

ТОПОЛОГИЯ ЯЗЫКА И СИГНАЛОВ МОЗГА

Кириллова Агата Андреевна

Студент

Факультет ВШП МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: kir_agata@mail.ru

Научный руководитель — Костикова О. И., Бернадотт А. К.

В настоящей работе проводится анализ топологических свойств крупномасштабной структуры естественного русского и английского языков и электромагнитных сигналов мозга (магнитоэнцефалография – МЭГ и электроэнцефалография – ЭЭГ) на предмет выявления устойчивых гомотопий первого порядка в семантическом пространстве языка и пространстве временных рядов оцифрованных сигналов мозга и на предмет построения гомеоморфизма между топологией языка и топологией сигналов мозга.

Исследование когомологий первого порядка языковых фрактальных структур проводилось для $n = 1$ (слова); $n = 2$ (биграммы); $n = 3$ (триграммы) на материале русского и английского языков. Все вычислительные эксперименты проводились над корпусами национальной литературы русского и английского языков. При выборе языков для сравнительного анализа мы руководствовались следующими соображениями: (1) корпуса каждого из языков насчитывают значительное количество; (2) языки существенно отличаются друг от друга. Сигналы мозга во время чтения слов, биграмм и триграмм записывались на аппарате с регистрацией МЭГ и ЭЭГ (на нейроинтерфейсе Гермес от ООО “Ребис” [1]).

Для заданного множества текстов естественного языка $\mathfrak{Z} = (\Omega_1, \dots, \Omega_N)$ строили множество всех d -мерных эмбедингов n -грамм рассматриваемого естественного языка $\mathfrak{R}_n(d)$, $n \in \{1, N\}$, $d \in \{1, D\}$.

Для данного n по множествам $\mathfrak{R}_n(d)$, $n \in \{1, N\}$, $d \in \{1, D\}$ строили множество персистентных гомотопий первого порядка $\{v_j\} = f(\{\mathfrak{R}_n(d)\})$ и определяли их контура; здесь предполагается, что диаметры $d_j = \text{diam} v_j > d_{\max}$, где $d_{\max}$ – максимальное расстояние внутри групп синонимов n -грамм в данном языке. Все найденные таким образом пустоты в семантическом пространстве являются элементами крупномасштабной топологической структуры языка. Для построения элементов семантического пространства (эмбедингов n -грамм) мы использовали модель CBOW и модель Громова [2][3].

Для n -грамм ввели следующую функцию расстояния – усреднён-

ное минимальное косинусное расстояние между парами слов из двух n -грамм:

$$p((a_1, \dots, a_n), (b_1, \dots, b_n)) = \frac{1}{n} \sum_{i=1}^n \min \cos(a_i, b_i) \quad (1)$$

Для поиска в семантическом пространстве исследуемого языка и сигналов мозга производился поиск гомотопий первого порядка (пустоты в семантическом пространстве); гомотопии нулевого порядка – компоненты связности.

Опираясь на модель сигналов мозга Бернадотт о представлении сигнала мозга набором затухающих осцилляторов:

$$g(t) = \sum_{i=1}^k c_i e^{\lambda_i t} \sin(\omega_i t + \varphi_i), \quad (2)$$

где $i \in \overline{1, k}$, $c_1, c_2, \dots, c_k, \lambda_1, \lambda_2, \dots, \lambda_k$ – константы, $\lambda_i \neq \lambda_j$ и лемму о размерности пространства данного сигнала вычисляли размерность пространства сигналов мозга [4], представленных развёртками временных рядов Бухштабера [5].

В результате исследования было обнаружено наличие общих персистентных гомотопий в языке и сигналах мозга, что говорит о возможности построения гомеоморфизма между семантическим пространством n -грамм языка и сигналами мозга носителя данного языка. В результате исследования было подтверждено утверждение Бернадотт-Громова об общей размерности языка и сигналов мозга.

Литература

1. Страница производителя ООО “Ребис”: <https://rebis.site/>
2. Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space // arXiv preprint arXiv:1301.3781, 2013.
3. Gromov V. A., Borodin N. S., Yerbolova A. S. A language and its dimensions: intrinsic dimensions of language fractal structures // Complexity, 8863360, Hindawi, 2024.
4. Bernadotte A., Buchstaber V.. Method for Evaluating the Number of Signal Sources and Application to Non-invasive Brain-computer Interface // arXiv preprint arXiv:2410.11844, 2024.
5. Buchstaber V. M. Time Series Analysis and Grassmannians // Applied problems of Radon transform, Amer. Math. Soc. Transl. 1994. Vol 2 (162), P. 1–17.