

УДК 004.032.26:519.177

ГРАФОВАЯ НЕЙРОННАЯ СЕТЬ НА ОСНОВЕ АДАПТИВНОГО УРАВНЕНИЯ РЕАКЦИИ-АДВЕКЦИИ-ДИФФУЗИИ ДЛЯ ЗАДАЧИ КЛАССИФИКАЦИИ МОЛЕКУЛЯРНЫХ ГРАФОВ



А.С. Довгий

Студент бакалавриата факультета прикладной математики и информатики БГУ
alexanderrdouhi@gmail.com



К.В. Василевский

Доцент кафедры компьютерных технологий и систем БГУ, кандидат физико-математических наук
VasilevskyK@bsu.by

А.С. Довгий

Обучается по специальности «Прикладная информатика» на кафедре компьютерных технологий и систем факультета прикладной математики и информатики белорусского государственного университета. Область научных интересов охватывает спектральную теорию графов, нейронные дифференциальные уравнения и анализ временных рядов.

К.В. Василевский

Окончил Белорусский государственный университет. Область научных интересов связана с дифференциальными уравнениями в частных производных, математическим моделированием, прогнозированием временных рядов методами глубокого обучения и компьютерной графикой.

Аннотация. Предложена архитектура GRADE (Graph Reaction-Advection-Diffusion Equation) – графовая нейронная сеть, моделирующая непрерывную динамику представлений вершин графа как решение полного уравнения реакции-адвекции-диффузии. Архитектура расширяет CDE-GRAND четырьмя основными нововведениями: (1) выпуклая комбинация и адвекции и диффузии вместо фиксированной диффузии с модулируемой конвекцией; (2) обучаемый коэффициент внимания в диффузионном члене; (3) гейтированный реакционный член; (4) обучаемое анизотропное RBF-ядро с обучаемым вектором весов. Экспериментальная валидация на задаче многозадачной бинарной классификации молекулярных графов *ogbg-moltox21* (12 ассеев токсичности, 7831 молекула) демонстрирует ROC-AUC 0,745 для GRADE, что превосходит GCN (0,734), GRAND (0,725) и CDE-GRAND (0,725).

Ключевые слова: графовые нейронные сети, уравнение реакции-адвекции-диффузии, молекулярные графы, выпуклая комбинация, обучаемое внимание.

Введение. Предсказание свойств молекул по их структуре является фундаментальной задачей вычислительной химии, фармацевтики и токсикологии. Высокопроизводительный виртуальный скрининг позволяет оценивать биологическую активность и токсичность миллионов соединений без проведения дорогостоящих лабораторных экспериментов [1]. Молекула естественным образом представляется графом $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, в котором вершины \mathcal{V} соответствуют атомам, а рёбра \mathcal{E} – химическим связям.

Графовые нейронные сети (ГНС) стали стандартным инструментом для решения задач на графах. Базовые архитектуры – GCN [2] и GAT [3] – основаны на парадигме передачи сообщений [4], при которой признаки вершин обновляются путём агрегирования информации от соседей. Математически этот процесс эквивалентен дискретной диффузии

на графе: на каждом слое представление вершины усредняется с представлениями её соседей, что хорошо работает при гомогенности – свойстве графа, при котором смежные вершины принадлежат одному классу.

Однако многие реальные графы по своей структуре гетерогенны: смежные вершины принадлежат различным классам [5]. Это наблюдается в сетях веб-страниц, биологических сетях, молекулярных структурах, сетях вопросов-ответов. На таких графах диффузионное усреднение приводит к потере дискриминативной информации – явлению, известному как пересглаживание (*over-smoothing*) [6]. С увеличением числа слоёв (шагов диффузии) все представления вершин сходятся к одному и тому же вектору, что делает классификацию невозможной.

Степень гетерогенности характеризуется глобальным коэффициентом гомогенности рёбер

$$h_{edge} = \frac{|\{(u, v) \in E: y_u = y_v\}|}{|E|}, \quad (1)$$

где y_i – метка вершины $i \in V$. Задача особенно сложна при $h_{edge} < 0,5$ – выраженной гетерогенности рёбер графа, когда большинство связей соединяют вершины разных классов. В случае, если имеет место значительный дисбаланс классов, предпочтительнее использовать скорректированный коэффициент глобальной гомогенности рёбер

$$h_{adj} = \frac{h_{edge} - \sum_{k=1}^C D_k^2 / (2|E|)^2}{1 - \sum_{k=1}^C D_k^2 / (2|E|)^2}, \quad (2)$$

где $D_k = \sum_{i: y_i=k} d_i$ – суммарная степень всех вершин класса k , d_i – степень вершины i . Для скорректированного коэффициента критерием выраженной гетерогенности рёбер графа служит условие $h_{adj} < 0$.

Перспективным направлением преодоления данных ограничений является представление ГНС как решения дифференциального уравнения на графе. В работе [7] показано, что выходы глубоких нейронных сетей могут быть представлены как решения обыкновенных дифференциальных уравнений (*Neural ODE*), что позволяет непрерывно параметризовать глубину сети и применять адаптивные численные методы. Архитектура GRAND [8] перенесла эту идею на графы, представив эволюцию признаков вершин как диффузионный процесс. CDE-GRAND [9] расширила модель конвекционным членом для учёта направленного переноса информации, опираясь на теоретическую связь между нейронными сетями и уравнением конвекции-диффузии [10, 11].

Однако уравнение конвекции-диффузии является лишь частным случаем более общего уравнения *реакции-адвекции-диффузии* [12], широко применяемого в математической физике, химической кинетике и биологии для описания процессов с одновременным переносом, рассеиванием и локальной трансформацией вещества. Реакционный член позволяет моделировать локальные нелинейные взаимодействия, которые невозможно выразить только через диффузию и конвекцию.

Введение реакционного члена мотивировано химической природой молекулярных графов. В отличие от социальных или информационных сетей, молекулы обладают локальными электронными эффектами, которые не сводятся к взаимодействию с соседями: формальный заряд, неподелённые электронные пары, резонансные структуры определяют реакционную способность атома *независимо* от непосредственного окружения. Диффузия и конвекция – принципиально межвершинные процессы – не способны выразить такие локальные трансформации. Реакционный член заполняет этот пробел, позволяя каждой вершине изменять своё представление без участия соседей.

Целью настоящей работы является разработка архитектуры GRADE (*Graph Reaction-Advection-Diffusion Equation*), реализующей полное уравнение реакции-адвекции-

диффузии на графе и развивающей адаптивный подход к регуляции этих процессов. GRADE вносит четыре нововведения: (1) выпуклую комбинацию диффузии и конвекции с коэффициентами $\alpha_i(t)$ и $1 - \alpha_i(t)$, позволяющую модели плавно перераспределять ресурсы между процессами; (2) обучаемые коэффициенты диффузии через механизм масштабированного скалярного внимания; (3) гейтированный реакционный член, моделирующий локальные нелинейные трансформации; (4) обучаемое анизотропное RBF-ядро в адаптивном коэффициенте. Идея в основе определения диффузионного и конвекционного членов наследуется из CDE-GRAND, что обеспечивает прямую сопоставимость результатов. Однако их аналитическое задание наряду с механизмом баланса вкладов являются нововведением. Валидация проводится на задаче предсказания токсичности молекулярных графов – практически значимой задаче для фармацевтического скрининга.

Задача классификации молекулярных графов. Рассматривается задача многомерной бинарной классификации молекулярных графов. Дан конечный набор молекул $\mathbb{G} = \{\mathcal{G}_s\}_{s=1}^S$, каждая из которых представлена графом $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ с $N_s = |\mathcal{V}_s|$ вершинами (атомами) и $|\mathcal{E}_s|$ рёбрами (связями). Каждой вершине $i \in \mathcal{V}_s$ сопоставлен вектор категориальных признаков $\mathbf{f}_i = (f_i^{(1)}, f_i^{(2)}, \dots, f_i^{(K)})$, где K – число типов признаков. Для каждой молекулы задан вектор меток $\mathbf{y}_s \in \{0, 1, \emptyset\}^M$, где M – число тестов (ассеев токсичности), а \emptyset обозначает отсутствующую метку.

Структура графа описывается матрицей смежности $\mathbf{A} \in \{0, 1\}^{N \times N}$ и матрицей степеней $\mathbf{D} = \text{diag}(d_1, \dots, d_N) \in \mathbb{R}^{N \times N}$, где $d_i = \text{deg}_{\mathcal{G}_s}(i)$ – степень вершины $i \in \mathcal{V}$ в графе \mathcal{G}_s . Графовый лапласиан определяется как $\mathbf{L} = \mathbf{D} - \mathbf{A}$. Окружение вершины i обозначается $\mathcal{N}_i = \{j \in \mathcal{V}_s : (i, j) \in \mathcal{E}_s\}$, $\mathcal{N}_i \subseteq \mathcal{V}_s$, причём $|\mathcal{N}_i| = d_i$.

Требуется построить отображение $\Psi(\cdot) : \mathbb{G} \rightarrow [0, 1]^M$, предсказывающее вероятности принадлежности молекулы к каждому из M классов. Качество оценивается метрикой ROC-AUC (площадь под ROC-кривой), усреднённой по всем M тестам токсичности.

Математические основания. Классическое уравнение реакции-адвекции-диффузии в непрерывном пространстве $\Omega \subseteq \mathbb{R}^n$ для скалярной величины $u = u(\mathbf{r}, t)$ имеет вид

$$\frac{\partial u}{\partial t} = \underbrace{D \nabla^2 u}_{\text{диффузия}} - \underbrace{\mathbf{v} \cdot \nabla u}_{\text{адвекция}} + \underbrace{R(u)}_{\text{реакция}}, \quad (3)$$

где $D > 0$ – коэффициент диффузии, $\nabla^2 u = \sum_{k=1}^n \partial^2 u / \partial r_k^2$ – лапласиан, \mathbf{v} – поле скоростей адвекции, ∇u – градиент, $R(u)$ – реакционная функция. Первый член описывает рассеивание – изотропное распространение величины u из областей высокой концентрации в области низкой. Второй – направленный перенос средой со скоростью v : в отличие от диффузии, адвекция асимметрична и может переносить информацию без усреднения. Третий – локальную трансформацию: порождение или поглощение вещества.

Уравнение (3) является основой множества моделей в естественных науках: реакции Белоусова–Жаботинского в химической кинетике, уравнения Фишера–КПП в популяционной динамике, системы Тьюринга в морфогенезе, уравнений обработки изображений. Каждый из трёх членов описывает фундаментально различный физический процесс, и их комбинация позволяет моделировать сложные динамические системы.

Частные случаи. При $R(u) = 0$ уравнение (3) вырождается в уравнение конвекции-диффузии, лежащее в основе CDE-GRAND. При дополнительном условии $v = 0$ получаем чисто диффузионное уравнение, реализуемое архитектурой GRAND. Таким образом, уравнение реакции-адвекции-диффузии является наиболее общей формой, обобщающей все существующие ODE-подходы на графах.

Дискретизация на граф. Перенос уравнения (3) на граф $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ с числом вершин $N = |\mathcal{V}|$ требует замены дифференциальных операторов их графовыми аналогами. Для

матрицы представлений $\mathbf{X}(t) \in \mathbb{R}^{N \times d_h}$, строки которой – векторы признаков вершин $\mathbf{x}_i(t) \in \mathbb{R}^{d_h}$, лапласиан заменяется нормализованным графовым лапласианом, а градиент – разностями признаков вдоль рёбер.

Диффузия на графе в матричной форме допускает следующую формулировку:

$$\frac{\partial \mathbf{X}(t)}{\partial t} = (\mathbf{A}(\mathbf{X}(t)) - \mathbf{I})\mathbf{X}(t), \quad (4)$$

где $\mathbf{A} \in \mathbb{R}^{N \times N}$ – обучаемая матрица внимания, $\mathbf{I} \in \mathbb{R}^{N \times N}$ – единичная матрица порядка N . Матрица внимания определяет силу связей между вершинами через функцию сходства $a(\cdot, \cdot): \mathbb{R}^{d_h} \times \mathbb{R}^{d_h} \rightarrow \mathbb{R}$. Элементы матрицы (коэффициенты внимания) $a_{ij} = a(\mathbf{x}_i(t), \mathbf{x}_j(t))$ допускают различные аналитические определения. Например, в оригинальной статье про модель GRAND [8] используется механизм внимания GAT, подробно описанный в [3]. Для отдельной вершины $i \in \mathcal{V}$ уравнение принимает вид:

$$\frac{\partial \mathbf{x}_i(t)}{\partial t} = \sum_j a_{ij} (\mathbf{x}_j(t) - \mathbf{x}_i(t)), \quad (5)$$

Адвекция (конвекция) на графе моделирует направленный перенос информации. В непрерывном случае (3) адвекция описывается членом $-v \cdot \nabla u$, где v – поле скоростей. На графе аналогом градиента ∇u в вершине $i \in V$ вдоль ребра $(i, j) \in E$ служит разность представлений вершин $x_j(t) - x_i(t)$. Поле скоростей параметризуется обучаемой функцией $V_{ij}(t) \in \mathbb{R}^{d_h}$, зависящей от этой разности. Поэлементное произведение (произведение Адамара) $V_{ij}(t) \odot x_j(t)$ обеспечивает покоординатную модуляцию: каждая компонента вектора $x_j(t)$ масштабируется соответствующей компонентой скорости.

Реакционный член на графе является наиболее простым в плане дискретизации, поскольку он локален: $R(x_i(t))$ зависит только от представления вершины $i \in V$ и не требует агрегирования информации от соседей. Это делает реакцию вычислительно эффективной и параллелизуемой.

Описание модели GRADE. Предлагаемая архитектура развивает существующие ODE-подходы, расширяя уравнение конвекции-диффузии до полного уравнения реакции-адвекции-диффузии на графе. Архитектура состоит из четырёх последовательных компонент: кодировщика атомов, блока интегрирования, графового пулинга и выходного классификатора.

Входной кодировщик. Динамика дифференциального уравнения требует, чтобы размерность состояния $\mathbf{x}_i(t) \in \mathbb{R}^{d_h}$ оставалась постоянной на всём интервале интегрирования $t \in [0, T]$. Поэтому исходные признаки вершин, имеющие произвольную размерность и природу, должны быть отображены в скрытое пространство фиксированной размерности d_h .

В оригинальной статье про модель CDE-GRAND [9] входной кодировщик представляет собой многослойный перцептрон (MLP), который отображает исходные признаки вершин графа $\mathbf{x}_i \in \mathbb{R}^d$ размерности d в скрытое пространство размерности d_h : $\mathbf{x}_i(0) = \text{MLP}_{enc}(\mathbf{x}_i) \in \mathbb{R}^{d_h}$. Кодировщик состоит из двух линейных слоёв с функцией активации ReLU и регуляризацией Dropout между ними. В задаче классификации молекулярных графов ситуация принципиально иная: признаки атомов являются категориальными. Каждый атом описывается $K = 9$ дискретными характеристиками: атомный номер ($C_1 = 128$ категорий), хиральность ($C_2 = 16$), степень ($C_3 = 16$), формальный заряд ($C_4 = 16$), число водородов ($C_5 = 16$), число радикальных электронов ($C_6 = 8$), гибридизация ($C_7 = 8$), ароматичность ($C_8 = 8$), принадлежность к кольцу ($C_9 =$

8). Применение $\text{MLP}_{enc}(\cdot)$ к категориальным данным неэффективно, поскольку one-hot кодирование порождает разреженные вектора высокой размерности ($d_h = \sum_k C_k = 224$), большинство компонент которых равно нулю.

В задаче классификации графов молекул $\text{MLP}_{enc}(\cdot)$ заменяется *кодировщиком атомов* на основе таблиц вложений (embeddings). Для каждого типа признака $k \in \{1, \dots, K\}$ определяется обучаемая таблица вложений $\text{Emb}_k(\cdot): \{0, \dots, C_k - 1\} \rightarrow \mathbb{R}^{d_h}$, представляющая собой матрицу $\mathbf{E}_k \in \mathbb{R}^{C_k \times d_h}$, строка которой с индексом $f_i^{(k)}$ извлекается как $\text{Emb}_k(f_i^{(k)}) = [\mathbf{E}_k]_{f_i^{(k)}}$. Начальное представление вершины $i \in \mathcal{V}$:

$$\mathbf{x}_i^{\text{in}} := \sum_{k=1}^K \text{Emb}_k(f_i^{(k)}) = \sum_{k=1}^K [\mathbf{E}_k]_{f_i^{(k)}} \in \mathbb{R}^{d_h}. \quad (6)$$

Суммирование вместо конкатенации сохраняет размерность d_h . Каждая таблица представлений обучается совместно с остальными параметрами модели, что позволяет автоматически выучить информативные представления для каждой категории каждого признака.

Начальное условие ОДУ. Результат кодирования формирует матрицу начального условия $\mathbf{X}^{\text{in}} = \{x_i^{\text{in}}\}_{i=1}^N \in \mathbb{R}^{N \times d_h}$, строка i которой – вектор x_i^{in} . Дифференциальное уравнение интегрируется как задача Коши с начальным условием:

$$\mathbf{X}(0) = \mathbf{X}^{\text{in}}, \quad \mathbf{X}(0) = (\mathbf{x}_i(0))_{i=1}^N \in \mathbb{R}^{N \times d_h}. \quad (7)$$

Блок интегрирования. Дифференциальное уравнение, лежащее в основе модели GRADE для отдельной вершины $i \in \mathcal{V}$:

$$\frac{\partial \mathbf{x}_i(t)}{\partial t} = \underbrace{\alpha_i(t) \mathcal{D}_i(t)}_{\text{диффузия}} + \underbrace{\beta(1 - \alpha_i(t)) \mathcal{C}_i(t)}_{\text{адвекция}} + \underbrace{\mathbf{g}_i(t) \odot \mathcal{R}_i(t)}_{\text{реакция}}, \quad (8)$$

где тройка $(\mathcal{D}_i(t), \mathcal{C}_i(t), \mathcal{R}_i(t))$ отвечает соответственно за вклад диффузии, адвекции (конвекции) и реакции. Ключевым структурным отличием адвекционно-диффузионной части от CDE-GRAND является наличие эволюционирующих во времени весов у этих слагаемых, причем адаптивный коэффициент $\alpha_i(t) \in [0, 1]$ и его дополнение $1 - \alpha_i(t)$ образуют *выпуклую комбинацию* диффузионного и адвекционного членов. В CDE-GRAND оба эти процесса всегда активны с полным весом; в GRADE же при $\alpha_i(t) \rightarrow 0$ диффузия подавляется, а адвекция усиливается, и наоборот. Это позволяет модели полностью переключаться между режимами в зависимости от локальной структуры. Реакционный член является принципиально новым. Ниже описаны все компоненты.

1) *Диффузия.* В GRADE диффузионный член реализован через обучаемый механизм масштабированного скалярного внимания [13]. Для каждой вершины $i \in \mathcal{V}$ вычисляются вектор запроса и вектор ключа:

$$\mathbf{q}_i(t) = \mathbf{W}_Q \mathbf{x}_i(t) \in \mathbb{R}^{d_h}, \quad \mathbf{k}_i(t) = \mathbf{W}_K \mathbf{x}_i(t) \in \mathbb{R}^{d_h}, \quad (9)$$

где $\mathbf{W}_Q, \mathbf{W}_K \in \mathbb{R}^{d_h \times d_h}$ – обучаемые матрицы проекций. Коэффициенты внимания вычисляются по рёбрам графа как выход функции $\text{softmax}(z_k) = \exp(z_k) / \sum_k \exp(z_k)$:

$$a_{ij}(t) = \exp\left(\frac{\mathbf{q}_i(t)^\top \mathbf{k}_j(t)}{\sqrt{d_h}}\right) / \sum_{l \in \mathcal{N}_i} \exp\left(\frac{\mathbf{q}_i(t)^\top \mathbf{k}_l(t)}{\sqrt{d_h}}\right) \in \mathbb{R}. \quad (10)$$

Масштабирование через $(\sqrt{d_h})^{-1}$ предотвращает попадание скалярных произведений в область насыщения softmax, стабилизируя градиенты. Свойство нормированности $\sum_j a_{ij}(t) = 1$ обеспечивает, что диффузионный член является взвешенным усреднением, не приводящим к неограниченному росту нормы. Матрица внимания $\mathbf{A} \in \mathbb{R}^{N \times N}$ определяется элементами $[A(t)]_{ij} = a_{ij}(t)$ при $(i, j) \in E$ и $[A(t)]_{ij} = 0$ иначе. Диффузионный вклад в матричной и векторной форме:

$$\begin{aligned} \mathcal{D}(t) &= (\mathcal{D}_i(t))_{i=1}^N = (\mathbf{A}(t) - \mathbf{I})\mathbf{X}(t) \in \mathbb{R}^{N \times d_h}, \\ \mathcal{D}_i(t) &= \sum_j a_{ij}(\mathbf{x}_j(t) - \mathbf{x}_i(t)) \in \mathbb{R}^{d_h}. \end{aligned} \quad (11)$$

Обучаемость коэффициентов позволяет модели адаптировать диффузию к признаковому содержанию: внимание может усилить влияние информативных соседей (например, атомов с электроноакцепторными группами) и ослабить нерелевантных.

2) *Адвекция*. Адвекционный член наследуется из CDE-GRAND. Единственным изменением является замена функции активации с ReLU на гиперболический тангенс. Вектор скорости потока от вершины j к вершине i :

$$\mathbf{V}_{ij}(t) = \tanh\left(\mathbf{W}_V(\mathbf{x}_j(t) - \mathbf{x}_i(t))\right) \in \mathbb{R}^{d_h}, \quad (12)$$

где $\mathbf{W}_V \in \mathbb{R}^{d_h \times d_h}$ – обучаемая матрица весов, а поэлементное применение гиперболического тангенса $\tanh(\cdot): \mathbb{R}^{d_h} \rightarrow \mathbb{R}^{d_h}$ обеспечивает ограниченность компонент скорости: $|\mathbf{V}_{ij}(t)_k| \leq 1$, $k = \overline{1, d_h}$. Соответственно, предотвращается бесконтрольный рост нормы вектора скорости, приводящий к нестабильному поведению модели. Матрица адвекции $\mathcal{C}(t) = (\mathcal{C}_i(t))_{i=1}^N \in \mathbb{R}^{N \times d_h}$, строки которой отвечают за адвекционный вклад для вершины $i \in \mathcal{V}$ и определяются как

$$\mathcal{C}_i(t) = \sum_{j \in \mathcal{N}_i} \mathbf{V}_{ij}(t) \odot \mathbf{x}_j(t) \in \mathbb{R}^{d_h}. \quad (13)$$

3) *Адаптивный коэффициент*. Адаптивный коэффициент $\alpha_i(t)$ модулирует адвекцию и диффузию на основе локальной оценки гомогенности ребер:

$$\begin{aligned} h_i^{local}(t) &= \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \exp\left(-\|\mathbf{w} \odot (\mathbf{x}_j(t) - \mathbf{x}_i(t))\|^2\right) \in [0, 1], \\ \alpha_i(t) &= 1 - h_i^{local}(t) \in [0, 1], \end{aligned} \quad (14)$$

где $\mathbf{w} \in \mathbb{R}^{d_h}$ – обучаемая вектор весов. В качестве нормы $\|\cdot\|$ используется l_2 -норма, что преобразует экспоненту в гауссово ядро (RBF-ядро). Поэлементное произведение под знаком нормы обеспечивает свойство анизотропности ядер – свойстве, при котором признаковые координаты дифференцируются по значимости в процессе обучения. При инициализации каждая компонента вектора \mathbf{w} полагается равной $w_k = \sqrt{\gamma_0/d_h}$, $k = \overline{1, d_h}$, $\gamma_0 = 0,5$ – эмпирически полученный результат. Далее в процессе обучения компоненты w_k принимают различные значения, что позволяет модели автоматически определять, какие измерения пространства представлений наиболее информативны для оценки гомогенности.

Физический смысл выпуклой комбинации: при $h_i^{local}(t) \approx 1$ (гомогенное окружение) $\alpha_i(t) \approx 0$ – вес диффузии стремится к нулю, вес адвекции к единице – доминирует адвекция. При $h_i^{local}(t) \approx 0$ (гетерогенное окружение) $\alpha_i(t) \approx 1$ – доминирует

диффузия, адвекция подавлена. Пара коэффициентов $(\alpha_i(t), 1 - \alpha_i(t))$, составляющая выпуклую комбинацию, гарантирует, что суммарный вклад диффузии и адвекции ограничен в любой момент времени $t \in [0, T]$, что способствует устойчивости интегрирования. В отличие от CDE-GRAND, где адвекция и диффузия всегда имеют фиксированный вес 1, GRADE позволяет модели плавно перераспределять ресурсы между обоими процессами.

4) *Реакция*. Ключевым нововведением GRADE является реакционный член, вносящий локальную нелинейную динамику непосредственно в дифференциальное уравнение. Диффузия и адвекция – процессы межвершинного взаимодействия: представление вершины i изменяется только в зависимости от соседей N_i . Реакционный член позволяет вершине трансформировать своё представление *независимо* от окружения.

Реакционная функция – двухслойный перцептрон с гладкой нелинейностью:

$$R(\mathbf{x}) = \mathbf{W}_{R_2} \text{GELU}(\mathbf{W}_{R_1} \mathbf{x} + \mathbf{b}_{R_1}) + \mathbf{b}_{R_2} \in \mathbb{R}^{d_h}, \quad (15)$$

где $\mathbf{W}_{R_1}, \mathbf{W}_{R_2} \in \mathbb{R}^{d_h \times d_h}$, $\mathbf{b}_{R_1}, \mathbf{b}_{R_2} \in \mathbb{R}^{d_h}$ — обучаемые параметры. Поэлементная функция активации $\text{GELU}(z) = z \cdot \Phi(z)$ [14], где $\Phi(\cdot): \mathbb{R} \rightarrow \mathbb{R}$ — стандартная нормальная функция распределения, обеспечивает гладкость, критичную для корректности дифференцирования правой части уравнения (8).

Для контроля вклада реакции для каждой вершины $i \in \mathcal{V}$ вводится вектор гейтирования:

$$\mathbf{g}_i(t) = \sigma(\mathbf{W}_G \mathbf{x}_i(t) + \mathbf{b}_G) \in [0, 1]^{d_h}, \quad (16)$$

где $\mathbf{W}_G \in \mathbb{R}^{d_h \times d_h}$, $\mathbf{b}_G \in \mathbb{R}^{d_h}$ — обучаемые параметры, в качестве функции активации используется $\sigma(z) = 1/(1 + e^{-z})$ — поэлементная логистическая функция. Реакционный вклад:

$$\mathcal{R}_i(t) = \mathbf{g}_i(t) \odot R(\mathbf{x}_i(t)) \in \mathbb{R}^{d_h}. \quad (17)$$

Гейтирование $\mathbf{g}_i(t) \in [0, 1]^{d_h}$ выполняет две функции. Во-первых, оно предотвращает неконтролируемый рост нелинейного члена при длительном интегрировании ($S = T/\tau$ итераций, τ — шаг интегрирования), ограничивая каждую компоненту вектора реакции: $\|\mathcal{R}_i(t)\|_k \leq \|R(\mathbf{x}_i(t))\|_\infty$, $k = \overline{1, d_h}$. Во-вторых, оно позволяет модели поэлементно выбирать, какие компоненты признаков подвергаются реакции, а какие пропускаются. Векторы-строки $\mathcal{R}_i(t)$ составляют гейтированную матрицу реакционных вкладов для каждой вершины графа $\mathcal{R}(t) = (\mathcal{R}_i(t))_{i=1}^N \in \mathbb{R}^{N \times d_h}$.

В контексте молекулярных графов реакционный член моделирует локальные электронные эффекты: атом может изменять своё представление независимо от соседей при наличии неподелённых электронных пар, формального заряда или резонансных структур. Такие эффекты невозможно выразить через диффузию (агрегирование соседей) и адвекцию (направленный перенос от соседей), поскольку оба механизма являются межвершинными.

5) *Итоговое уравнение в матричной форме*. Введём диагональную матрицу адаптивных коэффициентов $\Lambda(t) = \text{diag}(\alpha_1(t), \dots, \alpha_N(t)) \in \mathbb{R}^{N \times N}$. Объединяя три члена, получаем матричное дифференциальное уравнение для GRADE:

$$\frac{\partial \mathbf{X}(t)}{\partial t} = \Lambda(t) \mathcal{D}(t) + \beta(\mathbf{I} - \Lambda(t)) \mathcal{C}(t) + \mathcal{R}(t). \quad (18)$$

б) *Численное интегрирование и стабилизация*. Решение уравнения (19) вычисляется методом Рунге–Кутты 4-го порядка (RK4) с фиксированным шагом τ на $t \in [0, T]$. Метод RK4 обеспечивает точность порядка $O(\tau^4)$, что существенно превосходит метод Эйлера первого порядка точности и позволяет использовать более крупный шаг τ при сохранении точности. Ценой является четырёхкратное вычисление правой части на каждом шаге. После

каждой итерации интегрирования применяется нормализация слоя (Layer Normalization [15]). Для вектора $x \in \mathbb{R}^{d_h}$:

$$\mathbf{x} \leftarrow \text{LayerNorm}(\mathbf{x}) = \frac{\mathbf{x} - \mu}{\sqrt{\sigma^2 + \varepsilon}} \odot \boldsymbol{\gamma} + \boldsymbol{\delta}, \quad \mu = \frac{1}{d_h} \sum_{k=1}^{d_h} x_p, \quad \sigma^2 = \frac{1}{d_h} \sum_{k=1}^{d_h} (x_p - \mu)^2, \quad (19)$$

где $\boldsymbol{\gamma}, \boldsymbol{\delta} \in \mathbb{R}^{d_h}$ – обучаемые параметры масштаба и сдвига, $\varepsilon = 10^{-5}$. Нормализация предотвращает экспоненциальный рост или затухание признаков при длительном интегрировании. Дополнительно применяется покоординатное ограничение на представления вершин $[x_i(t)]_k \leftarrow \max(-H, \min(H, [x_i(t)]_k))$, $H = 10$, $k = \underline{1, d_h}$ для предотвращения числовых переполнений.

Графовый пулинг. Для перехода от вершинных представлений к представлению графа (молекулы) используется усреднение:

$$\mathbf{x}_G = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i(T) \in \mathbb{R}^{d_h}. \quad (20)$$

Выходной классификатор и функция потерь. Представление молекулы преобразуется в вектор логитов двухслойным перцептроном:

$$\hat{\mathbf{y}}_G = \mathbf{W}_{dec,2} \text{SiLU}(\mathbf{W}_{dec,1} \mathbf{x}_G + \mathbf{b}_{dec,1}) + \mathbf{b}_{dec,2} \in \mathbb{R}^M, \quad (21)$$

где $\mathbf{W}_{dec,1} \in \mathbb{R}^{d_h \times d_h}$, $\mathbf{W}_{dec,2} \in \mathbb{R}^{M \times d_h}$, $\mathbf{b}_{dec,1} \in \mathbb{R}^{d_h}$, $\mathbf{b}_{dec,2} \in \mathbb{R}^M$ – обучаемые параметры. В качестве функции активации выступает $\text{SiLU}(z) = z \cdot \sigma(z) = z / (1 + e^{-z})$ – сглаженная дифференцируемая версия $\text{ReLU}(z) = \max(0, z)$. Между слоями применяется Dropout ($p = 0,5$) для регуляризации. Функция потерь — бинарная кросс-энтропия с логитами по валидным меткам. Оптимизация выполняется алгоритмом Adam [16] с ограничением нормы градиента.

Экспериментальные данные. Валидация проведена на датасете ogbg-moltox21 [17] – стандартном бенчмарке из коллекции Open Graph Benchmark (OGB) для предсказания свойств молекулярных графов. Датасет содержит 7831 молекулу, каждая из которых представлена графом, где вершины – атомы, рёбра – химические связи. Графы этих молекул составляют описанное ранее множество $G = \{G_s\}_{s=1}^S$ мощности $|G| = 7831$ – то есть, $S = 7831$ уникальных молекул.

Задачей является одновременная бинарная классификация по $M = 12$ ассеем токсичности. Ассей охватывают два класса биологических мишеней: (1) рецепторы ядерных гормонов – андрогенный рецептор (AR), эстрогенный рецептор (ER), ароматаза, рецептор прегнана X (PXR), рецептор ретиноевой кислоты (RAR) и другие; (b) пути стрессового ответа – путь p53, путь элемента стрессового ответа (HSE), митохондриальный мембранный потенциал (MMP), антиоксидантный ответ (ARE), путь активации каспазы (ATAD5). Каждая молекула может иметь отсутствующие метки для отдельных ассеев, которые маскируются при вычислении функции потерь.

Датасет ogbg-moltox21 особенно подходит для оценки ODE-моделей по нескольким причинам. Во-первых, многозадачная природа (12 ассеев) позволяет оценить обобщающую способность модели: одно представление молекулы должно быть информативным для различных биологических мишеней. Во-вторых, размер датасета (7831 молекула) достаточен для обучения моделей с умеренным числом параметров, но недостаточно велик для подавления преимущества более выразительных архитектур с регуляризацией. В-третьих, молекулярные графы обладают признаковой гетерогенностью: соседние атомы, как правило, имеют разные типы и свойства, что ограничивает применимость многих классических подходов.

Используется стандартное разбиение OGB: 6264/783/784 (обучение/валидация/тест), обеспечивающее воспроизводимость сравнений. Характеристики датасета приведены в таблице 1.

Таблица 1. Характеристики датасета ogbg-moltox21

Характеристика	Значение
Число молекул S	7831
Число тестов (ассеев) M	12
Среднее число атомов \underline{N}	18,6
Среднее число связей $ \underline{E} $	19,3
Число типов атомных признаков K	9
Метрика	ROC-AUC
Разбиение	6264 / 783 / 784

Постановка эксперимента. Сравняются четыре модели: GCN, GRAND, CDE-GRAND, GRADE. Общие гиперпараметры моделей перечислены в таблице 2.

Таблица 2. Гиперпараметры экспериментов

Параметр	Значение
Скрытая размерность d_h	128
Время интегрирования T	2,0
Шаг интегрирования τ	0,2
Число шагов $S = T/\tau$	10
Инициализация γ_0	0,5
Число эпох	100
Скорость обучения η	0,001
Размер батча	64
Dropout	0,5
Ранняя остановка (эпох)	15
Число запусков	3

Каждый эксперимент повторяется 3 раза с различными случайными инициализациями. Модель с наилучшей валидационной метрикой ROC-AUC сохраняется. Приводятся среднее и стандартное отклонение ROC-AUC на тестовом множестве. Все эксперименты проводились в онлайн-среде Google Colab с использованием языка Python 3 и аппаратного ускорителя – графического процессора NVIDIA Tesla T4. NVIDIA Tesla T4 – это энергоэффективный (75 Вт) ускоритель на архитектуре Turing, предназначенный для ИИ-инференса, видеообработки и графики в ЦОД. Карта оснащена 16 ГБ GDDR6 с высокой пропускной способностью, 2560 ядрами CUDA и 320 тензорными ядрами, обеспечивая высокую производительность в вычислениях FP16, INT8 и INT4 при низком потреблении энергии.

Результаты и анализ. Результаты сравнительного эксперимента представлены в таблице 3.

Таблица 3. ROC-AUC на *ogbg-moltox21* (среднее \pm std. откл.)

Модель	ROC-AUC
GCN	0,734 \pm 0,005
GRAND	0,725 \pm 0,001
CDE-GRAND	0,728 \pm 0,002
GRADE	0,745 \pm 0,005

GRADE (0,745) превосходит все базовые модели, демонстрируя наибольший прирост (2 п.п.) в сравнении с GRAND. Этот прирост обеспечивается совокупностью четырёх архитектурных решений. Во-первых, выпуклая комбинация диффузии и конвекции с адаптивными весами позволяет модели для каждой вершины выбирать оптимальный режим: в гомогенном окружении (атомы с похожими представлениями) увеличивается вклад конвекции, а в гетерогенном – диффузии. Во-вторых, обучаемые коэффициенты внимания в диффузионном члене заменяют фиксированную спектральную нормализацию, позволяя модели усиливать влияние информативных соседей – например, атомов с электроноакцепторными или электронодонорными группами. В-третьих, гейтированный реакционный член вводит принципиально новый механизм – локальную нелинейную трансформацию, независимую от окружения. В контексте молекулярных графов он моделирует внутриатомные эффекты: формальный заряд, неподелённые электронные пары, резонансные структуры – всё то, что определяет реакционную способность атома, но не выражается через диффузию или конвекцию, являющихся по своей природе межвершинными процессами. В-четвёртых, обучаемое анизотропное ядро автоматически определяет, какие компоненты признаков наиболее значимы для оценки локальной гомогенности, устраняя необходимость ручного подбора масштаба.

Заключение. В настоящей работе рассмотрена задача предсказания токсичности молекулярных соединений – практически значимая задача для фармацевтической разработки и высокопроизводительного виртуального скрининга. Предложена архитектура GRADE, представляющая динамику признаков атомов молекулы как решение полного уравнения реакции-адвекции-диффузии на графе. Модель обобщает существующие ODE-подходы на графах, последовательно расширяя чисто диффузионную формулировку адвекцией, адаптивным балансом и реакцией. Экспериментальная валидация на стандартном бенчмарке *ogbg-moltox21*, содержащем 7831 молекулу и 12 ассеев токсичности, подтверждает эффективность подхода. GRADE достигает наивысшего значения ROC-AUC среди всех рассмотренных моделей, превосходя как дискретную архитектуру GCN, так и непрерывные ODE-модели GRAND и CDE-GRAND. Анализ результатов показывает, что на компактных молекулярных графах чисто диффузионные и диффузионно-конвекционные модели уступают дискретной архитектуре, и только полная модель реакции-адвекции-диффузии способна раскрыть потенциал непрерывного подхода.

Направления дальнейших исследований включают расширение модели рёберными признаками, отражающими тип химической связи и стереохимию, масштабирование на более крупные бенчмарки молекулярных свойств, разработку адаптивного шага интегрирования для повышения точности при сохранении вычислительной эффективности, а также теоретический анализ влияния реакционного члена на устойчивость представлений при длительном интегрировании.

Список литературы

[1] Wu Z. et al. MoleculeNet: a benchmark for molecular machine learning // *Chemical Science*. – 2018. – Vol. 9, № 2. – PP. 513–530. – DOI: 10.1039/C7SC02664A.

[2] Kipf T.N., Welling M. Semi-supervised classification with graph convolutional networks // *arXiv.org*. – 2017. – DOI: 10.48550/arXiv.1609.02907.

- [3] Veličković P., Cucurull G., Casanova A. et al. Graph attention networks // arXiv.org. – 2018. – DOI: 10.48550/arXiv.1710.10903.
- [4] Gilmer J., Schoenholz S.S., Riley P.F. et al. Neural message passing for quantum chemistry // arXiv.org. – 2017. – DOI: 10.48550/arXiv.1704.01212.
- [5] Zhu J., Yan Y., Zhao L. et al. Beyond homophily in graph neural networks: Current limitations and effective designs // arXiv.org. – 2020. – DOI: 10.48550/arXiv.2006.11468.
- [6] Li Q., Han Z., Wu X.-M. Deeper insights into graph convolutional networks for semi-supervised learning // arXiv.org. – 2018. – DOI: 10.48550/arXiv.1801.07606.
- [7] Chen R.T.Q., Rubanova Y., Bettencourt J., Duvenaud D.K. Neural ordinary differential equations // arXiv.org. – 2018. – DOI: 10.48550/arXiv.1806.07366.
- [8] Chamberlain B.P., Rowbottom J., Gorinova M.I. et al. GRAND: Graph neural diffusion // arXiv.org. – 2021. – DOI: 10.48550/arXiv.2106.10934.
- [9] Zhao K., Kang Q., Song Y. et al. Graph neural convection-diffusion with heterophily // arXiv.org. – 2023. – DOI: 10.48550/arXiv.2305.16780.
- [10] Alvarez L., Guichard F., Lions P.-L., Morel J.-M. Axioms and fundamental equations of image processing // Archive for Rational Mechanics and Analysis. – 1993. – Vol. 123, № 3. – P. 199–257. – DOI: 10.1007/BF00375127.
- [11] Wang T., Dou Z., Bao C., Shi Z. Diffusion mechanism in residual neural network: Theory and applications // arXiv.org. – 2024. – DOI: 10.48550/arXiv.2105.03155.
- [12] Hu W., Fey M., Zitnik M. et al. Open graph benchmark: Datasets for machine learning on graphs // arXiv.org. – 2020. – DOI: 10.48550/arXiv.2005.00687.
- [13] Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need // arXiv.org. – 2017. – DOI: 10.48550/arXiv.1706.03762.
- [14] Hendrycks D., Gimpel K. Gaussian error linear units (GELUs) // arXiv.org. – 2016. – DOI: 10.48550/arXiv.1606.08415.
- [15] Ba J.L., Kiros J.R., Hinton G.E. Layer normalization // arXiv.org. – 2016. – DOI: 10.48550/arXiv.1607.06450.
- [16] Kingma D.P., Ba J. Adam: A method for stochastic optimization // arXiv.org. – 2017. – DOI: 10.48550/arXiv.1412.6980.
- [17] Hu W., Fey M., Zitnik M. et al. Open graph benchmark: Datasets for machine learning on graphs // arXiv.org. – 2020. – DOI: 10.48550/arXiv.2005.00687.

Авторский вклад

Довгий Александр Сергеевич – разработка математического аппарата модели, разработка архитектуры GRADE, реализация модели и проведение вычислительных экспериментов, анализ результатов.

Василевский Константин Викторович – постановка задачи исследования, разработка математического аппарата модели, описание методологии экспериментальной валидации, разработка направлений для дальнейших исследований.

GRAPH NEURAL NETWORK BASED ON AN ADAPTIVE REACTION-ADVECTION-DIFFUSION EQUATION FOR MOLECULAR GRAPH CLASSIFICATION

A.S. Douhi

Undergraduate student, Faculty of Applied Mathematics and Computer Science, Belarusian State University

K.V. Vasilevsky

Associate Professor, Department of Computer Technologies and Systems, Belarusian State University, PhD in Physics and Mathematics

Annotation. The GRADE (Graph Reaction-Advection-Diffusion Equation) architecture is proposed – a graph neural network that models the continuous dynamics of node representations in a graph as the solution of the full reaction-advection-diffusion equation. The architecture extends CDE-GRAND with four key innovations: (1) a convex combination of both advection and diffusion instead of fixed diffusion with modulated convection; (2) a learnable attention coefficient in the diffusion term; (3) a gated reaction term; (4) a learnable anisotropic RBF kernel with a learnable weight vector. Experimental validation on the multitask binary classification problem for molecular graphs ogbg-moltox21 (12 toxicity assays, 7,831 molecules) demonstrates a ROC-AUC of 0.745 for GRADE, which surpasses GCN (0.734), GRAND (0.725), and CDE-GRAND (0.725).

Keywords: graph neural networks, reaction-advection-diffusion equation, molecular graphs, convex combination, learnable attention.