

УДК 004.852

CROSS-SOURCE MULTI-VIEW LEARNING FOR PERSONALITY PROFILING

Ян Ци (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – PhD, философские науки, Фарсеев А.И.

(Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Введение.

Личностные качества являются основной движущей силой наших решений и ежедневно влияют на наши жизненные пути. Выводы о таких чертах личности, как тип личности Майерса-Бриггса, а также понимание зависимости между личностными качествами и поведением пользователя в различных социальных сетях имеют решающее значение для современных исследований и промышленных приложений. Появление разнообразных и универсальных каналов социальных сетей позволило автоматически и эффективно проводить анализ личности пользователя на основе мультимодальных данных из множества источников данных. Тем не менее, относительно мало было проведено исследований по анализу личности на основе данных мультиплексных социальных сетей, и степень влияния различных данных социальных сетей на эффективность машинного обучения не была всесторонне оценена. Данное исследование является одной из первых попыток восполнить этот важный пробел в исследованиях. В частности, в этой работе мы делаем вывод об источниках данных индикаторов типа личности Майерса-Бриггса, применяя новую систему слияния нескольких видов под названием "MMPIS" и сравнивая результаты работы различных социальных сетей. Наши экспериментальные результаты демонстрируют способность MMPIS выполнять личностный анализ на основе многопросмотренных данных, эффективно используя значительные различия в данных из различных социальных мультимедийных источников. Мы также обнаружили, что выбор методов машинного обучения имеет решающее значение при выборе источников данных из социальных сетей, и что люди склонны раскрывать множество аспектов своей личности в различных каналах социальных мультимедиа.

В этой работе мы ищем ответы на три вопроса для исследования:

1. Могут ли черты личности пользователей быть в большом масштабе выведены из данных в их учетных записях в социальных сетях?
2. Помогают ли многомодальные данные из учетных записей пользователей в социальных сетях повысить эффективность выводов о личностных особенностях пользователей?
3. Чем отличаются предполагаемые личности в социальных сетях?

Мы предлагаем новую систему профилирования личности, которая называется "MMPIS", для решения вышеописанных вопросов исследования. Мы внедряем методы сбора и представления данных из нескольких социальных сетей (Facebook, Twitter и PersonalityCafe). Мы разделили процесс профилирования личности MBTI на четыре задачи бинарной классификации. Для помощи в дальнейших научных исследованиях мы выпустили наш многоисточниковый мультимодальный набор данных с маркировкой личности.

Чтобы ответить на вопрос нашего исследования, мы сравнили производительность MMPIS, обученных с использованием одного источника и комбинации модальностей. Для оценки производительности MMPIS-фреймворка, набор данных был равномерно разделен на тренировочный и тестовый набор в соотношении 85:15. Для поддержания одинакового распределения личностных этикеток в реальном мире мы сохраняем одинаковое распределение образцов в тренировочном и тестовом наборах. Для полной оценки MMPIS мы выбрали следующие исходные параметры:

1. Отдельное обучение классификаторов с использованием единой модели.

2. Early Fusion. Ранний классификатор слияния применяется к особенностям текста и изображений путем первого каскадирования особенностей текста и изображений.
3. Early Fusion (PCA200). Сначала мы соединяем текстовые элементы и изображения в новый обучающий набор и применяем PCA к ним, чтобы уменьшить их размер до 200.

Выводы.

Несмотря на то, что MMPIS превосходит базовые показатели в отдельных категориях личностей, достигнутая производительность МВТІ прогнозирования личностей пока не может быть использована в реальных приложениях. Это выдвигает прогнозирование личности на первый план как сложную проблему. Мы должны признать, что все еще существует множество способов, которые мы не использовали, например, аватар пользователя, биография, период деятельности и т. д. С ростом количества модальностей возникает необходимость в лучшей стратегии слияния, которая может более эффективно интегрировать мультимодальные данные. В этой работе мы только сравнили разницу в предсказании личности из разных источников. Но поведение одного и того же пользователя на разных платформах социальных сетей остается неясным.

В этой работе мы представили одно из первых исследований по сравнению производительности модели машинного обучения при профилировании личностных характеристик между различными источниками (Facebook, Twitter и PersonalityCafe) с предложенной нами моделью под названием "MMPIS". Экспериментальные результаты по трем источникам показывают, что MMPIS способен повысить производительность при прогнозировании личностных характеристик пользователей за счет агрегирования многомодальных данных, генерируемых пользователем из его аккаунта в социальных сетях. Мы выпустили мультимодальный набор данных для профилирования личности, чтобы помочь исследовательскому сообществу в дальнейших исследованиях в похожих областях, таких как профилирование пользователей и система рекомендатель с улучшенными личностными характеристиками.

В наших будущих работах мы планируем глубже изучить влияние различных источников данных на производительность моделей слияния и различное поведение пользователей, действующих в межъязыковой и межрегиональной среде.

Ян Ци (автор)

Подпись

Фарсеев А.И. (научный руководитель)

Подпись