

ИССЛЕДОВАНИЕ УДОВЛЕТВОРЕННОСТИ ПОТРЕБИТЕЛЕЙ В БАНКОВСКОЙ СФЕРЕ НА ОСНОВЕ АНАЛИЗА ТЕКСТОВЫХ ОТЗЫВОВ

М. В. Бойко

maxim.boyko87@gmail.com

ФГБОУ ВПО «Уфимский государственный авиационный технический университет» (УГАТУ)

Поступила в редакцию 10 июня 2014 г.

Аннотация. Предлагается подход к исследованию удовлетворенности потребителей. В качестве источника данных об удовлетворенности используются потребительские отзывы о продукции на естественном языке, публикуемые в сети Интернет. Для измерения удовлетворенности предлагается использовать оценки лексической тональности потребительских отзывов. Для обработки и анализа данных используется сентимент-анализ и извлечение аспектов продукции с применением методов машинного обучения. На основе предложенного подхода проведено количественное и качественное исследование удовлетворенности клиентов нескольких российских банков.

Ключевые слова: удовлетворенность потребителей; анализ текста; сентимент-анализ; интеллектуальный анализ данных.

Отличительной чертой современного рынка товаров и услуг является превышение предложения над спросом. Каждое предприятие стремится повысить конкурентоспособность своей продукции, достичь высоких показателей деятельности, завоевать как можно большую долю рынка.

Низкое качество продукции и услуг способствует снижению удовлетворенности потребителей. С другой стороны, в условиях жесткой конкуренции для потребителя не существует преград для смены поставщика товаров и услуг. Все это приводит к потере клиентов и снижению показателей эффективности предприятия. Поэтому одной из задач эффективного управления качеством продукции является оценка и мониторинг удовлетворенности потребителей.

Существующие методы измерения и мониторинга удовлетворенности потребителей продукцией требуют больших денежных вложений. Это вызвано тем, что сбор информации и её анализ, как правило, проводится вручную. К тому же такое исследование может занимать достаточно большое количество времени. Не каждое предприятие может позволить себе проведение подобных исследований. Поэтому разработка более эффективного подхода к исследованию удовлетворенности потребителей и его программная реализация является актуальной задачей.

На основе анализа существующих методов сбора данных об удовлетворенности потребителей и методах ее обработки и анализа сформулированы основные требования к информационной системе по сбору и анализу данных об удовлетворенности потребителей:

- 1) возможность ведения непрерывного мониторинга удовлетворенности потребителей;
- 2) минимальное участие человека, задействование специалистов только на этапе настройки системы;
- 3) автоматический сбор и обработка неструктурированных и слабоструктурированных данных;
- 4) возможность выработки рекомендаций для лица, принимающего решения, по повышению удовлетворенности потребителей и качества продукции на основе анализа данных об удовлетворенности потребителей.

В данной работе предлагается подход к исследованию удовлетворенности потребителей на основе сентимент-анализа потребительских отзывов, размещаемых в сети Интернет. Предложенный подход был использован для количественного и качественного исследования удовлетворенности потребителей в российской банковской сфере.

ПРЕДЛАГАЕМЫЙ ПОДХОД

Преодолеть существующие проблемы автоматизации процессов мониторинга и оценки удовлетворенности потребителей возможно путем использования альтернативных источников данных потребительских мнений и методов их сбора. Сегодня таким источником данных является Интернет, который содержит самую разнородную информацию, касающуюся всех сфер человеческой деятельности. Не являются исключением потребительские мнения о товарах и услугах, которые содержатся в статьях, отзывах, сообщениях и комментариях интернет-пользователей.

Регулярный сбор и анализ данных из интернета дает возможность предприятиям вести непрерывный мониторинг удовлетворенности потребителей. Однако здесь возникает проблема, связанная с обработкой неструктурированных данных (тексты, видеофайлы, музыка, рисунки), размещаемых в интернете. Отзывы, представленные в виде текстов на естественном языке, в чистом виде не поддаются статистическим методам анализа. Также существует проблема так называемых «больших данных» (Big Data), которая затрагивает вопросы, связанные с тенденцией роста и накопления разнородных данных в мире и способами их обработки, анализа и хранения.

Разрешение проблемы автоматической обработки огромных массивов текстовых данных возможно при помощи инструментов интеллектуального анализа данных (ИАД, Data Mining) и анализа текста (Text Mining). Данные инструменты основываются на технологиях искусственного интеллекта и используются в предлагаемом подходе к исследованию удовлетворенности потребителей.

Предлагаемый в данной работе подход к исследованию удовлетворенности потребителей можно представить в виде процесса, состоящего из четырех основных этапов (рис. 1). Модели и методы реализуемые в подходе подробно изложены в [1, 2].

ЭТАП СБОРА ПОТРЕБИТЕЛЬСКИХ ОТЗЫВОВ

На первом этапе производится сбор потребительских отзывов из сети Интернет. В качестве источников потребительских отзывов предлагается использование специализированных интернет-ресурсов. На сегодня существует большое количество сайтов, на которых пользователи могут оставлять свои отзывы о товарах и услугах. Одни специализируются на конкрет-

ных сферах услуг, другие могут охватывать широкий спектр товаров и услуг. Наиболее популярными примерами являются tophotels.ru (635 тыс. отзывов), kinopoisk.ru (467 тыс. отзывов). К зарубежным аналогам относятся yelp.com (53 млн. отзыва), tripadvisor.com (путешествия, 130 млн. отзывов). Подобные ресурсы продолжают набирать популярность.

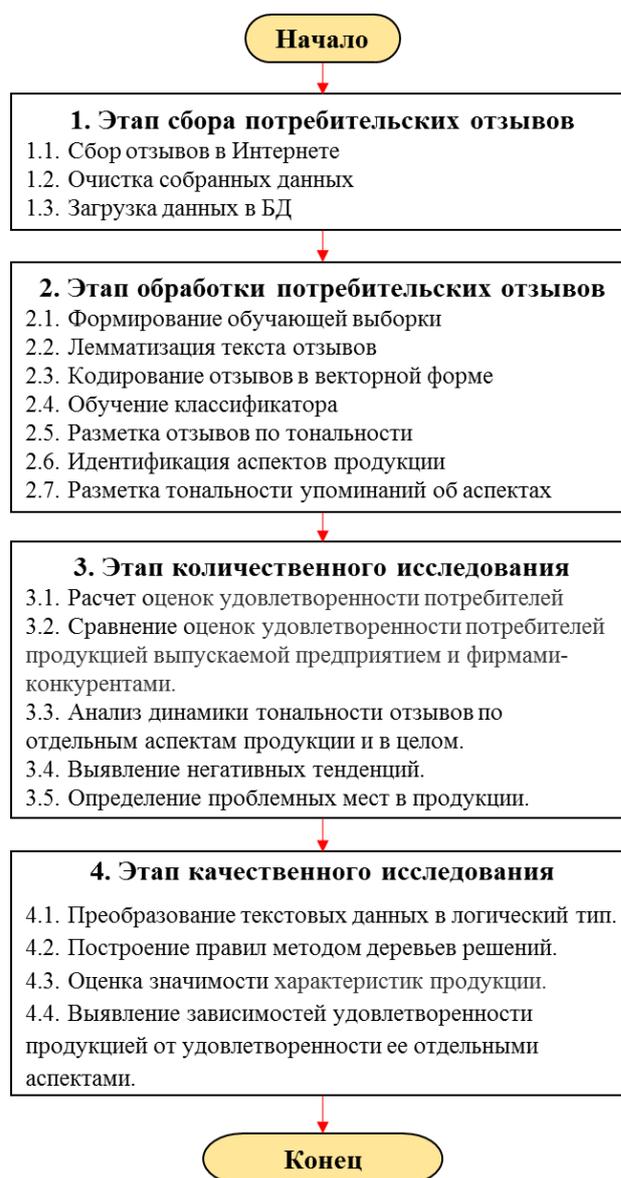


Рис. 1. Процесс исследования удовлетворенности потребителей

Задача сбора данных из интернета выглядит следующим образом. Дана интернет-страница S , содержащая множество неявных объектов (например, текст отзыва, наименование объекта отзыва, дата публикации отзыва). Необходимо определить отображение W , с помощью которого осуществлялось бы наполнение хранилища данных R объектами из S . Отображение W

должно выполняться и для других интернет-страниц S' подобных S . Под подобием в данном случае понимается использование однотипной структуры данных на множестве страниц одного интернет-ресурса.

В данной работе для сбора данных используется подход на основе анализа структуры HTML интернет страниц, содержащих потребительские отзывы. Для сбора данных в интернете используется специальная программа, называемая web-парсером, которая автоматически собирает данные из задаваемых иерархических блоков со страниц одного интернет-ресурса на множестве ссылок с одинаковой HTML структурой. Сбор потребительских отзывов проводится как в отношении продукции выпускаемой предприятием, так и в отношении продукции фирм-конкурентов.

ЭТАП ОБРАБОТКИ ПОТРЕБИТЕЛЬСКИХ ОТЗЫВОВ

На втором этапе производится обработка потребительских отзывов с применением инструментов анализа текста. Для оценки удовлетворенности потребителей продукцией в работе используется sentiment-анализ потребительских отзывов, представленных в виде текстовых документов на естественном языке [3, 4]. Под тональностью в данном контексте понимается эмоциональная оценка мнения потребителя по отношению к объекту тональности, т. е. к товару или услуге, а также их отдельных аспектов.

Задачу оценки удовлетворенности потребителей можно представить в виде задачи классификации потребительских отзывов по тональности. В данной работе рассматривается бинарная классификация отзывов на позитивные и негативные классы отзывов.

Для решения данной задачи в работе используются методы машинного обучения с учителем – наивный байесовский классификатор и метод опорных векторов. Они достаточно просты в программной реализации, не требуют построения лингвистических анализаторов или словарей тональности. Оценка тональности текста может выражаться количественно. Для применения данных методов необходима обучающая выборка. Для описания признаков пространства используется векторное представление текста отзывов с помощью модели «мешка слов» (bag-of-words). В качестве признаков рассматриваются бинарные векторы – наличие или отсутствие слова в тексте отзыва и частотные векторы – число вхождений слова в текст отзыва. Перед кодированием текста может приме-

няться процедура лемматизации, приводящая все слова к начальной форме.

Анализ тональности отзывов позволяет оценить общую удовлетворенность потребителей продукцией или предприятием. Для того чтобы определить удовлетворенность потребителей отдельными аспектами продукции необходимо проведение аспектного sentiment-анализа. Под аспектом продукции понимаются характеристики, атрибуты, качества, свойства, характеризующие продукцию, например, аккумулятор телефона или время доставки и т. п.

Аспектный sentiment-анализ потребительских отзывов является более сложной задачей и состоит из двух этапов: идентификации аспектов и определения тональности высказывания о них. Для решения задачи аспектного sentiment-анализа разработан простой и эффективный алгоритм.

Для идентификации аспектов используется частотный словарь, с помощью которого производится сравнение полученных частот со словарными частотами. Существительные с наибольшими частотными отклонения являются претендентами на включение в аспектные группы. Разбивку множества существительных на аспектные группы можно производить экспертным путем.

В качестве объектов аспектного sentiment-анализа рассматриваются предложения, в которых упоминаются извлеченные слова-аспекты. Такие предложения размечаются по тональности с помощью классификатора тональности. Предложения с ярко выраженной позитивной или негативной тональностью и содержащие упоминания аспектов продукции помечаются как мнения потребителей и используются при последующем анализе удовлетворенности потребителей как отдельными аспектами продукции, так и в целом продукцией.

ЭТАП КОЛИЧЕСТВЕННОГО ИССЛЕДОВАНИЯ УДОВЛЕТВОРЕННОСТИ ПОТРЕБИТЕЛЕЙ

На третьем этапе проводится количественное исследование удовлетворенности потребителей на основе sentiment-анализа потребительских отзывов. Оно проводится для измерения степени удовлетворенности потребителей, позволяет оценить текущую ситуацию и выявить тенденции в удовлетворенности.

На данном этапе производится оценка удовлетворенности потребителей. Оценка удовлетворенности потребителей продукцией в целом представляет собой долю позитивных отзывов

относительно всех отзывов по данной продукции и рассчитывается по формуле:

$$УП = \frac{N^{pos}}{N^{pos} + N^{neg}} \cdot 100\% , \quad (1)$$

где N^{pos} – число позитивных отзывов о продукции; N^{neg} – число негативных отзывов о продукции.

Оценка удовлетворенности потребителей j -й аспектной группой представляет собой долю отзывов, в которых имеются позитивные предложения с упоминанием слов j -й аспектной группы относительно числа отзывов, содержащих позитивные и негативные предложения с упоминанием слов j -й аспектной группы и рассчитывается по формуле:

$$УП_j = \frac{n_j^{pos}}{n_j^{pos} + n_j^{neg}} \cdot 100\% , \quad (2)$$

где n_j^{pos} – число отзывов содержащих позитивные предложения с упоминанием слов j -й аспектной группы; n_j^{neg} – число отзывов содержащих негативные предложения с упоминанием слов j -й аспектной группы; j – номер аспектной группы продукции.

Далее проводится сравнительный анализ полученных оценок удовлетворенности потребителей продукцией, выпускаемой самим предприятием и оценками удовлетворенности продукцией фирм-конкурентов. Проводится анализ динамики тональности как по отдельным аспектам продукции, так и продукцией в целом. На данном этапе выявляются негативные тенденции и определяются проблемные места в качестве продукции. Результатом количественного анализа является информация о текущей удовлетворенности потребителей в сравнении с продукцией конкурентов, а также оценки удовлетворенности по отдельным аспектам продукции.

ЭТАП КАЧЕСТВЕННОГО ИССЛЕДОВАНИЯ УДОВЛЕТВОРЕННОСТИ ПОТРЕБИТЕЛЕЙ

Качественные исследования удовлетворенности потребителей проводятся для выявления отдельных аспектов продукции и их значимости для потребителя с точки зрения удовлетворения его требованиям, способствуют пониманию отдельных восприятий и раскрытию идей и вопросов.

Задачей качественного анализа удовлетворенности потребителей является установление взаимосвязи между потребительской удовле-

творенностью продукцией и отдельными ее аспектами, а также оценка значимости продуктовых аспектов для потребителя. Для проведения качественного исследования удовлетворенности необходимо преобразование результатов сентимент-анализа к логическому типу данных с их последующим анализом с помощью построения деревьев решений. Метод деревьев решений прост в понимании и интерпретации результатов, а также способен объяснять ситуации при помощи булевой логик и визуализации полученного дерева.

Для построения дерева решений в качестве зависимых переменных выступает тональность потребительского отзыва – позитив или негатив. В качестве независимых переменных рассматривается присутствие или отсутствие в отзыве ярко выраженных тональных предложений с упоминанием об отдельных аспектах продукции.

Пример модели дерева решений удовлетворенности потребителя представлен на рис. 2. Узлами дерева являются аспектные переменные, т. е. возможное наличие в отзыве тональных предложений (позитивных или негативных), содержащих слова из какой-либо аспектной группы. Ребрами дерева являются значения аспектных переменных. Каждая аспектная переменная принимает два значения – 0 или 1, т. е. тональные предложения с упоминанием слов из аспектной группы могут либо присутствовать в отзыве, либо отсутствовать. Листьями дерева являются классы тональности отзыва, т.е. каждая ветвь приводит либо к положительному отзыву, либо к негативному отзыву, что соответствует удовлетворенности или неудовлетворенности потребителя выпускаемой продукцией.

Каждая ветвь построенного дерева решения представляет собой логическую конструкцию и называются правилами. Каждое правило можно представить в виде булевских функций в ДНФ или на естественном языке. Так, дерево, представленное на рис. 2, состоит из 4 правил:

$$\left\{ \begin{array}{l} 1) \overline{Нег.асп.гр.2} \rightarrow Поз. отзыв; \\ 2) Нег.асп.гр.2 \cap \overline{Поз.асп.гр.3} \rightarrow Нег. отзыв; \\ 3) Нег.асп.гр.2 \cap Поз.асп.гр.3 \cap \overline{Поз.асп.гр.1} \rightarrow Нег. отзыв; \\ 4) Нег.асп.гр.2 \cap Поз.асп.гр.3 \cap Поз.асп.гр.1 \rightarrow Поз. отзыв. \end{array} \right. \quad (3)$$

Значимость аспектной группы показывает, насколько сильно тональность отзывов определяется позитивной или негативной тональностью данной аспектной группы. Значимость рассчитывается после построения дерева клас-

сификационных правил. Пусть число переменных равно $2m$. Тогда формула для расчета значимости¹ q -й аспектной переменной будет иметь вид:

$$Z_q = \frac{\sum_{j=1}^{k_q} \left(E_{q,j} - \sum_{i=1}^{n_{q,j}} E_{q,j,i} \cdot \frac{N_{q,j,i}}{N_{q,j}} \right)}{\sum_{l=1}^{2m} \sum_{j=1}^{k_l} \left(E_{l,j} - \sum_{i=1}^{n_{l,j}} E_{l,j,i} \cdot \frac{N_{l,j,i}}{N_{l,j}} \right)} \cdot 100\% , (4)$$

где Z_q – значимость q -й аспектной переменной; k_l – количество узлов, которые были разбиты по аспектной переменной l ; $E_{l,j}$ – энтропия родительского узла, разбитого по аспектной переменной l ; $E_{l,j,i}$ – дочерний узел для j -го узла, который был разбит по аспектной переменной l ; $N_{l,j}$, $N_{l,j,i}$ – количество примеров в соответствующих узлах; $n_{l,j}$ – количество дочерних узлов для j -го родительского узла. Значимость аспектных переменных измеряется в процентах. Сумма значимостей по всем аспектным переменным составляет 100%.

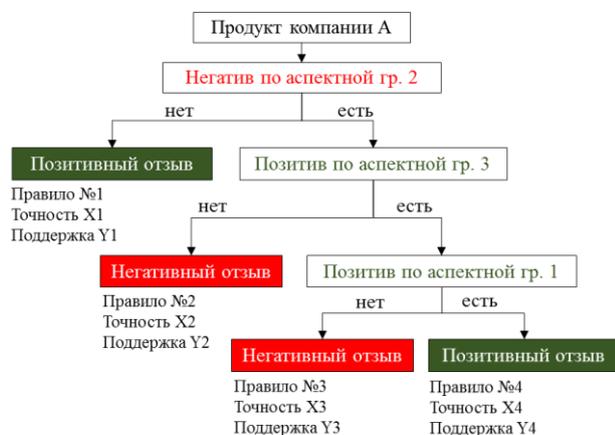


Рис. 2. Пример модели дерева решений удовлетворенности потребителей

ЭКСПЕРИМЕНТ

Для проведения вычислительного эксперимента было собрано 1153 потребительских отзыва о российских банках с сайта banki.ru. Выборка состоит из 304 положительных отзывов и 849 негативных отзывов на русском языке. Для проведения sentiment-анализа и аспектного sentiment-анализа была разработана программа

на языке C#, в которой были реализованы методы машинного обучения.

Нужно отметить, что в данных наблюдается перекоп в сторону негативных отзывов. Это объясняется тем, что клиенты чаще оставляют отзыв в случае недовольства банком. В случае отсутствия недовольства клиенты не имеют мотива оставлять отзыв. Исследования, проведенные в отношении гостинично-курортной деятельности, напротив, показывают перекоп в сторону позитивных отзывов [5].

Для оценки, обобщающей способности методов машинного обучения, использовался скользящий контроль или кросс-проверка. Оценка точности классификации рассчитывается как доля верно классифицированных позитивных и негативных отзывов относительно их общего количества. Результаты представлены в табл. 1.

Таблица 1

Оценки точности классификаторов

Методы машинного обучения	Векторное отображение отзыва	Точность
NB*	Бинарное	86,5%
NB	Частотное	86,8%
SVM**	Бинарное	87,7%
SVM	Частотное	85,0%
NB (тэги 'не_' 'ни_')	Частотное	88,0%

* NB – наивный байесовский классификатор

** SVM – метод опорных векторов

Оценки точности классификации методов машинного обучения получились сопоставимыми. Для проведения аспектного sentiment-анализа был выбран наивный байесовский классификатор с частотным векторным отображением текста отзывов. Также в текстах отзывов были помечены отрицательные частицы 'не_' и 'ни_' для учета их присутствия перед определенными словами и оборачивания смысла слова на противоположное значение. Данная техника позволила повысить точность классификации до 88 %.

На множестве отзывов были извлечены различные аспекты банковской деятельности. Извлеченные аспекты, сгруппированные по аспектным группам, представлены в табл. 2. Полученные аспекты использовались для проведения аспектного sentiment-анализа.

Было проведено количественное и качественное исследование удовлетворенности потребителей в отношении четырех российских банков: ВТБ24 (120 отзывов), Альфа-банка (131 отзыв), Сбербанк (232 отзыва), Банка Русский Стандарт (86 отзывов). Результаты количест-

¹ Официальный сайт компании BaseGroup Labs. Deductor. Руководство по алгоритмам. Версия 5.2.0. http://www.basegroup.ru/download/guides/guide_algorithm_5.2.0.

венного исследования удовлетворенности потребителей представлены в табл. 3. Среди рассматриваемых банков наиболее удовлетворены клиенты банка ВТБ24 (35 % положительных отзывов). Далее следует Альфа-банк и Сбербанк (26,7 % и 19,8 % соответственно). На последнем месте находится Банк Русский Стандарт (14 %). Полученные оценки удовлетворенности соответствуют оценкам «народного рейтинга» банков.

Таблица 2

Извлеченные аспекты	
Аспектная группа	Слова-аспекты
Персонал	Девушка, администратор, заведующая, инспектор, компетентность, консультант, менеджер, некомпетентность, обслуживание, оператор, операционист, операционистка, охранник, персонал, работник, руководство, сотрудник, сотрудница, специалист.
Кредит	Анкета, автокредит, выдача, долг, задолженность, заемщик, заявка, заявление, ипотека, кредит, кредитование, лимит, одобрение, отказ, погашение, просрочка, рассмотрение.
Депозит	Вкладчик, вклад, депозит.
Карта	Блокировка, заблокировать, банкомат, выпуск, выпускать, держатель, карта, карточка, кредитка, перевыпуск.
Расчетно-кассовое обслуж.	Касса, банкинг, кассир, квитанция, комиссия, перевод, платеж, поручение, списание, списать, счет, терминал, транзакция.

Таблица 3

Оценки удовлетворенности потребителей

	ВТБ24	Альфа-банк	Сбербанк	Банк Русский Стандарт
Персонал	54%	48%	42%	38%
Кредиты	40%	25%	29%	28%
Депозиты	100%	100%	69%	50%
Карты	41%	48%	39%	29%
РКО	42%	40%	31%	29%
Общая удовл.	35,0%	26,7%	19,8%	14,0%
Народный рейтинг*	38,6	37,1	32,7	31,6

* рейтинг банков, составленный на основе оценок клиентов, выставляемых на интернет-ресурсе banki.ru

Аспектный sentiment-анализ позволяет оценить удовлетворенность потребителей отдельными аспектами банковского обслуживания. Например, несмотря на первое место по общей удовлетворенности, ВТБ24 имеет более низкую удовлетворенность по карточным продуктам, чем у Альфа-банка. У Альфа-банка в свою очередь имеются проблемы с кредитными продуктами. Так, клиенты жалуются на необоснован-

ные случаи просрочки и возникновение задолженности по кредитам.

На основе результатов sentiment-анализа отзывов и аспектного sentiment-анализа с помощью аналитической программы Deductor Academic было построено дерево решений (рис. 3).



Рис. 3. Дерево решений

Решающие правила, формирующие полученное дерево и имеющие достоверность более 75 %, представлены в табл. 4. Правило 1 означает, что отсутствие в отзыве позитивных упоминаний о персонале и картах приводит к негативному отзыву с достоверностью 92,9 %. Данное правило содержится в 60,2 % всех отзывов. Позитивные упоминания о персонале и отсутствие негативных упоминаний о персонале и РКО приводят к позитивному отзыву с достоверностью 79,4 % (правило 2).

Таблица 4

Извлеченные правила

№	Правила	П*	Д**
1	$\text{Персонал}^+ \cap \text{Карты}^+ = \text{Нег. отзыв}$	60,2%	92,9%
2	$\text{Персонал}^+ \cap \text{Персонал}^- \cap \text{РКО}^- = \text{Поз. отзыв}$	20,2%	79,4%
3	$\text{Персонал}^+ \cap \text{Персонал}^- = \text{Нег. отзыв}$	7,4%	85,9%
4	$\text{Персонал}^+ \cap \text{Карты}^+ \cap \text{Карты}^- = \text{Нег. отзыв}$	4,1%	87,2%

* П – поддержка показывает долю отзывов от исходной выборки, содержащих данное правило.

** Д – достоверность показывает какая доля отзывов, содержащих правило, имеют данную тональность.

В табл. 5 представлены рассчитанные значения оценок значимостей аспектных групп для клиентов банков. Наибольшее влияние на удовлетворенность потребителей оказывают негативные и позитивные упоминания о персонале банков. Далее по значимости аспектных групп следуют карточные продукты и РКО. Остальные аспектные группы имеют оценки значимости близкие к нулевым.

Таблица 5
Оценки значимостей

№	Тональные упоминания по аспектным группам	Оценка значимости, %
1	Позитив по персоналу	54,7
2	Негатив по персоналу	19,5
3	Позитив по картам	14,9
4	Негатив по картам	6,3
5	Негатив по РКО	4,6
6	Негатив по депозитам	0
7	Позитив по РКО	0
8	Позитив по кредитам	0
9	Негатив по кредитам	0
10	Позитив по депозитам	0

ЗАКЛЮЧЕНИЕ

Предложенный подход к исследованию удовлетворенности потребителей на основе анализа потребительских отзывов с использованием инструментов анализа текста позволяет в автоматизированном режиме проводить количественное и качественное исследование удовлетворенности потребителей. Данный подход позволяет значительно снизить трудоемкость исследования удовлетворенности потребителей, что делает его доступным для применения широким кругом предприятий.

На базе предложенного подхода было разработано программное обеспечение для проведения сентимент-анализа. Проведенный эксперимент показал эффективность подхода для решения реальных задач исследования удовлетворенности потребителей, удовлетворительную точность методов машинного обучения, логичность полученных результатов.

В рамках дальнейших исследований представляет интерес приложения извлекаемых правил и оценок значимостей для различных клиентских сегментов [6], а также автоматическое аннотирование текстовых документов и извлечение полезной и уникальной информации.

СПИСОК ЛИТЕРАТУРЫ

1. Юсупова Н. И., Богданова Д. Р., Бойко М. В. Алгоритмическое и программное обеспечение для анализа тональности текстовых сообщений с использованием машинного обучения // Вестник УГАТУ. 2012. Т. 16, № 6 (51). С. 91–99. [N. I. Yussupova, D. R. Bogdanova, and M. V. Boyko, "Algorithmic and software support for sentiment analysis of text messages using machine learning," (in Russian), *Vestnik UGATU*, vol. 16, no. 6 (51), pp. 91-99, 2012.]
2. Yussupova N. I., Bogdanova D. R., Boyko M. V. Applying of sentiment analysis for texts in russian based on machine learning approach // Proc. 2nd Int. Conf. on Advances in Information Mining and Management IMMM2012 (Venice, Italy, October 21-26, 2012). Venice: Xpert Publishing Services, 2012. P. 8–14. [N. I. Yussupova, D. R. Bogdanova, M. V. Boyko, "Applying of sentiment analysis for texts in russian based on machine learning approach," in *Proc. the Second International*

Conference on Advances in Information Mining and Management (IMMM2012), Venice, Italy, 2012, pp. 8-14.]

3. Pang B., Lee L. Opinion mining and sentiment analysis // *Foundations and Trends in Information Retrieval*. 2008. Vol. 2, no. 1–2. P. 1–135. [B. Pang, L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, Vol. 2, No. 1–2, pp. 1-135, 2008.]

4. Pang B., Lee L. Thumbs up? Sentiment classification using machine learning techniques // *Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP)*. Philadelphia, 2002. P. 79–86. [B. Pang, L. Lee, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, Philadelphia, 2002. pp. 79-86.]

5. Юсупова Н. И., Богданова Д. Р., Бойко М. В. Математическое обеспечение для поддержки принятия решений при управлении качеством продукции на основе анализа текстовой информации // *Современные проблемы науки и образования*. 2014. № 3. [Электронный ресурс]. URL: www.science-education.ru/117-13024 (дата обращения: 22.06.2014). [N. I. Yussupova, D. R. Bogdanova, and M. V. Boyko, "Mathematical support for decision making in quality management based on text mining," (in Russian), *Modern problems of science and education*, no. 3, 2014. Available: www.science-education.ru/117-13024]

6. Юсупова Н. И., Ахметова Ю. Ф., Богданова Д. Р. Классификация клиентов на основе нечеткой информации // Вестник УГАТУ. 2013. Т. 17, № 5 (58). С. 93–100. [N. I. Yussupova, Yu. F. Akhmetova, D. R. Bogdanova, "Client classification based on fuzzy information," *Vestnik UGATU*, vol. 17, no. 5 (58), pp. 93-100, 2014.]

ОБ АВТОРАХ

БОЙКО Максим Викторович, асп. каф. выч. математики и кибернетики. Дипл. мат.-экон. (УГАТУ, 2011).

METADATA

Title: Customer satisfaction surveys in the banking sector based on the analysis of text messages.

Authors: M. V. Boyko.

Affiliation:

Ufa State Aviation Technical University (UGATU), Russia.

Email: maxim.boyko87@gmail.com.

Language: Russian.

Source: *Vestnik UGATU* (scientific journal of Ufa State Aviation Technical University), vol. 18, no. 5 (66), pp. 135-141, 2014. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

Abstract: The paper proposes an approach to the study of consumer satisfaction. As a source of satisfaction data is used consumer product reviews on natural language published on the Internet. To measure the satisfaction is suggested using sentiment analysis of consumer reviews. For data processing and analysis used sentiment analysis and retrieval aspects of products using machine learning methods. This approach applied for quantitative and qualitative surveys of customer satisfaction of several Russian banks.

Key words: customer satisfaction; text mining; sentiment analysis; data mining.

About authors:

BOYKO, Maxim Victorovich, PhD Student, Chair of computational mathematics and cybernetics. Dipl. Mathematician-economist (USATU, 2011).