



12 апреля, 2017

Извлечение симптомов заболеваний и побочных эффектов из отзывов на основе машинного обучения

Тутубалина Елена, Мифтахутдинов Зульфат, Алимова Ильсеяр

Казанский (Приволжский) федеральный университет

Тропша Александр, Университет Северной Каролины в Чапел Хилл

Введение

Идея

Информация о свойствах лекарств может быть получена из социальных сетей.

Цель работы: разработка методов извлечения информации о новых эффектах лекарственных препаратов для построения гипотез о перепрофилировании лекарств.



Качественные
жалобы



Клинические
испытания



Благоприятные
эффекты



Побочные
эффекты



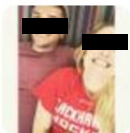
Вопросы
пациентов

Извлечение информации

Объект исследования: мнения пользователей о лекарственных препаратах, представленные в виде неструктурированных текстов на естественном языке.

Задачи

- 1) извлечение из текста всех упоминаний лекарств, показаний к применению, реакций организма, и проч;
- 2) приведение извлеченных фраз в формальные термины медицинской направленности;
- 3) выявление эмоционального тона автора к результату лечения и терапии.



@[redacted]



+ Follow



Albuterol has me feeling all sorts of dizzy and weak this morning 🥵👉😊



6:21 AM - 17 [redacted] 15



Asthenia
MedDRA: 10003549



Dizziness
MedDRA: 10013573

visual snow dont want to eat cross vision lost my appetite
apetite surpressed googley eyed no appetitey doublevision
blind appetite is nonexistent blurry vision 🙄 miss feeling hungry
anorexic visión doble #notevenhungry blindness cant eat
couldn't see making me eat like a mouse vision change blurry
lost appetite lost their eyesight didn't get hungry seeing double
can't eat killed my apetite googly eyed lost teh appetite
sin hambre crosseved killed my appetite seeing weird couldnt see

Visual impairment

MedDRA 10047571

Visual impairment

SNOMED 397540003

Decreased appetite

MedDRA 10061428

Loss of appetite

SNOMED 79890006

Данные и методы

visual snow dont want to eat cross vision lost my appetite
apetite surpressed googley eyed no appetitey doublevision
blind Varied Spelling rry vision Typo ling hu
anorexia vision double #noteverhungry business cant eat
uldn't see making me eat like a mouse vision change blurry
st a Emoticons Languages
can't eat killed my apetite googly eyed lost teh appetite
sin hambre Invented words and hashtags
cross eyed seeing weird colour stomach small changes in vis
lack of appetite never want to eat seeing weird color never h

Данные

- Мы собрали 217,485 отзывов с сайта webmd.com;
- Было размечено 699 отзывов о 20 наиболее комментируемых диагнозах;
- аннотации: *Medication* (Лекарство) и *Disease* (Болезнь).

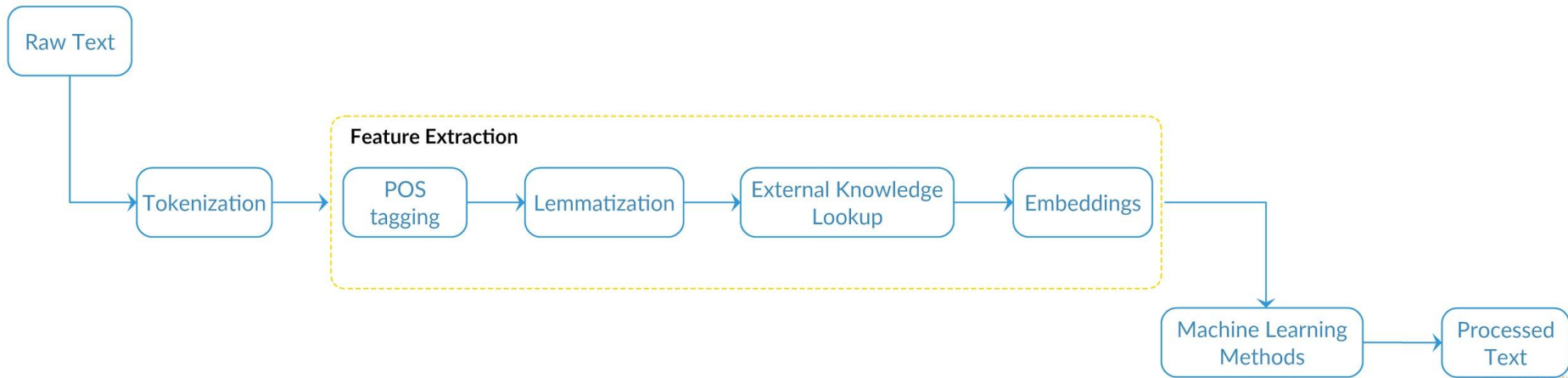
1	ID:10028
2	URL: http://www.webmd.com/drugs/drugreview-9824-Xanax+oral.aspx?conditionFilter=-1&drugid=9824&drugname=Xanax+oral
3	TITLE:
4	User Reviews & Ratings - Medication (Drugname) Xanax oral
5	TEXT
6	I was diagnosed with Disease [redacted] GAD when I was 16.
7	I had frequent episodes of anxiety attacks. Disease [redacted]
8	I've seen psychiatrists and therapists the help me deal with my anxiety. Disease [redacted]
9	I only take .25mg as needed because my doctor doesn't want me to get addicted considering I'm only 19. Disease [redacted]
10	I usually only have to take it to get myself to calm down so I can go to sleep. Disease [redacted] Disease [redacted]
11	RATING:5
12	Condition: Anxious
13	NOTE

# of reviews	699
# of uniq. Drugs	32
# of tokens	67,597
# of uniq. tokens	4,834
# of sentences	3,931
# of Medication annotations	924
# of Disease annotations	3,695
# of 1-gram Disease ann.	1,264
# of 2-gram Disease ann.	1,000

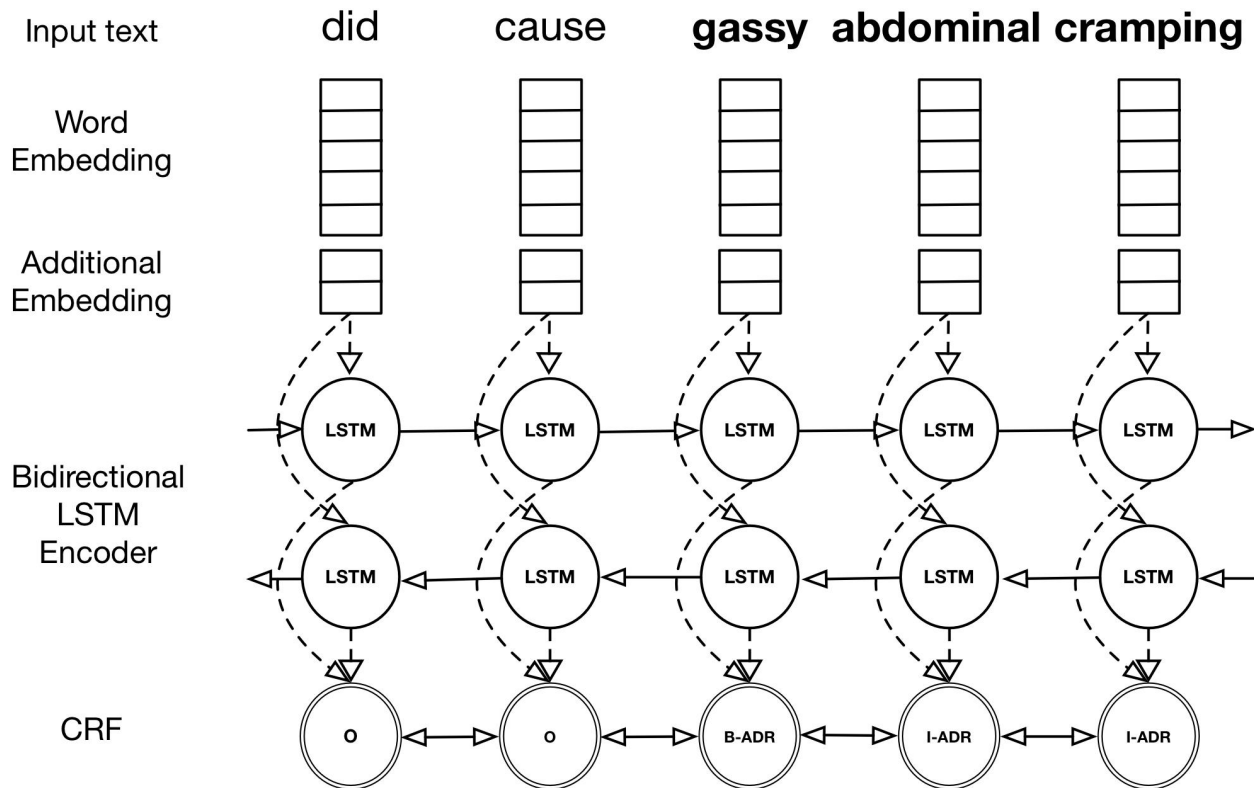
Предложенный метод

Для извлечения информации использованы методы машинного обучения:

- Условные случайные поля (Conditional Random Fields);
- Рекуррентные сети (LSTM and GRU);



Комбинирование методов



Эксперименты

Сравнение алгоритмов

- CRF: точность = 72%, полнота = 70%, F1 = 71% (точное)
точность = 89%, полнота = 81%, F1 = 83% (частичное)
- GRU: точность = 76%, полнота = 66%, F1 = 71% (точное)
точность = 83%, полнота = 91%, F1 = 85% (частичное)
- LSTM+CRF: точность = 74%, полнота = 77%, F1 = 75% (точное)
точность = 87%, полнота = 88%, F1 = 87% (частичное)

Применимость метода для перепрофилирования лекарств

- Три proof-of-concept примера: Abilify, Vyvanse, and Yaz;
- Мы собрали отзывы с webmd.com, askapatient.com, and drugs.com;

Пример 1. Abilify (Aripiprazole)

Показания к применению: 2002 -- schizophrenia. 2004 -- bipolar disorder.
2014 -- Tourette's disorder.

Из постов: *cleared my tics, tourette's* (2006). *can't stay still, twitching, tics, right hand don't function normally* (2007).

В науке: The first paper about the treatment of Tourette's Disorder with Aripiprazole was published in September 2005.

Пример 2. Vyvance (Lisdexamfetamine Dimesylate)

Показания к применению: 2007 -- Attention-Deficit/Hyperactivity Disorder. 2015 -- Moderate to Severe Binge Eating Disorder (BED).

Из постов: *decrease in appetite* (2007). *appetite decreased* , *appetite suppression*, *no appetite* (2008)

В науке: The first clinical study of Lisdexamfetamine in Binge Eating Disorder was started in January 2010.

Пример 3. Yaz (Drospirenone; Ethinyl Estradiol)

Показания к применению: 2006 -- prevention of pregnancy. 2011 -- treatment of symptoms of premenstrual dysphoric disorder (PMDD) and moderate acne vulgaris.

Из постов: *premenstrual disorder, loss of interest in hobbies, no more headaches, acne cleared up, mood stabilized, cleared up my face* (2007).

В науке: The first paper about the treatment of PMDD with a novel formulation of Drospirenone and Ethinyl Estradiol was published in 2007.

Заключение

- Из комментариев в социальных сетях можно извлекать новую информацию о свойствах лекарств.
- Используются методы обработки естественного языка и алгоритмы машинного обучения.
- Анализ комментариев подтверждает наличие новых связей между лекарством и заболеваниями, не упомянутыми в инструкции по применению лекарства.

Спасибо за внимание