

УДК 004.896

ОПРЕДЕЛЕНИЕ ВРЕДНОСНЫХ ПОЛЬЗОВАТЕЛЕЙ ЧЕРЕЗ АНАЛИЗ ЭГО-ГРАФОВ С ИСПОЛЬЗОВАНИЕМ GNN МОДЕЛЕЙ.

Мурашов И.Г. (Университет ИТМО), Терентьева В.В. (Университет ИТМО)
Научный руководитель – кандидат технических наук, доцент Муравьев С.Б.
(Университет ИТМО)

Введение. Социальные сети являются одними из наиболее популярных медиа-ресурсов, жизнь миллионов людей каждый день связана с социальными сетями. Из-за этого множество людей используют свои аккаунты для мошеннической или опасной активности в сети, поэтому задача определения спамеров и других вредоносных пользователей очень актуальна для социальных сетей. Такие аккаунты портят пользовательский опыт и вносят шум в данные, что пагубно сказывается на рекомендательных системах.

Задача определения вредоносных пользователей является задачей классификации. Существуют несколько подходов к определению вредоносных аккаунтов, основанных на анализе поведения пользователей, однако большинство из таких подходов мошенники научились обманывать, используя паттерны поведения, делающие их неотличимыми от обычных пользователей.

В настоящее время активно развивается направление по решению различных ML-задач на графах с использованием графовых нейронных сетей. Детекторы аномалий на основе графовых нейронных сетей обладают лучшей производительностью для обнаружения аномалий в социальных структурах, т.к. учитывают информацию о связях пользователя в социальном графе, то есть его друзьях.

Основная часть.

Данные для решения задачи были предоставлены компанией ВКонтакте в формате csv. Табличный файл содержит в себе данные о нескольких эго-графах, причем на каждой строчке файла указаны данные об отдельной дружеской связи, в формате строчки:

ego_id - идентификатор графа

u - идентификатор первого друга

v - идентификатор второго друга

t — целое число дней, прошедшее с возникновения дружбы между каждой парой пользователей

x1, x2, x3 — разные показатели активности дружбы

Графовые нейронные сети — это особый тип нейронных сетей, которые способны работать со структурой данных графа. GNN используются для прогнозирования узлов, ребер и свойств данных на основе набора графов. Входной граф передается в GNN как набор ребер. Структура входного графа преобразуется во эмбединг графа, что позволяет нам сохранять информацию об узлах, ребрах и глобальном контексте. Графовые нейронные сети используют механизм распространения информации для обработки графов. Набор модулей взаимодействует друг с другом в соответствии с графовыми связями и связан с узлами графа. В процессе обучения модули обновляют свои состояния и обмениваются информацией до достижения устойчивого равновесия. Выходные данные GNN вычисляются на основе состояния каждого модуля на узлах графа. [1]

В качестве базового решения был взято решение с выделением признаков и дальнейшей обработкой датасета логистической регрессией.

После этого будут протестированы решения классификационных задач на графах на основе доступных научных работ.[2] Также мы предложим свое решение на основе GCN, чтобы улучшить результаты относительно предыдущих доступных.

Выводы. В рамках работы будет проведено сравнение различных GNN классификаторов на графах, и будет предложен новый, их превосходящий, улучшающий

точность определения спамеров в социальной сети.

Список использованных источников:

1. Hwan Kim, Byung Suk Lee, Won-Yong Shin, Senior Member, IEEE, and Sungsu Lim, Member, IEEE: Graph Anomaly Detection with Graph Neural Networks: Current Status and Challenges. // IEEE Access – 2022
2. Zhiyuan Liu, Chunjie Cao, Jingzhang Sun: Mul-GAD: a semi-supervised graph anomaly detection framework via aggregating multi-view information. // Sci. Adv. – 2022
3. Yangyang Li; Yipeng Ji; Shaoning Li; Shulong He; Yinhao Cao; Yifeng Liu; Hong Liu; Xiong Li; Jun Shi; Yangchao Yang: Relevance-Aware Anomalous Users Detection in Social Network via Graph Neural Network // IEEE Access - 2021

Автор _____ Мурашов И.Г.
Автор _____ Терентьева В.В.
Научный руководитель _____ Муравьев С.Б.