

## ВЕБ-ПРИЛОЖЕНИЕ ДЛЯ ОЦЕНКИ ЭМОЦИОНАЛЬНОЙ ТОНАЛЬНОСТИ ОТЗЫВОВ НА ФИЛЬМЫ

При принятии того или иного решения человек часто обращается к советам, рекомендациям и мнениям других людей. Если еще несколько десятков лет назад люди обращались за традиционным советом к своим знакомым, то сейчас все изменилось. С бурным развитием информационно-вычислительных технологий и глобальной сети Интернет, процесс получения рекомендаций значительно упростился. Но вместе с увеличением числа пользователей сети Интернет, возрастает и количество генерируемого ими контента. Объемы информации в Интернете на сегодняшний день так велики, что необходимые данные попросту теряются в море бесполезных данных.

Исходя из вышесказанного, в настоящее время чрезвычайно востребованной является задача автоматического анализа мнений пользователей о товарах или услугах.

Тональность текста – это эмоциональная оценка, выраженная в тексте. Анализ тональности текста (или сентимент-анализ) – это область компьютерной лингвистики, которая занимается извлечением из текста эмоционально окрашенной лексики и получением представления об эмоциональном отношении авторов к объектам, о которых идёт речь в тексте, что влияет на формирование определённого образа у реципиентов [1].

Задача анализа тональности текста возникла сравнительно недавно, поэтому исследования в этой области проводятся в настоящее время, и на данный момент все еще актуальна задача построения системы извлечения мнений и анализа тональностей.

Целью данной работы является создание веб-приложения, при помощи которого можно отыскать интересующий фильм, после просмотреть его детальную информацию и отзывы на него, а также оценку эмоциональной тональности для каждого из отзывов, вычисленную с помощью обученного байесовского классификатора или веб-сервиса *IBMWatsonNaturalLanguageUnderstanding*.

Рассмотрим подробнее оценку эмоциональной тональности с помощью наивного Байесовского классификатора.

Байесовский классификатор – один из самых часто используемых классификаторов, из-за сравнительной простоты в имплементации и тестировании. В то же время, наивный Байесовский классифи-

катор демонстрирует не худшие результаты по сравнению с другими, более сложными классификаторами.

В основе наивного Байесовского классификатора лежит теорема (или формула) Байеса [2].

Основная идея байесовской стратегии заключается в оценке вероятности некоторого вывода с учетом фактов, подтверждающих или опровергающих этот вывод. Формулировка теоремы Байеса, известная из теории вероятностей, следующая.

Пусть имеется  $n$  несовместных событий  $H_1, H_2, \dots, H_n$ . Известны вероятности этих событий:  $P(H_1), P(H_2), \dots, P(H_n)$ , причем  $P(H_1) + P(H_2) + \dots + P(H_n) = 1$ . Это означает, что события  $H_1, H_2, \dots, H_n$  образуют полную группу событий, т.е. одно из них происходит обязательно. С событиями  $H_1, H_2, \dots, H_n$  связано некоторое событие  $E$ . Известны вероятности события  $E$  при условии того, что какое-либо из событий  $H_1, H_2, \dots, H_n$  произошло:  $P(E/H_1), P(E/H_2), \dots, P(E/H_n)$ . Пусть событие  $E$  произошло. Тогда вероятность того, что при этом какое-либо из событий  $H_i$  ( $i=1, \dots, n$ ) произошло, можно найти по следующей формуле (формула Байеса):

$$P(H_i/E) = \frac{P(E/H_i)P(H_i)}{P(E/H_1)P(H_1) + P(E/H_2)P(H_2) + \dots + P(E/H_n)P(H_n)} = \frac{P(EH_i)}{P(E)}. \quad (1)$$

События  $H_1, H_2, \dots, H_n$  называются гипотезами, а событие  $E$  – свидетельством. Вероятности гипотез  $P(H_i)$  без учета свидетельства (т.е. без учета того, произошло событие  $E$  или нет) называются доопытными (априорными), а вероятности  $P(H_i/E)$  – послеопытными (апостериорными). Величина  $P(EH_i)$  – совместная вероятность событий  $E$  и  $H_i$ , т.е. вероятность того, что произойдут оба события вместе. Величина  $P(E)$  – полная (безусловная) вероятность события  $E$ .

Пусть необходимо определить принадлежность отзыва к позитивному отзыву  $D_1$  либо негативному отзыву  $D_2$ . Для этих двух типов отзывов известны условные вероятности  $P(S|D_j)$ , характеризующие появление набора слов  $S$  в отзыве  $D_j$ . Здесь  $S = \{S_1, S_2, \dots, S_k\}$ , где  $S_k$  – различные слова. Пусть далее  $P(D_j)$  – априорная вероятность отзыва  $D_j$ . Тогда задача определения, является ли отзыв положительным, может быть сведена к статистической задаче выбора гипотез, решение которой основывается на использовании теоремы Байеса:

$$P(D_j | S) = \frac{P(D_j) \cdot P(S|D_j)}{\sum_{j=1}^n P(D_j) \cdot P(S|D_j)}, \quad (j = 1, \dots, n). \quad (2)$$

Если для какого-нибудь сообщения  $D_j$  рассчитанная вероятность принадлежности отзыва к положительному по формуле (2) значитель-

но превосходит значение вероятности принадлежности отзыва к негативному, то оптимальное правило приписывает это сообщение к позитивному.

Использование формулы (2) встречается со следующими трудностями: условные вероятности  $P(S|D_j)$  можно сравнительно легко вычислить лишь при небольшом числе признаков  $S_1, S_2, \dots, S_k$ . Во-первых, это может быть связано с тем, что число различных комбинаций признаков может быть очень большим. Например, если каждое слово имеет только два различных значения (позитивное/негативное), а число слов равно 30, то число всех возможных вариантов может исчисляться величиной  $2^{30}$ . Поэтому будем считать слова в отзыве независимыми друг от друга.

В веб-приложении для создания и обучения модели Байесовского классификатора использовался фреймворк для машинного обучения *Accord.NET*.

Вторым способом оценки эмоциональной тональности отзывов является использование веб-сервиса *IBMWatsonNaturalLanguageUnderstanding*.

*IBMWatsonNaturalLanguageUnderstanding* – это набор *API*, которые предлагают анализ текста посредством обработки естественного языка. Этот набор *API* может анализировать текст, чтобы помочь вам понять его концепции, сущности, ключевые слова, настроения и многое другое.

На рисунке 1 представлен пользовательский интерфейс разработанного веб-приложения. На данном представлении можно найти фильм по названию или выбрать интересующие жанры, год выпуска и порядок сортировки.

На рисунке 2 представлено модульное окно для просмотра детальной информации о фильме, а также продемонстрированы отзывы и их оценки эмоциональной тональности.

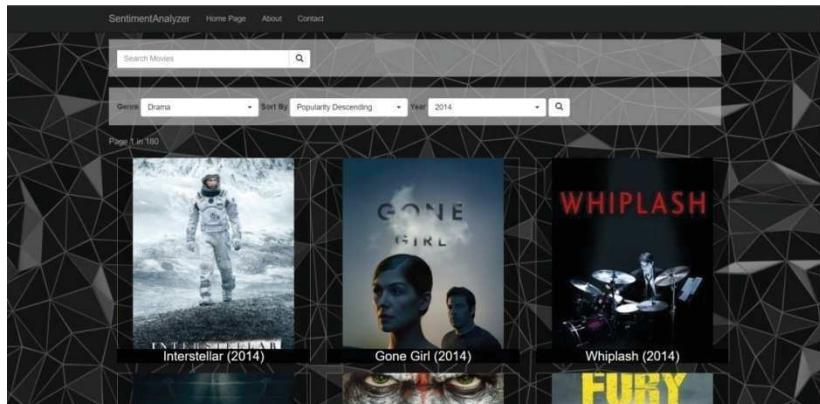


Рисунок 1 – Поиск фильмов

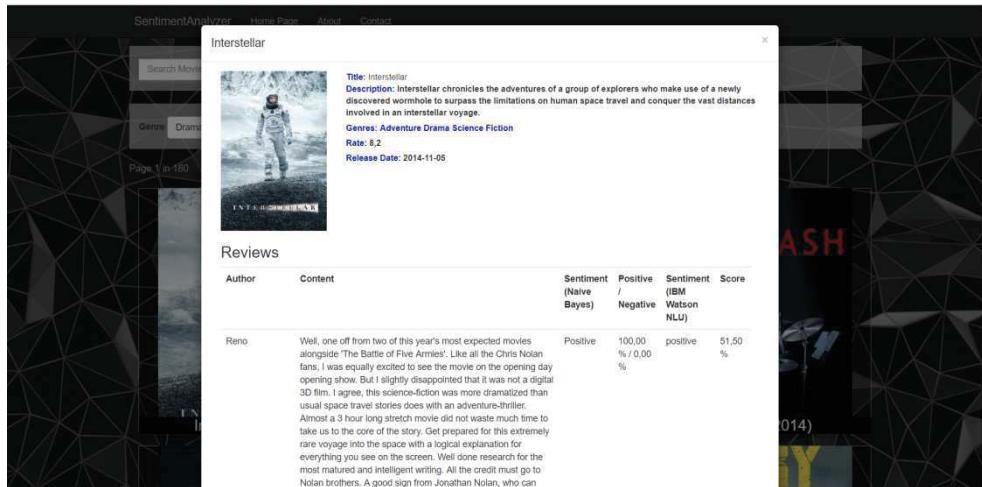


Рисунок 2 – Окно для просмотра информации о фильме

Таким образом, благодаря данному веб-приложению можно быстро и легко оценить рейтинг фильма исходя из полученных тональностей отзывов. Соответственно при использовании продемонстрированного веб-приложения можно значительно сэкономить время, особенно, если фильм имеет большое количество отзывов.

## ЛИТЕРАТУРА

1. Pang B. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval / Б. Панг, Л. Ли. – Москва: Вильямс, 2008. – 235 с.

2. Теорема Байеса [Электронный ресурс] – Режим доступа: <http://www.nsu.ru/mmf/tvims/chernova/tv/lec/node15.html>. Дата доступа: 19.02.2019.

УДК 004.056

Студ. Н. И. Борсук  
Науч. рук. доц. Д. В. Шиман  
(кафедра программной инженерии, БГТУ)

## СЕРВИС ПОДБОРА НЕДВИЖИМОСТИ ДЛЯ ДОЛГОВРЕМЕННОЙ АРЕНДЫ

Первым делом при подборе недвижимости для долговременной аренды человек идет за помощью к сервисам в интернете для подбора жилья. Современные сервисы предлагают возможности подобрать жилье по существующим объявлениям. Сервис для размещения предложений аренды недвижимости должен выполнять следующие основные задачи:

- создание объявления об аренде;
- поиск объявлений по параметрам;