

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

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Щуцкий Н.А.

Борисенко О.Ф. - кандидат физ.-матем.наук, доцент

В данной научной работе были исследованы процессы и результаты, относящиеся к генерации текста посредством цепей Маркова первого порядка и рекуррентных нейронных сетей. Марковские цепи позволяют установить статистическую закономерность связи тех или иных слов, в результате имея алгоритм, способный к использованию только уже бывших в статистике паттернов. В то же время нейронная сеть на основе статистики позволяет установить закономерности построения предложений, то есть частично восстановить синтаксические правила языка, что позволяет ей работать значительно эффективнее на новых данных. На основе данного исследования было решено заняться модификацией данного процесса, что в последствии привело к появлению математико-синтаксической теории.

Основные постулаты, справедливые в контексте данной теории таковы:

1. Слово - некоторым образом закодированная информация, представимая в виде набора чисел.
2. Лексика - группа лексических аргументов(слов), для которой справедливы аксиомы теории групп.
3. Синтаксис языка – множество разрешённых в данном языке операций склеивания.
4. При склеивании нескольких лексических аргументов происходит объединение в один лексический аргумент, не представимый словом, при это порядок аргумента повышается на единицу.
5. Порядок аргумента – число операций склеивания, уже произведённых над данным аргументом.

Для простоты применения данной теории введём абстрактный синтаксис, включающий 4 части речи:

1. Объекты – лексические аргументы, являющиеся некоторыми сущностями или телами.
2. Действие – лексические аргументы, являющиеся взаимодействиями объектов.
3. Свойства – лексические аргументы, описывающие параметры объектов или действий.
4. Конструкции – лексические аргументы, связывающие между собой все остальные части речи.

Множество допустимых операций склеивания может быть определено рядом правил:

1. Аргументы 0-ого порядка одной части речи склеиваются в аргумент той же части речи.
2. Аргументы 1-ого порядка склеиваются в объект или действие 2-ого порядка.
3. Аргументы 2-ого порядка склеиваются в аргумент 3-его порядка – простое предложение.
4. Аргументы 3-его порядка склеиваются в аргумент 4 порядка – сложное предложение.
5. Аргументы 3-его и 4-ого порядков склеиваются в аргумент 5 порядка – текст.

Оптимизация формирования синтаксиса 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 конкретного языка могут быть построены синтаксически корректные и связанные тексты, однако их смысловая наполненности всё ещё будет целиком зависеть от некоторой обучающей выборки из конкретных слов, на основе которой будут определены возможные связи. Таким образом задача оптимизации сводится к формированию полного набора операций синтаксиса языка за наименьшее время, что происходит в результате действий, обратных аппроксимации, по отношению к результатам каждой из предыдущих этапов обучения, начиная с теоретически воссозданного абстрактного стартового синтаксиса, что в последствии уменьшает число требуемых вычислений до регулируемого уровня.

Список использованной литературы:

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

ФРАКТАЛЫ В ПРИРОДЕ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Смирнов А.Д., Даукшис В.В.

Новик Ю.Ф. – магистр физико-математических наук

Человек с самого начала своего существования наблюдал за уникальными природными явлениями, такими как молния, облака, горы, снежинки, морозные узоры на окне. У него возникали вопросы, как это устроено, ведь все это выходит за пределы евклидовой геометрии. Невозможно описать камень или облако с помощью треугольников, кружков и