
Сравнительный анализ графовых и гиперграфовых нейронных сетей и их применение в задачах Retrieval-Augmented Generation

A Preprint

Аристеев О. А.
МГУ имени М. В. Ломоносова
Москва, Россия

Прокофьев П. А.
МГУ имени М. В. Ломоносова
Москва, Россия

Abstract

Графовые и гиперграфовые нейронные сети (GNN и HGNN) становятся ключевыми инструментами для анализа данных со сложной структурой. В то время как GNN зарекомендовали себя как фундаментальный метод в обработке графов, HGNN позволяют моделировать более сложные n -арные отношения, выходя за рамки бинарных связей. Настоящее исследование имеет три направления: (1) сравнительный анализ возможностей GNN и HGNN, (2) включение HGNN в рамки Geometric Deep Learning как обобщённой парадигмы обучения на структурированных данных, (3) применение гиперграфов в построении Retrieval-Augmented Generation (RAG) для работы с графами знаний. В работе систематизируются существующие методы, обсуждаются их сильные и слабые стороны, а также выявляются перспективные направления, связанные с интеграцией HGNN и LLM через гиперграфовые представления знаний.

Keywords графовые нейронные сети · гиперграфовые нейронные сети · Geometric Deep Learning · Retrieval-Augmented Generation · LLM

1 Введение

Современные приложения искусственного интеллекта требуют методов, способных обрабатывать сложные структуры данных: от социальных сетей и биологических систем до баз знаний для больших языковых моделей. Классические графовые нейронные сети (GNN) показали высокую эффективность, однако они ограничены моделированием бинарных связей. Гиперграфовые нейронные сети (HGNN) открывают новые возможности, позволяя работать с n -арными отношениями и более богатыми структурными зависимостями.

Одним из подходов является развитие методов HGNN и их сопоставление с GNN по качеству, вычислительной эффективности и устойчивости. Параллельно HGNN могут быть встроены в более широкую теоретическую рамку Geometric Deep Learning, которая рассматривает нейросетевые архитектуры как обобщение принципов симметрий и инвариантностей. Ещё одно направление связано с практическими приложениями: графы знаний, являясь важным источником структурированной информации, могут быть представлены в виде гиперграфов и использоваться в Retrieval-Augmented Generation (RAG). Такой подход позволяет моделировать не только простые пары сущностей, но и сложные факты с несколькими участниками. Сравнительное изучение этих направлений даёт основу для создания более выразительных и интерпретируемых моделей.

В рамках работы предполагается провести сравнительный анализ GNN и HGNN с опорой на современные исследования и бенчмарки. Кроме того, будет рассмотрено место HGNN в общей картине Geometric Deep Learning. Особое внимание планируется уделить применению гиперграфов для представления знаний и их интеграции в RAG, что может расширить возможности LLM за счёт более богатых структурных

представлений. Такой комплексный подход позволит выявить сильные и слабые стороны текущих методов и очертить направления для дальнейших исследований.

2 Постановка задачи

В данном разделе мы формализуем задачу генерации ответов, дополненной извлечением данных (Retrieval-Augmented Generation, RAG), и последовательно вводим математический аппарат для ее решения с использованием графовых и гиперграфовых структур знаний.

2.1 Стандартная модель RAG

Задача RAG заключается в генерации ответа y на заданный вопрос q с опорой на обширную базу знаний K , представленную в виде набора текстовых документов. Стандартный подход моделирует этот процесс вероятностно. Сначала из базы K извлекается подмножество релевантных документов d , а затем на их основе генерируется ответ. Формально, условная вероятность ответа y при заданном вопросе q определяется как:

$$P(y|q) = \sum_d P(y|q, d) \cdot P(d|q, K) \quad (1)$$

Здесь $P(d|q, K)$ представляет собой модель извлечения (retriever), которая оценивает релевантность документа d для вопроса q , а $P(y|q, d)$ — модель генерации (generator), которая создает ответ, основываясь как на вопросе, так и на извлеченном контексте. Несмотря на свою эффективность, данный подход рассматривает базу знаний как «плоский» набор текстов, игнорируя сложные взаимосвязи между сущностями, описанными в них.

2.2 Переход к графовому представлению знаний

Для преодоления ограничений стандартного RAG, был предложен подход, основанный на графовом представлении знаний (GraphRAG). В этой парадигме база знаний K структурируется в виде графа $G = (V, E)$, где V — это множество вершин, представляющих сущности, а E — множество ребер, представляющих бинарные отношения между ними.

Процесс генерации ответа в таком случае модифицируется для работы со структурированными фактами F , извлеченными из графа, а не с текстовыми фрагментами. Математическая постановка задачи принимает следующий вид:

$$P(y|q) = \sum_F P(y|q, F) \cdot P(F|q, G) \quad (2)$$

Такой подход позволяет моделировать явные связи между сущностями, однако он ограничен бинарной природой отношений, так как ребро в обычном графе может соединять только две вершины. Это затрудняет представление сложных фактов реального мира, в которых часто участвуют три и более сущности (n-арные отношения).

2.3 Формализм RAG на основе гиперграфов

Чтобы адекватно моделировать n-арные отношения ($n \geq 2$), мы вводим понятие гиперграфа знаний.

Определение 1 (Гиперграф Знаний). Гиперграф знаний — это пара $G_H = (V, E_H)$, где V — это множество вершин (сущностей), а E_H — множество гиперребер. Каждое гиперребро $e_H \in E_H$ представляет собой n-арное отношение и соединяет подмножество вершин $V_{e_H} \subseteq V$, где мощность $|V_{e_H}| \geq 2$.

В рамках нашей задачи, каждый факт F_n из базы знаний представляется в виде $F_n = (e_H, V_{e_H})$, где e_H — это семантическое описание отношения, а V_{e_H} — набор вовлеченных в него сущностей.

Процесс генерации ответа, дополненный извлечением из гиперграфа (HyperGraphRAG), можно разбить на три формализованных этапа:

1. Построение гиперграфа знаний. Из исходного корпуса документов K извлекается множество парных фактов F_n , которые формируют гиперграф G_H . Процесс извлечения с помощью языковой модели π можно описать как:

$$\{F_n\} \sim \pi(p_{ext}, K)$$

где p_{ext} — это промпт, направляющий модель на извлечение фактов.

2. Стратегия извлечения из гиперграфа. При поступлении запроса q выполняется поиск релевантного подграфа. Этот процесс включает два шага:

- Извлечение сущностей (R_V): Идентификация в G_H множества сущностей V_q^* , релевантных запросу q .

$$V_q^* = R_V(q, G_H)$$

- Извлечение гиперребер (R_H): Идентификация множества гиперребер E_{Hq}^* , семантически близких к запросу q .

$$E_{Hq}^* = R_H(q, G_H)$$

Совокупность извлеченных знаний K_H определяется как объединение фактов, связанных с V_q^* и E_{Hq}^* .

3. Генерация ответа. Финальный ответ y генерируется языковой моделью π с учетом запроса q и извлеченного из гиперграфа контекста K_H . Таким образом, итоговая постановка задачи для HyperGraphRAG формализуется следующим образом:

$$P(y|q) = \sum_{K_H} P(y|q, K_H) \cdot P(K_H|q, G_H) \quad (3)$$

Здесь $P(K_H|q, G_H)$ представляет собой нашу усовершенствованную модель извлечения, оперирующую на гиперграфе, а $P(y|q, K_H)$ — модель генерации, способную интерпретировать структурированные парные факты для синтеза точного и полного ответа. Такой система позволяет не просто находить релевантные фрагменты текста, но и понимать и использовать сложные, многоаспектные связи между сущностями для логического вывода и генерации.

3 Литературный обзор

Классические работы по GNN [??] заложили основу для современных моделей, использующих свёртки, агрегацию признаков и механизмы внимания. С другой стороны, HGNN [?] позволили расширить эти идеи на гиперграфовые структуры, что открыло путь к обработке парных связей. Современные обзоры [?] и бенчмарки [?] подтверждают рост интереса к HGNN, предлагая классификации методов и стандартизированные наборы данных для их оценки.

Geometric Deep Learning [?] представляет собой общую рамку, объединяющую CNN, GNN, трансформеры и их обобщения через симметрии и инвариантности. HGNN занимают в этой картине особое место, так как они естественно связаны с более высокопорядковыми структурами. Таким образом, HGNN рассматриваются как связующее звено между GNN и более общими методами топологического машинного обучения.

Графы знаний традиционно используются в NLP для представления фактов и связей между сущностями [?]. С появлением RAG [?] возникла новая парадигма — использование внешних структурированных знаний для улучшения генерации. Современные исследования расширяют эту идею: от GraphRAG [?] до HyperGraphRAG [?], где гиперграфовые структуры позволяют интегрировать сложные отношения. Дополнительные работы [??] показывают, что графы знаний могут стать динамическим компонентом памяти LLM-агентов. В этой перспективе HGNN могут стать естественным инструментом для представления и обработки гиперграфов знаний.

Список литературы