

цифровой
прорыв 

сезон: ИИ

КЕЙС

Предсказание интенсивности взаимодействия между
друзьями в социальной сети ВКонтакте



Министерство
экономического развития
Российской Федерации

РОССИЯ –
СТРАНА
ВОЗМОЖНОСТЕЙ

Кейсодержатель



01 Сфера деятельности

Информационные технологии

02

Краткое описание кейса

Участникам хакатона предлагается предсказать интенсивность взаимодействия между пользователями, основываясь на информации об интенсивности других связей и пользовательской информации.



Сайт организации

<https://vk.company/ru/>

Постановка задачи

Посмотрите на социальный граф наших пользователей! Пара пользователей соединена ребром в графе, если они добавили друг друга в друзья.

Эго-графами называют графы социальных связей конкретного пользователя и его друзей, каждое ребро в которых означает связь (например взаимное добавление в друзья). Мы рассказывали об эго-графах в статье на Хабре.

В этой задаче мы предлагаем вам не огромный социальный графа всей социальной сети ВКонтакте, а несколько эго-графов отдельных пользователей. Все они не пересекаются. В каждом графе эго-вершина (вершина, связанная со всеми остальными) имеет ID 0. У каждого пользователя будет не более 300 смежных вершин в эго-графе. На каждом ребре будет указан момент появления связи между пользователями t — целое число дней, прошедшее с возникновения дружбы между каждой парой пользователей (-1, если дружбы нет). На части вершин будут также указаны коэффициенты активности взаимодействия. Чем они больше, тем взаимодействие между пользователями в этой паре интенсивнее.

Кроме того, вы получите вспомогательную информацию о пользователях: возраст, идентификатор города, пол, идентификаторы указанной в профиле школы и университета.

Предскажите интенсивность взаимодействия x между пользователями по информации об интенсивности других связей и информации о пользователях. Для оценки вашего решения мы будем использовать метрику RMSE

Проблематика

Улучшаем качество социального взаимодействия между пользователями ВКонтакте – почти всей аудиторией рунета.

Решение

Текстовый файл с предсказанными целевыми значениями функции интенсивности взаимодействия между пользователями.
Исходный код решения, оформленный как последовательность действий в colab или jupyter notebook, с возможной ручной обработкой данных

Стек технологий, обязательных к использованию

Необходимые данные, дополнения, пояснения, уточнения

01

Python, Colab, Jupyter notebook,

02

Ознакомьтесь с тем, как устроены эго-
графы, в нашей статье на хабре

<https://habr.com/ru/companies/vk/articles/552162>

Оценка

- Для оценки решений применяется метод экспертных оценок и автоматизированные средства оценивания.
- Жюри состоит из технических экспертов и представителей кейсодержателя.
- На основании описанных ниже характеристик, жюри выставляет оценки 0-3 балла.
- Итоговая оценка определяется как сумма баллов экспертов: технического и/или представителя кейсодержателя, как значение, выданное автоматизированными средствами оценивания, либо как итоговый балл жюри, умноженный на оценку автоматизированной системы.

Технический эксперт оценивает решение по следующим критериям:

01

Запускаемость кода

02

Обоснованность
выбранного метода
(описание подходов к
решению, их обоснование
и релевантность задаче)

03

Точность работы
алгоритма
(возможность оценить
формальной метрикой с
обоснованием выбора)

04

Адаптивность/
Масштабируемость

Автоматизированные средства оценивания точности
работы предложенных участниками алгоритмов
(решений) выставляют оценку в диапазоне 0-1, где 1
равно 100% точности работы решения.

Итоговая оценка определяется как итоговый балл жюри,
умноженный на оценку автоматизированной системы.

цифровой
прорыв



сезон: III



Министерство
экономического развития
Российской Федерации

РОССИЯ –
СТРАНА
ВОЗМОЖНОСТЕЙ

