

# СОВРЕМЕННЫЕ МЕТОДЫ И ИНСТРУМЕНТЫ РАБОТЫ С ГРАФАМИ. ПРИМЕНИМОСТЬ РАБОТЫ С ГРАФАМИ В БАНКОВСКИХ ПРОЦЕССАХ И ПОДРАЗДЕЛЕНИЯХ БЕЗОПАСНОСТИ БАНКОВ

**Носков Алексей Александрович**

Аспирант, АНО ВО «Российский новый университет»

alexeinoskov91@gmail.com

## MODERN METHODS AND TOOLS FOR WORKING WITH GRAPHS. THE APPLICABILITY OF WORKING WITH GRAPHS IN BANKING PROCESSES AND SECURITY DEPARTMENTS OF BANKS

**A. Noskov**

*Summary.* Graph data representation is growing and becoming increasingly important because it helps solve complex problems involving large amounts of information. In this representation, data is structured in a way that reflects the relationships between different elements. This can be useful for social network analysis, web search, route optimization, and many other applications. One of the main reasons graph data is important is its ability to handle uncertainty and incompleteness of information. Unlike traditional relational databases, where data is stored in tables with a rigidly defined structure, graphs can contain information about different types of relationships between objects. This makes them ideal for working with data that is not always fully defined or has a complex structure. In addition, graph data representation also helps solve problems involving the analysis of large amounts of information. For example, when working with social networks, graphs can be used to identify groups of users with similar interests or behavior. Similarly, when searching the web, graphs can help determine the most relevant results based on the relationships between web pages. Thus, graph representation of data plays a key role in solving complex problems related to processing large amounts of information. It provides a flexible tool for data analysis and helps to find new patterns and relationships between different elements. Speaking about the applicability of graphs to bank data, we can say that the existing connections and relationships of bank clients can also be represented as a graph of client relationships. This approach to solving problems allows you to open a new perspective on solving classic problems of security departments in banks.

*Keywords:* graphs, big data, machine learning, data science, data science, deep learning.

*Аннотация.* Графовое представление данных развивается, и становится всё более важным, поскольку оно позволяет решать сложные задачи, связанные с большими объёмами информации. В этом представлении данные структурируются таким образом, чтобы отражать связи между различными элементами. Это может быть полезно для анализа социальных сетей, поиска информации в интернете, оптимизации маршрутов и многих других приложений. Одной из главных причин важности графового представления данных является его способность обрабатывать неопределённость и неполноту информации. В отличие от традиционных реляционных баз данных, где данные хранятся в таблицах с жёстко определённой структурой, графы могут содержать информацию о различных типах связей между объектами. Это делает их идеальными для работы с данными, которые не всегда полностью определены или имеют сложную структуру. Кроме того, графовое представление данных также позволяет решать задачи, связанные с анализом больших объёмов информации. Например, при работе с социальными сетями графы могут использоваться для определения групп пользователей со схожими интересами или поведением. Аналогично, при поиске информации в интернете графы могут помочь определить наиболее релевантные результаты на основе связей между веб-страницами. Таким образом, графовое представление данных играет ключевую роль в решении сложных задач, связанных с обработкой больших объёмов информации. Оно предоставляет гибкий инструмент для анализа данных и помогает находить новые закономерности и связи между различными элементами. Говоря о применимости графов к данным банка, можно сказать, что имеющиеся связи и отношения клиентов банка можно также представить в виде графа клиентских связей. Такой подход к решению задач позволяет открыть новый взгляд на решение классических проблематик подразделений безопасности в банках.

*Ключевые слова:* графы, большие данные, машинное обучение, дата саенс, глубокое обучение.

Помимо сравнительно стандартных по подходу к решению задач в области классического машинного и глубокого обучения — классификация, регрессия (на табличных данных), кластеризация, NLP (natural language processing), CV (computer vision), сейчас также набирают популярность алгоритмы и модели, связанные

с задачами на графах, такие, например, как GNN (Graph neural network).

GNN (Graph neural network) — это класс искусственных нейронных сетей для обработки данных, которые могут быть представлены в виде графов.

Графы использовались для обозначения данных из различных областей, от социальных наук, лингвистики до химии, биологии и физики.

Поскольку данные из реальных прикладных задач имеют очень разнообразные формы, от матриц и тензоров до последовательностей и временных рядов, возникает естественный вопрос: почему мы пытаемся представить данные в виде графов? Есть два основных мотива. Во-первых, графы обеспечивают универсальное представление данных. Данные из многих систем в различных областях могут быть явно обозначены в виде графов, таких как социальные сети, транспортные сети, сети взаимодействия белок-белок, графы знаний и сети мозга. Между тем, многие другие типы данных могут быть преобразованы в форму графов [1]. Во-вторых, значительное количество реальных проблем можно рассматривать как вычислительные задачи на графах и, соответственно, эффективно их решать с помощью графового представления.

Следующие классы задач, имеющих огромное прикладное значение, можно обобщить как проблему классификации узлов [2]:

- Обнаружение аномальных узлов (например, спамеров или террористов).
- Предложение новых связей или контактов людям на основе поиска других людей со схожими интересами, демографией или опытом.
- Разработка систем рекомендаций, предлагающие объекты (музыку, фильмы, занятия) на основе интересов других людей с пересекающимися характеристиками.
- Разработка систем ответов на вопросы, которые направляют вопросы тем, у кого больше всего опыта в ответе на данный вопрос.
- Разработка рекламных систем, которые показывают рекламу тем людям, которые, скорее всего, будут заинтересованы и восприимчивы к рекламе по определенной теме.
- Идентификация генов, соответствующих заболеваниям, а также предложение лекарств пациентам.
- Социологическое исследование сообществ, например, степень, в которой сообщества формируются вокруг определенных интересов или принадлежности.
- Исследование того, как идеи и «мемы» распространяются в сообществах с течением времени.

Рекомендации друзей в социальных сетях можно также решать как задачу прогнозирования связи (link prediction), к такой же проблематике, по существу могут быть отнесены рекомендации, предсказание побочных эффектов полипрагмазии, идентификация взаимодействия лекарственного средства с целью (drug-target

interaction, при этом целью может быть конкретный белок, фермент, рецептор, нуклеиновая кислота или другие биомолекулы, участвующие в болезненных процессах или нормальных физиологических функциях) и заполнение графа знаний [3, 4]. Прогнозирование белкового интерфейса можно рассматривать как задачу классификации графов [5].

Узлы на графах по своей сути связаны, что предполагает, что узлы не являются независимыми. Однако традиционные методы машинного обучения часто предполагают, что данные независимы и одинаково распределены. Таким образом, они не подходят для непосредственного решения вычислительных задач на графах. Существует два основных направления разработки решений. Можно использовать классификацию узлов в качестве наглядного примера для обсуждения этих двух направлений. Одним из направлений является создание нового механизма, специфичного для графов. Задача классификации, разработанная для графов известна как коллективная классификация [6]. В отличие от традиционной классификации для узла, коллективная классификация рассматривает не только сопоставление между его признаками и его меткой, но также и сопоставление его окрестности. Другое направление — сгладить граф, создав набор признаков для обозначения его узлов, где можно применить традиционные методы классификации. Это направление может использовать преимущества традиционных методов машинного обучения; таким образом, он становится все более популярным и доминирующим. Ключом к успеху этого направления является то, как построить набор признаков для узлов (или представлений узлов). Было доказано, что глубокое обучение эффективно в обучении представлениям (representation learning), которое значительно продвинуло различные области, такие как компьютерное зрение, распознавание речи и обработка естественного языка. Таким образом, сочетание глубокого обучения с графами открывает беспрецедентные возможности. Однако глубокое обучение на графах также сталкивается с огромными проблемами. Во-первых, традиционное глубокое обучение было разработано для регулярных структурированных данных, таких как изображения и последовательности, в то время как графы нерегулярны, когда узлы в графе неупорядочены и могут иметь различные окрестности. Во-вторых, структурная информация для обычных данных проста; в то время как для графов это сложно, особенно учитывая, что существуют различные типы сложных графов, а узлы и ребра могут ассоциироваться с богатой дополнительной информацией; таким образом, традиционного глубокого обучения недостаточно для сбора такой богатой информации. Была создана новая область исследований — глубокое обучение на графах, охватывающее беспрецедентные возможности и огромные проблемы.

Чтобы лучше использовать преимущества современных моделей машинного обучения для этих вычислительных задач, эффективное представление графов играет ключевую роль. Существует два основных способа извлечения признаков (features) для представления графов, включая разработку признаков (feature engineering) и обучение представлениям (representation learning). Разработка признаков (feature engineering), как правило, осуществляется специалистом по данным вручную, что отнимает много времени и часто не является оптимальным для определенных последующих задач. В то время как обучение представлениям (representation learning) заключается в автоматическом изучении признаков (features), что требует минимальных человеческих усилий и адаптируется к поставленным нижестоящим задачам. Таким образом, обучение представлениям (representation learning) на графах было широко изучено. Область обучения представлениям (representation learning) на графах было значительно развито за последние десятилетия, и ее можно условно разделить на три поколения, включая традиционные графовые эмбединги, современное графовые эмбединги и глубокое обучение на графах. Как первое поколение обучения представлению графа, традиционные графовые эмбединги были исследованы в контексте классических методов уменьшения размерности на графах, таких как IsoMap, LLE и eigenmap.

Метод Word2vec предназначен для изучения представлений различных объектов, которые встречаются вместе, например, слов из большого корпуса текста, в виде вектора с определенной размерностью. При обучении таких векторов учитывается в контексте каких других ближайших объектов встречается данный. Сгенерированные представления слов расширили многие задачи обработки естественного языка. Успешные расширения word2vec для области графов положили начало второму поколению обучения представлениям на графах, то есть современным графовым эмбедингам.

Учитывая огромный успех методов глубокого обучения в обучении представлениям (representation learning) в областях изображений и текста, были предприняты усилия по их обобщению на графы, которые открыли новую главу обучения представлению графов, то есть глубокое обучение на графах. Все больше и больше свидетельств демонстрируют, что третье поколение обучения представлению графов, особенно графовых нейронных сетей (GNN), значительно облегчило вычислительные задачи на графах, включая задачи, ориентированные как на узлы, так и на графы. Революционные достижения, принесенные GNN, также внесли огромный вклад в глубину и широту внедрения обучения графовому представлению в реальных приложениях. Для классических областей применения обучения представлению графов, таких как рекомендательные системы и анализ социаль-

ных сетей, GNN обеспечивают современную производительность и выводят их на новые рубежи. Тем временем постоянно появляются новые области применения GNN, такие как комбинационная оптимизация, физика и здравоохранение. Такое широкое применение GNN обеспечивает разнообразный вклад и точки зрения из разных дисциплин и делает эту область исследований действительно междисциплинарной.

Поскольку графовые нейронные сети применяются для графовых данных, то, например, в задаче классификации GNN должна классифицировать узлы графа: сами узлы должны представлять собой метку некоторого класса. Поэтому данные, которые представляют собой графовую структуру (как, например, связи в социальных сетях либо любых иных сетях взаимодействия различных объектов), могут быть использованы для обучения GNN.

В более общем плане «геометрического глубокого обучения» некоторые существующие архитектуры нейронных сетей можно интерпретировать как GNN, работающие на соответствующим образом определенных графах. Сверточные нейронные сети в контексте компьютерного зрения можно рассматривать как GNN, применяемую к графам, структурированным как сетки пикселей. Преобразователи в контексте обработки естественного языка можно рассматривать как GNN, применяемые к полным графам, узлами которых являются слова в предложении.

Ключевым элементом дизайна GNN является использование попарной передачи сообщений, так что узлы графа итеративно обновляют свои представления, обмениваясь информацией со своими соседями. С момента их создания было предложено несколько различных архитектур GNN, которые реализуют различные виды передачи сообщений, начиная с рекурсивных или сверточных конструктивных подходов. По состоянию на 2022 год, возможно ли определить архитектуры GNN, «выходящие за рамки» передачи сообщений, или может ли каждая GNN быть построена на передаче сообщений по соответствующим образом определенным графам, является открытым исследовательским вопросом.

Говоря о применимости графов в автоматизации функций банка можно сказать, что имеющиеся связи и отношения клиентов банка можно также представить в виде графа клиентских связей. Так, например, в виде ребра графа можно представить транзакционную связь между клиентами банка — вершинами графа.

В свою работу с окружением клиентов (соседями вершины) — предсказание связи (link prediction), анализ окружения и множество других задач получают новое переосмысление в контексте работы подразделений

безопасности по оценке деловой репутации, расследований мошеннических и иных кейсов и многих других прикладных задач служб безопасности банков. Так, например, с помощью решения задачи предсказания связи (link prediction) можно детектировать потенциальные мошеннические операции — осуществлять мэчнинг предсказанных возможных связей и фактически совершаемой. Также с помощью решения задачи предсказания связи (link prediction) можно устанавливать связь с другими фигурантами расследуемых дел, например, от дропов к фактическим организаторам и бенефициарам преступных схем. С помощью анализа окружения

можно предсказывать возможные мошеннические клиентские профили, или клиентские профили, созданные на несуществующих физических лиц, так называемые мертвые души.

В заключение стоит отметить, что тема работы с графами в современном мире очень актуальна, так как многие классические структуры данных могут быть представлены в виде графа, в том числе данные банков. А с помощью графового представления можно эффективно решать самые распространенные и сложные задачи подразделений безопасности банков.

---

#### ЛИТЕРАТУРА

1. Representing Big Data as Networks: New Methods and Insights / Xu, Jian, 2017, arXiv preprint arXiv:1712.09648.
2. Node classification in social networks / Bhagat, Smriti, Cormode, Graham, and Muthukrishnan, S., 2011, Pages 115–148 of: Social network data analytics. Springer.
3. The link-prediction problem for social networks / Liben-Nowell, David, and Kleinberg, Jon, 2007., Journal of the American society for information science and technology, 58(7), 1019–1031.
4. Novel drug-target interactions via link prediction and network embedding / E. Amiri Souiri, R. Laddach, S.N. Karagiannis, L.G. Papageorgiou & S. Tsoka, 04 April 2022, BMC Bioinformatics (23, Article number: 121).
5. Protein interface prediction using graph convolutional networks / Alex M. Fout, Jonathon Byrd, Basir Shariat, A. Ben-Hur, 13 November 2017, NIPS, 2-2.
6. Collective classification in network data / Sen, Prithviraj, Namata, Galileo, Bilgic, Mustafa, Getoor, Lise, Galligher, Brian, and Eliassi-Rad, Tina, 2008, AI magazine, 29(3), 93–93.

---

© Носков Алексей Александрович (alexeynoskov91@gmail.com)  
Журнал «Современная наука: актуальные проблемы теории и практики»