

ОЦЕНКА РЕЛЕВАНТНОСТИ ТЕКСТОВЫХ ОТЗЫВОВ В ПЕРСОНАЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ СРЕДЕ

RELEVANCE SCORE OF FEEDBACK IN THE PERSONAL LEARNING ENVIRONMENT

Алексей Александрович Даничев **Aleksey Aleksandrovich Danichev**

кандидат технических наук, доцент
aleksey-d@inbox.ru

**Александр Константинович
Погребников** **Alexander Konstantinovich
Pogrebnikov**

старший преподаватель
apogrebnikov@sfu-kras.ru

Юрий Юрьевич Якунин **Yuri Yurievich Yakunin**

кандидат технических наук, доцент
yakuninyu@mail.ru

ФГАОУ ВО «Сибирский федеральный
университет», Красноярск, Россия

Siberian Federal University, Krasnoyarsk, Russia

Аннотация. Рассмотрен анализ текстовых ответов при анкетировании студентов в персональной образовательной среде. Изложены основные аспекты автоматического определения тематики и выявления эмоциональных оттенков отзывов.

Abstract. The article discusses the analysis of textual answers to students' questionnaires in the personal learning environment. The main aspects of the automatic determination of the topic and the identification of emotional shades of reviews are described.

Ключевые слова: персональная образовательная среда, семантический анализ, тональность текстов.

Keywords: personal learning environment, semantic analysis, sentiment analysis of text.

Систематическая обработка и анализ результатов опроса студентов об их отношении к образовательному процессу в организации, где они проходят обучение, требуют много ресурсов, тем более зачастую ответы респондентов не формализованы (представлены в свободной форме). Авторами было проведено исследование, направленное на создание методики автоматического синтеза справочника смысловых форм и их тональностей [1, 2].

Исследование выполнено на основе множества собираемых и хранимых в персональной образовательной среде отзывов студентов Института космических и информационных технологий (ИКИТ) Сибирского федерального университета [3]. Начиная с 2015 г. после каждой промежуточной аттестации регулярно проводится анкетирование (исходный корпус состоит из 4 900 отзывов студентов). Анкета включает 16 вопросов, сгруппированных по

трех аспектам образовательной среды, касающимся отношений студентов:

- к работе выпускающей кафедры;
- к работе учебно-организационного отдела (деканата);
- к преподаванию дисциплин и обеспечивающих их материалов.

Как правило, студенты оставляют отрицательные текстовые отзывы, а для положительных оценок ограничиваются только баллами (рис. 1).

Не все оценки респондентов и ответы на вопросы анкет можно использовать как достоверную информацию. Проблема релевантности оценок тесно связана с эмоциональным фактором, субъективным отношением.

Анализ текстовых ответов позволяет наиболее эффективно осуществлять обратную связь на отзыв студента, но такой анализ реализуем только вручную. Лишь в некоторой части этот процесс можно выполнить автоматически, что позволяет применить дополнительные критерии оценки «релевантность ответов студентов» наравне со статистическим анализом балльных оценок [3].

Для автоматизации анализа отзывов необходимо получать численные оценки текста. Для



Рис. 1. Отзывы при анкетировании студентов ИКИТ

текстовых данных авторами было предложено два критерия оценки релевантности ответа [1]:

- соответствие текста отзыва словарю вопроса;
- соответствие эмоциональных оттенков отзыва числовой оценке.

Оценка соответствия выполняется с помощью классификатора. При поступлении нового ответа из него выделяются значимые наборы слов и вычисляются степени соответствия этих наборов слов перечню вопросов. Другой классификатор сопоставляет наборы слов с частотными тональными словарями аспектов.

Автоматически можно выделить только случаи, когда набор слов ответа студента не соответствует типичным словосочетаниям для данного вопроса или балльная оценка ответа не соответствует эмоциональной окраске текста.

Для машинного обучения необходимо сформировать обучающие выборки (словари) с информацией о соответствии слов и вопросов, а также указать для ответов тональность (положительную, отрицательную, нейтральную).

Формирование словарей. Как правило, для автоматической классификации применяется *стемминг*. Он не требователен к ресурсам, но может совершенно разные по смыслу слова привести к одной форме. Сами словари в этом случае представляют собой наборы букв и не позволяют интерпретировать результаты. Альтернативой является *лемматизация текста* (приведение различных словоформ к базовому виду с использованием морфологических словарей и алгоритмов контекстного анализа). В общем случае процесс лемматизации крайне ресурсоемкий. Так как все ответы рассматриваются в одном контексте (в сфере образования), то для данной задачи удалось составить единый словарь, который почти безошибочно позволяет определять базовую форму слов.

Лемматизация базового набора признаков выполнена с помощью библиотеки морфологического анализа *phpMorphy* [4]. Для однозначных случаев сформировано словарное ядро, для прочих морфем определяется лемма, наиболее близкая к леммам ядра. В качестве мер близости используются следующие параметры:

- частота встречаемости леммы в ядре;
- частота встречаемости леммы в опорных текстах;

Классификатор. Словари аспектов

Предложения по улучшению преподавания дисциплины		Удобство графика работы со студентами	
Словосочетание	Условная частота	Словосочетание	Условная частота
предложение	54	пара	12
нет ~ предложение	17	прием	8
нормально	14	неудобный	4
тест	12	обед	4

Таблица 2

Классификатор. Тональный словарь.
Отрицательная тональность

Предложения по улучшению преподавания дисциплины		Удобство графика работы со студентами	
Словосочетание	Условная частота	Словосочетание	Условная частота
должный	8	прием	8
узкий	7	неудобный	4
экзамен	7	вечер	3
преподаватель ~ курс	6	график	3

Таблица 3

Классификатор. Тональный словарь.
Положительная тональность

Предложения по улучшению преподавания дисциплины			
Словосочетание	Условная частота	Словосочетание	Условная частота
нормально	14	все супер	4
комментарий	11	отличный курс	4
все норма	4	прекрасный преподаватель	4

позволяет отсеять шуточные ответы студентов. В табл. 4 приведен пример двух отзывов с низкой релевантностью, тем не менее являющихся корректными. Всего таких ошибочных идентификаций было найдено 19 из 1792 текстовых ответов.

У многих работы были выполнены, но очереди на сдачу не дождешься

Желтый: понятие

Подчеркнуто: отношение

Курсив: неважные фразы

Красный: фраза атрибута отрицания

Рис. 2. Ключевые понятия и связи

Критерий соответствия текста отзыва словарю вопроса

Вопрос	Отзыв	Прогноз	Критерий соответствия
Предложения по улучшению преподавания дисциплины	Если ты болел и принес справку, она не учитывается и все пропуски приходится отрабатывать	Уровень сервиса по работе с заявлениями студентов, выдачи справок	0,55
	Хотелось бы еще и заниматься говорением, а не работать только в рамках электронного образовательного курса	Полнота и качество электронного образовательного ресурса	0,29

В результате, система обработки отзывов в персональной образовательной среде была дополнена двумя критериями релевантности текстовых ответов. Инструмент InterSystems iKnow позволяет быстро получать признаки для анализа текста. При наработанном сло-

варе (из 4000 пар «лексема — лемма» в сфере образования) процедура лемматизации также не требовательна к вычислительным ресурсам. При получении новых ответов переобучение классификатора аспектов происходит автоматически.

Список литературы

1. Даничев, А. А. Аспектный анализ тональности отзывов в образовательной среде / А. А. Даничев, Ю. Ю. Якунин. Текст: непосредственный // Информатизация образования и методика электронного обучения: материалы 3-й Международной научной конференции: в 2 частях / под общ. ред. М. В. Носкова. Красноярск, 24–27 сент., 2019 г. Красноярск: Изд-во Сиб. фед. ун-та, 2019. Ч. 1. С. 61–65.
2. Даничев, А. А. Анализ неструктурированных данных обратной связи в персональной образовательной среде / А. А. Даничев, Ю. Ю. Якунин. Текст: непосредственный // Наука. Информатизация. Технологии. Образование: материалы 12-й Международной научно-практической конференции. Екатеринбург, 25 февр. – 1 марта, 2019 г. Екатеринбург: Изд-во Рос. гос. проф.-пед. ун-та, 2019. С. 263–271.
3. Погребников, А. К. Персональная образовательная среда в системе управления институтом / А. К. Погребников, Ю. Ю. Якунин. Текст: непосредственный // Информатика и образование. 2017. № 2. С. 55–60.
4. Библиотека морфологического анализа phpMorphy URL: <http://phpmorphy.sourceforge.net> (дата обращения: 31.01.2020). Текст: электронный.
5. Онлайн-тезаурус русского языка «Карта слов» URL: <https://github.com/dkulagin/kartaslov> (дата обращения: 31.01.2020). Текст: электронный.
6. Русская семантика. URL: <http://русскаясемантика.рф/dictionary> (дата обращения: 31.01.2020). Текст: электронный.
7. Толковый словарь, систематизированный по классам слов и значений. / под общ. ред. Н. Ю. Шведовой. Москва: Изд-во Ин-та русского языка Рос. акад. наук, 2007. Т. 1–4. Текст: непосредственный.
8. InterSystems Documentation. URL: <https://docs.intersystems.com> (дата обращения: 31.01.2020). Text: electronic.