

УДК 004.8

## АНАЛИЗ ИНТЕРЕСОВ И КУЛЬТУРНЫХ ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЕЙ И СООБЩЕСТВ В ЦИФРОВЫХ СОЦИАЛЬНЫХ СИСТЕМАХ

Щепина Е.К.<sup>1</sup>, Егорова Е.В.<sup>2</sup>, Федотов П.А.<sup>3</sup>

Научный руководитель – Суриков А.Г.<sup>4</sup>, к.т.н., научный сотрудник НЦКР

<sup>1</sup>Университет ИТМО, Факультет цифровых трансформаций

<sup>2</sup>Университет ИТМО, Факультет цифровых трансформаций

<sup>3</sup>Университет ИТМО, Факультет цифровых трансформаций

<sup>4</sup>Университет ИТМО, Национальный центр когнитивных разработок

В настоящей работе представлена мультимодальная модель анализа гетерогенных цифровых следов пользователя, осуществляющая представление пользователей и групп в едином пространстве интересов и культурных предпочтений. Данная модель позволяет комплексно анализировать интересы и предпочтения пользователей, определять степень их близости друг к другу по данному критерию, выделять кластеры, наблюдать за развитием интересов отдельных пользователей и за динамикой развития групп и сообществ.

### Введение.

Пользователи оставляют большое количество цифровых следов в цифровых социальных системах, к примеру, каждую минуту отправляется 41 миллион сообщений (данные на конец 2020 года). Значительная часть цифровых следов может быть представлена в топологии взаимосвязей: подписки на группы, музыкальные подписки и пр. В данной работе мы предпринимаем попытку объединить выводы по гетерогенным цифровым следам пользователей, собрав результаты их анализа в едином мультимодальном пространстве интересов и культурных предпочтений. Цель данной работы заключается в получении комплексного вывода об интересах пользователей, представляющего как самостоятельную ценность (например, для решения задач социальных исследований), так и в качестве промежуточного предиктора (эмбединга) для улучшения качества вывода других моделей (например, моделей психотипирования пользователей по их цифровым следам).

**Основная часть.** Музыкальные интересы, подписки на группы и пр. размеченные данные социальных систем могут дать представление о пользователях и их интересах в унимодальных структурах, таких, например, как взвешенные графы, вершинами которых являются точки интереса (музыкальные исполнители, сообщества и пр.), а ребрами – их связанность подписками пользователей. Кластеризация этих структур позволяет выделить интерпретируемые области интересов и культурных предпочтений в рамках отдельных модальностей. Далее мы рассматриваем выделенные кластеры различных модальностей как основу для построения мультимодальной структуры – графа, вершинами которого являются выделенные области интересов и культурных предпочтений в рамках отдельных модальностей, ребра – степень их связанности подписками пользователей. Таким образом, мы объединяем гетерогенные данные в единую гомогенную структуру, в которой мы также можем выделить кластеры интересов, на этот раз уже мультимодально, и сделать финальный вывод об интересах и культурных предпочтениях пользователя. Модель обучается первично на ограниченном наборе данных и, в дальнейшем, дообучается итеративно на данных новых пользователей.

**Выводы.** Предложенная модель позволяет объединить гетерогенные цифровые следы пользователей социальных систем в едином гомогенном пространстве и сделать комплексный вывод по их интересам и культурным предпочтениям. Данные результаты будут основой для дальнейшего исследования взаимосвязи культурных интересов и психоэмоциональных черт пользователей.