

Сентимент талдау есебін шешу

Болатбек М.А.

Сентимент талдау анықтамасы

Мәтіннің тоналдығын талдау (сентимент-талдау, ағылш. Sentiment analysis, ағылш. Opinion mining) — мәтіндерде эмоционалды түсті лексиканы автоматты түрде анықтауға және мәтінде айтылатын объектілерге қатысты авторлардың (пікірлердің) эмоционалды бағалауына арналған компьютерлік лингвистикадағы мазмұнды талдау әдістерінің класы. Тоналдылық-бұл мәтінде көрсетілген белгілі бір объектіге (нақты әлем объектісі, оқиға, процесс немесе олардың қасиеттері/атрибуттары) сөйлем авторының эмоционалды қатынасы. Лексема немесе коммуникативті фрагмент деңгейінде көрсетілген эмоционалды компонент лексикалық тональдылық (немесе лексикалық эмоция) деп аталады. Бүкіл мәтіннің тоналдылығын тұтастай алғанда оны құрайтын бірліктердің (сөйлемдердің) лексикалық кілттерінің функциясы (қарапайым жағдайда қосындысы) және оларды біріктіру ережелері ретінде анықтауға болады.

Тоналдылықты талдау (сентимент-анализ) —

Сентимент талдау анықтамасы

Введите текст для поиска...

Войти

прикольная такая психоделика

Neidux
24 мая 2018 19:03
Гости

Крутой сериал, посмотрел первый сезон, остались только положительные эмоции от просмотра. Интересный и неожиданный сюжет. Рекомендую! Сайт тоже очень крутой, реклама вообще не появляется, хорошее качество картинки и озвучки. Спасибо)

Карыч
25 мая 2018 18:22
Гости

Сериал крутой, всем советую. Но вот реклама... Превышает весь просмотр!!! Останавливает серию и все, не идет дальше. Перезагрузил страницу все заработало, казалось бы, но потом опять, и все по кругу!!!

ねこちゃん
15 июня 2018 06:49
Гости

Посмотрела первый сезон, оставило хорошее впечатление. Очень не плохой такой психоделический ужастик. Озвучка и качество на высшем уровне! Не знаю что у других, но у меня ни разу реклама не появлялась. К тому же сайт сам по себе очень хорош. Огромное спасибо!

Беркут
31 июля 2018 20:03
Гости

Сериал Классный, Спасибо администрации что нету рекламы)

дэн
13 августа 2018 00:23

ОПИСАНИЕ ОТЗЫВЫ (70) НАЛИЧИЕ В МАГАЗИНАХ ДОСТАВКА И ОПЛАТА

Юлия
Москва, Москва
Возраст: 15
Тон кожи: Нейтральный
Цвет глаз: Карий
Тип кожи: Комбинированная

ОЧЕНЬ ХОРОШАЯ ТУШЬ
★★★★★ 20.07.2018
Cabaret Premiere даёт невероятный объём. Однозначно буду брать её ещё!

Яра

★★★★★ 03.06.2018
Понравилась тушь! Надо будет еще такую же водостойкую купить для моря!)))

ОЧЕНЬ ХОРОШО ЗА СВОЮ ЦЕНУ
★★★★★ 07.02.2018
Обратила внимание на эту тушь в первую очередь потому, что она есть в коричневом цвете (не так просто блондинкам подобрать тушь).
[Читать дальше](#)

★★★★★ 09.03.2017
Прочитала отзывы на сайте, и пошла покупать, до нее я брала только Max Factor, но эта тушь произвела на меня огромное впечатление.
[Читать дальше](#)

В НАЧАЛО < 1 2 3 4 5 > В КОНЕЦ

бұл мәтіндік құжаттардағы
пікірлер мен эмоцияларды
зерттейтін компьютерлік
лингвистиканың саласы.

Автоматты тоналдылықты талдау жүйелері не үшін қажет?

Тоналдылықты талдау және пікір алу жүйелері келесі салаларда практикалық қолдануды табады:

- әлеуметтану: халықтың діни көзқарастары туралы мәліметтер;
- саясаттану: халықтың саяси көзқарастарының мониторингі;
- маркетинг: ноутбуктің қай моделі үлкен сұранысқа ие екендігі туралы Twitter хабарламаларын талдау;
- медицина және психология: сентименталды талдауды әлеуметтік медиа пайдаланушыларындағы депрессияны анықтау үшін қолдануға болады;
- қаржы саласы: қор және валюта нарықтарындағы трендтерді анықтау үшін қаржылық есептер мен қаржылық жаңалықтардың тоналдылығын талдау;
- пікірлерде спам іздеу;

СА не себепті маңызды?

Клиенттер өз ойлары мен сезімдерін бұрынғыдан да ашық білдіретіндіктен, көңіл-күйді талдау осы көңіл-күйді бақылау мен түсінудің маңызды құралына айналады. Клиенттердің пікірлерін автоматты түрде талдау, мысалы, сауалнамалар мен әлеуметтік медиа сөйлесулеріндегі пікірлер, брендтерге тұтынушылар тұтынушыларының қажеттіліктерін қанағаттандыру үшін өнімдер мен қызметтерді бейімдей алатындай етіп клиенттерді не бақытты немесе көңілсіз ететінін білуге мүмкіндік береді.

Мысалы, 4000-нан астам тұтынушының сіздің өніміңізге қанағаттану сауалнамасын автоматты түрде талдау үшін көңіл-күй талдауын пайдалану клиенттердің сіздің жоспарларыңызға және клиенттерге қызмет көрсетуге қанағаттанғанын анықтауға көмектеседі. Мүмкін сіз әлеуметтік желілердегі брендтің көңіл-күйін нақты уақытта және уақыт өте келе бағалағыңыз келуі мүмкін, осылайша сіз қанағаттанбаған клиенттерді бірден анықтап, мүмкіндігінше тезірек жауап бере аласыз.

- Мәтіннің тоналдылығын талдау (немесе сентимент-талдау) – Data Science мамандары жұмыс істейтін міндеттердің бірі. Осындай талдаудың көмегімен сіз хабарламалар мен басқа да деректердің массивін зерттеп, олардың эмоционалды түрде қалай боялғанын – оң, теріс немесе бейтарап екенін анықтай аласыз.

Тоналдылықты талдаудың негізгі мақсаты-мәтіндегі пікірлерді табу және олардың қасиеттерін анықтау. Қандай қасиеттер зерттелетіні тапсырмаға байланысты. Мысалы, талдаудың мақсаты автор, яғни пікірге ие адам болуы мүмкін. Пікірлер екі түрге бөлінеді:

1. тікелей пікір;
2. Салыстыру

Тікелей пікірде автордың бір объект туралы мәлімдемесі бар. Тікелей пікірдің ресми анықтамасы келесідей: "тікелей пікір бес элементтен тұрады (e, f, op, h, t), мұнда:

- (entity, feature) — e кілтінің нысаны (автор айтқан нысан) немесе оның қасиеттері f (атрибууттар, объектінің бөліктері) ;
- orientation немесе polarity-тоналды бағалау (автордың аталған тақырыпқа қатысты эмоционалды позициясы);
- ұстаушы-тоналдылық тақырыбы (автор, яғни бұл пікір кімге тиесілі);
- уақыты, пікір қалдырылған кезде.

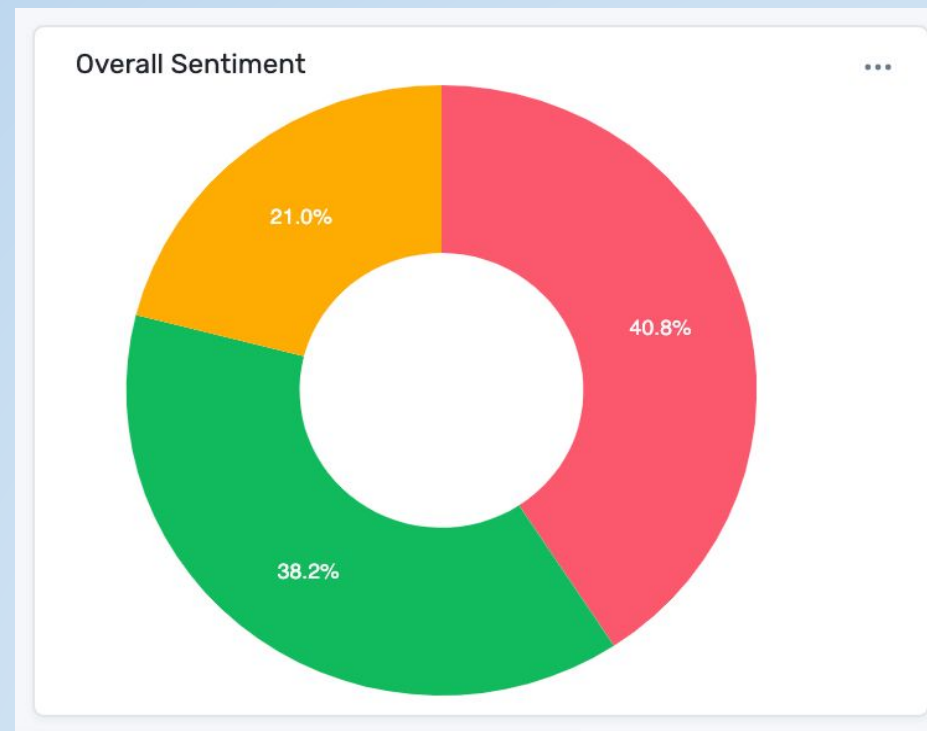
Тоналды бағалау мысалдары:

1. оң;
2. теріс;
3. Бейтарап

"Бейтарап" дегеніміз мәтінде эмоционалды бояу жоқ дегенді білдіреді. Басқа тоналды бағалау да болуы мүмкін.

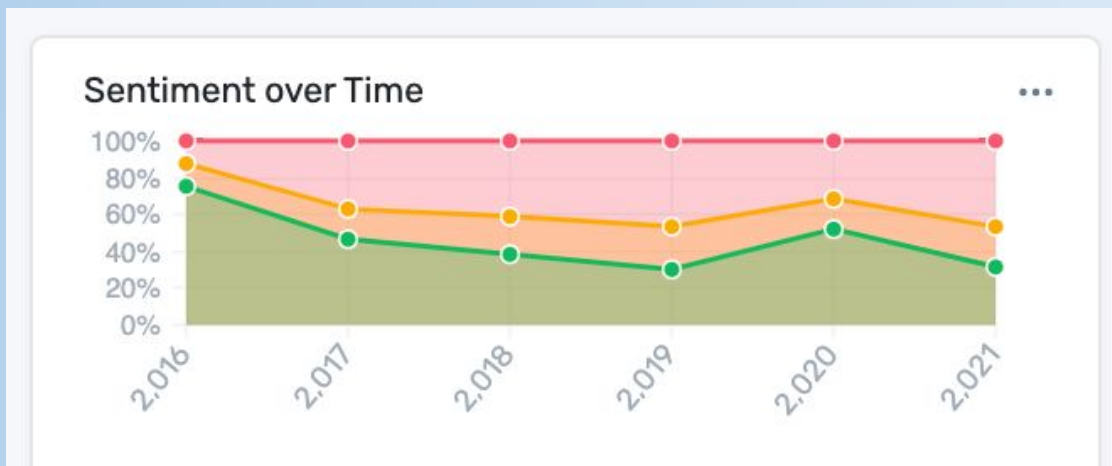
Жалпы көңіл-күй

Біз жоғарыдағы бақылау тақтасынан тиісті графиканы шығарудан бастаймыз. Сіз бұл нәтижелер Trustpilot шолуынан мүлдем өзгеше екенін байқайсыз (82% өте жақсы және т.б.). Себебі, MonkeyLearn көңіл-күйді талдау AI әр сөйлемді, сөзді, сөзді талдау арқылы көңіл-күйді кеңейтілген талдауды жүзеге асырады. Сізге қалған нәрсе- клиенттер жазған барлық нәрсені дәл бағалау. Бұл талдау сізге қарастырылуы тиіс нүктелерді дәлірек және толығырақ көрсете алады.



Уақыт бойынша көңіл-күй

Деректерді визуализациялаудың бұл үлгісі-классикалық уақыттық диаграмма, нәтижелерді бақылайтын және оларды белгілі бір уақыт аралығында көрсететін диаграмма түрі. Бұл кесте біздің көңіл - күй туралы жалпы мәліметтерімізді кеңейтеді-ол 2016 жылдан 2021 жылға дейінгі шолулардағы оң, бейтарап және теріс пікірлердің жалпы үлесін бақылайды. Бұл график осы бес жылдық кезең ішінде олардың жазбаша шолуларының мазмұнының біртіндеп өзгеруін көрсетеді. Мысалы, теріс жауаптар 2019-2020 жылдардан бастап төмендеді, содан кейін 2021 жылы бұрынғы деңгейге көтерілді.



Рейтинг бойынша көңіл-күй

Енді біз мәтіндік пікірімізді алдыңғы Trustpilot нәтижелеріне байланыстыратын нәрсеге көшеміз. TrustPilot - тың әр санатын 1-ден нашар-5-ке дейін алып, жазбаша пікірлердің мәтінін бағалауға бөлу арқылы Сіз жоғарыдағы кестені ала аласыз.

Нәтижелерге қарап, пікірлерді терең зерттеу арқылы көңіл-күйді талдау арқылы біз бірден бірнеше қызықты қорытынды жасай аламыз. Trustpilots нәтижелері пайдасыз емес-шолулар неғұрлым жақсы болса, оң көңіл-күйдің үлесі соғұрлым жоғары болады, ал шолулар неғұрлым нашар болса, теріс көңіл-күй соғұрлым көп болады.

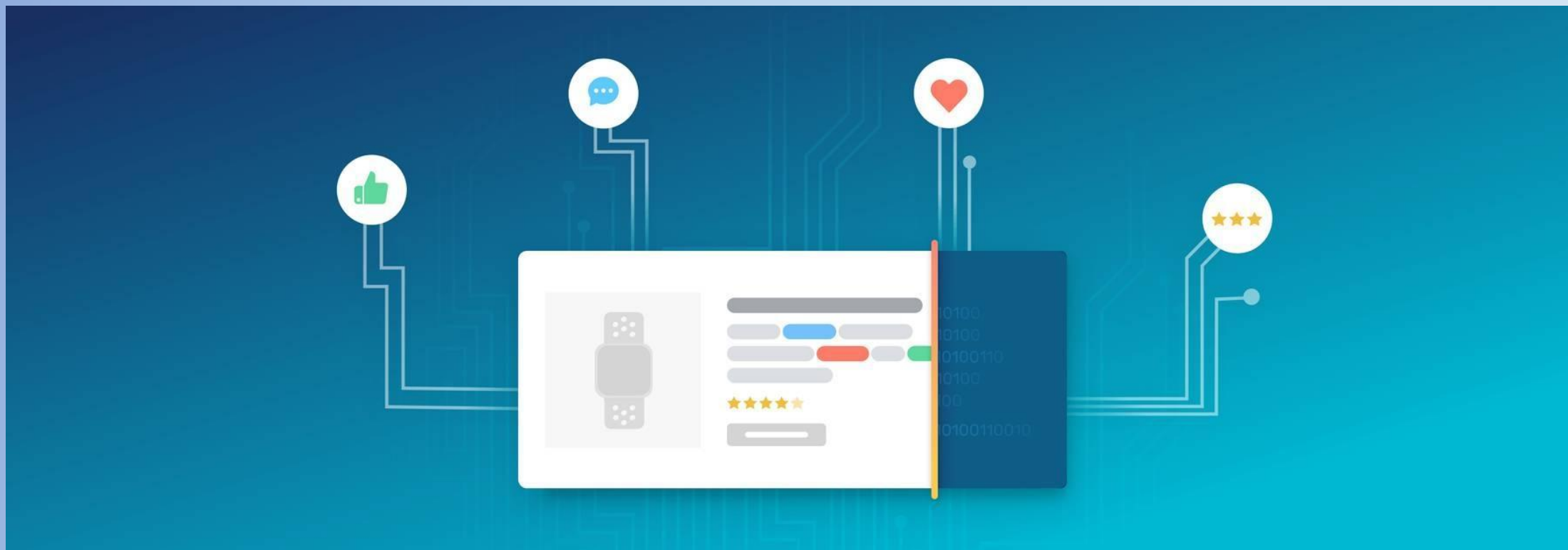
Бірақ барлық шолуларда көңіл-күйдің барлық түрлері бар - біз шолуларымызда нюанстар бар екенін білдік, сондықтан біз үшін одан да жасырын түсінік болуы мүмкін! Біздің пікірлеріміз полярланған. Олар мөлшерде 5 және 1 жағына ауысады. Бұл жылдам тұжырымдар бізге қосымша талдау жасау үшін алтын кен орындарын көрсетеді. Атап айтқанда, теріс пікірлердегі оң пікірлердің бөлімдері және оң пікірлердің теріс бөлімі, сондай-ақ 2-4 шолу (неге олар өздерін осылай сезінеді, біз олардың бағаларын қалай жақсарта аламыз?).



Тақырып бойынша көңіл-күй

Қорытындылай келе, көңіл-күйді талдау бізді деректерімізге қалай тереңдете алатындығын суреттеу үшін тақырып бойынша көңіл-күйді қарастыра аламыз. Жоғарыда келтірілген диаграммада берілген көңіл-күйді өнімнің/қызметтің нақты функцияларымен салыстыру үшін көңіл-күйді талдаудан басқа, өніммен байланысты мәтінді жіктеу қолданылады, бұл аспектілерге негізделген көңіл-күйді талдау ретінде белгілі. Бұл дегеніміз, біз белгілі бір ауырсыну нүктелерін немесе проблемаларын шешуге және жоюға көмектесу арқылы клиенттеріміздің не туралы екенін біле аламыз. Мұның бәрі көңіл-күйді талдаудың құндылығын көрнекі түрде көрсетуге арналған керемет сілтемелер, бірақ олар тек оның шынайы күшінің көрсетеді.

СА қалай жұмыс істейді?



Пікірді талдау деп аталатын көңіл-күйді талдау табиғи тілді өңдеу (NLP) және интернеттегі сөйлесулердің эмоционалды үнін автоматты түрде анықтау үшін машинаны оқыту алгоритмдері арқылы жұмыс істейді. Көңіл-күйді талдау модельдерінде сіз қанша деректерді талдау керек және модель қаншалықты дәл болуы керек екеніне байланысты әртүрлі алгоритмдер бар. Төменде біз олардың кейбірін толығырақ қарастырамыз.

Көңіл-күйді талдау алгоритмдері үш типтегі әдістердің біріне түседі:

1. Ережеге негізделген: бұл жүйелер автоматты түрде қолмен жасалған ережелер жиынтығы негізінде көңіл-күйді талдайды.
2. Автоматты: жүйелер мәліметтер негізінде оқыту үшін машиналық оқыту әдістеріне сүйенеді.
3. Гибридті жүйелер ережелерге негізделген және автоматты тәсілдерді біріктіреді.

Ережеге негізделген әдіс

Әдетте ережеге негізделген жүйе субъективтілікті, полярлықты немесе пікірдің тақырыбын анықтауға көмектесу үшін адам жасаған ережелер жиынтығын қолданады.

Бұл ережелер компьютерлік Лингвистикада жасалған NLP әдістерін қамтуы мүмкін, мысалы:

- Стемминг, токенизация, сөйлем мүшелері бойынша талдау және талдау.
- Лексикондар (яғни сөздер мен сөз тіркестерінің тізімі).

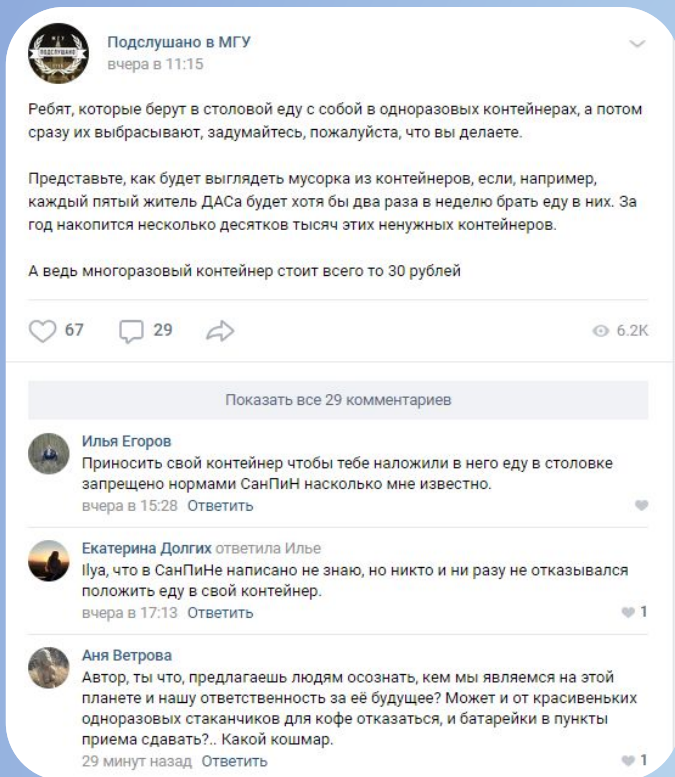
Ережеге негізделген жүйенің қалай жұмыс істейтіні туралы қарапайым мысал:Полярланған сөздердің екі тізімін анықтайды (мысалы, жаман, нашар және т.б. сияқты теріс сөздер және жақсы, керемет, әдемі және т. б. сияқты жағымды сөздер).

Осы мәтінде пайда болатын оң және теріс сөздердің санын есептейді.Егер сөздердің оң көріністерінің саны сөздердің теріс көріністерінің санынан көп болса, жүйе оң көңіл-күйді қайтарады және керісінше. Егер сандар жұп болса, жүйе бейтарап қатынасты қайтарады.

Ережеге негізделген жүйелер өте қарапайым, өйткені олар сөздердің қалай үйлесетінін ескермейді. Әрине, өңдеудің жетілдірілген әдістерін қолдануға болады және жаңа өрнектер мен лексиканы қолдау үшін жаңа ережелер қосуға болады. Алайда, жаңа ережелерді қосу алдыңғы нәтижелерге әсер етуі мүмкін және бүкіл жүйе өте күрделі болуы мүмкін. Ережеге негізделген жүйелер көбінесе дәл баптауды және техникалық қызмет көрсетуді қажет ететіндіктен, оларға тұрақты инвестициялар қажет болады.

Корпус жинау. Алдын ала өңдеу

Бұл мәтіндер корпусын қолмен жинауға болады немесе осы мақсат үшін арнайы бағдарламаларды қолдануға болады: Webometric Analyst, Datacol, VKComment Parser және т. б.



Подслушано в МГУ
вчера в 11:15

Ребята, которые берут в столовой еду с собой в одноразовых контейнерах, а потом сразу их выбрасывают, задумайтесь, пожалуйста, что вы делаете.

Представьте, как будет выглядеть мусорка из контейнеров, если, например, каждый пятый житель ДАСа будет хотя бы два раза в неделю брать еду в них. За год накопится несколько десятков тысяч этих ненужных контейнеров.

А ведь многоразовый контейнер стоит всего то 30 рублей

67 29 6.2K

Показать все 29 комментариев

Илья Егоров
Приносить свой контейнер чтобы тебе наложили в него еду в столовке запрещено нормами СанПиН насколько мне известно.
вчера в 15:28 Ответить

Екатерина Долгих ответила Илье
Иуа, что в СанПиНе написано не знаю, но никто и ни разу не отказывался положить еду в свой контейнер.
вчера в 17:13 Ответить 1

Аня Ветрова
Автор, ты что, предлагаешь людям осознать, кем мы являемся на этой планете и нашу ответственность за её будущее? Может и от красивеньких одноразовых стаканчиков для кофе отказаться, и батарейки в пункты приема сдавать?.. Какой кошмар.
29 минут назад Ответить 1



Бұл бағдарламалар мәтінді келесі өңдеуді жүзеге асыра алады:

- лемматизация-сөздерді қалыпты (сөздік) формаларға келтіру процесі (бұдан әрі оларды сөздіктерден іздеу, шаблондарды таңдау және т. б. ыңғайлы болады);

"такой интересной книги я давно не читала" =>

=>"такой интересный книга я давно не

читать"

Алдын ала өңдеу

- **стемминг** – қосымшалардан ажырату:
«**мал**юесенький экран»));
- **стоп-сөздерді өшіру** – жиі кездесетін, бірақ мағыналық күштері жоқ сөздер
 - **предлогтар** (**в, на, под**);
 - **кейбір есімдіктер** (жіктік: **оның, менің және т.б.**);
 - **кейбір шылаулар** (**және, мен**);
және т.б.

Алдын ала өңдеу

- төменгі регистрге келтіру (кейбір жағдайларда эмоционалды екпін де жоғалады):
- **«Приобретение данного товара было ОШИБКОЙ» =>**
=> «Приобретение данного товара было ошибкой»
- морфологическая разметка (в текстовый корпус вставляются метаданные для обозначения частей речи и др.);

Маңыздарды ерекшелеу

Сентименталды талдау кезінде келесі компоненттерді бөліп алу керек:

- 1) тоналдылық субъекті - пікірдің қайнар көзі, хабарламаның авторы;
- 2) тоналдылық нысаны-мәтінде (фильм, ноутбук моделі) айтылып жатқан нәрсе;
- 3) тоналдылық аспектісі — объектінің сипаттамасы (мысалы, фильм үшін бұл актерлердің ойыны, арнайы эффектілер, сюжет, музыкалық қатар және т. б. болуы мүмкін);
- 4) тоналдылықты бағалау — пікірдің түрі, автордың жеке аспектісіне немесе жалпы объектіге қатынасы).

Сентимент талдау кезіндегі классификация

Мәтін: **оң/ теріс**

Комментарий:

- **мұңды**
- **қуанышты**
- **ашулы**

Пікір:

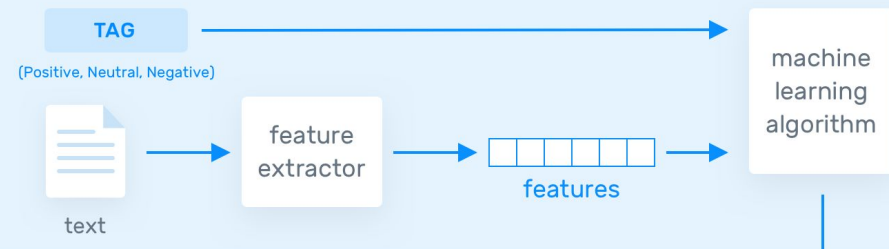
- ◆ **оң**
- ◆ **нейтралды**
- ◆ **теріс**

Автоматты СА тәсілдері

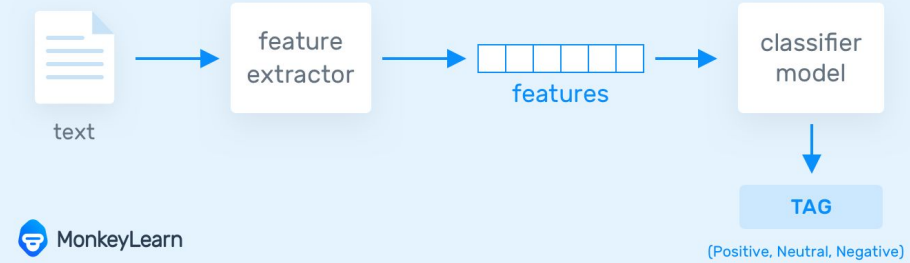
Автоматты әдістер, ережеге негізделген жүйелерден айырмашылығы, қолмен жасалған ережелерге емес, машинаны оқыту әдістеріне сүйенеді. Көңіл-күйді талдау міндеті, әдетте, жіктеу тапсырмасы ретінде модельденеді, онда классификатор мәтін алады және категорияны қайтарады, мысалы, оң, теріс немесе бейтарап.

How Does Sentiment Analysis Work?

(a) Training



(b) Prediction



Оқыту және болжау процестері

Оқу процесінде (а) біздің модель оқыту үшін қолданылатын тест үлгілеріне негізделген белгілі бір кірісті (яғни мәтінді) тиісті тұжырыммен (тегпен) байланыстыруды үйренеді. Нысандарды шығару құралы енгізілген мәтінді объектілер векторына өткізеді. Модель құру үшін машинаны оқыту алгоритміне белгілер мен тег векторларының жұптары (мысалы, оң, теріс немесе бейтарап) енгізіледі. Болжау процесінде (b) объект экстракторы көрінбейтін мәтіндік кірістерді объект векторларына түрлендіру үшін қолданылады. Содан кейін бұл белгілер векторлары болжамды тегтерді құратын модельге енгізіледі (қайтадан оң, теріс немесе бейтарап).

Мәтіннен функцияларды шығару

Машиналық оқыту мәтінін жіктеудегі алғашқы қадам - мәтінді шығаруды түрлендіру немесе мәтінді векторлау, ал классикалық тәсіл-бұл сөздер жиынтығы немесе жиілігі бар нграмм пакеті. Жақында сөздерді ендіруге негізделген объектілерді алудың жаңа әдістері қолданылды (сөз векторлары деп те аталады). Мұндай көріністер ұқсас мағынасы бар сөздерге ұқсас көрініске ие болуға мүмкіндік береді, бұл классификаторлардың жұмысын жақсарта алады.

Жіктеу алгоритмдері

Жіктеу кезеңі әдетте статистикалық модельді қамтиды, мысалы, аңғал Байес, логистикалық регрессия, тірек векторлық машиналар немесе нейрондық желілер:

Аңқау Байес: Мәтін санатын болжау үшін Байес теоремасын қолданатын ықтималды алгоритмдер тобы.

Сызықтық регрессия: статистикадағы өте танымал алгоритм, белгілер жиынтығын (X) ескере отырып, белгілі бір мәнді (Y) болжау үшін қолданылады.

Тірек векторлық машиналар: көп өлшемді кеңістіктегі нүктелер түрінде мәтіндік мысалдардың көрінісін қолданатын ықтималды емес модель. Әр түрлі категориялардың (көңіл-күйлердің) мысалдары осы кеңістіктегі әртүрлі аймақтармен салыстырылады. Содан кейін жаңа мәтіндерге қолданыстағы мәтіндерге және олар салыстырылатын аймақтарға ұқсастық негізінде санат беріледі.

Терең оқыту: деректерді өңдеу үшін жасанды нейрондық желілерді қолдана отырып, адам миын модельдеуге тырысатын әртүрлі алгоритмдер жиынтығы.

Автоматты тоналдылықты талдау тәсілдері

Мәтіннің тоналдылығын автоматты түрде анықтаудың негізгі тәсілдерін 2 үлкен топқа бөлуге болады:

- ережелер, үлгілер мен сөздіктерге негізделген лингвистикалық алгоритмдер;
- Машиналық оқыту әдістерін қолданатын алгоритмдер. Көптеген коммерциялық жүйелер бірінші тәсілді ең дәл деп санайды.

Бағалау лексикасының сөздіктері

Бағалау лексикасының сөздігі-сөздер мен n-компоненттік тізбектер сақталатын мәліметтер базасы-N-граммдар (мысалы, фразеологизмдер және әртүрлі тұрақты тіркестер ("екі езуі екі құлағында")), әр осындай бірлікке эмоционалды бағалау деңгейі беріледі.

Сөздіктер:

- әр түрлі бағалау шкалаларын қолданады
- тізімдерді автоматты түрде толтырады.

Сөздіктердің түрлері

Сөздерді тек бір тоналды бағалауға жатқызуға болады – полярлықтың сандық мәні (нөлден үлкен сан – оң сентимент, нөлден кіші сан – теріс сентимент).

Сөз/сөз тіркесі	Эмоционалдық баға деңгейі
отвратительный	-5
с гилькин нос	-2
модный	+3
усталый	-2

Ағылшын тіліне арналған сөздік:

AFINN

Кейбір басқа жүйелерде (мысалы, SentiStrength) сөз топтары бір емес, екі тоналды бағалауды алады (оң және теріс).

Сөздіктердің түрлері

Сөздерге әр түрлі эмоционалды категориялар жататын лексикондар бар, оларға NRC Word-Emotion Association Lexicon кіреді. Мұнда әр сөзге 2 тоналды бағалау және 8 эмоция сәйкес келеді: "ашу", "қорқыныш", "ескерту", "сенім", "таңдану", "қайғы", "жиіркеніш", "қуаныш". Бұл сөздіктің тізімі бірнеше ондаған тілдерге аударылды, олардың арасында орыс тілі де бар.

Сөз	Эмоционалдық немесе тоналды бағасы	Мәні (1 – бар; 0 – жоқ)
откровенный	гнев	0
	страх	0
	предчувствие	0
	вера	1
	удивление	0
	грусть	0
	отвращение	0
	радость	0
	положительная	1
	отрицательная	0

Сөздіктердің түрлері

Эмоциональная метка	Пример
Эмоция (emotion)	сущ. гнев#1, гл. бояться#1 (fear)
Настроение (mood)	сущ. враждебность#1 (animosity), прил. любезный#1J (amiable)
Особенность (trait)	сущ. агрессивность#1 (aggressiveness), прил. соперничающий#1 (competitive)
Когнитивное состояние (cognitive state)	сущ. замешательство#2 (confusion), прил. потрясенный#2 (dazed)
Физическое состояние (physical state)	сущ. хворь#1 (illness), прил. выдохнувшийся#1 (all in)
Гедонический сигнал (hedonic signal)	сущ. боль#3(hurt), сущ. страдание#4 (suffering)
Ситуации, вызывающие эмоции (emotion-eliciting situation)	сущ. неловкость#3 (awkwardness), сущ. безопасность#1 (out of danger)
Эмоциональные отклики (emotional responses)	сущ. холодный пот#1 (cold sweat), гл. дрожать#2 (tremble)
Поступки (behaviour)	сущ. преступление#1 (offense), прил. заторможенный#1 (inhibited)
Отношение, позиция (attitude)	сущ. нетерпимость#1 (intolerance), сущ. оборонительная позиция#1 (defensive)
Чувство (sensation)	сущ. холод#1 (coldness), гл. чувствовать#3 (feel)

WordNet-Affect тезаурусында эмоционалды категорияны көрсететін белгілермен ("ашу", "қорқыныш", "таңдану", "қайғы", "жиіркеніш", "қуаныш") және валенттіліктермен (оң, теріс, белгісіз, бейтарап), сөздік бірліктермен – синсеттермен, синонимдік қатарлармен – эмоцияларды сипаттайтын белгілер салыстырылды: "физикалық жағдай", "көңіл-күй", "мінез-құлық", "көзқарас", "сезім" және т.б. бұл тезаурус ағылшын тілінен аударылды орыс және румын тілдеріне аударылған.

Лексикалық тәсіл

Шаблондар: **<сын есім зат есім>**, **<сын есім сын есім>**

Шаблондар бойынша мәтіннен N-грамм алынады. Олардың тоналдылығы сөздіктің көмегімен де, ережелер арқылы да анықталады.

Бүкіл мәтіннің тоналдылығы сөйлемдердің тоналдылығынан, ал сөйлемдердің тоналдылығы сөздердің тоналдылығынан тұрады. Соңғы бояуды алу үшін таразының жалпы сомасын белгілі бір шешімді жасаушылар құрайтын формула бойынша есептеу керек, әмбебап формула жоқ.

Лингвистикалық тәсілдегі ережелердің

МЫСАЛЫ

1. «Егер ..., онда» моделі бойынша құрастырылған ережелер

Егер тізбекте тізімдегі етістік болса ("жақсы көру", "ұнату", және т.б.) және басқа тізімнен етістік болмаса ("қорқынышты", "жиіркенішті" және т. б.) немесе бас тарту болса, онда оның тоналдылығы оң болады.

2. Сөздерді олардың модификаторларымен қатар өңдейтін ережелер

Модификаторлар:

- бастапқы тоналдылықты күшейту ("өте", "көп") ;
- бастапқы тоналдылықты төмендету ("тым", "аз") ;
- кері ("Жоқ", «емес») бастапқы тоналдылыққа айналдыру.

Кейбір коэффициенттер тональды модификаторларға жатады, олар тиісті бағалау сөзінің априорлық полярлығына қатысты факторлар ретінде қарастырылады.

Лингвистикалық тәсілдегі ережелердің мысалы

3. Сөздерді коннотациямен өңдеу ережелері.

Коннотациялар-бұл сөздермен байланысты бағалау бірлестіктері. Мәтінде оң немесе теріс коннотациялары бар сөздердің пайда болуы мәтінде көрсетілген тиісті бағалармен байланысты. Сонымен, фильмдердегі шолуларда әдетте белгілі актерлердің есімдері оң коннотациялары бар сөздер болып табылады. Орыс тіліндегі мейрамханалар туралы шолуларда "майонез" және "майлық" сияқты сөздер теріс коннотацияға ие.

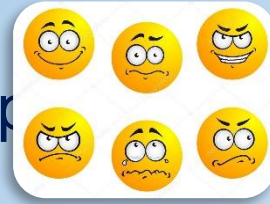
Егер бұл сөздер шолуда пайда болса, әдетте бұл жерде теріс бағалау көрсетіледі.

«Кәдімгі дастарханның орнына майлы шүберек...»

«Әкелген салаттарда майонез көп болды!»

UGC (user-generated content) мәтіндердің ерекшеліктері

Әлеуметтік медиа мәтіндерінің ерекшеліктері:



- Эмотикондар мен смайликтер
- Қате теру;
- Неологизмдер («пичалька» - қате жазылды ма, жоқ па?);
- Окказионализмдер – жеке авторлық неологизмдер;
- Эмоционалды боялған аббревиатуралар («omg!..»);

және т.б.

Смайл	Тональность	Смайл	Тональность
:-)	положительный	:o)	положительный
:-D	положительный	;)	положительный
;-)	положительный	;v)	положительный
xD	положительный	:D	положительный
;-P	положительный	:^D	положительный
:-p	положительный	:/	негативный
8-)	положительный	:-b	положительный
B-)	положительный	=^*	положительный
:-{	негативный	:-x	положительный
;-]	положительный	8-]	положительный
3(негативный	>:-(негативный
:!(негативный	>:[негативный
:_(негативный	:-0	негативный
:{(негативный	:-o	негативный
:o	негативный	;-{	негативный
3-)	положительный	:!-(негативный
O:)	положительный	;-)	положительный
;o	негативный	:<	негативный
<3	положительный	:-	негативный
:-*	положительный	D;	негативный

Лингвистикалық тәсілдің кемшіліктері

Артықшылығы: жоғары дәлдік

Кемшіліктері:

- ережелер жүйесін құру өте қиын міндет;
- ережелер мен сөздіктер әдісі әмбебап емес (пәндік аймаққа тәуелділік бар)

Гибридті Тәсілдер

Гибридті жүйелер ережелер мен автоматты әдістерге негізделген қажетті элементтерді бір жүйеге біріктіреді. Бұл жүйелердің үлкен артықшылықтарының бірі-нәтижелер көбінесе дәлірек болады.

Сентимент талдаудың артықшылықтары

1) Деректерді масштабта сұрыптау. Сіз мыңдаған твиттерді, тұтынушыларды қолдау сөйлесулерін немесе сауалнамаларды қолмен сұрыптауды елестете аласыз ба? Қолмен өңдеуге арналған бизнес деректері өте көп. Көңіл-күйді талдау компанияларға құрылымданбаған деректердің үлкен көлемін тиімді және үнемді өңдеуге көмектеседі.

2) Нақты Уақыттағы Талдау. Көңіл-күйді талдау нақты уақыттағы маңызды мәселелерді анықтай алады, мысалы, әлеуметтік желілердегі PR дағдарысы күшейе ме? Мүмкін ашулы клиент жарылып кетуі мүмкін бе? Көңіл-күйді талдау модельдері сізге осындай жағдайларды тез арада анықтауға көмектеседі, осылайша сіз дереу әрекет ете аласыз.

3) Келісілген критерийлер. Белгілі бір мәтіннің көңіл-күйін анықтауда адамдар тек 60-65% жағдайда келіседі деп есептеледі. Мәтінді сезіммен белгілеу өте субъективті, оған жеке тәжірибелер, ойлар мен сенімдер әсер етеді. Көңіл-күйді талдаудың орталықтандырылған жүйесін қолдана отырып, компаниялар барлық мәліметтерге бірдей өлшемдерді қолдана алады, бұл олардың дәлдігін арттыруға және толық түсінік алуға көмектеседі.

Сентимент талдаудың жалпы мәселелері

Тоналдылықты автоматты түрде анықтаудың кез келген жүйесінде 2 үлкен мәселе кездеседі:

□ Қалжың және сарказм сөйлемдерін өңдеу;

«Было скучно. Давно не смотрела фильмов с настолько интригующим сюжетом»

□ Қолданушы метафораларын өңдеу;

«школа как второй дом»

□ Пікірде бірден бірнеше нысан жайлы айту кезінде болатын қиындықтар;

«Huawei впервые обогнала Apple по продажам смартфонов в России».

SentiStrength

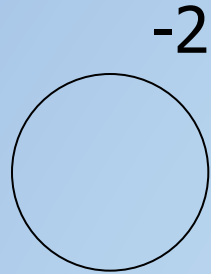
- Қысқа бейресми мәтіндегі оң және теріс көңіл-күйдің күшін анықтау
- Стандартты грамматика мен емленің болмауына байланысты жұмыс жасау
- MySpace немесе СМС (мысалы :-) немесе һаарrrrrууу үшін ерекше эмоцияны білдіру формаларын қолдану)
- Бір уақытта оң 1-5 және теріс 1-5 көңіл-күйді жіктеу

SentiStrength

- 2489 терминдердің оң және теріс мағыналарының тізімі және күшті жақтары (1-ден 5-ке дейін), мысалы
- ache = -2, dislike = -3, hate=-4, excruciating -5
 - encourage = 2, coolest = 3, lover = 4

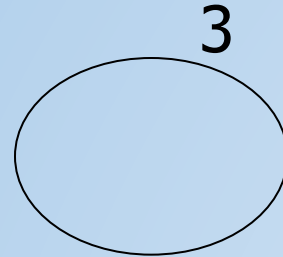
positive, negative

- My legs ache.



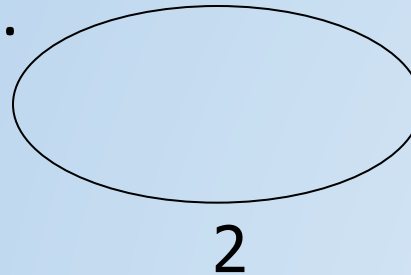
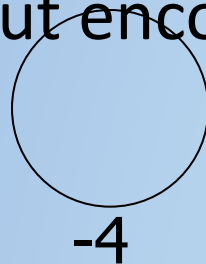
1, -2

- You are the coolest.



3, -1

- I hate Paul but encourage him.



2, -4

Экстра сентименттік әдістер

- **spelling correction** **nicce -> nice**
- **booster words** alter strength **very happy**
- **negating words** flip emotions **not nice**
- **repeated letters** boost sentiment/+ve **niiiice**
- **emoticon list** **:) =+2**
- **exclamation marks** count as +2 unless -ve **hi!**
- **repeated punctuation** boosts sentiment **good!!!**
- **negative emotion ignored in questions** **u h8 me?**
- **Sentiment idiom list** **shock horror = -2**

SentiStrength
күші адамдармен
бір-бірімен
келіскендей
үйлеседі

Data set	Positive scores -correlation with humans	Negative scores -correlation with humans
YouTube	0.589	0.521
MySpace	0.647	0.599
Twitter	0.541	0.499
Sports forum	0.567	0.541
Digg.com news	0.352	0.552
BBC forums	0.296	0.591
All 6 data sets	0.556	0.565

1-тамаша келісім, 0-кездейсоқ келісім

СТ қиындықтар:

- Ирония, сарказм:
 - David Cameron must be **very happy** that I have lost my job.
 - It is **really interesting** that David Cameron and most of his ministers are millionaires.
 - Your argument is a **joke**.

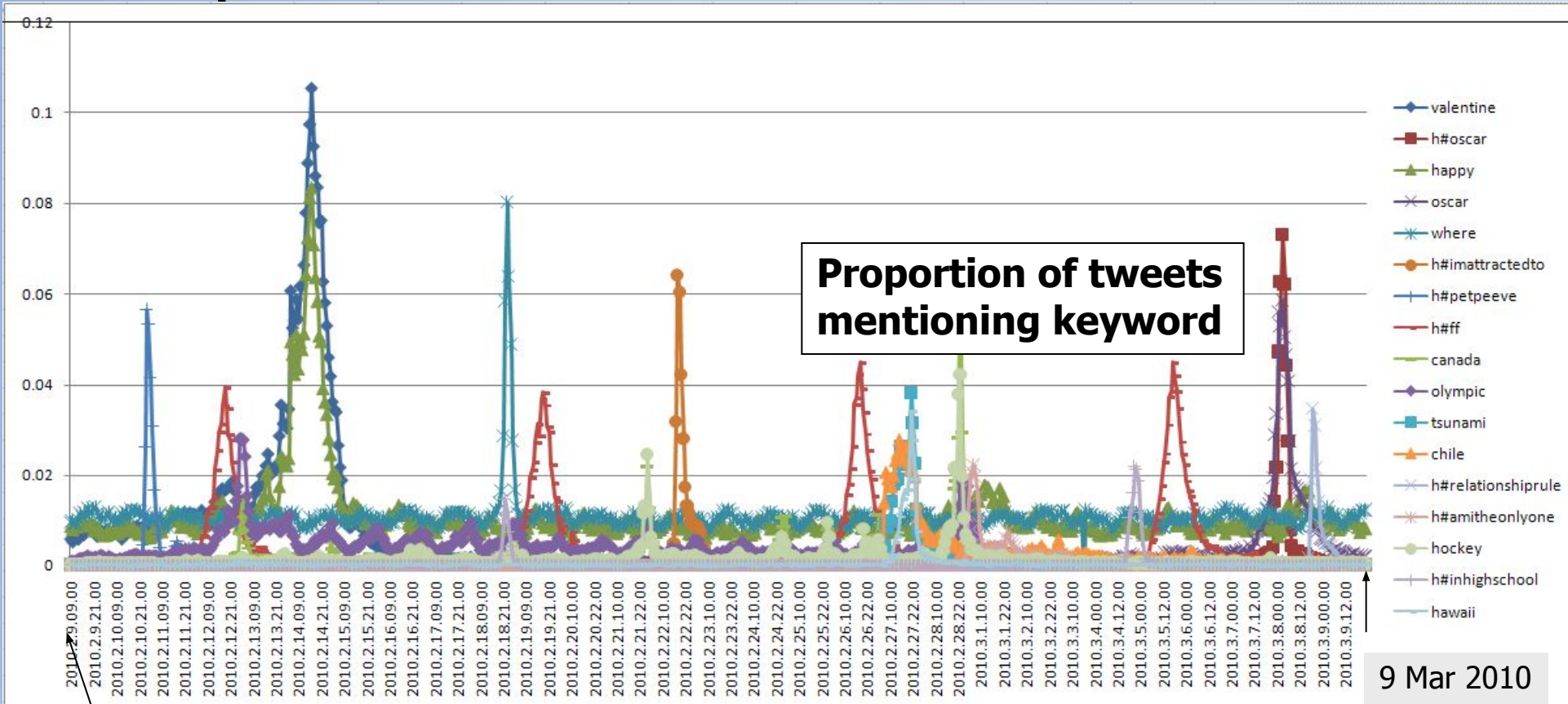


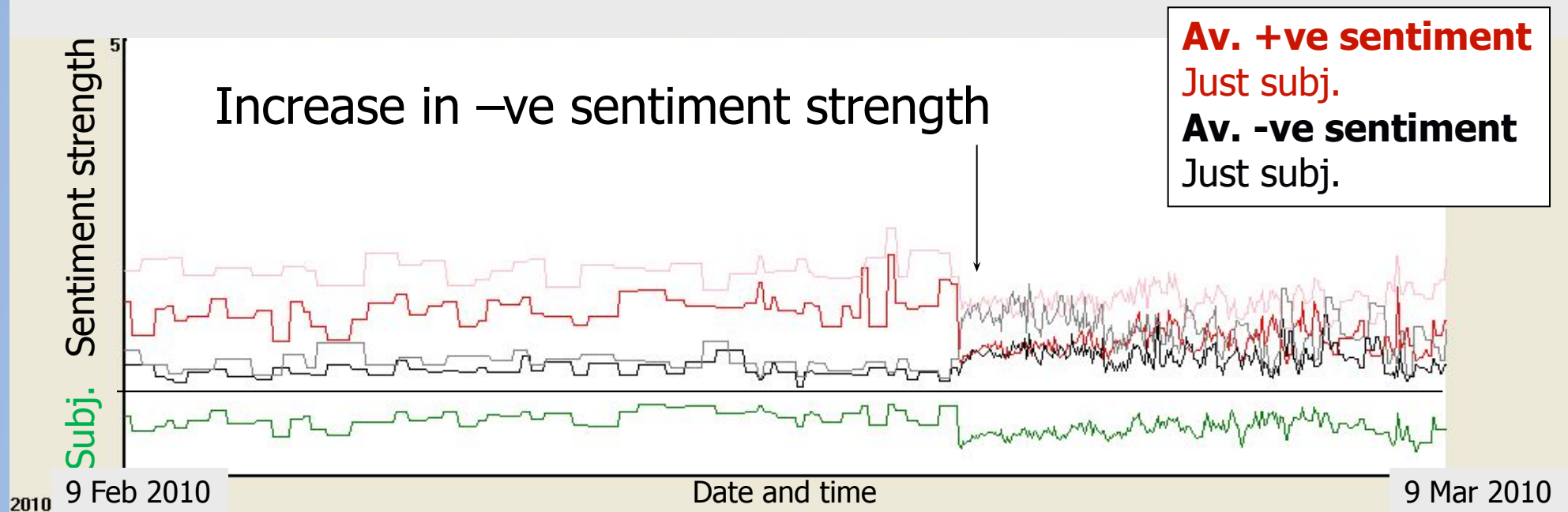
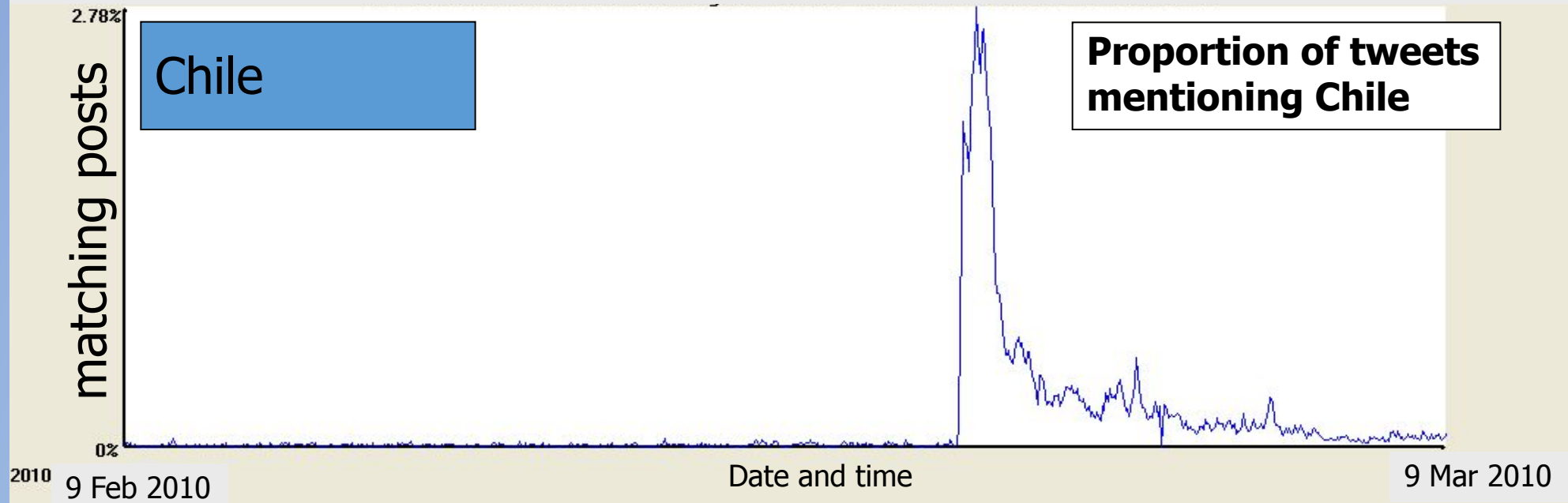
Мысал-бұқаралық ақпарат құралдарындағы ірі іс-шаралардағы көңіл-күй

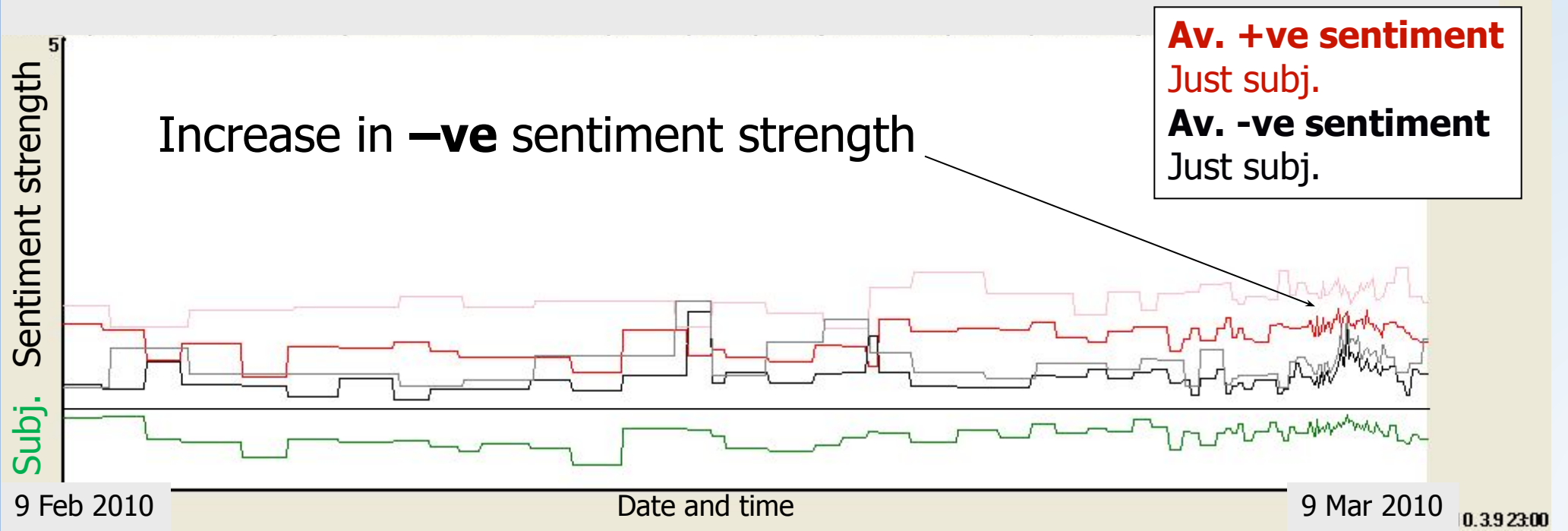
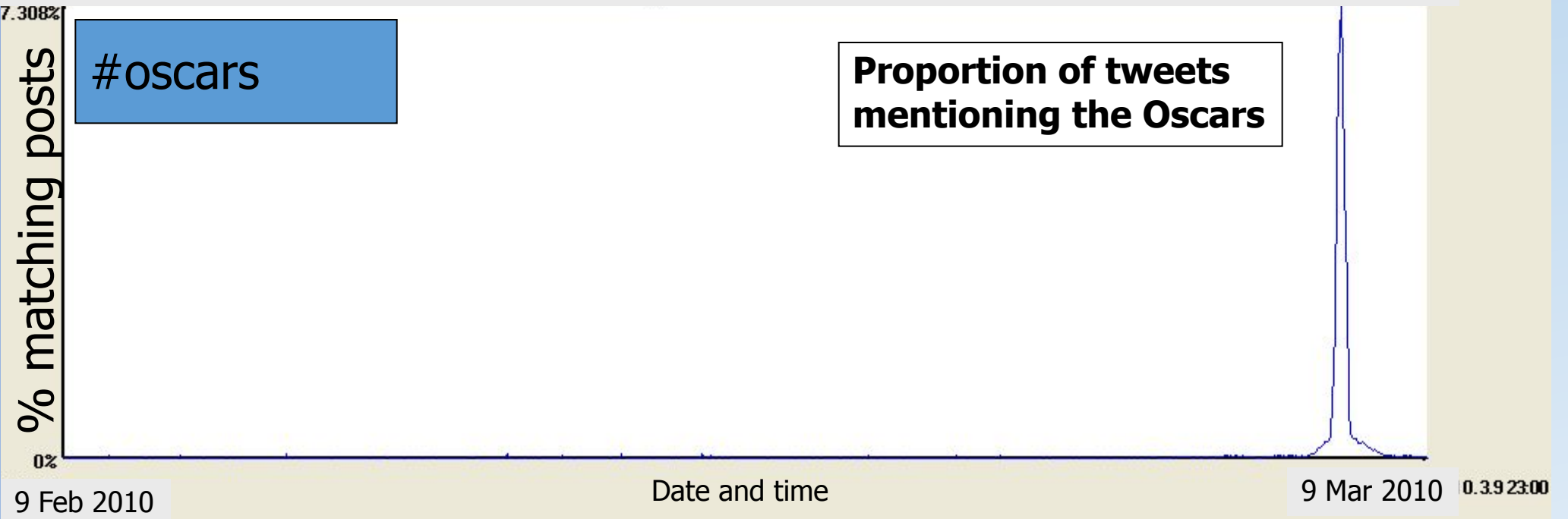
- 1 айдағы ағылшын тіліндегі Твиттердегі хабарламалар корпусын талдау (35 миллион, 2,7 миллион аккаунттан)
- Жарылыстарды (оқиғаларды)автоматты түрде анықтау
- Бұқаралық ақпарат құралдарындағы ірі оқиғалар кезінде көңіл-күйдің өзгеретінін бағалау



Твиттердегі автоматты түрде анықталатын оқиғалар







Көңіл-күйді талдау табиғи тілді өңдеудегі ең қиын міндеттердің бірі болып табылады, өйткені адамдарға көңіл-күйді дәл талдау қиынға соғады. Деректерді өңдеу мамандары көңіл-күйдің нақты жіктеуіштерін құруда жетілдірілуде, бірақ әлі ұзақ жол бар. Машиналық көңіл-күйді талдаудың негізгі мәселелерін егжей-тегжейлі қарастырайық:

- Субъективтілік және Тон
- Контекст және полярлық
- Ирония және Сарказм
- Салыстыру
- Эмодзи
- Адам аннотаторының дәлдігі

Назарларыңызға рақмет!