

МЕТОДИКА ПОИСКА ЗАКОНОМЕРНОСТЕЙ В РЕЗУЛЬТАТАХ АНКЕТИРОВАНИЯ

Себко Е.С.,

Московский государственный университет им. М.В. Ломоносова, г. Москва

Научный руководитель: к.ф.-м.н. Абрамов В.Г.,

Московский государственный университет им. М.В. Ломоносова, г. Москва

Введение. Анкетирование как метод исследования сегодня широко используется в различных областях деятельности, например, в демографии, социологии. Для проведения количественных исследований применяется массовое анкетирование, в котором могут принимать участие от нескольких сотен до нескольких тысяч респондентов.

Для проведения массового анкетирования исследователям необходимо подготовить вопросы анкеты. Часто требуется составить большое количество вопросов, отвечающих исследовательской задаче. Составление такого опросника – длительный, трудоемкий процесс, требующий объемной ручной работы экспертов. Таким образом, актуальной является задача автоматизированного добавления новых вопросов в анкету. При этом полученный расширенный опросник должен по-прежнему соответствовать тематике исследования и уточнять результаты опроса.

Также при проведении крупных исследований количество вопросов в анкете может быть настолько большим, что прохождение опроса становится для респондента утомительным и это искажает результаты анкетирования. На практике был бы полезен инструмент, который позволит подбирать вопросы для каждого участника в режиме онлайн на основе его предыдущих ответов. В итоговый список вопросов, на которые придется ответить респонденту, не войдут те вопросы, ответы на которые можно предугадать на основе предыдущих ответов.

Цель работы. Предложить методику выявления закономерностей в данных анкетирования для автоматизации процесса формирования массовых и персональных опросников в заданной предметной области.

Базовые положения исследования. Наборы вопросов и ответов, результаты их анализа существенно разнятся в зависимости от предметной области, в которой проводится исследование. В работе в качестве исходных данных рассматриваются материалы мониторинга РМЭЗ. Данные представляют собой вопросы анкеты и соответствующие варианты ответов респондентов. В работе предлагается построить разбиение вопросов на группы (разделы) так, чтобы в одной группе оказались близкие семантически вопросы. Затем рассматриваются ответы на вопросы анкеты в соответствии с разделами вопросов. Внутри каждого раздела предлагается выделить кластеры (категории) ответов. В каждую категорию должны попасть похожие наборы ответов. Полученные категории внутри каждого раздела фиксируются. Для каждого респондента формируется набор категорий, в которые попадают его ответы. Затем выделяются статистические взаимосвязи между наборами категорий. В результате построено разбиение вопросов на разделы, ответов – на категории внутри разделов. Также для каждого респондента получен набор категорий, в которые попали его ответы, и выявлены некоторые закономерности среди наборов категорий. Предлагаются алгоритмы формирования индивидуальной анкеты и расширения массового опросника на основе полученных статистических данных.

Промежуточные результаты. В рамках работы исследованы статистические подходы к анализу данных анкетирования, изучены методы семантического анализа коротких текстов, классификации текстовых данных, кластеризации дискретных данных. Предлагаемая методика опробована на реальных данных РМЭЗ, проанализирована результативность выбранных подходов для решения поставленной задачи.

Основной результат. Разработана программная система, реализующая предложенную методику обработки данных анкетирования и формирования массового и индивидуального опросников.

Список литературы:

1. Pang B., Lee L., Vaithyanathan S. Thumbs up? Sentiment Classification Using Machine Learning Techniques. - Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP'02), 2002.
2. Olson D. L., Dursun D. Advanced Data Mining Techniques. - Berlin: Springer-Verlag Berlin Heidelberg, 2008
3. Ершов К., Романова Т. Анализ и классификация алгоритмов кластеризации. - М.: МГТУ им. Н.Э. Баумана, 6 с.
4. Скворцова, Е. В. Измерение и шкалирование в маркетинговых исследованиях / Е.В. Скворцова // Маркетинг: идеи и технологии. – 2013
5. Дюран Б., Одедд П. Кластерный анализ / Пер. с англ. М.: УРСС, 1977.
6. Рубаков С.В. Современные методы анализа данных // Наука. Инновации. Образование. 2008
7. «Российский мониторинг экономического положения и здоровья населения НИУ-ВШЭ (RLMS-HSE)», проводимый Национальным исследовательским университетом "Высшая школа экономики" и ООО «Демоскоп» при участии Центра народонаселения Университета Северной Каролины в Чапел Хилле и Института социологии Федерального научно-исследовательского социологического центра РАН. (Сайты обследования RLMS-HSE: <http://www.cpc.unc.edu/projects/rlms> и <http://www.hse.ru/rlms>)