

Генерация комментариев по содержанию поста

Калашников Дмитрий
Горбунова Ольга
Сыромятников Александр



Задача

- ★ Генерировать комментарий по содержанию поста социальной сети
- Seq2Seq
- Выход: реакция на содержание входного текста - не совсем суммаризация
- Правдоподобность: связность, разговорный стиль, пунктуация



Лентач
24 мая в 8:00

Мужчина в Москве загрузил в банкомат более миллиона рублей из «банка приколов» и обналичил деньги

<http://news.lenta.ch/b41m>



3 509



165



3 177

300K

Сначала интересные ▾



Сергей Фролов

Это что за банкомат такой? У меня банкоматы порой реальные деньги не с первого раза принимают 😊

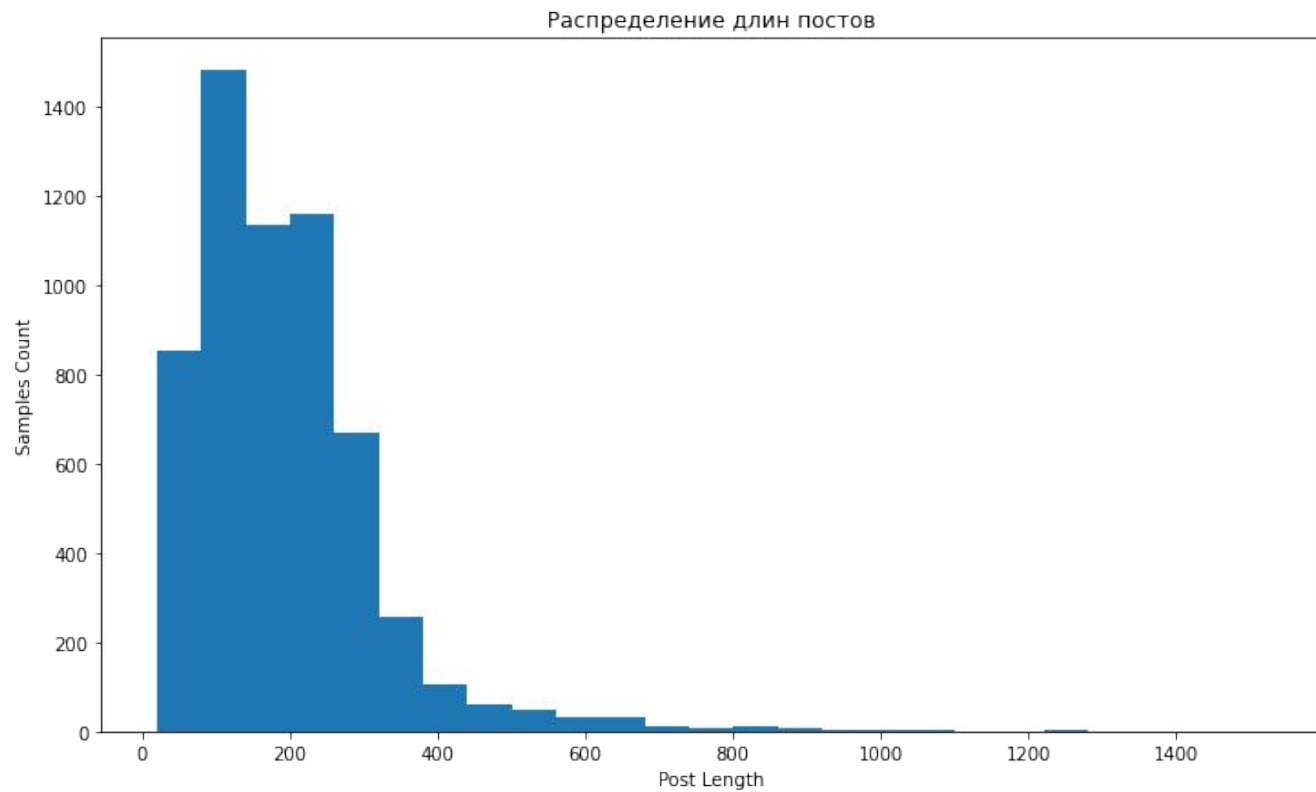
24 мая в 8:02 Ответить

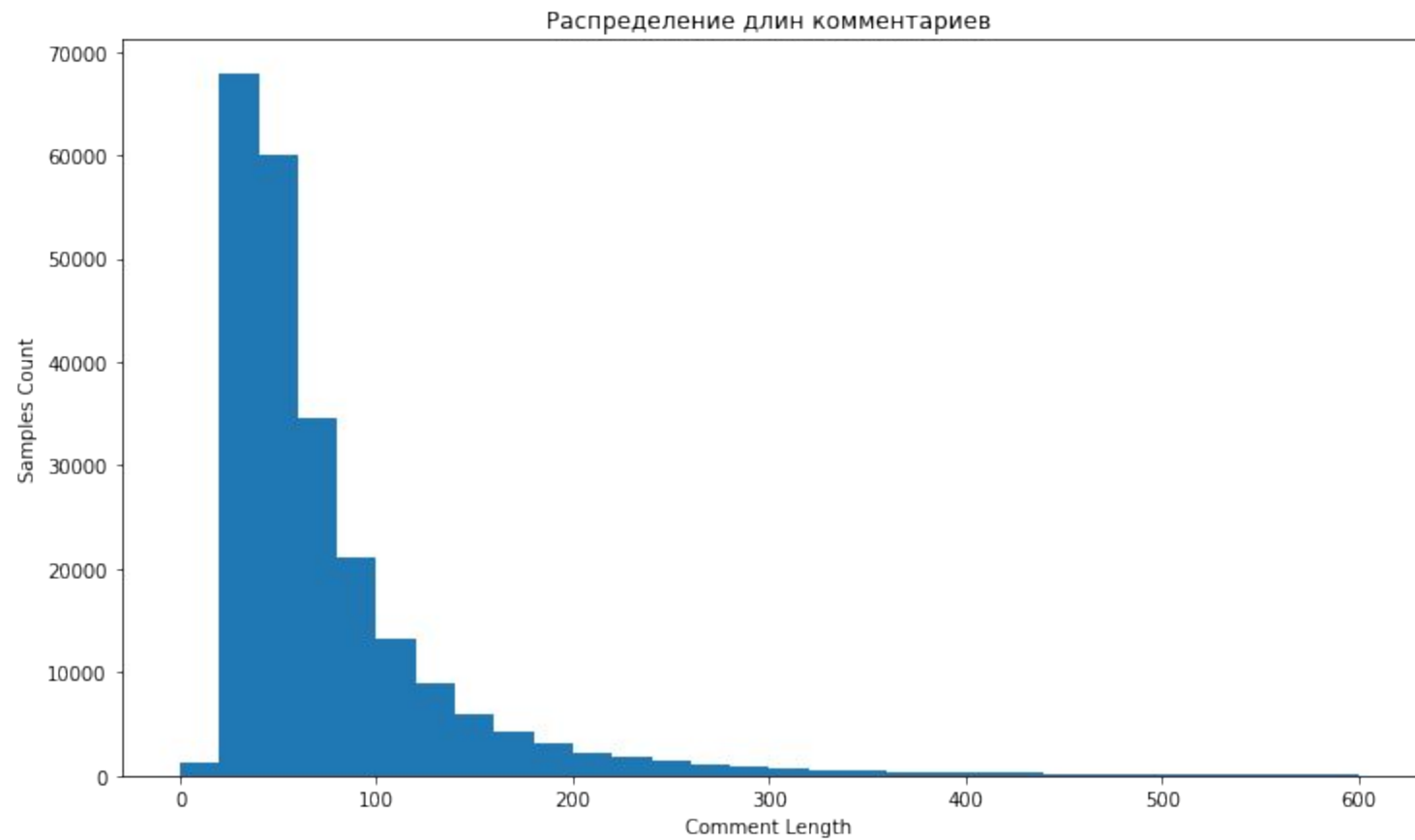
1446

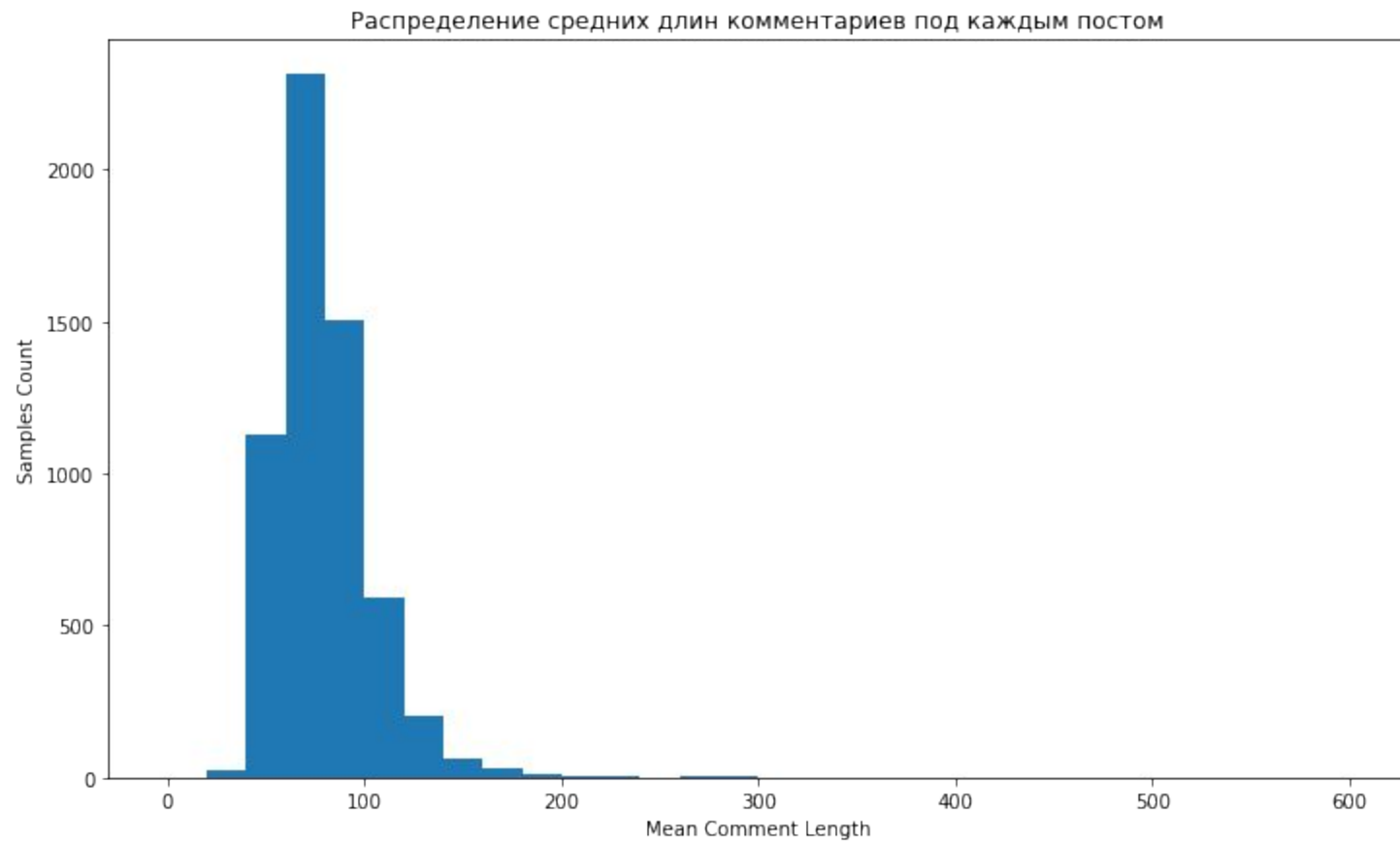
Датасет



- Социальная сеть: **VK**
- Собрано с помощью VK API.
- Паблик: Лентач
- Обработано 20000 последних постов, для каждого - 100 комментариев
- Нижний регистр, хештеги, кодировки, специфичные для паблики элементы
- Ограничение на длины поста и комментария снизу
- Оставлены только посты и комментарии с допустимыми символами
- Ручная чистка
- 5880 постов, около 100 комментариев для каждого.







Трансформеры

- Предобученный Russian GPT-3 model(ruGPT3Small) от Сбера
- Fine-tuning
- 231118 пар “пост-комментарий”, 1 эпоха
- 25 часов суммарно
- Google Colab
- Скрипт для генерации комментария по посту:

<https://github.com/dmkalash/Comment-Generator>



Что было до нас

- Комментарии к новостям
- Bi-LSTM + Attention
- Персонализированные комментарии
- Датасет на китайском языке
- 2017-2019 года

- ❑ “Automatic Generation of Personalized Comment Based on User Profile”, Wenhuan Zeng et al.
- ❑ “Automatic Generation of News Comments Based on Gated Attention Neural Networks”, HAI-TAO ZHENG et al.

标题: 中共第十九次全国代表大会10月18日在北京召开
Title: The Ninth National Congress of the Communist Party of China will be held in Beijing on October 18th

热门跟贴 (跟贴114条 有10063人参与)



天行寶 [网易四川省成都市手机网友]

2017-08-31 18:05:16

祝贺十九大胜利召开!
Congratulations to the 19th!

顶[1345] 回复 收藏 分享 复制



头像确实牛逼 [网易山东省手机网友]

2017-08-31 18:06:10

喜迎十九大
Welcome to the 19th

顶[581] 回复 收藏 分享 复制



杯子PL17 [网易广东省深圳市手机网友]

2017-08-31 18:04:55

这个是大事
This is a big deal

顶[712] 回复 收藏 分享 复制



有态度网友06jHSa [网易上海市手机网友]

2017-08-31 18:05:45

预祝中共十九大胜利召开!
I wish the 19th a victory!

顶[439] 回复 收藏 分享 复制

Примеры работы

Р: власти петербурга не знают, кто занимался демонтажом скк, в результате чего произошло обрушение здания.

С: и что происходит с людьми.

Р: аrosalyptica – музыкальная группа из финляндии, стиль которой чаще всего характеризуют как симфонический метал и ли виолончельный метал традиционный радиолентач

С: Музыка, которая не похожа на другие группы, но все равно очень нравится.

Р: boris brejcha– немецкий диджей и продюсер. описывает свой музыкальный стиль как "high-tech minimal". традиционны й радиолентач.

С: как всегда. а как же музыка из детства?

Р: бывший замглавы фсин, задержанный по делу о злоупотреблении должностными полномочиями, назвал питание в столично м сизо "лефортово" лучшим из всех тюрем европы

С: а вот что касается еды в сизо, то это уже не тот случай, когда еда в сизо лучше всего. а то что там не кормят в сизо – это уже не новость

Примеры работы

Р: пока улицы питера вдруг начали заполонять коммунальщики и бюджетники с лопатами, с крыш, которые не чистили всю зиму, начали падать таящие ледяные глыбы. по всему городу сейчас вот так:

С: а может просто я не настолько хороший, чтобы платить налоги? может мне просто не хватает опыта в этой области? а может я просто не знаю, что мне нужно?...

Р: пересядь с иглы интернет-одобрения на биржу труда

С: и работа пойдет быстрее. я не против. только не надо будет искать работу через интернет.я и так работаю удаленно

Р: россия не будет запрещать экспорт леса в китаи; латинская америка может стать полностью безвизовой для россиян; полиция нью-йорка потребовала у google не показывать данные о патрульных постах и другие новости к этому часу

С: я не против, но не хочу чтобы это стало поводом для новых санкций против кого-то из наших.я против таких санкций , которые не могут быть санкционированы никакими санкционными мерами.

Р: россия не будет запрещать экспорт леса в китаи; латинская америка может стать полностью безвизовой для россиян; полиция нью-йорка потребовала у google не показывать данные о патрульных постах и другие новости к этому часу

С: я не против запретить оружие в китае, но мне кажется что это не решит проблему. китаи это не только китаи, но и российская экономика. и это не мешает нам покупать лес в китае.

Спасибо за
внимание :)

Отлично!
Работаем дальше

