

Анализ и предсказание решения суда с помощью методов машинного обучения

Геллер Марк Аркадьевич
научный руководитель: А.А. Шпильман

СПб АУ НОЦНТ РАН

12 июня 2017 г.

Решения суда можно предсказывать автоматически ¹.

Мотивировочная часть решения - обстоятельства и доказательства по делу, представленные сторонами и иными лицами участвующими в деле.

Разметка мотивировочной части может быть использована юристами при анализе документов.

¹N. Altetras et Al. - 2016

Цель: разработать систему, автоматически определяющую важные части мотивировочной части при принятии решения судом.

Задачи:

- Сбор данных
- Выбор алгоритма
- Произвести обучение
- Оценка качества
- Интерпретация модели
- Визуализация

Данные для обучения были взяты с kad.arbitr.ru.

Около 10000 дел, которые состоят из:

- Вида спора
- Категории спора
- Мотивировочной части решения суда
- Вердикта
- Других данных

Похожая задача — Sentiment Analysis.

Методы, позволяющие определить влияние слов/предложений:

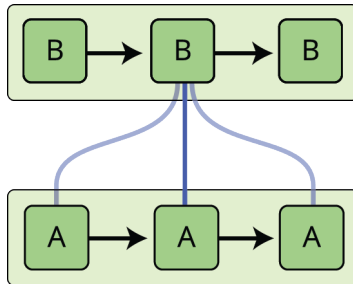
- **Naive Bayessian Model**
- **Bag of Words + Word2Vec**
- **Attention Networks**

Подход из статьи о предсказании решений ЕСПЧ:

- **SVM + n-gram**

Attention - механизм, распределяющий "Внимание" в зависимости от контекста.

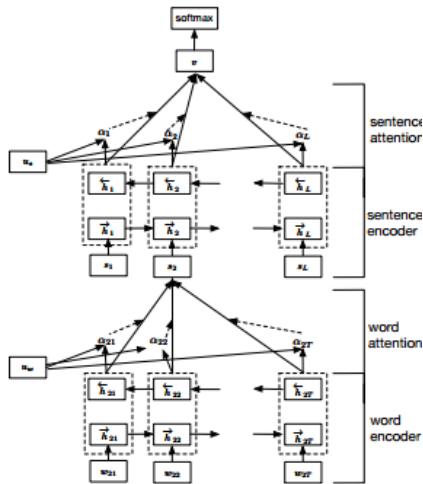
Идея заключается в генерации весов для подсчета взвешенной суммы входа.



Hierarchical Attention Networks

Hierarchical Attention Networks (HAN)¹ - двух-уровневая система дву-направленных LSTM + Attention

Наилучший результат для решения задачи sentiment analysis на нескольких наборах данных



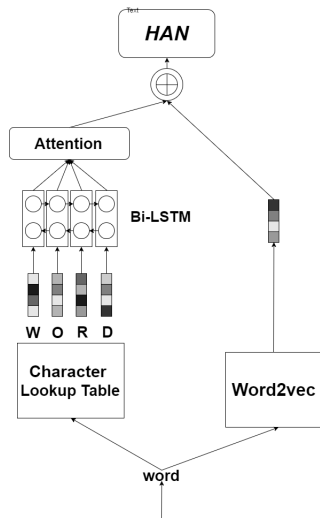
¹Yang et al. — 2016

Финальная Архитектура (Слова + Предложения)

Особенности:

- Малое количество данных
- Язык с богатой морфологией

Character to Word¹ - получение представления слов на уровне символов



¹Ling et al. — 2015

Размер выборки:

- Train - 8000 дел (3 пары вид/категория)
- Validate - 1000 дел (3 пары вид/категория)
- Test (Same) - 1000 дел (3 пары вид/категория)
- Test (Other) - 600 дел (другая пара вид/категория)

Регуляризация - *Dropout*

Метрика качества - *Accuracy*

Сравниваемые методы:

- NB - Naive Bayes (TF-IDF)
- BoW - Average Word2Vec
- SVM + n-gram
- LSTM - Два двунаправленных LSTM слоя
- HAN - Hierarchical Attention Network
- WC2V-HAN - предлагаемая архитектура

	NB	BoW	SVM	LSTM	HAN	WC2V-HAN
Same	74.46	84.12	82.50	92.61	95.69	97.17
Other	66.12	73.95	77.28	88.87	91.31	93.92

Веса предложений (W_s)/слов (W_w):

- *Полные Веса* — вклад в решение сети (W_{sf}/W_{wf})
- *Положительные веса* — вклад в "удовлетворение иска" (W_{sp}/W_{wp})
- *Отрицательные веса* — вклад в "отклонение иска" (W_{sn}/W_{wn})

Ключевые слова — слова с большими абсолютными весами ($W_{sf} \cdot W_{wf}$)

Получение весов — две сети с одним нейроном на выходе. Веса представлены внутренними весами Attention слоя

Истцом представлены доказательства направления уточнения в адрес ответчика. Weights -- Positive: 0.085, Negative: 0.013.

(0.069)	(0.066)	(0.075)	(0.075)	(0.075)	(0.062)
(0.075)	(0.075)				

Сведений опровергающих доводы истца ответчик суду также не представил. Weights -- Positive: 0.042, Negative: 0.127.

(0.093 0.089)	(0.053 0.086)	(0.092 0.090)	(0.085 0.099)	(0.096 0.035)	(0.013 0.013)
(0.013 0.013)	(0.013 0.013)	(0.013 0.013)			

Наличие долга подтверждено материалами дела ответчиком данный факт не оспаривается. Weights -- Positive: 0.008, Negative: 0.055.

(0.076)	(0.076)	(0.076)	(0.076)	(0.076)	(0.076)	(0.076)
(0.039)	(0.010)	(0.010)				

Определением от 28 07 2016 рассмотрение дела назначено в порядке упрощенного производства.

Создан инструмент, выделяющий важные части в мотивировочной части решения.

Получены оптимальные результаты в задаче классификации собранных данных.

Проведен экспертный анализ размеченных документов.

Ссылка на материалы: <https://yadi.sk/d/se9a17K23K337X>