

УДК 004.932.2

МЕТОД СОПОСТАВЛЕНИЯ АККАУНТОВ ПОЛЬЗОВАТЕЛЯ В РАЗЛИЧНЫХ СОЦИАЛЬНЫХ СЕТЯХ НА ОСНОВЕ ПУБЛИКУЕМЫХ ФОТОГРАФИЙ

Т. Р. Сохин

Научный руководитель к.т.н., старший научный сотрудник Н. А. Бутаков
ФГАОУ ВО «Санкт-Петербургский национальный исследовательский университет
ИТМО», Санкт-Петербург

Введение. В настоящее время обычной практикой является ежедневное использование нескольких социальных сетей с использованием разного контента, разного стиля общения. Если мы намерены понимать поведение человека через его цифровое отображение, сочетание данных из нескольких социальных сетей может лучше описать этого пользователя.

Для этого мы предлагаем новый подход к сопоставлению профилей пользователей, основанный на общедоступных изображениях пользователей и идентификации лиц, и проводим экспериментальное исследование его эффективности. Предложенный подход не зависит от упомянутых выше изменений в содержании, стиле общения и другом поведении, которое является специфическим для определенной социальной сети.

Цель работы. Оценка эффективности предложенного метода сопоставления профилей социальных сетей на пример социальных сетей ВКонтакте и Instagram.

Базовые положения исследования. Предложенный подход основывается на методах компьютерного зрения. При этом алгоритм сопоставления состоит из следующих этапов:

- обработка всех изображений профиля пользователя социальной сети с целью извлечения изображений лиц и получения их векторного представления в евклидовом пространстве;
- фильтрация векторных представлений путем оценки качества исходного изображения (в количестве пикселей), удаления векторных представлений лиц детей;
- определение лиц владельца профиля путем кластеризации векторных представлений всех лиц, формирование одного определяющего вектора для профиля;
- сопоставление профилей путем сравнения определяющих векторов.

Промежуточные результаты. Были проведены экспериментальные исследования предложенного метода на наборе пар профилей социальных сетей ВКонтакте и Instagram, принадлежащих одним владельцам. Точность и полнота составили 0.994 и 0.769 соответственно на наборе 3021 пары профилей. Дополнительные исследования проводились в условиях искусственно повышенной неопределенности с уменьшением доли доступных фотографий с профиля пользователя и уменьшением количества существующих пар пользователей в исследуемой выборке.

Результаты показывают сохранение точности на значении 0.802 на доле доступных фотографий в 10% со снижением полноты до 0.182. При значении доли парных профилей в 10% точность снижается до 0.61, полнота сохраняется на уровне 0.75-0.76.

Основной результат. В ходе выполнения работы был разработан и исследован метод сопоставления профилей пользователей в разных социальных сетях. Даная работа имеет ценность в исследовании возможности использования контента, публикуемого пользователем социальных сетей, для определения единого представления этого пользователя и последующего использования этого представления.

Метод обладает высокой точностью при различных условиях доступности информации и состоянии социального графа пользователей, что позволяет эффективно объединять профили социальных сетей для дальнейшего изучения поведения пользователей на большем наборе данных.

(подпись)

(подпись)

/ _____ /
(фамилия, инициалы)

/ _____ /
(фамилия, инициалы)