

Реконструкция образа автора как когнитивный процесс

В.В. Рыков¹, В.А. Колчина²

¹Московский физико-технический институт (государственный университет)

²ООО Техносерв консалтинг

Знание принципиально отличается от информации. При выполнении информационных процессов информационная система (ИС) находит заданную ей по запросу информацию. При выполнении когнитивных процессов (КП) генерируется новое знание. С когнитивной информатикой (КИ) связаны большие надежды на то, что при помощи ее моделей можно будет находить решения различных проблем, недоступных человеку в силу самых разных причин. Начиная от тривиальных, но необходимых, – например, обсчет алгоритмов эвакуации из большого строящегося здания и до нахождения решений проблем глобальной экологии.

Нахождение решений различных когнитивных проблем осуществляется обычно в рамках определенных когнитивных моделей (КМ). В данной статье рассматривается модель автоматического нахождения или генерации образа автора (ОА) из некоторого корпуса авторских публикаций. Образ автора – риторический термин. Эта проблема является во многих отношениях новой и актуальной. Представить себе автора текста интересно многим читателям. Но сейчас человек давно стал частью многих глобальных ИС (например, соцсетей) и поэтому также важно знать хотя бы некоторые качества их участников [1].

В настоящее время уже созданы КМ автоматического нахождения или генерации ОА их доступных авторских текстов. В данной работе рассматриваются конкретно две когнитивные модели фирмы IBM – Персональный взгляд (Personality Insights) и Анализатор тона (Tone Analyzer). Обе модели или сервиса генерируют новую и неочевидную информацию об авторе. Однако между этими моделями есть различия т. к. они предназначены для получения несколько разных образов автора. Модель Персональный взгляд строит статичный образ автора (ОА), а Анализатор тона строит его конкретный тон или динамическую характеристику, реализованных при написании данного корпуса исходных текстов и даже выявить эмоциональное состояние автора [1,2].

Итак, поставленную задачу можно декомпозировать на две более простые подзадачи:

- Обработка текстов на естественном языке и анализ полученного текста, вычленение потенциальных характеристик личности автора
- Построение модели личности автора по полученным характеристикам.

Если первая подзадача решается математическими или иными подходами строго формальными, то решение второй подзадачи лежит в большей степени в плоскости психологии. Одной из важных моделей, используемых для таких целей, является так называемая “Большая пятерка”. “Большая пятерка” это диспозиционная модель личности человека. В соответствии с названием, модель предполагает, что личность включает в себя пять общих и относительно независимых черт (диспозиций):

- Экстраверсия (общительность)
- Доброжелательность (дружелюбие, способность прийти к согласию)
- Добросовестность (сознательность)
- Нейротизм (противоположный полюс – эмоциональная стабильность)
- Открытость опыту (интеллект).

Модель выводится эмпирически, с использованием данных самоотчетов (вопросники, шкалы прилагательных), экспертных оценок (внешних наблюдателей за поведением) и данных поведенческих, получаемых в рамках проведенных исследований. Основным статистическим средством извлечения модели является факторный анализ. Таким образом, в эмпирических исследованиях 5 черт чаще всего предстают в виде сравнительно автономных факторов. Данная модель активно используется для принятия решений о характеристиках личности в выбранных для анализа сервисах [1,2].

Модель Персональный взгляд (Personality Insights) сервис IBM, который способен извлекать и анализировать персональные характеристики личности, помогая больше узнать о человеке и

настроить с ним более персонализированное взаимодействие. Сервис делит выводимые характеристики на 3 группы: пятерка личностных характеристик, ценности и нужды. Для работы Personality Insights требуется минимум 3500 слов любого текста, написанного данным человеком. Модель Анализатор тона использует лингвистический анализ для поиска трех основных характеристик: эмоции, социальные тенденции и стиль написания. Все эмоции делятся на: злость, страх, радость, грусть и раздражение. Социальные тенденции включают элементы из Большой пятерки персональных характеристик используемые некоторыми психологами. Стили написания делятся на: *уверенный, аналитический и осторожный*. Сервис может анализировать электронные письма и любые другие тексты. Результат работы сервиса можно использовать для определения того, несет ли письмо нужную заказчику эмоциональную окраску или другую характеристику. Сервис использует для анализа модель Большая пятерка.

Вот что было получено для портрета папы Римского:

Экспрессивный и уверенный в себе человек. Он легко заводит знакомства и уверенно чувствует себя в компании. Данный человек не очень консервативен и имеет собственную точку зрения по всем вопросам. Он стремится к самосовершенствованию и пытается проявлять себя профессионалом в нужный момент.

Можно ли проанализировать полученный результат? С этой точки зрения, результат работы сервиса можно охарактеризовать как предельно верный, но достаточно расплывчатый. Можно предположить, что любой известный в каких-либо кругах человек обладает уверенностью в себе, общителен и безусловно разбирается в выбранной области [1,2].

Модель Анализатор тона также предоставляет несомненный практический интерес. Например, диалог разговор агента сервис центра с клиентом (Customer service chat). Запустить данный пример можно на официальной странице сервиса в разделе демоверсия. В данном диалоге клиент связывается с агентом сервис центра по вопросу удаления аккаунта, зарегистрированного на его почтовый адрес, но не принадлежащего ему. Агент сервис центра долго не может решить проблему, предлагая порой недопустимые решения. Данный диалог носит сердито-извинительный характер. В результате, видно, что главные эмоции в этом разговоре гнев и страх. Стил общения преимущественно аналитический, строгий, формальный [1,2].

Представленные когнитивные модели генерации образа автора представляют собой несомненный научную и практическую значимость. В докладе будут представлены другие интересные примеры и предполагаются дальнейшие исследования в этой области. Были исследованы тексты речей Джона Кеннеди, Маска и других политических деятелей. Полученные результаты не только хорошо согласуются со здравым смыслом, но и дают новые знания об особенностях образов этих людей. Более подробно об этом можно прочитать здесь - <https://drive.google.com/open?id=0B-AvMZiuNAQQd1J3UHJDRWp5R28>

Проведенное исследование позволило прийти к следующим выводам:

- Реконструкция образа автора текста есть важная и продуктивная когнитивная модель.
- Эта модель имеет широкую и важную практическую значимость.
- Необходимы дальнейшие исследования в этой области.

Литература

1. Рыков В.В. Обработка нечисловой информации. Управление знаниями. – М.: МФТИ, 2008.
2. Родина С.В. Автоматическое формирование образа автора. Курсовой проект. - М.: МФТИ, 2016.