печать~~ ~~запрос~~ ~~и~~ ~~документ~~

