

Исследование психометрического распределения (кластеризации) пользователей социальной сети Я исследую пользователя vk И его друзей.

Цель исследования: На основании лексической информации о пользователе, сделать вывод о его психо-типе, сгруппировать друзей пользователя по этому параметру. Отнести самого пользователя к той или иной группе. Рекомендовать пользователю друзей.

Обоснование исследования: Ключевым подходом к решению задачи, определяющим именно психометрическую типизацию пользователей, является выбор источников информации. Я опираюсь на то, что пользователь пишет о себе(осознаваемый манифест) во-первых и на скрытый лексический контекст(неосознаваемый манифест) в таких действиях пользователя как: Выбор сообществ и Выбор музыки. Такую активность(со скрытым манифестом) я описываю текстовым формальным описанием Групп(сообществ) и жанров музыки.

Метод исследования: Группирую пользователей по эмбедингам (Word2Vec) на основе полученной по вышеописанному принципу текстовой информации.

Примечание: Я группирую пользователей по сообществам, рассматривая сообщество не как единицу вектора некоего эмбединга пользователя, а строю для каждого сообщества свой эмбединг, таким образом два схожих сообщества будут иметь схожий эмбединг. И уже по этим эмбедингам строю эмбединг пользователя, таким образом схожие сообщества не рассматриваются как две разных единицы измерения вектора, а имеют конгруэнтное влияние.

Выводы. С практической точки зрения я оцениваю результат как удовлетворительный. Гипотеза о наличии психологической метрики, обуславливающей стиль текстов пользователя его характером требует более тщательного исследования.

Я искал корреляцию между текстами, описывающими сообщества пользователя и текстом написанным собственноручно пользователем в качестве своего описания профиля.

На основе такой корреляции я хотел сгруппировать пользователей и сделать вывод о их стабильном внутри каждой группы психотипе.

Классификация пользователей по текстам описаний сообществ и описаний личных профилей оказалась неконгруэнтной.

Кластеризация же пользователей по эмбедингам сообществ показала достоверный результат. В частном случае группа пользователей из 100+ человек была разделена на две подгруппы с устойчивой корреляцией внутри группы.