

Модели, методы и алгоритмы, устойчивые к шуму, в задачах обработки естественного языка

Валентин Андреевич Малых

Специальность 05.13.11 – Математическое и программное обеспечение машин, комплексов и компьютерных сетей

Диссертация на соискание ученой степени кандидата технических наук

План

1. Моделирование шума
2. RoVe: устойчивые к шуму вектора слов
3. Прикладные задачи обработки текста
 - a. Устойчивая к шуму классификация текстов
 - b. Устойчивое к шуму распознавание именованных сущностей
 - c. Устойчивое к шуму извлечение аспектов

Актуальность исследования

- В мире существует множество шумных текстов (содержащих опечатки).
- Существующие решения по проверке орфографии не идеальны
 - для русского языка исправляется 85% ошибок
- Предлагается альтернативный подход - делать системы, которые изначально устойчивы к шуму.

Основные задачи

1. Разработать устойчивые к шуму векторные представления слов.
2. Исследовать устойчивость к шуму существующих моделей классификации текстов и предложить расширения, устойчивые к шуму.
3. Исследовать устойчивость к шуму существующих моделей распознавания именованных сущностей и предложить расширения, устойчивые к шуму.
4. Исследовать устойчивость к шуму существующих моделей извлечения аспектов и предложить расширения, устойчивые к шуму.

Научная новизна

1. Разработаны новые устойчивые к шуму векторные представления слов.
2. Предложена новая архитектура классификации текстов, устойчивая к шуму.
3. Предложена новая архитектура распознавания именованных сущностей, устойчивая к шуму.
4. Предложена новая архитектура извлечения аспектов, устойчивая к шуму.

План

1. Моделирование шума
2. RoVe: устойчивые к шуму вектора слов
3. Прикладные задачи обработки текста
 - а. Устойчивая к шуму классификация текстов
 - б. Устойчивое к шуму распознавание именованных сущностей
 - с. Устойчивое к шуму извлечение аспектов

Шум

- Опечатки, орфографические ошибки называются **шумом**.
- Для определения величины зашумленности слова берется расстояние Дамерау-Левенштейна (с перестановкой букв) от шумной словоформы до исходной.
- Под исходной словоформой понимается грамматически и орфографически корректная словоформа в данном контексте.
- Под шумной словоформой понимается, соответственно, любая отличающаяся от исходной словоформы.

Моделирование шума

- Вероятность шума в реальных текста по некоторым оценкам достигает 10%
- Модель шума выбрана по аналогии с моделями, используемыми в литературе по моделям проверки орфографии.
- Для исследуемых языков не существует открытых корпусов текстов с исправленными опечатками.

Моделирование шума

$B(1,p)$ - биномиальное распределение,

$U\{1,|A|\}$ - равномерное распределение,

$$p \in [0, 0.3]$$

$|A|$ - длина алфавита

Виды шума:

- удаление текущего символа с некоторой вероятностью **$B(1,p)$**
- добавление произвольного символа **$U\{1,|A|\}$** после текущего с некоторой вероятностью **$B(1,p)$**
- замена текущего символа на произвольный **$U\{1,|A|\}$** с некоторой вероятностью **$B(1,p)$**
- замена соседних букв друг на друга с некоторой вероятностью **$B(1,p)$**

План

1. Моделирование шума
2. RoVe: устойчивые к шуму вектора слов
3. Прикладные задачи обработки текста
 - а. Устойчивая к шуму классификация текстов
 - б. Устойчивое к шуму распознавание именованных сущностей
 - с. Устойчивое к шуму извлечение аспектов

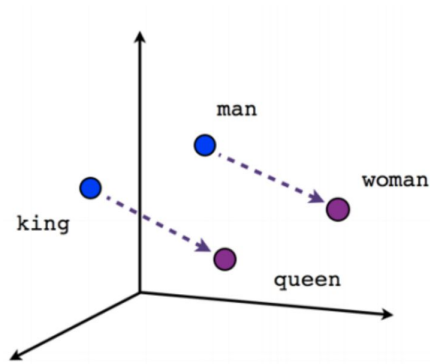
Векторные представления слов

- Слова не могут быть восприняты компьютером, как человеком. Требуется числовые представления.
- Простые представления слов на основе нумерации по словарю являются недостаточными, так как не учитывают многих важных аспектов, например, семантики. Требуется более сложные векторные представления. Примером может служить модель Word2Vec.

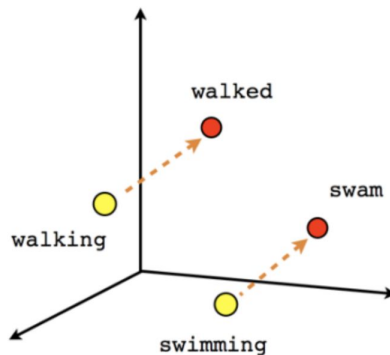
motel [0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0] AND
hotel [0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0] = 0

Векторные представления слов

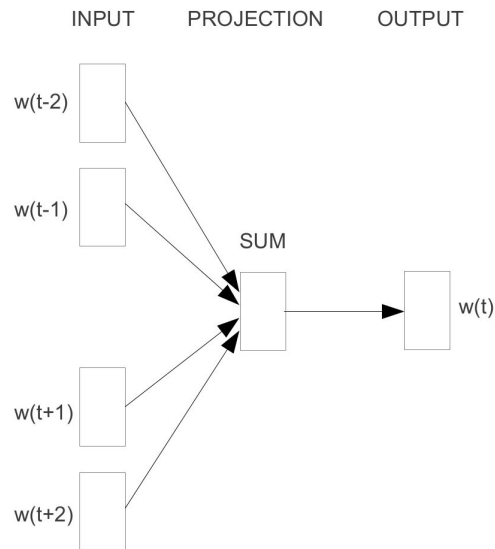
- Word2Vec - работает с фиксированным написанием слова, не учитывают контекст
- fastText - работает с n-граммным представлением слова в дополнение к фиксированному, не учитывает контекст



Male-Female



Verb tense



Устойчивые к шуму вектора слов

$\|$ - конкатенация векторов

$c_1 \dots c_k$ - вектора символов (в слове w),
состоящие из нулей и одной 1 по номеру
символа в алфавите

n_b - длина префикса, n_e - длина суффикса,
выбранные по предварительным
экспериментам

enc - некоторая функция, например, RNN, C_{left} и C_{right} - левый и правый
контексты

$$B(w) = c_1 \| \dots \| c_{n_b}$$

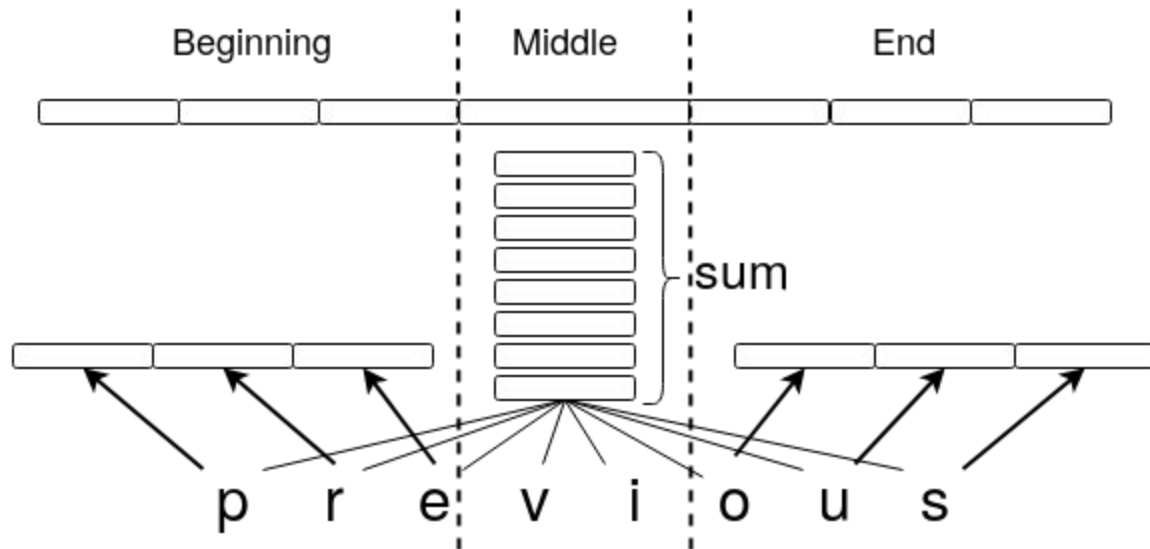
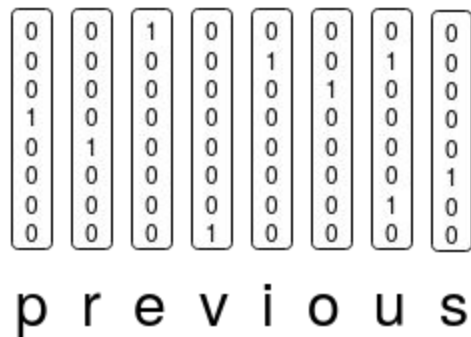
$$E(w) = c_{k-n_e} \| \dots \| c_k$$

$$M(w) = \sum_{i=1}^k c_i$$

$$BME(w) = B(w) \| M(w) \| E(w)$$

$$RoVe(w) = enc(BME(w); C_{left}, C_{right})$$

Устойчивые к шуму вектора слов

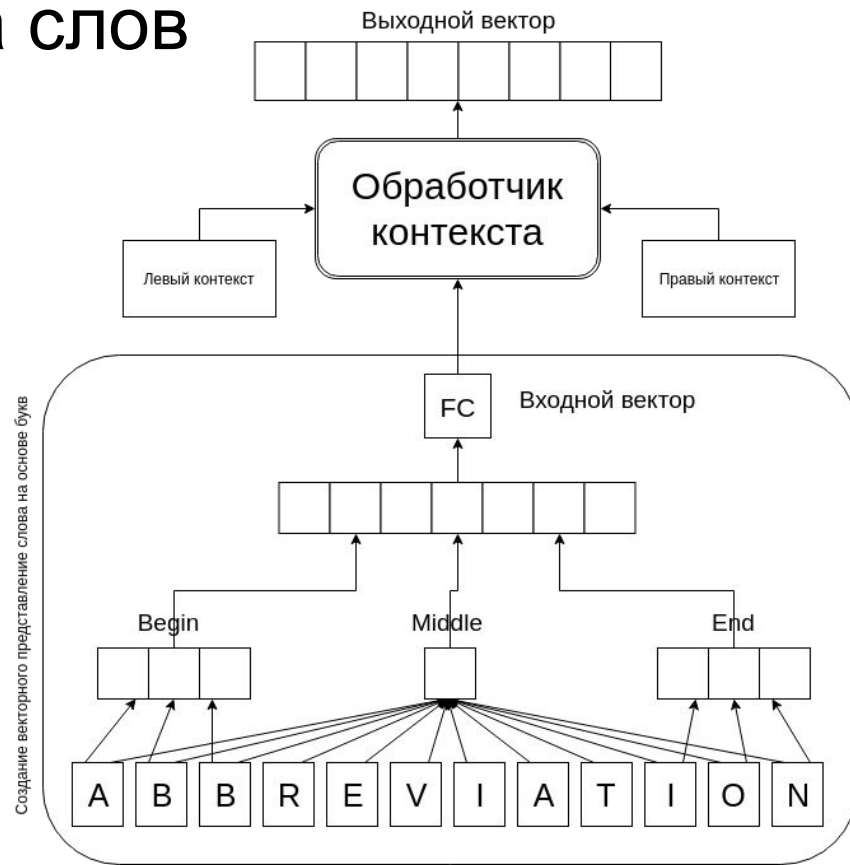


получение BME представления для слова “previous”

Устойчивые к шуму вектора слов

получение векторного представления
для слова “abbreviation” с учетом
контекста

левый и правый контексты - это
состояние обработчика контекста
после обработки
предыдущих/следующих слов



Обучение модели

Сэмплирование негативных примеров:

$$L(x) = \log\left(\sum_{i \in C} e^{-s(x, w_i)}\right) + \log\left(\sum_{j \notin C} e^{s(x, w_j)}\right)$$

$L(x)$ - функция потерь для позитивных примеров (слова в контексте должны иметь близкие вектора слов) $\log\left(\sum_{i \in C} e^{-s(x, w_i)}\right)$ и негативных примеров (слова *не* в контексте должны иметь далекие вектора слов) $\log\left(\sum_{j \notin C} e^{s(x, w_j)}\right)$

x - центральное слово, w_i - слово в контексте C для x , w_j - слово вне контекста C для x

$s(x, w)$ - функция сходства слов, например, косинусное расстояние между векторами

Тестирование модели RoVe: задачи и языки

Языки:

- *английский* (изолирующий язык),
- *русский* (флективный язык),
- *турецкий* (агглютинативный язык)

Задачи:

- Распознавание парафраз
- Анализ тональности
- Распознавание логического следования

Задача распознавания парафраз

Задача распознавания парафраз: на вход модели подается пара предложений, на выходе ожидается метка класса 0 - не парафраз, 1 - парафраз. Пример:

В Пакистане застрелили одного из лидеров группировки "Хаккани".

В Пакистане застрелили финансиста сети «Хаккани».

1 (парафраз)

Корпуса, на которых производилось тестирование:

- Microsoft Research Paraphrase Corpus: 5800 пар предложений, английский язык
- Paraphraser.ru: 4400 пар предложений, русский язык
 - нейтральный класс был исключен из рассмотрения
- Turkish Paraphrase Corpus: 700 пар предложений, турецкий язык

Задачи логической связи и анализа тональности

Задача логической связи: на вход подается пара предложений, на выходе ожидается метка класса 0 - логической связи нет, 1 - логическая связь есть.

Корпус:

- Stanford Natural Language Inference: 570152 пары предложений, английский язык

Задача анализа тональности: на вход подается текст, на выходе ожидается метка класса 0 - негативная тональность, 1 позитивная тональность.

Корпуса:

- Stanford Sentiment Treebank: 11855 документов, используется 10662, английский язык
 - нейтральный класс был исключен из рассмотрения
- Russian Twitter Sentiment: 114911 + 111923 документов, русский язык

Постановка эксперимента

- Цель эксперимента - проверить качество векторных представлений слов.
- Задачи распознавания парафраз и логического следования - модель без обучаемых параметров. Для каждого предложения считается среднее от векторов всех слов в предложении. В качестве предсказания используется косинусное произведение этих векторов
- Задача анализа тональности - предсказания производятся наивным Байесовским классификатором, признаки - среднее от векторов всех слов в предложении
- Качество измеряется метрикой ROC-AUC.

Базовые модели и модификации RoVe

- Word2Vec - предобученная модель Word2Vec, обученная на Google News
- fastText - предобученная модель fastText, обученная на CommonCrawl
- fastText+speller - предобученная модель fastText, обученная на CommonCrawl; вход модели обрабатывается системой исправления орфографии (использовался Yandex.Speller)

- SRU, biSRU и stackedLSTM - варианты модуля обработки контекста для модели RoVe

Результаты на корпусе Paraphraser.ru

noise (%)	0	10	20	30
Базовые модели				
Word2Vec	0.800	0.546	0.535	0.647
fastText	0.813	0.645	0.574	0.632
fastText + spell-checker	0.813	0.693	0.525	0.490
RoVe				
stackedLSTM	0.723	0.703	0.674	0.601
SRU	0.823	0.716	0.601	0.647
biSRU	0.841	0.741	0.641	0.718

на задаче определения парафраз для русского языка

Результаты на корпусе Stanford Natural Language Inference

noise (%)	0	10	20	30
Базовые модели				
Word2Vec	0.624	0.593	0.574	0.557
fastText	0.642	0.563	0.517	0.480
fastText+speller	0.642	0.498	0.481	0.482
RoVe				
biSRU	0.651	0.621	0.598	0.536

на задаче выявления логического следования для английского языка

Результаты на корпусе Turkish Paraphrase Corpus

noise	0	10	20	30
Базовые модели				
Word2Vec	0.647	0.586	0.534	0.475
fastText	0.632	0.595	0.514	0.503
RoVe				
biSRU	0.718	0.641	0.587	0.556

на задаче определения парафраз для турецкого языка

Выводы к разделу 2

- Существующие векторные представления слов не являются устойчивыми к шуму.
- Даже использование систем проверки орфографии не позволяет сделать векторные представления устойчивыми к шуму.
- Предложенные векторные представления являются более устойчивыми к шуму.
- Предложенные векторные представления являются независимыми от особенностей языка.

План

1. Моделирование шума
2. RoVe: устойчивые к шуму вектора слов
3. Прикладные задачи обработки текста
 - а. Устойчивая к шуму классификация текстов
 - б. Устойчивое к шуму распознавание именованных сущностей
 - с. Устойчивое к шуму извлечение аспектов

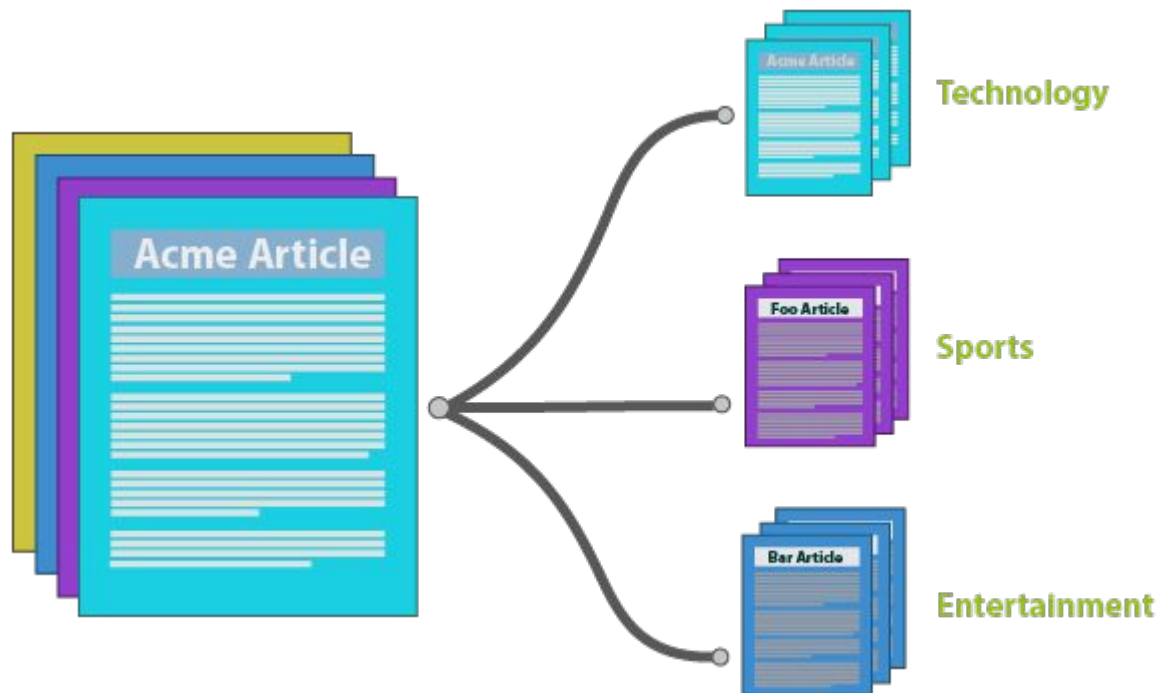
Прикладные задачи обработки текста

- Векторные произведения являются базовой задачей, напрямую не применимой на практике
- Произведено исследование 3 прикладных задач обработки текста, где предложены модели, устойчивые к шуму

План

1. Моделирование шума
2. RoVe: устойчивые к шуму вектора слов
3. Прикладные задачи обработки текста
 - а. Устойчивая к шуму классификация текстов
 - б. Устойчивое к шуму распознавание именованных сущностей
 - с. Устойчивое к шуму извлечение аспектов

Задача классификации текстов



Задача классификации текстов

Задача классификации: на вход модели подается текст, на выходе ожидается метка класса. Оценивание происходит по количеству угаданных меток классов по метрике F1.

Корпуса, на которых производилось тестирование:

- Twitter US Airline Sentiment
 - 14485 твитов
- Movie Review
 - 25000 позитивных и 25000 негативных отзывов

Постановки экспериментов

Задача экспериментов:

- сравнить устойчивость к шуму существующих архитектур для классификации текстов и их расширений;
- показать, что используемый искусственный шум близок к натуральному.

Список экспериментов:

1. Обучающая и тестовая выборки берутся без изменения.
2. Обучающая и тестовая выборки: выполняется проверка орфографии и накладывается искусственный шум.
3. Для обучающей выборки выполняется проверка орфографии и накладывается искусственный шум. Тестовая выборка берется без изменений.

Проверяемые модели

- CharCNN - модель, основанная на векторных представлениях символов и сверточной сети для создания скрытых представлений.
- CharCNN-WordRNN - модель, основанная на векторных представлениях символов и сверточной сети для создания векторных представлений слов. Векторные представления слов обрабатываются рекуррентной нейронной сетью.
- FastText - векторные представления слов порождаются моделью fastText. Векторные представления слов обрабатываются рекуррентной нейронной сетью GRU.
- RoVe - векторные представления слов порождаются моделью RoVe. Векторные представления слов обрабатываются рекуррентной нейронной сетью GRU.

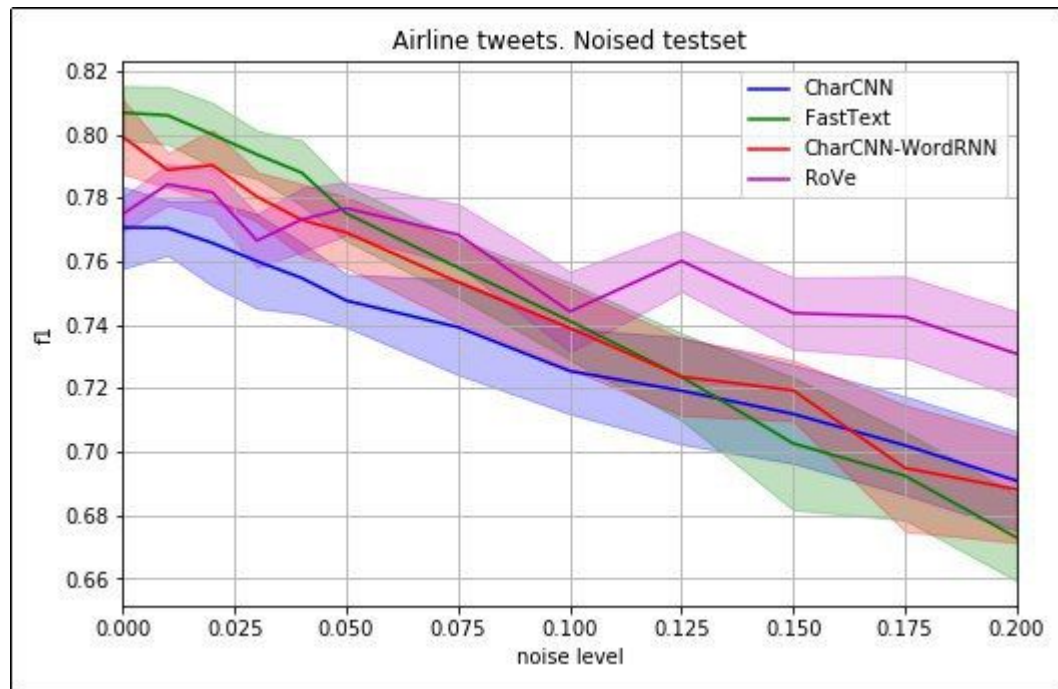
Результаты на исходных данных

Модель	Movie Review	Twitter Sentiment
CharCNN	0.74	0.77
FastText	0.84	0.76
CharCNN-WordRNN	0.80	0.81
RoVe	0.79	0.80

результаты тестирования моделей на исходных корпусах по метрике F1

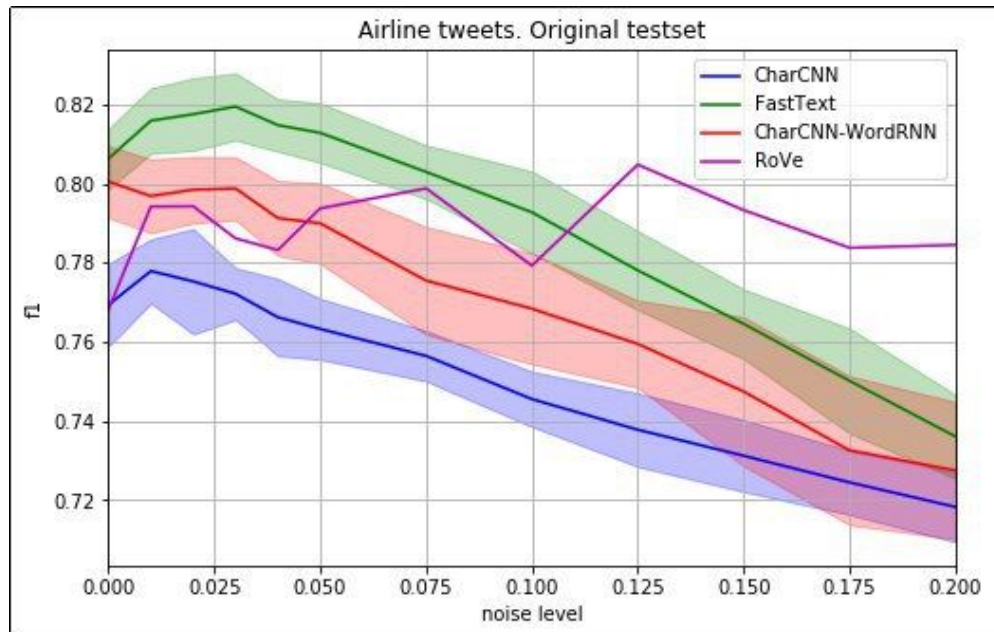
Результаты для Airline Twitter Sentiment (искусственный шум в тестовой выборке)

- модель FastText имеет более высокое качество на низких уровнях шума
- модель RoVe имеет более высокое качество на высоких уровнях шума



Результаты для Airline Twitter Sentiment (естественный шум в тестовой выборке)

- модель FastText имеет более высокое качество на низких уровнях шума
- модель RoVe имеет более высокое качество на высоких уровнях шума
- качество тестируемых систем уменьшается аналогичным предыдущему эксперименту образом



План

1. Моделирование шума
2. RoVe: устойчивые к шуму вектора слов
3. Прикладные задачи обработки текста
 - а. Устойчивая к шуму классификация текстов
 - б. Устойчивое к шуму распознавание именованных сущностей**
 - с. Устойчивое к шуму извлечение аспектов

Задача распознавания именованных сущностей

Input: Vancouver is a coastal seaport city on the mainland of British Columbia. The city's mayor is Gregor Robertson.

Location

Output: Vancouver is a coastal seaport city on the mainland of British Columbia. The city's mayor is Gregor Robertson.

Location

Person

Vancouver/B-LOC is/O a/O coastal/O seaport/O city/O on/O
the/O mainland/O of/O British/B-LOC Columbia/I-LOC.
The/O city's/O mayor/O is/O Gregor/B-PER Robertson/I-PER.

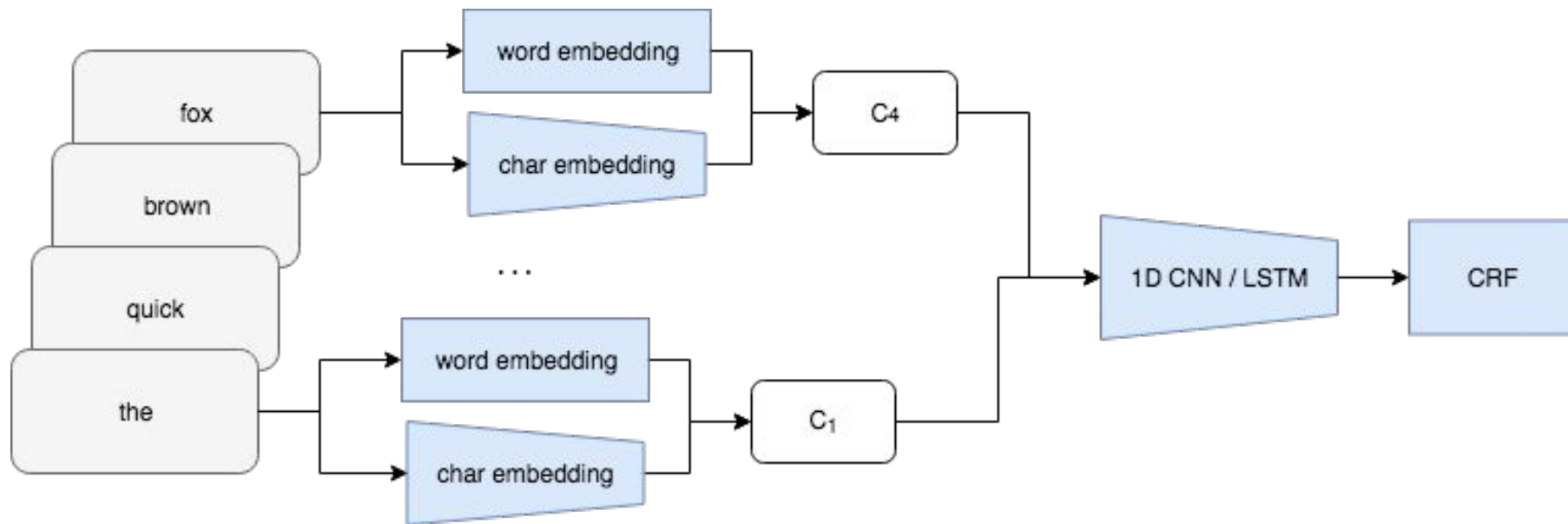
Задача распознавания именованных сущностей

На вход модели подается текст, на выходе для каждого слова ожидается тег. Качество подсчитывается по сущностям (последовательностям тегов).

Корпуса, на которых производилось тестирование:

- Collection 5 (доразмеченный Persons-1000)
 - 1000 документов, размечен 5 тегами LOC, PER, ORG, MEDIA, GEOPOLIT
- CoNLL'03 (английская часть)
 - 946+216+241 документ, размечен тегами LOC, PER, ORG
- CAp'2017 (французский твиттер)
 - 3000+3645 твитов, размечен 13 тегами

Используемая архитектура



применяется для английского и русского языков, впервые применена автором для французского языка

Варианты модели

Комбинация векторных представлений слов (word embeddings) и символов (char embeddings).

Векторные представления слов:

- Word2Vec - инициализация векторами из модели Word2Vec для матрицы векторных представлений слов
- fastText - аналогично для модели fastText
- EmbedMatrix - матрица векторных представлений выучивается в процессе обучения
- RandomEmbed - матрица векторных представлений задается случайно

Векторные представления слов на основе букв:

- nochar - без добавления признаков от побуквенного представления слова;
- CNN - сверточная сеть на уровне символов

Постановки экспериментов

Задача экспериментов:

- проверить устойчивость к шуму существующей лучшей архитектуры для распознавания именованных и ее расширений;
- показать, что используемый искусственный шум близок к натуральному.

Список экспериментов:

1. Обучающая и тестовая выборки берутся без изменения.
2. Обучающая и тестовая выборки: выполняется проверка орфографии и накладывается искусственный шум.
3. Для обучающей выборки выполняется проверка орфографии и накладывается искусственный шум. Тестовая выборка берется без изменений.

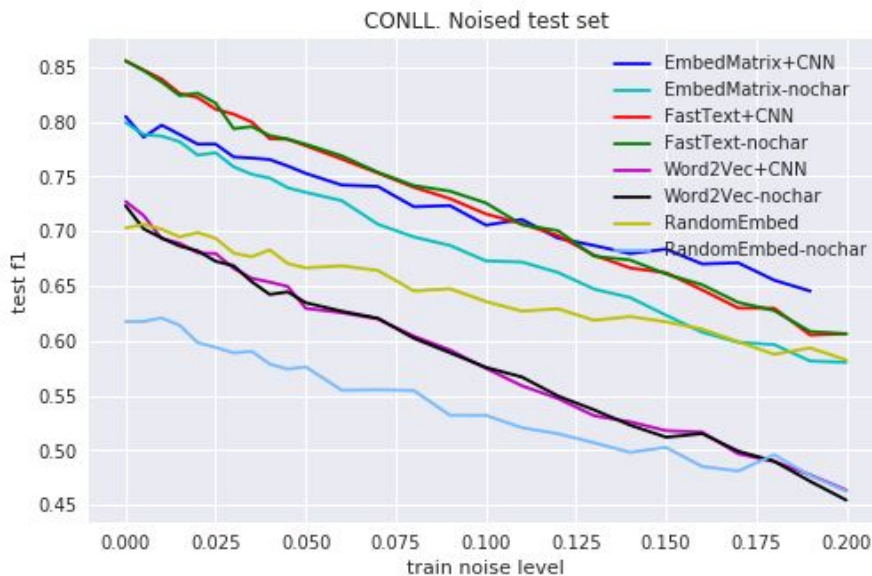
Результаты на исходных данных

Модель	CoNLL'03	Persons-1000	CAp'2017
EmbedMatrix+CNN	0.81	0.85	0.43
EmdebMatrix-nochar	0.80	0.81	0.44
RandomEmbed+CNN	0.69	0.77	0.31
RandomEmbed-nochar	0.61	0.48	0.22
FastText+CNN	0.86	0.69	0.41
FastText-nochar	0.86	0.69	0.41
Word2Vec+CNN	0.73	0.72	н/д
Word2Vec-nochar	0.72	0.72	н/д

тестирование вариантов модели на исходных данных по метрике F_1

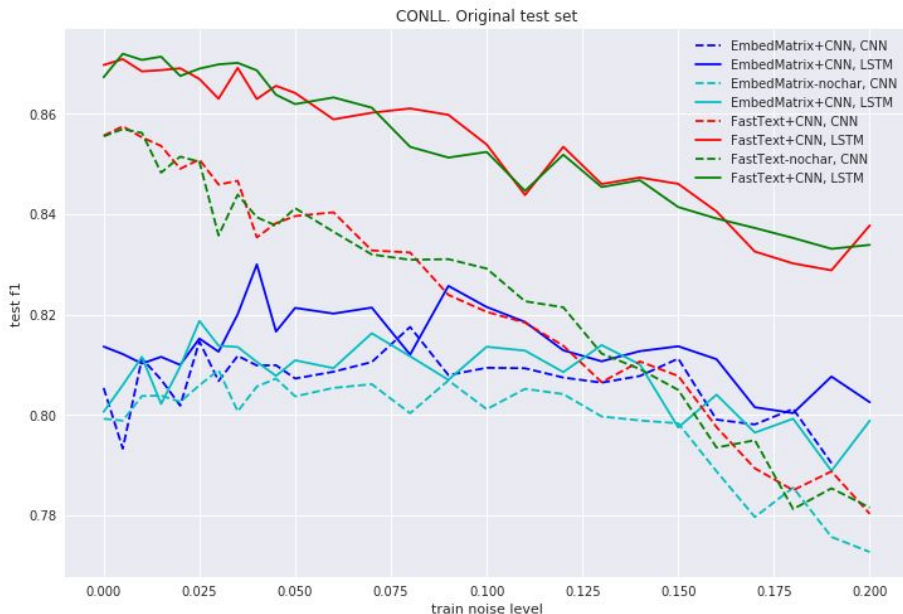
Результаты для английского языка (искусственный шум в тестовой выборке)

- лучше всего себя проявляют варианты модели FastText, за исключением высоких уровней шума, где лучшего всего себя показывает EmbedMatrix+CNN



Результаты для английского языка (естественный шум в тестовой выборке)

- результаты аналогичны предыдущему эксперименту для FastText и EmbedMatrix+CNN
- приведенные модели с использованием модуля обработки контекста LSTM показывают лучший результат для всех протестированных моделей



План

1. Моделирование шума
2. RoVe: устойчивые к шуму вектора слов
3. Прикладные задачи обработки текста
 - а. Устойчивая к шуму классификация текстов
 - б. Устойчивое к шуму распознавание именованных сущностей
 - с. Устойчивое к шуму извлечение аспектов

Извлечение аспектов

Aspect mining или Aspect Extraction:

- Пример: “the **call quality** of **this phone** is **amazing**”

Аспект “call quality” для сущности “this phone”.

Задача извлечения аспектов может решаться как задача тематического моделирования.

Тематическое моделирование

t_i - параметры
распределения для
темы i

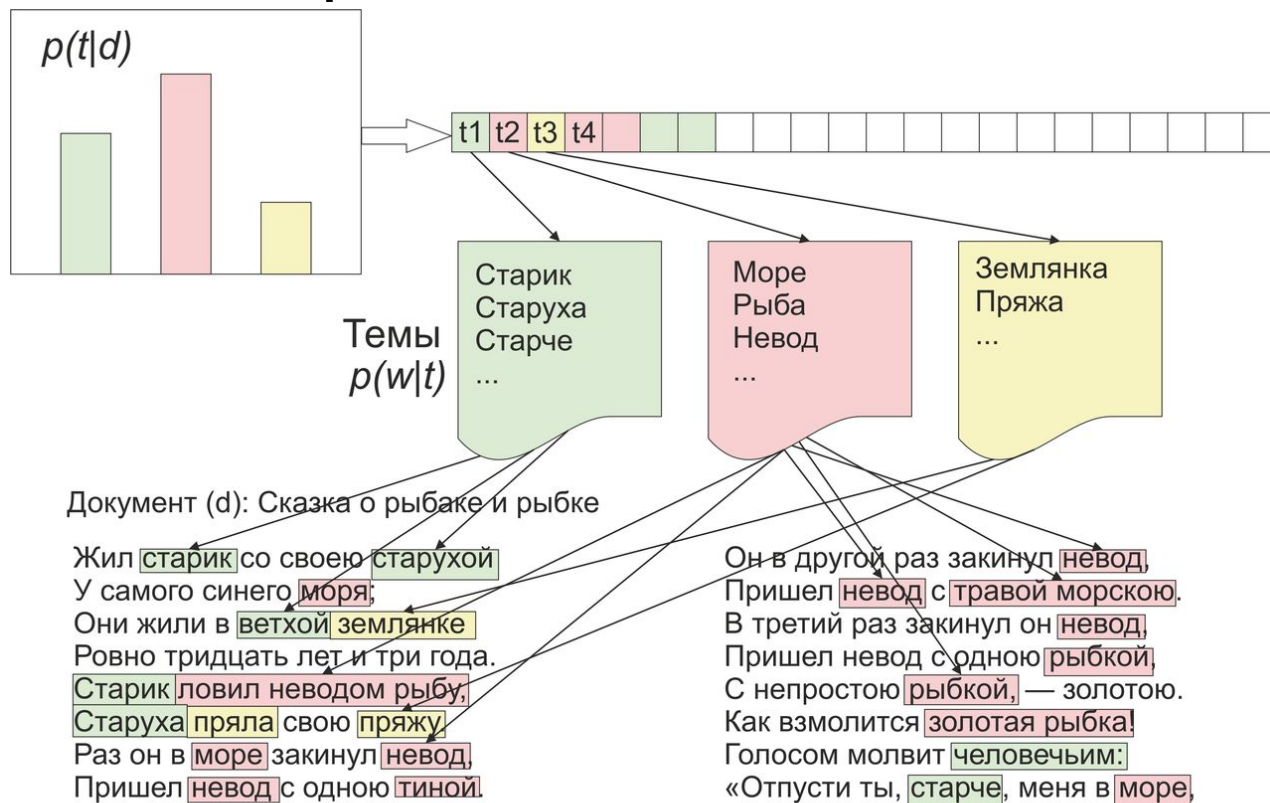
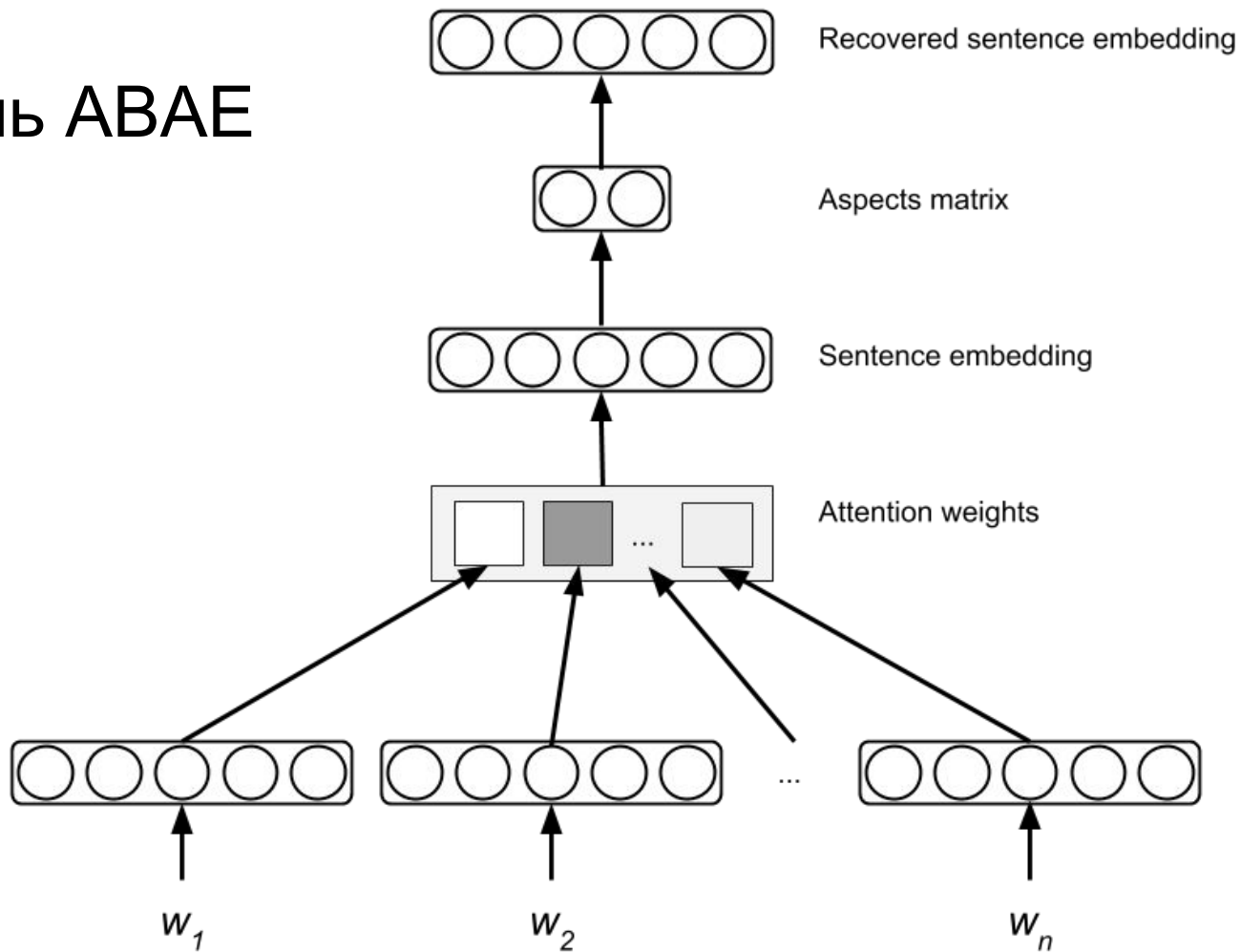


Иллюстрация принадлежит К.В.
Воронцову

Модель Attention-Based Aspect Extraction (ABAE)

- Цель модели - получить векторные представления аспектов.
- Каждый аспект представлен вектором, который близок к некоторому набору [векторных представлений] слов.
- Модель строит матрицу векторных представлений аспектов, инициализируемую случайно или на основе информации о корпусе.
- Задача модели породить векторное представление текста из векторных представлений слов и приближающую его линейную комбинацию векторных представлений аспектов.
- Функцией потерь для модели является разница между векторным представлением текста на основе векторных представлений слов и на основе матрицы аспектов.

Модель ABAE



Модель ABAE

s - предложение, z_s - векторное представление предложения

a_i - веса, получаемые из механизма внимания, y_s - промежуточное векторное представление предложения

e_w - векторное представление для слова w

A - матрица модели внимания

T - аспектная матрица

p_s - веса для суммирования векторов аспектов

r_s - “восстановленное” с помощью матрицы аспектов T векторное представление

$$a_i = \text{softmax}(e_{w_i}^T \cdot A \cdot y_s)$$

$$y_s = \sum_{i=1}^n e_{w_i}$$

$$z_s = \sum_{i=1}^n a_i e_{w_i}$$

$$p_s = \text{softmax}(W \cdot z_s + b)$$

$$r_s = T^T \cdot p_s$$

Корпус

Citysearch, содержит 50000 отзывов на рестораны Нью-Йорка

У отзывов выделены размечены категории аспектов:

- Food
- Price
- Service
- Ambience
- Anecdotes
- Miscellaneous

Постановка задачи

- был взят подкорпус Citysearch, состоящий отзывов, которые содержали только одну категорию аспектов
- выделяемые моделью аспекты вручную размечаются по категориям аспектов
- решается задача классификации, но без учителя
- соответствие категорий измеряется при помощи метрики F1

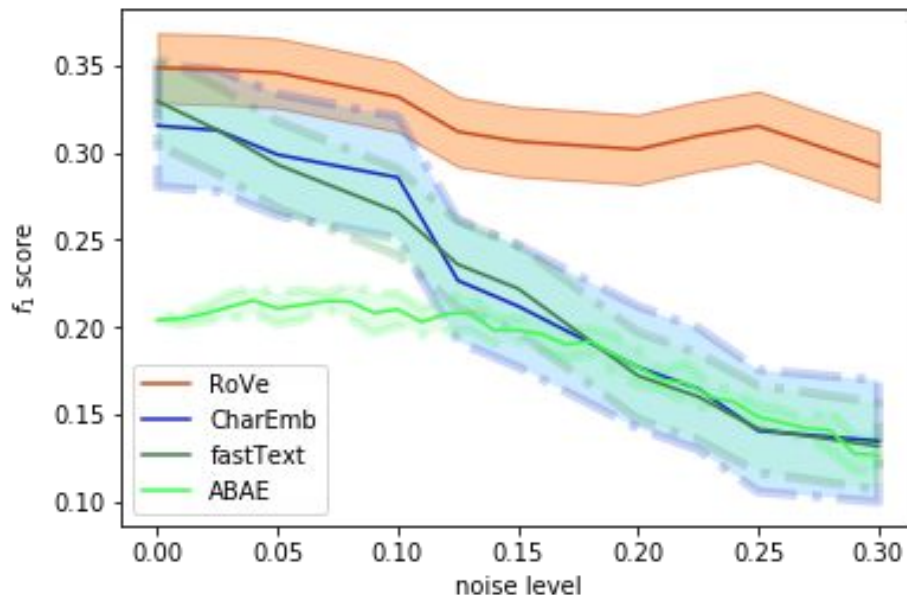
Предлагаемые расширения модели ABAE

- применение символьной информации для векторного представления слов (CharEmb)
- применение векторных представлений слов fastText (fastText)
- применение векторных представлений слов RoVe (RoVe)

Результаты для модели ABAE и предложенных расширений

тестирование моделей
производится по метрике F1

модель RoVe показывает лучшую
устойчивость к шуму



Выводы к разделу 3

- Продemonстрировано, что используемые в настоящий момент модели для прикладных задач не являются устойчивыми к шуму.
- Предложены варианты моделей, которые являются более устойчивым к шуму.
- Графики падения качества моделей в зависимости от уровня шума имеют сходную форму для естественного и искусственного шума. Следовательно, предложенная модель шума достаточно точно приближает естественный шум.

Основные положения

- Предложен и реализован новый алгоритм построения устойчивых к шуму векторных представлений слов.
- Предложен и реализован новый алгоритм устойчивой к шуму классификации текстов.
- Предложен и реализован новый алгоритм устойчивого к шуму распознавания именованных сущностей.
- Предложен и реализован новый алгоритм устойчивого к шуму извлечения аспектов.
- Предложенные методы проверены на эффективность.

Основные публикации по теме диссертации

1. Valentin Malykh. — «Robust word vectors for Russian language». — В: Proceedings of Artificial Intelligence and Natural Language AINL FRUCT 2016 Conference, Saint-Petersburg, Russia. — 2016, — С. 10—12.
2. Valentin Malykh и Vladislav Lyalin. — «Named Entity Recognition in Noisy Domains». — В: Записки научных семинаров ПОМИ. Серия “искусственный интеллект”. — 2018.
3. Valentin Malykh и Vladislav Lyalin. — «On Classification of Noisy Texts». — В: Записки научных семинаров ПОМИ. Серия “искусственный интеллект”. — 2018.
4. Valentin Malykh и Taras Khakhulin. — «Noise Robustness in Aspect Extraction Task». — В: Записки научных семинаров ПОМИ. Серия “искусственный интеллект”. — 2018.

Спасибо за внимание!

Соответствие паспорту специальности 05.13.11

Формула специальности:

- 05.13.11 – специальность, включающая задачи развития теории программирования, **создания и сопровождения программных средств различного назначения**

Области исследований:

- 7. **Модели и методы создания программ и программных систем для параллельной и распределенной обработки данных**, языки и инструментальные средства параллельного программирования.

Соответствие техническим наукам

На основе предложенных моделей и методов разработано программное обеспечение, получившее внедрение в библиотеку DeerPavlov, широко применяемую в области диалоговых систем, а также более общих задач обработки естественного языка. Проведены экспериментальные исследования, доказывающие значимый технический эффект при использовании данных результатов.

Литература

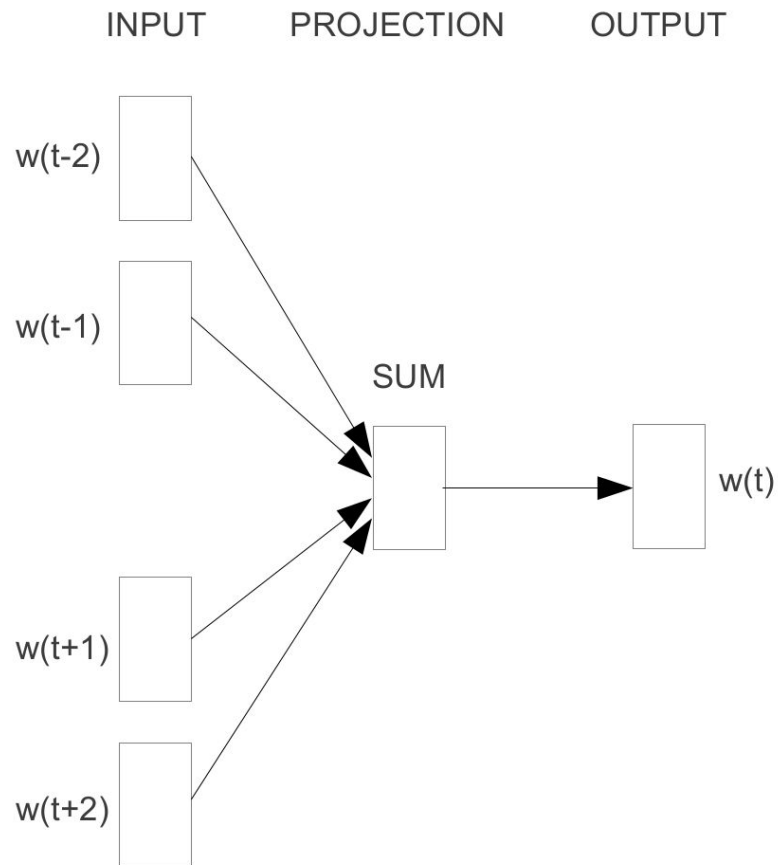
- Cucerzan, S. and Brill, E., 2004. Spelling correction as an iterative process that exploits the collective knowledge of web users. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing.
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems 2013 (pp. 3111-3119).
- Ю. В. Рубцова. Построение корпуса текстов для настройки тонового классификатора // Программные продукты и системы, 2015, №1(109), –С.72-78
- Kim Y. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882. 2014 Aug 25.
- He R, Lee WS, Ng HT, Dahlmeier D. An unsupervised neural attention model for aspect extraction. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) 2017 (Vol. 1, pp. 388-397).

Word2Vec

$$L = \frac{1}{N} \sum_i \ln(p(w_i | C(w_i))) \rightarrow \max$$

$$p(w_i | C(w_i)) = \text{softmax} \left(\sum_{w_k \in C(w_i)} v_{w_k}^\top u_{w_i} \right)$$

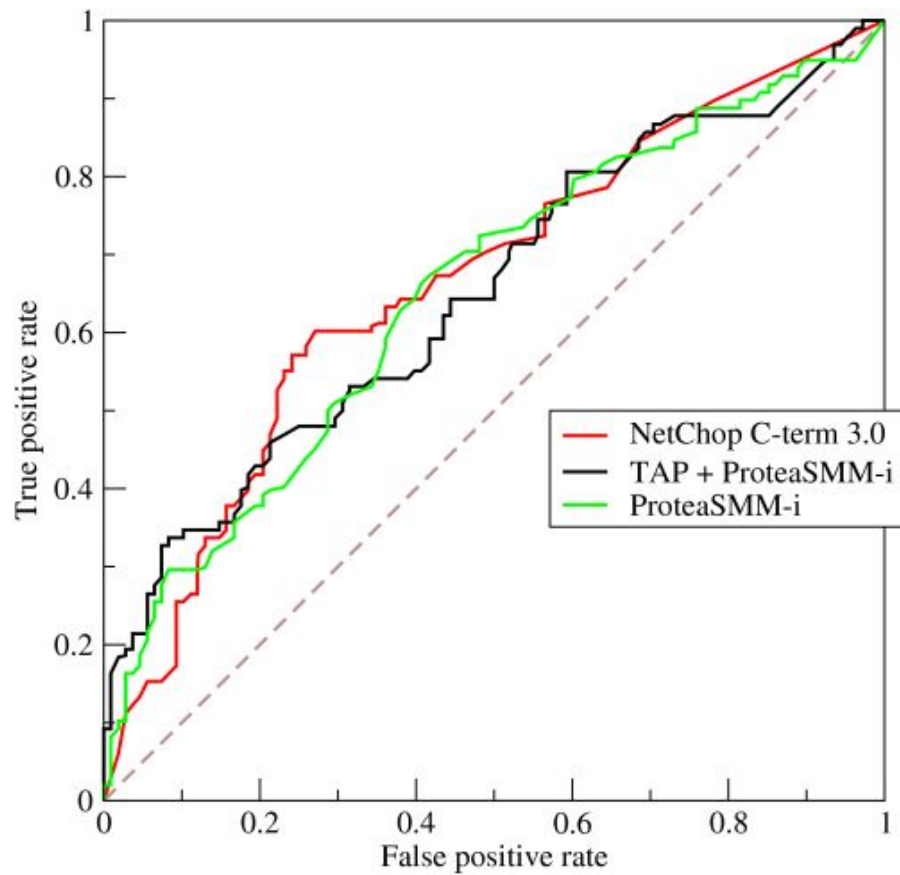
w_i - слово в контексте C ; v , u - вектора слов



CBOW

Area Under the Curve for Receiver Operating Characteristic (ROC-AUC)

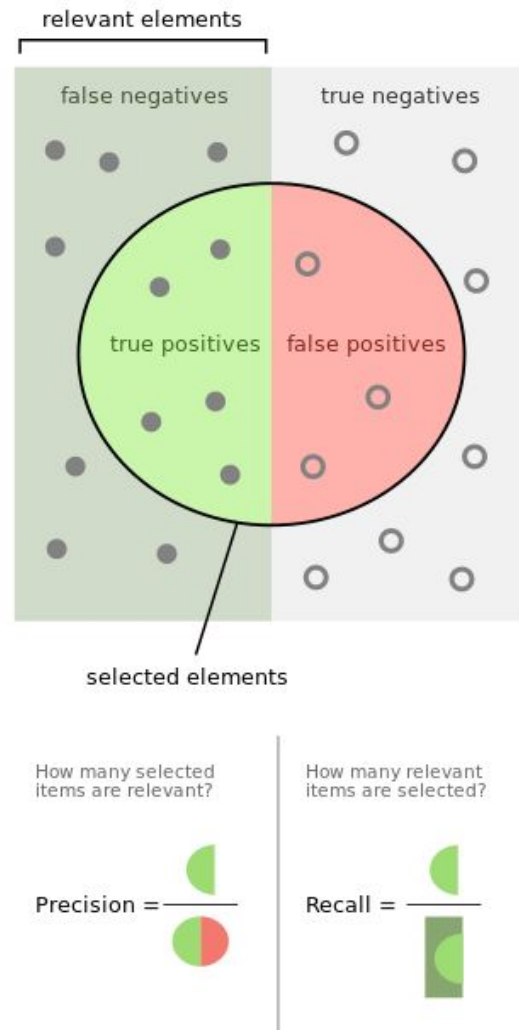
ROC-кривая - график, позволяющий оценить качество бинарной классификации, отображает соотношение между долей объектов от общего количества носителей признака, верно классифицированных как несущих признак, (true positive rate) и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущих признак (false positive rate) при варьировании порога решающего правила.



Метрика F_1

Метрика F_1 - гармоническое среднее между точностью (precision) и полнотой (recall) классификации

$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$



Результаты на MRPC

noise (%)	0	5	10	15	20	25	30
SRU	0.707	0.701	0.681	0.661	0.641	0.623	0.611
biSRU	0.715	0.691	0.687	0.662	0.644	0.637	0.631
CNN-1d	0.632	0.620	0.616	0.610	0.606	0.601	0.596
convLSTM	0.628	0.621	0.616	0.609	0.605	0.592	0.587
LSTM	0.598	0.593	0.583	0.541	0.549	0.513	0.518
biLSTM	0.680	0.650	0.643	0.638	0.627	0.594	0.568
stackedLSTM	0.672	0.658	0.637	0.629	0.606	0.595	0.554
Word2Vec	0.715	0.593	0.573	0.569	0.564	0.565	0.557
fastText	0.720	0.606	0.594	0.581	0.587	0.573	0.569
fastText+speller	0.720	0.638	0.598	0.579	0.585	0.563	0.543

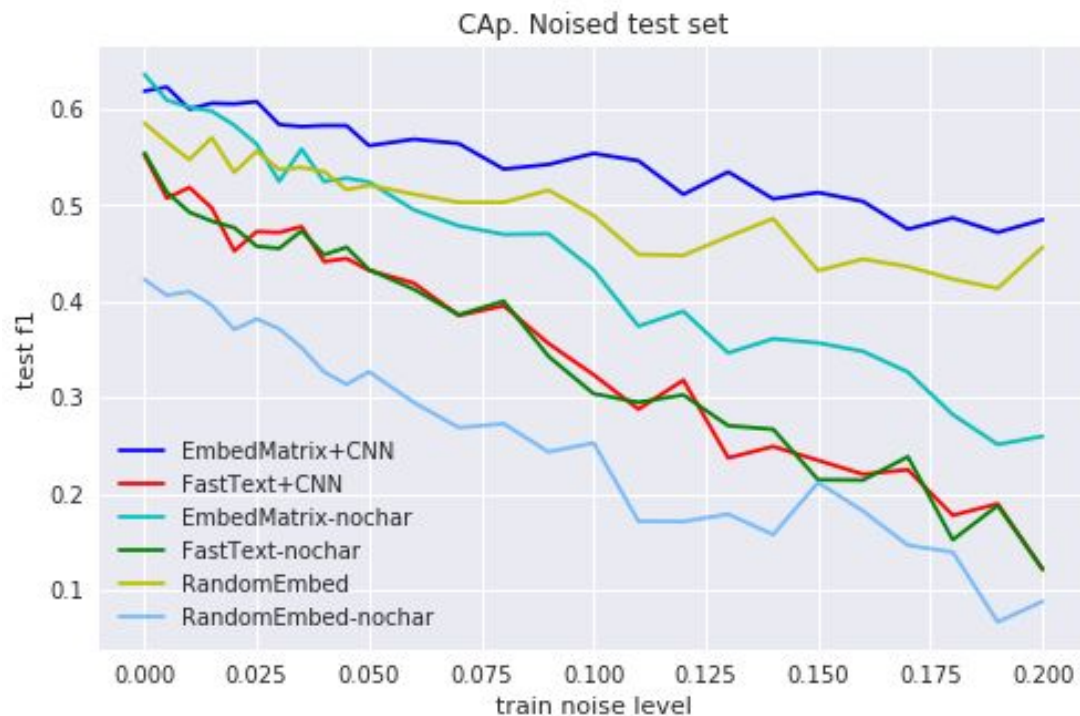
Результаты на Stanford Sentiment TreeBank

noise (%)	0	5	10	15	20	25	30
SRU	0.627	0.606	0.590	0.581	0.568	0.537	0.531
biSRU	0.656	0.640	0.621	0.612	0.598	0.586	0.536
CNN-1d	0.533	0.529	0.511	0.544	0.497	0.509	0.483
convLSTM	0.625	0.591	0.546	0.519	0.515	0.503	0.481
lstm	0.598	0.593	0.553	0.512	0.514	0.499	0.456
biLSTM	0.611	0.607	0.542	0.523	0.528	0.514	0.501
stackedLSTM	0.621	0.598	0.593	0.569	0.586	0.544	0.563
Word2Vec	0.649	0.607	0.611	0.583	0.554	0.513	0.457
fastText	0.662	0.639	0.615	0.557	0.524	0.521	0.523
fastText+speller	0.645	0.643	0.573	0.549	0.521	0.448	0.531

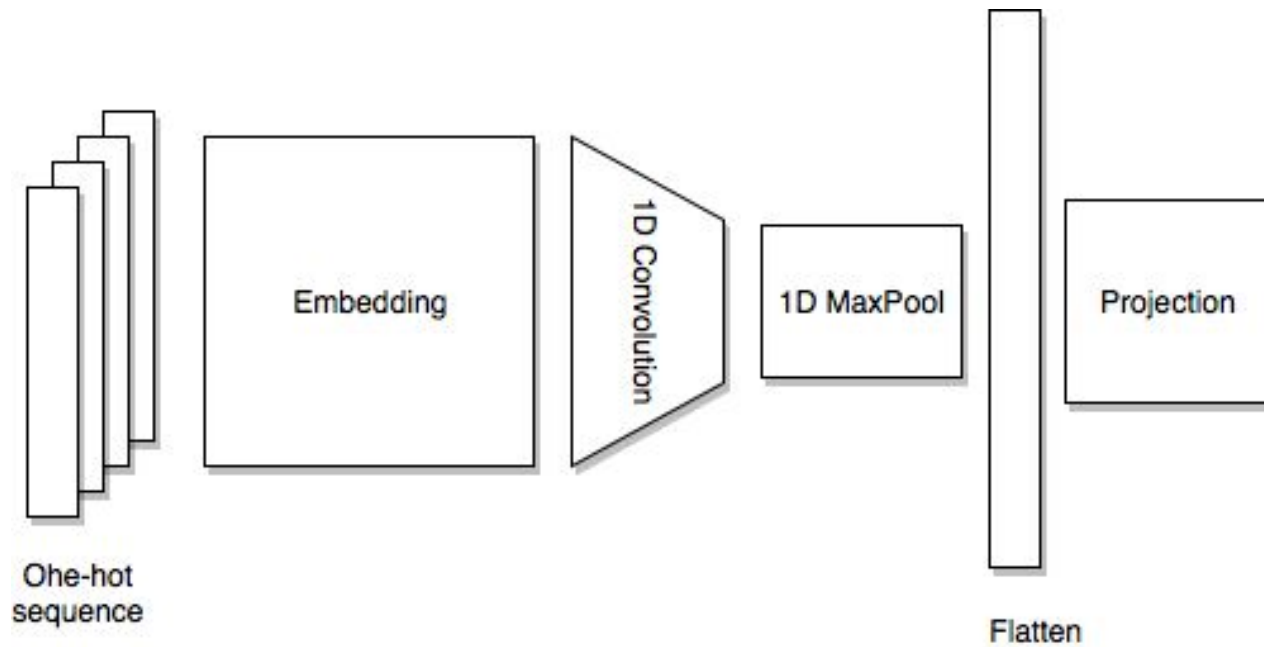
Результаты для французского языка

Model	orig.	sp.-ch.
EmbedMatrix+CNN, CNN	0.42	0.63
EmbedMatrix-nochar, CNN	0.44	0.64
EmbedMatrix+CNN, LSTM	0.39	0.59
EmbedMatrix-nochar, LSTM	0.38	0.59
FastText+CNN, LSTM	0.52	0.67
FastText-nochar, LSTM	0.53	0.69

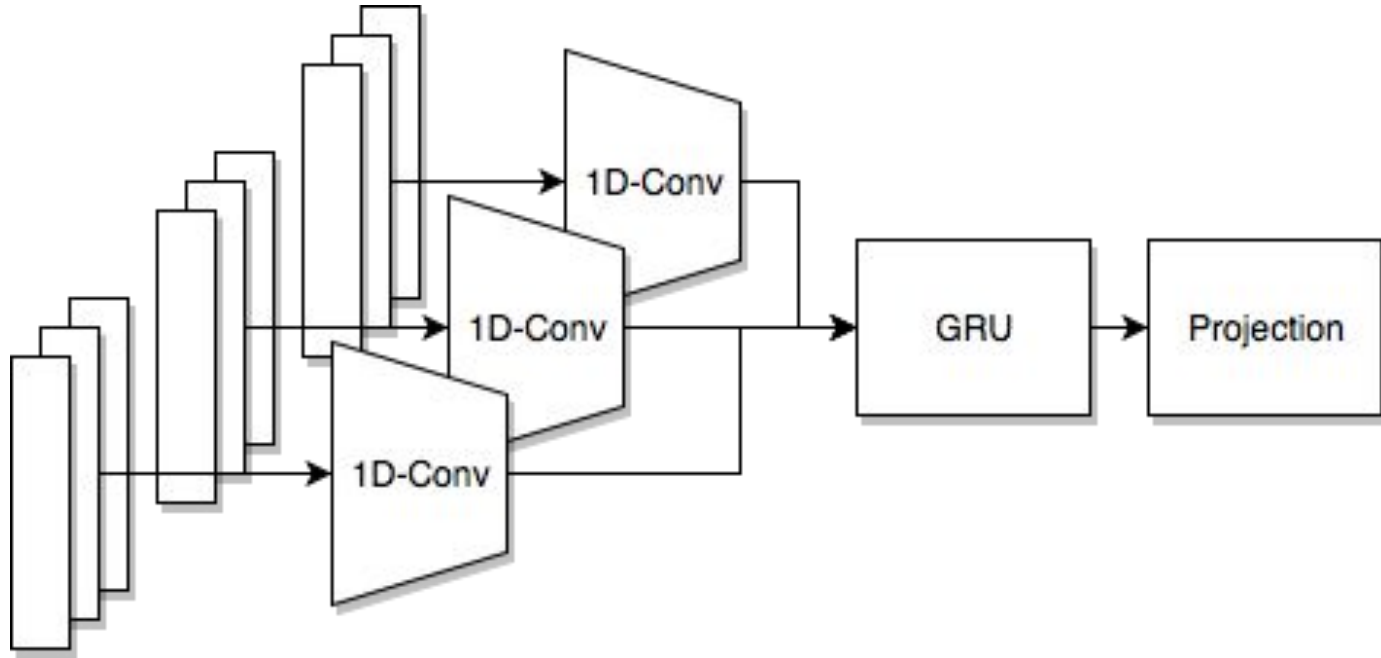
Результаты для французского языка



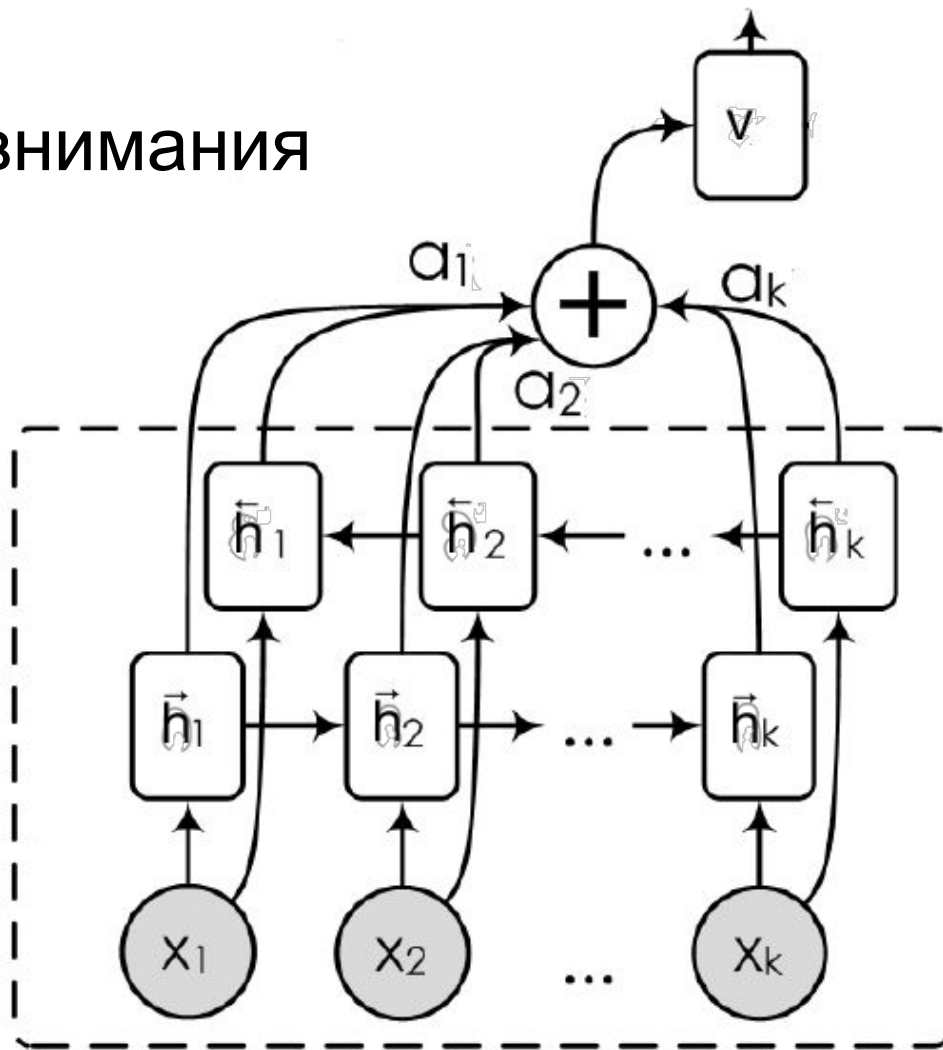
CharCNN



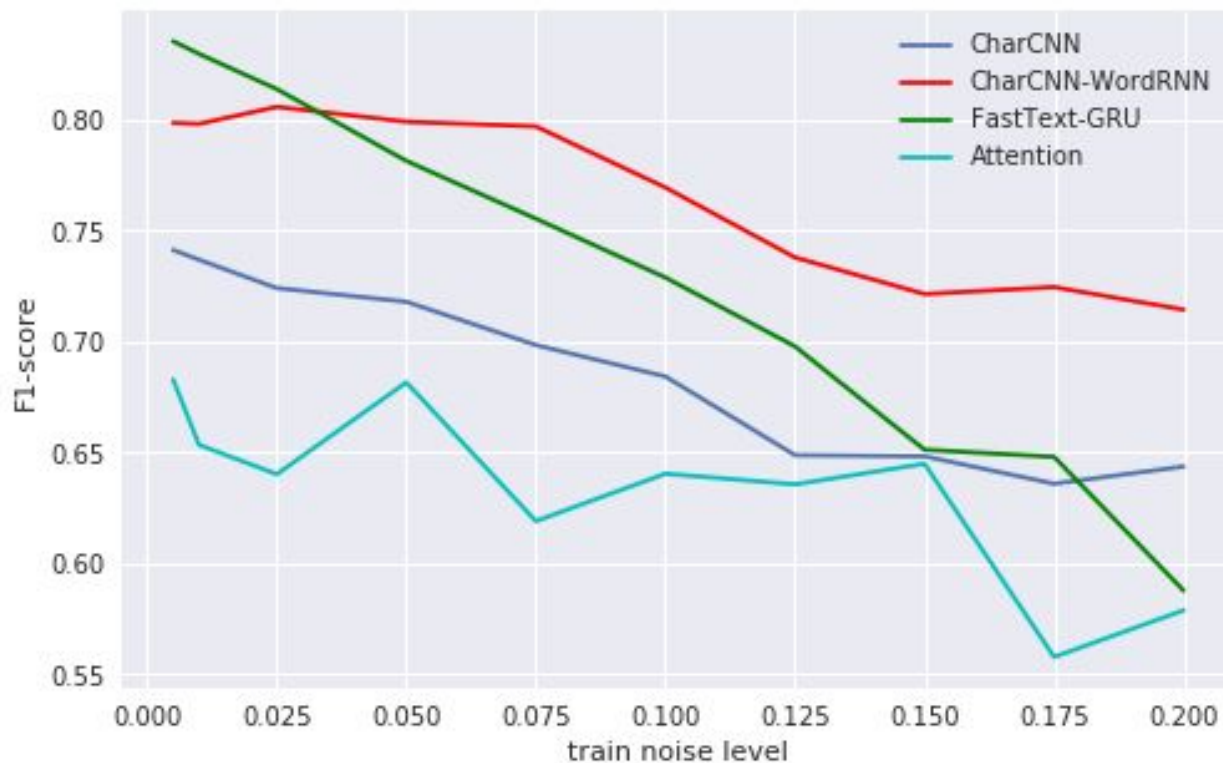
CharCNN-WordRNN



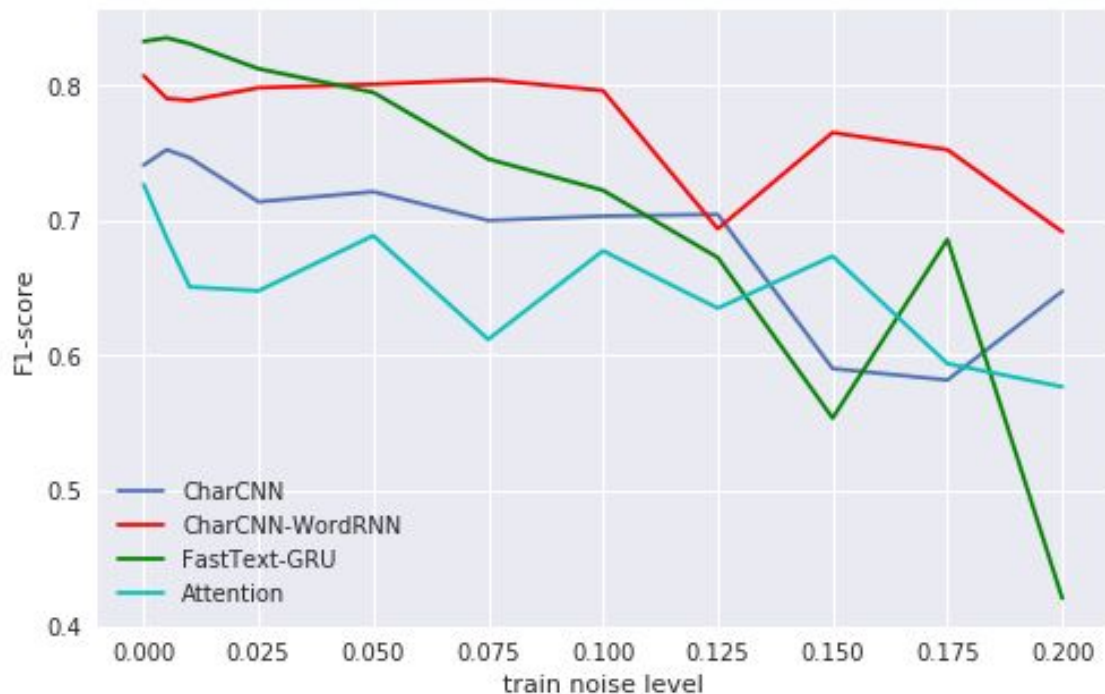
Механизм внимания



Результаты для английского языка (ИСК)



Результаты классификации для английского языка



Условные случайные поля (CRF)

суммарный потенциал для объекта x : $P(x) = \frac{1}{Z} \prod_k \varphi_k(x_{\{D\}})$

потенциальная функция: $\varphi_k(x_{\{k\}}) = \exp(\sum_k \lambda_k f_k(y_y, y_{t-1}, x_t))$

набор признаковов функций: $\Sigma\{f_k(y_t, y_{t-1}, x_t)\}$

$$p(y | x) = \frac{1}{Z(x)} \prod_k \exp(\sum_k \lambda_k f_k(y_y, y_{t-1}, x_t))$$

нормировочный коэффициент:

$$Z(x) = \sum_y \prod_k \exp(\sum_k \lambda_k f_k(y_y, y_{t-1}, x_t))$$

Результаты для оригинальной модели ABAE

