

Н. А. Щуцкий
(УО «БГУИР», Минск)

МАТЕМАТИКО-СИНТАКСИЧЕСКАЯ ТЕОРИЯ КАК СРЕДСТВО ОПТИМИЗАЦИИ АЛГОРИТМОВ РАБОТЫ С ЕСТЕСТВЕННЫМИ ЯЗЫКАМИ

В результате исследования работы рекуррентных нейронных сетей и цепей Маркова было установлено, что первые справляются с задачей генерации текста значительно лучше, однако затрачивают огромное количество ресурсов вычислительных систем [1]. Данная теория позволяет минимизировать затраты и повысить точность.

Теория постулирует, что лексика языка может быть представлена как группа, для которой определена бинарная не коммутативная операция склеивания [2]. Склеиванием является объединение двух лексических аргументов в один смысловой образ, который, может не иметь словесного представления в натуральном языке. Синтаксисом языка является множество всех допустимых операций склеивания этого языка [3].

Оптимизация формирования синтаксиса естественного языка на нейронных сетях происходит в 3 этапа обучения:

- 1) Заменяем лексические аргументы обучающей выборки на 4 части речи абстрактного языка.
- 2) На основе данной выборки обучаем нейронную сеть, дополняя стартовый абстрактный синтаксис.
- 3) Заменяем абстрактные части речи на естественные, проводим частичное переобучение.
- 4) Проводим обучение, заменяя естественные части речи на слова естественного языка.

В итоге на основе модели синтаксиса некоторого языка могут быть построены синтаксически корректные и связанные тексты, однако их смысловая наполненности всё ещё будет целиком зависеть от некоторой обучающей выборки из слов естественного языка.

Литература

- 1 Goodfellow, I. Deep learning / Ian Goodfellow, Yoshua Bengio, Aaron Courville. – Cambridge, MA: MIT press, 2017. – 373 с.
- 2 Gallian, J. A. Contemporary abstract algebra / Joseph A. Gallian. – Boston, MA; Brooks Cole, 2016. – 49 с.
- 3 Kornai, A. Mathematical linguistics / Andras Kornai. – Cambridge, MA: Springer, 2008. – 201 с.