

Замечено, что не только сущности, традиционно называемые искусственными и естественными языками, обладают структурой, для описания и изучения которой можно использовать инструменты теории формальных языков. В частности, структуры биологических цепочек, цепочек системных вызовов, химические соединения изучаются с применением конечных автоматов, n-грамм, грамматик. При этом важной особенностью такого рода объектов является то, что в последовательностях содержится достаточно большое количество “шумов” (например, мутации в биологических последовательностях). Это затрудняет построение точных моделей структур таких цепочек с использованием грамматик, а значит требуется привлечение вероятностных подходов. Таким образом, целью ВКР Полины Луниной является разработка метода для изучения структур цепочек с использованием формальных грамматик для описания структуры и искусственных нейронных сетей (ИНС) для построения вероятностной модели. Кроме разработки метода требовалось провести его апробацию, решив прикладную задачу.

Поставленные задачи выполнены на высоком уровне. Разработанный метод основан на следующих идеях. Во-первых, можно не пытаться описать с помощью грамматики структуру всей цепочки, а использовать грамматику для описания характерных особенностей, а синтаксический анализ для поиска этих особенностей. Во-вторых, представление результатов синтаксического анализа возможно в нескольких форматах, что даёт возможность использовать различные типы ИНС и выбирать наиболее подходящие для конкретной задачи. Стоит отметить, что разработанный метод слабо зависит от класса грамматик и алгоритмов синтаксического анализа. Для апробации предложенного метода была реализована система, использующая матричный алгоритм синтаксического анализа и пакеты Keras и TensorFlow для конструирования и обучения нейронных сетей. В качестве задачи была выбрана задача классификации транспортных РНК. Для её решения была разработана контекстно-свободная грамматика, описывающая шпильки и их композиции — базовые элементы вторичной структуры. Также было исследовано применение свёрточных и полносвязных ИНС для решения поставленной задачи.

В ходе работы Полина Лунина показала себя как хороший исследователь и инженер, продемонстрировала хорошие навыки постановки и проведения экспериментов, высокий уровень инженерной культуры.

Текст работы в достаточной мере раскрывает основные аспекты работы, однако некоторые моменты могли бы быть описаны более подробно. Изложение материала аккуратное и последовательное.

Результаты работы представлены на конференции BIOINFORMATICS-2019 и опубликованы в сборнике материалов конференции, и на конференции VIATA в виде постера. Несмотря на указанные недочёты, считаю, что работа Полины Луниной заслуживает оценки «отлично».