

# Построение карты связности функциональных групп в задаче декодирования сигналов головного мозга

*Вареник Н.В., Стрижов В.В.*

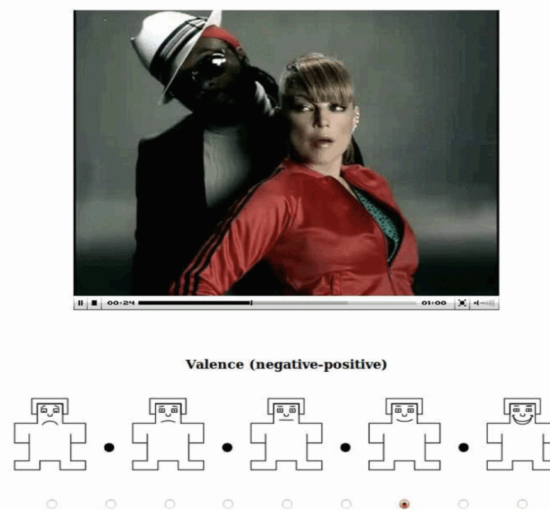
**Программа минимум:** Предлагается использовать графовое представление пространственной структуры размещения электродов для выявления эффективных взаимосвязей различных функциональных групп в пространстве при анализе активности головного мозга. Решается задача построения матрицы смежности для определения графовой структуры сигнала в выбранных задачах декодирования сигналов. Для оценки матрицы смежности рассматриваются детерминированные подходы определения связи между временными рядами на основе спектрального анализа, метода синхронизации фаз и метода задержек. Также предлагается оптимизационный подход для нахождения обучаемой матрицы смежности через оценку апостериорного максимума. **Программа максимум:** Graph Neural ODE, GRAND.

## 1 Данные

### Выборка DEAP.

Рассматривается общедоступный набор данных для анализа аффективных состояний человека, A Database for Emotion Analysis using Physiological Signals. В эксперименте приняли участие 32 здоровых человека (50% женщин) в возрасте 19-37 лет. Все испытуемые соответствовали необходимым требованиям по здоровью и состоянию, подписали согласие и были ознакомлены с протоколом эксперимента. Тщательно отобранные 40 музыкальных видеоклипов были выбраны в качестве визуальных стимулов, вызывающие у испытуемых различные эмоции. Длительность видео составляла 1 минуту. Перед каждым испытанием проводилась 5-секундная базовая стадия записи. В течении эксперимента были получены записи электроэнцефалограммы (EEG) и 13 периферийных физиологических сигналов участников, для 22 участников были дополнительно получены видеозаписи лица. Использовался накопитель EEG с 32 каналами, расположенными по системе 10-20. Участники оценивали каждое видео с точки зрения возбуждения, валентности, симпатии/антипатии, доминирования и знакомства. Для каждого измерения, кроме знакомства была принята непрерывная 9-балльная шкала для измерения уровня этих параметров. Для уровня знакомства принята 5-балльная дискретная шкала.

Еще одна выборка Найду и опишу позже



**Рис. 1** Визуализация шкалы для самооценки уровня валентности. Можно выбирать значения в промежутке между 9 отрисованными точками.

Рис. 1 Визуализация шкалы для самооценки уровня валентности. Можно выбирать значения в промежутке между 9 отрисованными точками.

## 2 Постановка задачи

### 2.1 Построение карты связности

Исходный сигнал EEG задан в виде матрицы  $\mathbf{X}_i \in \mathbb{R}^{T \times E}$ ,  $i \in \overline{1, M}$ , где  $T$  соответствует количеству отсчетов времени при измерении сигнала, а  $E$  - количеству электродов,  $M$  - количеству примеров в выборке. Также дана матрица координат электродов  $\mathbf{Z} \in \mathbb{R}^{E \times 2}$ , определяемая выбранным при снятии данных стандартом размещения. В текущей работе предлагается рассмотреть сигнал EEG в виде ненаправленного графа  $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, \mathbf{A}_{\mathbf{X}_i})$  для извлечения более сложных пространственных взаимосвязей между электродами сигнала. В качестве множества вершин графа  $\mathcal{V}_i$  рассматриваются электроды, значения сигнала на которых являются значениями вершин. Множество ребер  $\mathcal{E}_i$  графа задается матрицей смежности вершин  $\mathbf{A}_{\mathbf{X}_i}$ . Главной задачей работы является исследование различных методов построения матрицы смежности  $\mathbf{A}_{\mathbf{X}_i}$  вершин  $\mathcal{V}_i$ . Где  $\mathbf{A}_{\mathbf{X}_i}$  есть симметричная действительная матрица размера  $E \times E$ , в которой  $a_{ij}$  - вес ребра между вершинами  $i$  и  $j$  в случае взвешенного графа или  $a_{ij} \in \{1, 0\}$  - индикатор наличия ребра между вершинами, в случае невзвешенного графа.

Требуется найти функцию:

$$\mathbf{f}_A : (\mathbf{X}_i, \mathbf{Z}) \rightarrow \mathbf{A}_{\mathbf{X}_i, \mathbf{Z}}$$

### 2.2 Задача классификации сигнала

Дана выборка

$$\begin{aligned} \mathcal{D} &= (\mathbf{X}, \mathbf{Z}, \mathbf{y}), \text{ где } \mathbf{X} = \{\mathbf{X}_i\}_{i=1}^M, \\ \mathbf{X}_i &= \{\mathbf{x}_t\}_{t \in T}, \mathbf{x}_t \in \mathbb{R}^E, T = \{t_i\}_{i=1}^N, t_i \in \mathbb{R}, \\ \mathbf{Z} &= \{\mathbf{z}_j\}_{j=1}^E, \mathbf{z}_j \in \mathbb{R}^2, \\ \mathbf{y} &= \{y_i\}_{i=1}^M, y_i \in \{1, \dots, C\}, \\ E &= 32 - \text{количество электродов,} \\ N &= \text{количество наблюдений в одном отрезке сигнала.} \end{aligned}$$

Для решения задачи классификации рассматривается класс графовых нейронных сетей:

$$g_\theta : (\mathbf{X}, \mathbf{A}_{\mathbf{X}, \mathbf{Z}}) \rightarrow \mathbf{y}.$$

В качестве функции ошибки выбрана кросс-энтропия:

$$\mathcal{L}_{ce} = -\frac{1}{M} \sum_{i=1}^M \left[ \sum_{c=0}^C y_i^c \log(p_i^c) \right]$$

$p_i^c = g_\theta(\mathbf{X}_i, \mathbf{A}_{\mathbf{X}_i, \mathbf{Z}})$  - вероятность класса  $c$  для  $\mathbf{X}_i$  с матрицей смежности  $\mathbf{A}_{\mathbf{X}_i, \mathbf{Z}}$ .

Тогда получаемая оптимизационная задача примет вид:

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}_{ce}(\theta, \mathbf{X}, \mathbf{A}_{\mathbf{X}, \mathbf{Z}}).$$

## 3 Оценка матрицы смежности

### 3.1 Базовые методы

– Евклидово расстояние:

$$\begin{aligned} d_{ij} &= \|\mathbf{z}_i - \mathbf{z}_j\|_2^2, \\ a_{ij} &= \begin{cases} d_{ij}, & \text{если } d_{ij} \leq \rho_d \\ 0, & \text{иначе.} \end{cases} \end{aligned}$$

– Линейная корреляция Пирсона:

$$r_{ij} = \frac{\sum_{t=1}^T (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)}{(T-1)s_{x_i}s_{x_j}} = \frac{\sum_{t=1}^T (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)}{\sqrt{\sum_{t=1}^T (x_i(t) - \bar{x}_i)^2 \sum_{t=1}^T (x_j(t) - \bar{x}_j)^2}},$$

$$a_{ij} = \begin{cases} r_{ij}, & \text{если } r_{ij} \geq \rho_r \\ 0, & \text{иначе.} \end{cases}$$

### 3.2 Метод спектральной согласованности (Magnitude Squared Coherence)

Распространенный способ оценки линейной связи между двумя сигналами:

$$\gamma_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)},$$

где по теореме Винера-Хинчина определяется связь спектральной плотности с автокорреляцией через преобразование Фурье:

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-i2\pi f\tau} d\tau - \text{авто-спектральная функция плотности,}$$

$$S_{xy}(f) = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{-i2\pi f\tau} d\tau - \text{кросс-спектральная функция плотности,}$$

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t) \overline{y(t-\tau)} dt - \text{функция корреляции,}$$

$$\overline{x(t)} - \text{комплексное сопряжение.}$$

Для дискретного случая получаем:

$$S_{xx}(f) = \lim_{N \rightarrow \infty} \frac{(\Delta t)^2}{T} \left| \sum_{n=-N}^N x_n e^{-i2\pi f n \Delta t} \right|^2, \quad x_n = x(t_n) = (\mathbf{x})_n,$$

$$S_{xy}(f) = \sum_{n=-\infty}^{\infty} R_{xy}(\tau_n) e^{-i2\pi f n \tau_n}, \quad \tau_n - \text{временной лаг.}$$

Итого:

$$a_{ij} = \begin{cases} \gamma_{x_i x_j}, & \text{если } \gamma_{x_i x_j} \geq \rho_\gamma \\ 0, & \text{иначе.} \end{cases} \quad (1)$$

### 3.3 Метод частично направленной согласованности (Partial Directed Coherence)

Основное преимущество данного подхода оценки линейной связи заключается в способности выявлять дополнительную информацию о причинно-следственной связи между

наблюдениями. Концепция причинности по Грейнджеру (ссылка) основана на идее адекватности, согласно которой причины предшествуют своим последствиям во времени. В линейной структуре причинность по Грейнджеру обычно оценивается путем подбора векторных авторегрессионных моделей.

Пусть набор одновременно наблюдаемых временных рядов  $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$  адекватно описывается векторной регрессионной моделью порядка  $p$ :

$$\mathbf{x}(t) = \sum_{k=1}^p \mathbf{W}_k \mathbf{x}(t-k) + \mathbf{b}(t), \text{ где}$$

$$\mathbf{W}_k = \begin{bmatrix} w_{11}(k) & \dots & w_{1N}(k) \\ \vdots & \ddots & \vdots \\ w_{N1}(k) & \dots & w_{NN}(k) \end{bmatrix} - \text{матрица авторегрессионных коэффициентов лага } k,$$

$$\mathbf{b}(t) = [b_1(t) \dots b_N(t)]^T - \text{вектор с нулевым средним и матрицей ковариации } \Sigma_b.$$

Коэффициенты  $w_{ij}(k)$  отображают эффект влияния  $x_j(t-k)$  на  $x_i(t)$ . Чтобы описать причинность в частотной области вводится понятие частично направленной согласованности (Partial Directed Coherence).

Пусть

$$\mathbf{W}(\lambda) = \sum_{k=1}^p \mathbf{W}_k e^{-i2\pi\lambda k} - \text{преобразование Фурье матрицы коэффициентов,}$$

$$\lambda \in [-0.5, 0.5] - \text{частота.}$$

Тогда частично направленная согласованность (интенсивность информационного потока из канала  $j$  в канал  $i$  по частоте  $\lambda$ ) определяется как:

$$\pi_{i \leftarrow j}(\lambda) = \frac{\frac{1}{\sigma_i} |\overline{\mathbf{W}}_{ij}(\lambda)|}{\sqrt{\sum_{m=1}^p \frac{1}{\sigma_m^2} \overline{\mathbf{W}}_{mj}(\lambda) \overline{\mathbf{W}}_{mj}^H(\lambda)}}, \text{ где}$$

$$\overline{\mathbf{W}}(\lambda) = \mathbf{I} - \mathbf{W}(\lambda), \sigma_i^2 - \text{дисперсия процесса } b_i(t),$$

$H$  – эрмитово сопряжение.

### 3.4 Метод синхронизации фазы (Phase Locking Value)

Фазовая синхронизация представляет собой подход к анализу возможных нелинейных взаимозависимостей и фокусируется на фазах сигналов. Значение синхронизации фазы (PLV) является одним из наиболее часто используемых мер связи фаз. Предполагается, что две динамические системы могут иметь синхронизацию фаз, даже если их амплитуды независимы. Синхронизация фаз понимается как:

$$|\varphi_x(t) - \varphi_y(t)| = \text{const.}$$

Чтобы оценить фазу сигнала вычисляется аналитическое представление сигнала с использованием преобразования Гильберта:

$$H(t) = x(t) + i\tilde{x}(t), \text{ где}$$

$$\tilde{x}(t) = \frac{1}{\pi} v.p. \int_{-\infty}^{\infty} \frac{x(t')}{t - t'} dt' - \text{преобразование Гильберта сигнала } x(t),$$

*v.p.* — главное значение интеграла по Коши.

Фаза аналитического сигнала определяется как:

$$\varphi(t) = \arctan \left( \frac{\tilde{x}(t)}{x(t)} \right).$$

Таким образом, для двух сигналов  $x(t)$ ,  $y(t)$  равной продолжительности с фазами  $\varphi_x(t)$ ,  $\varphi_y(t)$  мера PLV задается уравнением:

$$PLV_{xy} = \left| \frac{1}{N} \sum_{j=0}^{N-1} \exp(i(\varphi_x(j\Delta t) - \varphi_y(j\Delta t))) \right|$$

где  $\Delta t$  — шаг по времени, а  $N$  — количество наблюдений сигнала.

### 3.5 Метод нелинейной синхронизации

По имеющимся двум временным рядам сигналов  $\mathbf{x} \in \mathbb{R}^N$  и  $\mathbf{y} \in \mathbb{R}^N$  строятся векторы задержек:

$$\begin{aligned} \mathbf{x}_n &= [x_n, x_{n-\tau}, \dots, x_{n-(d-1)\tau}]^T, \\ \mathbf{y}_n &= [y_n, y_{n-\tau}, \dots, y_{n-(d-1)\tau}]^T, \end{aligned}$$

где  $d$  — размерность вектора,  $\tau$  — временной сдвиг,  $n = 1, \dots, N$ .

Пусть  $r_{n,j}$ ,  $s_{n,j}$ ,  $j = 1, \dots, k$  — временные индексы ближайших  $k$  соседей по Евклидовому расстоянию для  $\mathbf{x}_n$  и  $\mathbf{y}_n$  соответственно.

Для каждого  $\mathbf{x}_n$  среднее квадратичное расстояние до  $k$  соседей:

$$R_n^{(k)}(\mathbf{x}) = \frac{1}{k} \sum_{j=1}^k \|\mathbf{x}_n - \mathbf{x}_{r_{n,j}}\|^2.$$

Обусловленное на  $\mathbf{y}$  среднее квадратичное расстояние определяется заменой индекса соседа для  $\mathbf{x}_n$  на соответствующий индекс соседа для  $\mathbf{y}_n$ :

$$R_n^{(k)}(\mathbf{x}|\mathbf{y}) = \frac{1}{k} \sum_{j=1}^k \|\mathbf{x}_n - \mathbf{x}_{s_{n,j}}\|^2.$$

Для облака точек  $\{\mathbf{x}_n\}$  определим средний квадратичный радиус:

$$R(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^N R_n^{(N-1)}(\mathbf{x}).$$

Тогда сигналы строго коррелированы, если  $R_n^{(k)}(\mathbf{x}|\mathbf{y}) \approx R_n^{(k)}(\mathbf{x}) \ll R(\mathbf{x})$ , если же  $R_n^{(k)}(\mathbf{x}|\mathbf{y}) \approx R(\mathbf{x}) \gg R_n^{(k)}(\mathbf{x})$ , то они независимы. Исходя из этого определим меру синхронизации двух сигналов:

$$S^{(k)}(\mathbf{x}|\mathbf{y}) = \frac{1}{N} \sum_{n=1}^N \frac{R_n^{(k)}(\mathbf{x})}{R_n^{(k)}(\mathbf{x}|\mathbf{y})}.$$

#### 4 Оценка матрицы смежности. Оптимизационный подход

Пусть  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$  - ненаправленный граф с множеством вершин  $\mathcal{V}$ ,  $|\mathcal{V}| = E$ , набором ребер  $\mathcal{E}$  и взвешенной матрицей смежности  $\mathbf{A}$ .  $\mathbf{A}$  есть симметричная матрица размера  $E \times E$ , где  $a_{ij}$  - вес ребра, соединяющего вершины  $i$  и  $j$ .

Определим дискретный оператор Лапласа для графа:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \quad \mathbf{D} = \text{diag}\{d_1, \dots, d_E\}, \quad d_i = \sum_{j=1}^E a_{ij} \quad (2)$$

Сигнал  $\mathbf{X} \in \mathbb{R}^{E \times N}$ , определенный на графе  $\mathcal{G}$  является гладким по отношению к  $\mathcal{G}$  если:

$$\mathcal{L}(\mathbf{X}, \mathbf{A}) = \text{tr}(\mathbf{X}^T \mathbf{L} \mathbf{X}) = \sum_{i=1}^E \sum_{j=1}^E a_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 < \varepsilon. \quad (3)$$

Данное неравенство требует, чтобы вершины со строго различающимися значениями имели малый вес соединяющего ребра  $a_{ij}$ . Если же вершины имеют схожие значения вес может быть большим. Таким образом, описанное условие вынуждает структуру графа  $\mathcal{G}$  отображать попарные сходства сигнала  $\mathbf{X}$ .

**МАР оценка матрицы смежности.**

Имея наблюдаемый сигнал  $\mathbf{X}$  требуется найти оптимальную матрицу смежности  $\hat{\mathbf{A}}$ :

$$\tilde{\mathbf{A}}_{MAP}(\mathbf{X}) = \arg \max_{\hat{\mathbf{A}}} f(\mathbf{X}|\hat{\mathbf{A}})g(\hat{\mathbf{A}}), \quad (4)$$

где  $f(\mathbf{X}|\hat{\mathbf{A}})$  - функция правдоподобия,  $g(\hat{\mathbf{A}})$  - априорное вероятностное распределение.

**Априорное распределение.**

Определяется из априорных знаний о  $\hat{\mathbf{A}}$ : 1) ограничение разреженности  $g_s(\hat{\mathbf{A}})$ , 2) ограничение корректно определенной матрицы смежности  $g_p(\hat{\mathbf{A}})$ .

Ввиду независимости ограничений имеем:

$$g(\hat{\mathbf{A}}) = g_s(\hat{\mathbf{A}})g_p(\hat{\mathbf{A}}). \quad (5)$$

**Ограничение разреженности.** Исходит из эмпирических наблюдений о том, что отношения попарного сходства часто разрежены. Кроме этого данное ограничение предотвращает переобучение модели из-за выявления сильной связи между вершинами:

$$g_s(\hat{\mathbf{A}}) = \exp(-\lambda_1 \|\hat{\mathbf{A}}\|_1). \quad (6)$$

**Ограничение корректной определенности.** Состоит из следующих свойств, которым должна удовлетворять подходящая матрица смежности:

- Симметричность. Рассматривается неориентированный граф, в котором любое ребро  $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{E}$  идентично  $(\mathbf{x}_j, \mathbf{x}_i) \in \mathcal{E}$ . Следовательно, матрица смежности является симметричной:  $\hat{\mathbf{A}}^T = \hat{\mathbf{A}}$ ;
- Нормализованность. Данное условие рассматривается для устранения вычислительной неустойчивости и проблем со взрывом и затуханием градиентов при обучении модели:  $\hat{\mathbf{A}}\mathbf{1} = \mathbf{1}$ ;
- Отсутствие петель. Специфика рассматриваемой задачи не требует рассмотрения графовой структуры с петлями:  $tr(\hat{\mathbf{A}}) = 0$ .

Соединяя рассмотренные свойства получаем априорное распределение:

$$g_v(\hat{\mathbf{A}}) = \exp(-\lambda_2 \|\hat{\mathbf{A}}^T - \hat{\mathbf{A}}\|_F^2 - \lambda_3 \|\hat{\mathbf{A}}\mathbf{1} - \mathbf{1}\|_F^2 - \lambda_4 |tr(tr(\hat{\mathbf{A}}))|^2). \quad (7)$$

### Функция правдоподобия.

В качестве правдоподобия предлагается рассмотреть функцию гладкости сигнала по отношению к графу. Минимизация такой функции способствует построению подходящей топологии графа, задаваемой  $\hat{\mathbf{A}}$  к структурным предположениям о данных.

$$f(\mathbf{X}|\hat{\mathbf{A}}) = \exp(-\lambda_0 tr(\mathbf{X}^T \mathbf{L} \mathbf{X})) = \exp(-\lambda_0 tr(\mathbf{X}^T (\mathbf{I} - \hat{\mathbf{A}}) \mathbf{X})) \quad (8)$$

### Окончательная оценка.

Объединяя (4) - (8) искомая оценка получается решением задачи оптимизации:

$$\max_{\hat{\mathbf{A}}} \exp(-\lambda_0 tr(\mathbf{X}^T (\mathbf{I} - \hat{\mathbf{A}}) \mathbf{X})) \cdot \exp(-\lambda_1 \|\hat{\mathbf{A}}\|_1) \cdot \exp(-\lambda_2 \|\hat{\mathbf{A}}^T - \hat{\mathbf{A}}\|_F^2 - \lambda_3 \|\hat{\mathbf{A}}\mathbf{1} - \mathbf{1}\|_F^2 - \lambda_4 |tr(tr(\hat{\mathbf{A}}))|^2) \quad (9)$$

Логарифмированием и домножением на  $-1$  получаем:

$$\min_{\hat{\mathbf{A}}} \lambda_0 tr(\mathbf{X}^T (\mathbf{I} - \hat{\mathbf{A}}) \mathbf{X}) + \lambda_1 \|\hat{\mathbf{A}}\|_1 + \lambda_2 \|\hat{\mathbf{A}}^T - \hat{\mathbf{A}}\|_F^2 + \lambda_3 \|\hat{\mathbf{A}}\mathbf{1} - \mathbf{1}\|_F^2 + \lambda_4 |tr(tr(\hat{\mathbf{A}}))|^2 \quad (10)$$

Итого, найденная оценка послужит дополнительным слагаемым к функции ошибки обучаемой модели, благодаря которому при обратном распространении градиентов будет построена оптимальная структура графа по отношению к входящему графовому сигналу:

$$\begin{aligned} \mathcal{L}_{gl} &= \mathcal{L}_{smoothness} + \mathcal{L}_{sparsity} + \mathcal{L}_{validity}, \\ \mathcal{L}_{smoothness} &= \lambda_0 tr(\mathbf{X}^T (\mathbf{I} - \hat{\mathbf{A}}) \mathbf{X}), \quad \mathcal{L}_{sparsity} = \lambda_1 \|\hat{\mathbf{A}}\|_1, \\ \mathcal{L}_{validity} &= \lambda_2 \|\hat{\mathbf{A}}^T - \hat{\mathbf{A}}\|_F^2 + \lambda_3 \|\hat{\mathbf{A}}\mathbf{1} - \mathbf{1}\|_F^2 + \lambda_4 |tr(tr(\hat{\mathbf{A}}))|^2. \end{aligned}$$

## 5 Graph RNN

Пусть  $\{\mathbf{x}_t\}_{t \in T}$  - последовательность  $N$ -мерных векторов  $\mathbf{x}_t \in \mathbb{R}^N$ . Рекуррентная нейронная сеть отображает извлекаемую последовательную информацию о данных в скрытом состоянии через нелинейное отображение входных данных и текущего скрытого состояния в новое скрытое состояние:

$$\mathbf{z}_t = \sigma(\mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{z}_{t-1}),$$

$\mathbf{A} \in \mathbb{R}^{N \times N}$ ,  $\mathbf{B} \in \mathbb{R}^{N \times N}$  - линейные операторы,  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  - нелинейная функция, применяемая поэлементно.

Пусть  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$  - граф, где  $\mathcal{V} = \{1, \dots, N\}$  - вершины,  $\mathcal{E} \subseteq N \times N$  - ребра,  $\mathcal{W} : \mathcal{E} \rightarrow \mathbb{R}$  - весовая функция ребер. Пусть  $\mathbf{x}$  является графовым сигналом, то есть его компоненты связаны графовой структурой. Для определения связи между графовым сигналом  $\mathbf{x}$  и графом  $\mathcal{G}$  задается оператор графового сдвига  $\mathbf{S} \in \mathbb{R}^{N \times N}$  - матрица, значения которой отражают влияние компонент сигнала друг на друга, для которой также выполнено  $s_{ij} \neq 0$  если  $i = j$  или  $(i, j) \in \mathcal{E}$ . Примером таких матриц являются матрица смежности, граф-Лапласиан, матрица случайных блужданий.

Обозначим через  $\mathbf{S}\mathbf{x}$  линейное локальное отображение графового сигнала:

$$[\mathbf{S}\mathbf{x}]_i = \sum_{j=1}^N [\mathbf{S}]_{ij} [\mathbf{x}]_j = \sum_{j \in \mathcal{N}_i} s_{ij} x_j,$$

где  $\mathcal{N}_i = \{j \in \mathcal{N} : (i, j) \in \mathcal{E}\}$  - множество непосредственных соседей  $i$ -й вершины.

Используем понятие графового сдвига для определения графовой свертки через взвешенную сумму сдвинутых версий сигнала:

$$\mathbf{A}(\mathbf{S})\mathbf{x} = \sum_{k=0}^{K-1} a_k \mathbf{S}^k \mathbf{x},$$

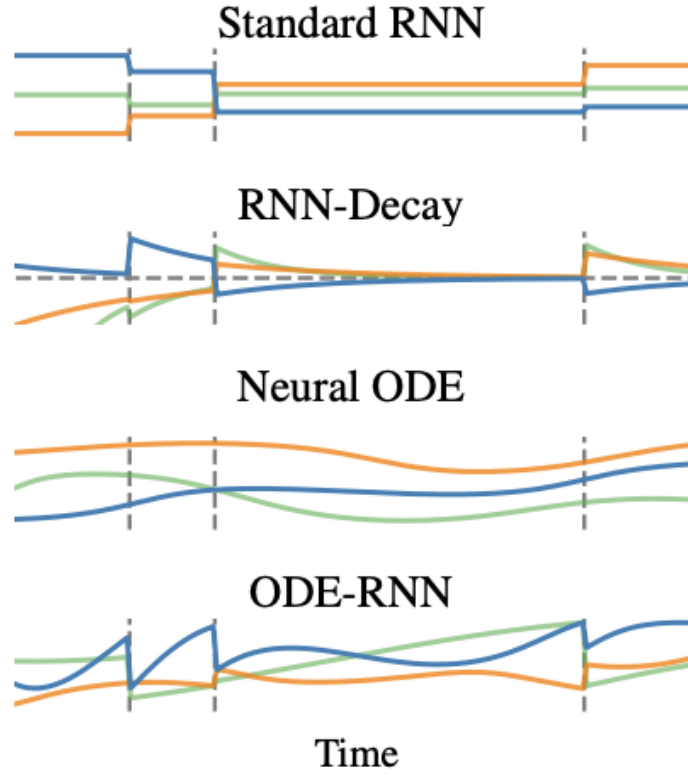
где  $\mathbf{S}^k \mathbf{x} = \mathbf{S}(\mathbf{S}^{k-1} \mathbf{x})$  - производит суммарную информацию, содержащуюся в  $k$ -hor соседстве каждой вершины. Для  $0 \leq k \leq K-1$  коэффициенты  $\mathbf{a} = [a_0, a_1, \dots, a_{K-1}] \in \mathbb{R}^K$  отражают важность информации, находящейся в  $k$ -hor соседстве.

Рассмотрим далее в качестве входных данных графовый процесс - последовательность  $\{\mathbf{x}_t\}_{t \in T}$  графовых сигналов  $\mathbf{x}_t \in \mathbb{R}^N$ , иными словами графовый процесс это меняющийся во времени графовый сигнал. Пусть скрытое состояние рекуррентной нейронной сети также является графовым сигналом, тогда обновление скрытого состояния графовой рекуррентной нейронной сети определяется как:

$$\mathbf{z}_t = \sigma(\mathbf{A}(\mathbf{S})\mathbf{x}_t + \mathbf{B}(\mathbf{S})\mathbf{z}_{t-1}).$$

## 6 Graph ODE RNN






---

**Algorithm 1** The ODE-RNN. The only difference, highlighted in **blue**, from standard RNNs is that the pre-activations  $h'$  evolve according to an ODE between observations, instead of being fixed.

---

**Input:** Data points and their timestamps  $\{(x_i, t_i)\}_{i=1..N}$

$h_0 = \mathbf{0}$

**for**  $i$  in  $1, 2, \dots, N$  **do**

$h'_i = \text{ODESolve}(f_\theta, h_{i-1}, (t_{i-1}, t_i))$

▷ Solve ODE to get state at  $t_i$

$h_i = \text{RNNCell}(h'_i, x_i)$

▷ Update hidden state given current observation  $x_i$

**end for**

$o_i = \text{OutputNN}(h_i)$  for all  $i = 1..N$

**Return:**  $\{o_i\}_{i=1..N}; h_N$

---