

The background is a dark blue gradient. In the top-left corner, there are two overlapping parallelogram shapes, one blue and one light green. In the bottom-left, there is a circular inset showing a detailed image of a circuit board. In the top-right, there is a faint, high-contrast image of a circuit board layout.

Graph Neural Network

СОДЕРЖАНИЕ

- Обзор
- Виды архитектур
- Message Passing Interface
- QSAR Датасеты

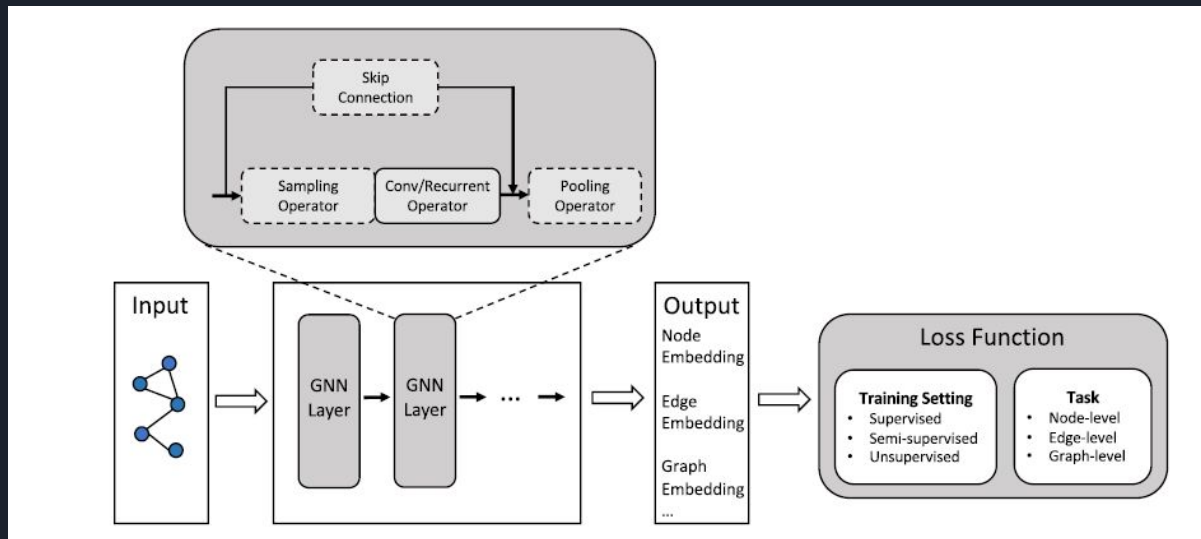


Обзор

GNNs - методы глубокого обучения, которые работают с графами.

Модули:

- 1) Propagation Module
- 2) Sampling Module
- 3) Pooling Module



Input и обозначения

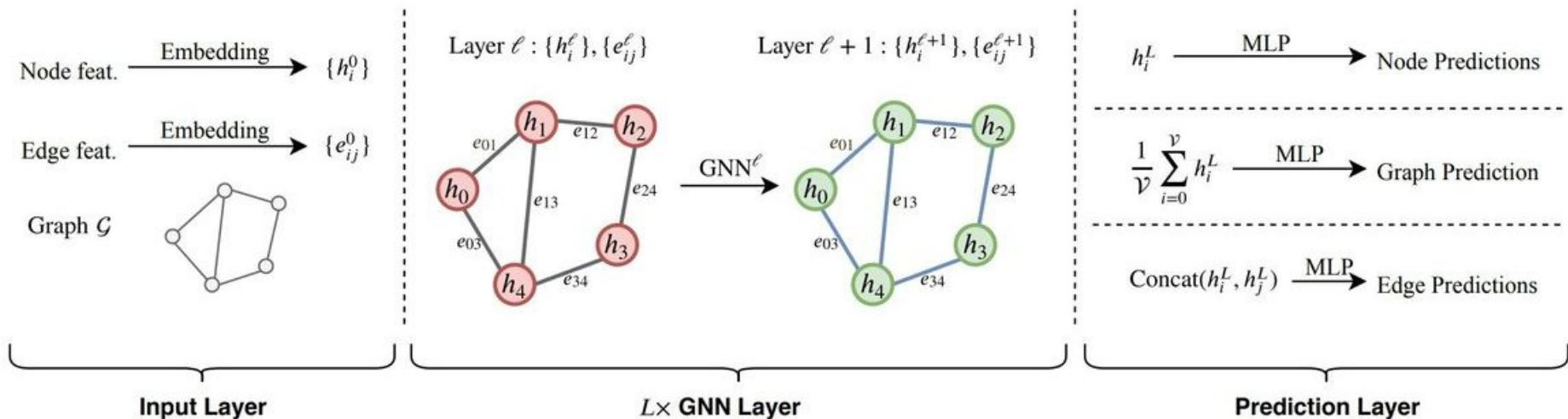
Input (граф):

- 1) $X(H)$ - матрица со свойствами вершин
- 2) A - матрица смежности
- 3) E - матрица свойств ребер

Обозначения:

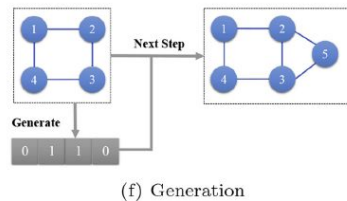
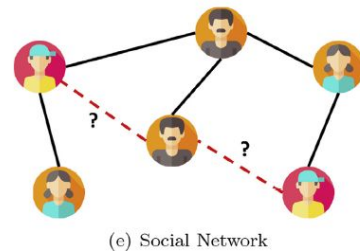
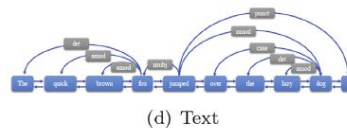
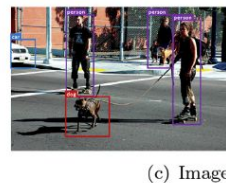
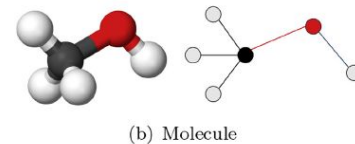
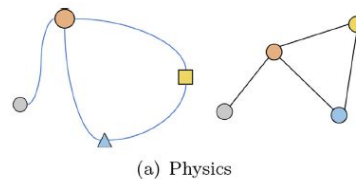
h_v (x_v) - вектор свойств вершины v

e_{vw} - вектор свойств ребра vw



Типы задач

- 1 Node - level.
Например, классификация вершины.
- 2 Edge - level
Например, предсказание связи
- 3 Graph - level
Например, классификация графа



Propagation Modules: Convolution

Spectral approaches

- Идея: 1) сделаем графовое преобразование Фурье для сигнала \mathbf{x}
2) Делаем свертку
3) Делаем обратное преобразование..

$$\mathcal{F}(\mathbf{x}) = \mathbf{U}^T \mathbf{x},$$
$$\mathcal{F}^{-1}(\mathbf{x}) = \mathbf{U} \mathbf{x}.$$

\mathbf{U} - нормализованный Лапласиан Графа или матрица собственных векторов матрицы \mathbf{L} .

$$\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$$

$$\mathbf{g} \star \mathbf{x} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{g}) \odot \mathcal{F}(\mathbf{x}))$$
$$= \mathbf{U}(\mathbf{U}^T \mathbf{g} \odot \mathbf{U}^T \mathbf{x}),$$

\mathbf{g}_w - диагональная



$$\mathbf{g}_w \star \mathbf{x} = \mathbf{U} \mathbf{g}_w \mathbf{U}^T \mathbf{x}.$$

Spectral Approaches

ChebNet: приближаем \mathbf{g}_w многочленами Чебышева до степени K .

$$\mathbf{g}_w \star \mathbf{x} \approx \sum_{k=0}^K w_k \mathbf{T}_k(\tilde{\mathbf{L}}) \mathbf{x},$$

GCN: $K=1$. Большая производительность.

$$\mathbf{g}_w \star \mathbf{x} \approx w \left(\mathbf{I}_N + \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \right) \mathbf{x}.$$



$$\mathbf{H} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X} \mathbf{W},$$

Формула обновления состояний

<https://arxiv.org/abs/1609.02907>



Spatial Approaches

Идея: будем агрегировать информацию в исходном пространстве.

1) Basic

$$\mathbf{t} = \mathbf{h}_v^t + \sum_{u \in \mathcal{N}_v} \mathbf{h}_u^t,$$

$$\mathbf{h}_v^{t+1} = \sigma\left(\mathbf{t} \mathbf{W}_{|\mathcal{N}_v|}^{t+1}\right),$$

2) GraphSage. Сэмплируем соседей.
AGG - mean pooling, LSTM и т.д

$$\begin{aligned} \mathbf{h}_{\mathcal{N}_v}^{t+1} &= \text{AGG}_{t+1}\left(\{\mathbf{h}_u^t, \forall u \in \mathcal{N}_v\}\right), \\ \mathbf{h}_v^{t+1} &= \sigma\left(\mathbf{W}^{t+1} \cdot \left[\mathbf{h}_v^t \parallel \mathbf{h}_{\mathcal{N}_v}^{t+1}\right]\right). \end{aligned}$$

3) GAT

$$\begin{aligned} \mathbf{h}_v^{t+1} &= \rho\left(\sum_{u \in \mathcal{N}_v} \alpha_{vu} \mathbf{W} \mathbf{h}_u^t\right), \\ \alpha_{vu} &= \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W} \mathbf{h}_v \parallel \mathbf{W} \mathbf{h}_u]))}{\sum_{k \in \mathcal{N}_v} \exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W} \mathbf{h}_v \parallel \mathbf{W} \mathbf{h}_k]))}, \end{aligned}$$

Reccurent Modules

Different variants of recurrent operators.

Variant	Aggregator	Updater
GGNN	$\mathbf{h}_{\mathcal{N}_v}^t = \sum_{k \in \mathcal{N}_v} \mathbf{h}_k^{t-1} + \mathbf{b}$	$\mathbf{z}_v^t = \sigma(\mathbf{W}^z \mathbf{h}_{\mathcal{N}_v}^t + \mathbf{U}^z \mathbf{h}_v^{t-1})$ $\mathbf{r}_v^t = \sigma(\mathbf{W}^r \mathbf{h}_{\mathcal{N}_v}^t + \mathbf{U}^r \mathbf{h}_v^{t-1})$ $\tilde{\mathbf{h}}_v^t = \tanh(\mathbf{W} \mathbf{h}_{\mathcal{N}_v}^t + \mathbf{U}(\mathbf{r}_v^t \odot \mathbf{h}_v^{t-1}))$ $\mathbf{h}_v^t = (1 - \mathbf{z}_v^t) \odot \mathbf{h}_v^{t-1} + \mathbf{z}_v^t \odot \tilde{\mathbf{h}}_v^t$
Tree LSTM (Child sum)	$\mathbf{h}_{\mathcal{N}_v}^{ti} = \sum_{k \in \mathcal{N}_v} \mathbf{U}^i \mathbf{h}_k^{t-1}$ $\mathbf{h}_{\mathcal{N}_v k}^{tf} = \mathbf{U}^f \mathbf{h}_k^{t-1}$ $\mathbf{h}_{\mathcal{N}_v}^{to} = \sum_{k \in \mathcal{N}_v} \mathbf{U}^o \mathbf{h}_k^{t-1}$ $\mathbf{h}_{\mathcal{N}_v}^{tu} = \sum_{k \in \mathcal{N}_v} \mathbf{U}^u \mathbf{h}_k^{t-1}$	$\mathbf{i}_v^t = \sigma(\mathbf{W}^i \mathbf{x}_v^t + \mathbf{h}_{\mathcal{N}_v}^{ti} + \mathbf{b}^i)$ $\mathbf{f}_{vk}^t = \sigma(\mathbf{W}^f \mathbf{x}_v^t + \mathbf{h}_{\mathcal{N}_v k}^{tf} + \mathbf{b}^f)$ $\mathbf{o}_v^t = \sigma(\mathbf{W}^o \mathbf{x}_v^t + \mathbf{h}_{\mathcal{N}_v}^{to} + \mathbf{b}^o)$ $\mathbf{u}_v^t = \tanh(\mathbf{W}^u \mathbf{x}_v^t + \mathbf{h}_{\mathcal{N}_v}^{tu} + \mathbf{b}^u)$ $\mathbf{c}_v^t = \mathbf{i}_v^t \odot \mathbf{u}_v^t + \sum_{k \in \mathcal{N}_v} \mathbf{f}_{vk}^t \odot \mathbf{c}_k^{t-1}$ $\mathbf{h}_v^t = \mathbf{o}_v^t \odot \tanh(\mathbf{c}_v^t)$

Tree LSTM (N-ary)

$$\mathbf{h}_{\mathcal{N}_v}^{ti} = \sum_{l=1}^K \mathbf{U}_l^i \mathbf{h}_{vl}^{t-1}$$

$$\mathbf{h}_{\mathcal{N}_v k}^{tf} = \sum_{l=1}^K \mathbf{U}_{kl}^f \mathbf{h}_{vl}^{t-1}$$

$$\mathbf{h}_{\mathcal{N}_v}^{to} = \sum_{l=1}^K \mathbf{U}_l^o \mathbf{h}_{vl}^{t-1}$$

$$\mathbf{h}_{\mathcal{N}_v}^{tu} = \sum_{l=1}^K \mathbf{U}_l^u \mathbf{h}_{vl}^{t-1}$$

Graph LSTM in
(Peng et al., 2017)

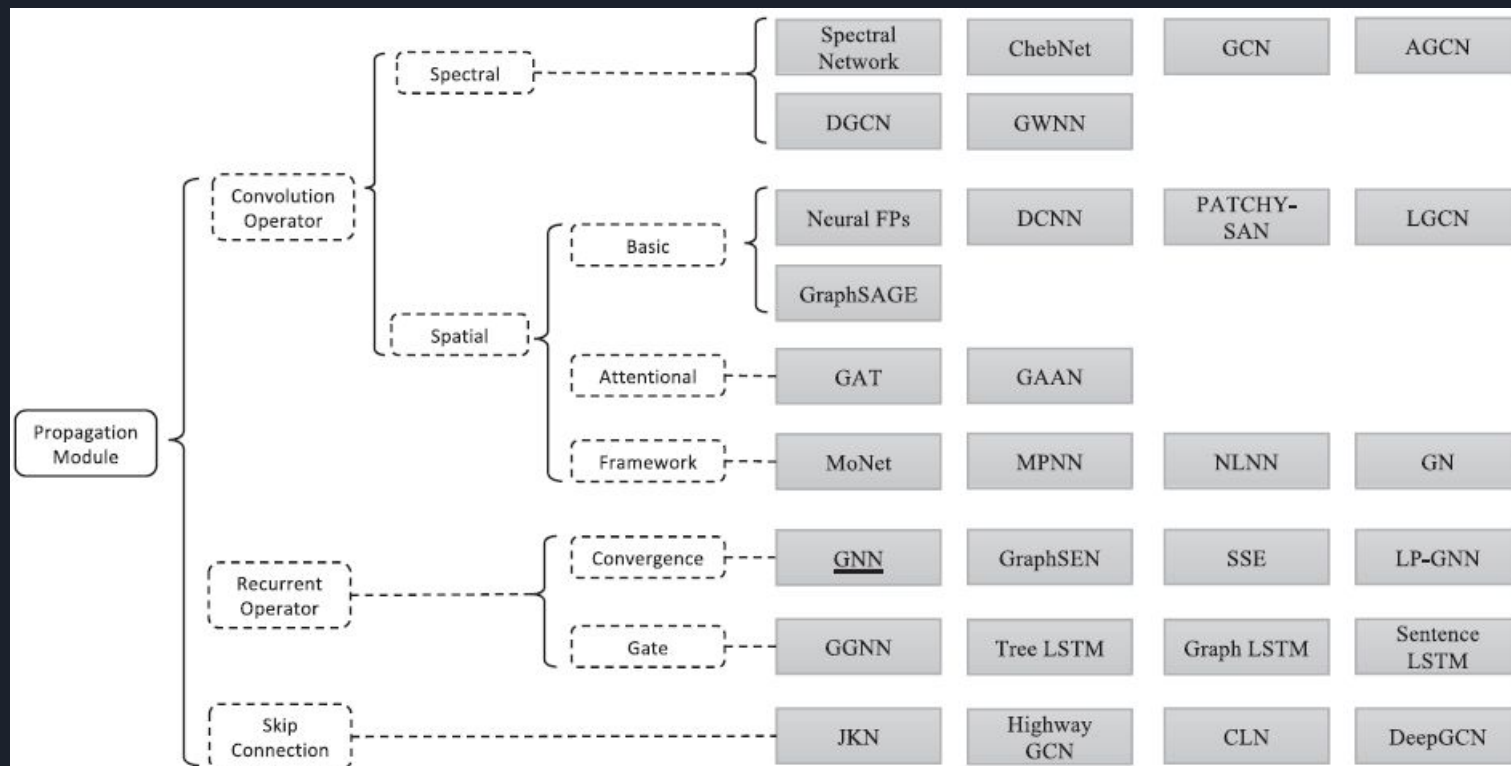
$$\mathbf{h}_{\mathcal{N}_v}^{ti} = \sum_{k \in \mathcal{N}_v} \mathbf{U}_{m(v,k)}^i \mathbf{h}_k^{t-1}$$

$$\mathbf{h}_{\mathcal{N}_v k}^{tf} = \mathbf{U}_{m(v,k)}^f \mathbf{h}_k^{t-1}$$

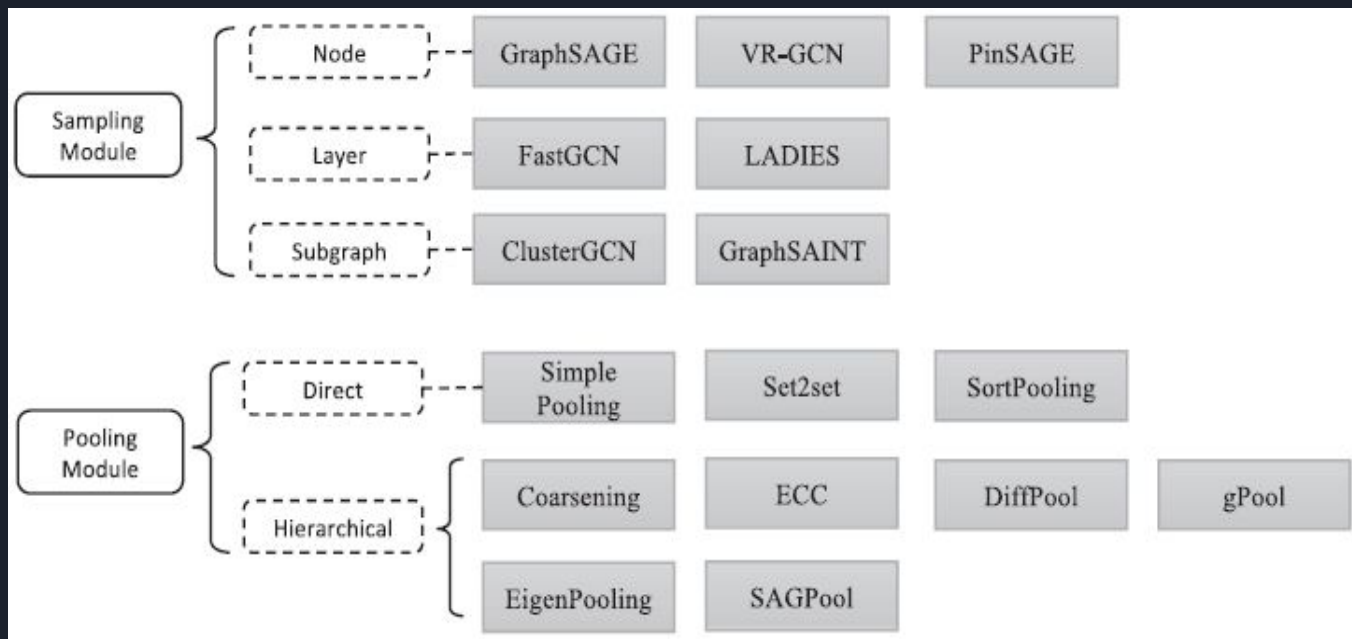
$$\mathbf{h}_{\mathcal{N}_v}^{to} = \sum_{k \in \mathcal{N}_v} \mathbf{U}_{m(v,k)}^o \mathbf{h}_k^{t-1}$$

$$\mathbf{h}_{\mathcal{N}_v}^{tu} = \sum_{k \in \mathcal{N}_v} \mathbf{U}_{m(v,k)}^u \mathbf{h}_k^{t-1}$$

Виды архитектур



Виды архитектур





Message Passing Interface

Общий фреймворк для некоторых spatial методов.

- 1) Используем функцию M_t для агрегации информации от соседей
- 2) Используем функцию U_t для обновления скрытых состояний
- 3) Используем R - readout функцию для агрегации информации со всего графа.

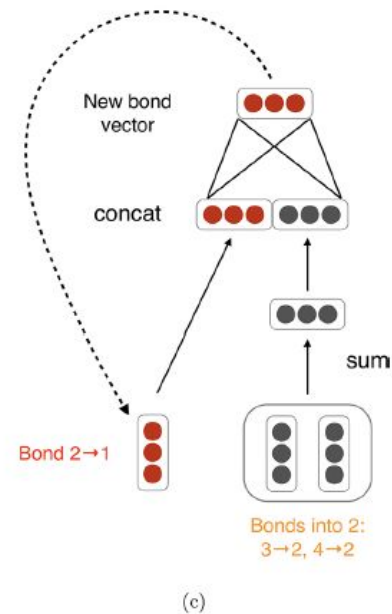
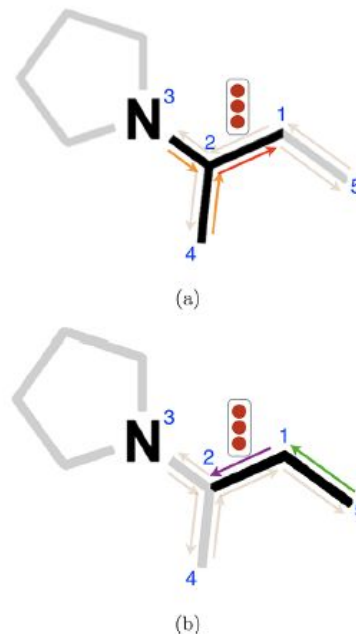
$$\mathbf{m}_v^{t+1} = \sum_{u \in \mathcal{N}_v} M_t(\mathbf{h}_v^t, \mathbf{h}_u^t, \mathbf{e}_{vu}),$$
$$\mathbf{h}_v^{t+1} = U_t(\mathbf{h}_v^t, \mathbf{m}_v^{t+1}).$$



$$\hat{\mathbf{y}} = R(\{\mathbf{h}_v^T \mid v \in G\}),$$

Идеи.

- 1) Заикливание информации может создавать шум
- 2) Перейдем к связям
- 3) Сделаем граф ориентуемым



Ход обучения

$$h_{vw}^0 = \tau(W_i \text{ cat}(x_v, e_{vw}))$$

$$m_{vw}^{t+1} = \sum_{k \in \{N(v) \setminus w\}} h_{kv}^t$$

$$h_{vw}^{t+1} = \tau(h_{vw}^0 + W_m m_{vw}^{t+1})$$

$$m_v = \sum_{w \in N(v)} h_{vw}^T$$

$$h_v = \tau(W_a \text{ cat}(x_v, m_v))$$

$$h = \sum_{v \in G} h_v$$

$$\hat{y} = f(h)$$

Инициализация

Table 1. Atom Features^a

feature	description	size
atom type	type of atom (ex. C, N, O), by atomic number	100
# bonds	number of bonds the atom is involved in	6
formal charge	integer electronic charge assigned to atom	5
chirality	unspecified, tetrahedral CW/CCW, or other	4
# Hs	number of bonded hydrogen atoms	5
hybridization	sp, sp2, sp3, sp3d, or sp3d2	5
aromaticity	whether this atom is part of an aromatic system	1
atomic mass	mass of the atom, divided by 100	1

^aAll features are one-hot encodings except for atomic mass, which is a real number scaled to be on the same order of magnitude.

Table 2. Bond Features^a

feature	description	size
bond type	single, double, triple, or aromatic	4
conjugated	whether the bond is conjugated	1
in ring	whether the bond is part of a ring	1
stereo	none, any, E/Z or cis/trans	6

^aAll features are one-hot encodings.

Некоторые результаты

1) Они применили модель глубокого обучения под названием Chemprop, которая прогнозировала антибиотические свойства молекул: подавление размножения кишечной палочки.

2) Хорошее качества на множестве публичных датасетов

Table 8. Number of Public Data Sets Where D-MPNN Is Statistically Significantly Better than, Equivalent to, or Worse than Each Baseline Model

baseline	D-MPNN is better	D-MPNN is the same	D-MPNN is worse	no. of data sets
MoleculeNet ²	5	3	2	10
Mayr et al. ¹²	8	10	1	19
RF on Morgan	14	0	1	15
FFN on Morgan	14	5	0	19
FFN on Morgan Counts	15	4	0	19
FFN on RDKit	8	5	4	19



QSAR Датасеты

data set	no. of tasks	task type	no. of compounds	metric
QM7	1	regression	6,830	MAE
QM8	12	regression	21,786	MAE
QM9	12	regression	133,885	MAE

- 1) QM9 - вычисленные геометрические, энергетические, электронные и термодинамические свойства 134 тыс. Стабильных малых органических молекул из CHONF.
- 2) QM8 - Электронные спектры



Возможные направления исследования

- 1) Graph Pretraining
- 2) Graph AutoEncoder
- 3) Попробовать улучшить какую-нибудь архитектуру, например, ChemProp