

*Серда А.А. — рецензент Киселев Г.Д.
УНК “ИПСА” НТУУ “КПИ”, Киев, Украина*

Применение нейронных сетей на разных этапах распознавания естественной речи

В настоящее время актуальной проблемой является распознавание естественной речи. Множество современных методов, используемых для ее решения, для своей работы требуют больших вычислительных ресурсов, объем которых часто бывает ограничен. В последнее время для некоторых этапов распознавания все чаще используют методы, основанные на нейронных сетях. Этот переход осуществляется благодаря тому, что нейронные сети просты в использовании, а также имеют достаточно высокую скорость работы.

Распознавание речи является многоступенчатым процессом, на разных этапах которого используют множество подходов и алгоритмов. Доклад включает описание уже существующих алгоритмов, а также их модификации, построенных на нейросетях.

В работе рассматривается подход к распознаванию, при котором сам процесс делится на пять основных этапов [1]. Рассмотрим их.

Первый этап – разбиение входного потока данных на отдельные слова. Процесс усложняется тем, что при произнесении любой фразы носителем языка, она звучит практически слитно, четкого разделения на слова не слышно. Тем не менее, данная задача довольно точно решается. Хорошие результаты показывают алгоритмы на нейронных сетях. В результате обработки входного потока данных, на выходе получаем звуковые отрезки, которые, предположительно, образуют одно слово.

Второй этап – разбиение слова на отдельные звуки. Данный этап является достаточно простым, так как длительность звуков при произнесении слова практически одинакова, и выделить цепочку звуков достаточно просто. К этому этапу можно также отнести и дальнейший процесс распознавания выделенных звуков. Для этого как нельзя лучше подходят нейронные сети, которые уже показали свою эффективность в распознавании образов из небольшого числа шаблонов. На выходе получаем возможные варианты звуков для каждой отдельной позиции.

На третьем этапе из полученных вариантов звуков строятся последовательности, которые просеиваются с помощью применения скрытых марковских моделей (СММ) [2]. Таким образом, для каждой полученной цепочки получаем вероятность её появления.

Четвертым этапом является получение буквенных последовательностей из цепочек звуков, их просеивание с помощью СММ и дальнейшая проверка по словарю.

Последний – пятый этап – предполагает выделение предложений и получение их интонационного окраса. Нейронные сети и с этой задачей справляются довольно хорошо, если не учитывать сложные предложения. На выходе данного этапа получаем уже готовый текст, который удалось выделить из входных данных.

Выводы. В результате применения предложенных алгоритмов на нейросетях, предполагается получить систему, которая будет давать приемлемые результаты работы и, при этом, потреблять небольшое количество ресурсов.

Литература

1. Ямов С.И., Кабак И.С., Курочкин С.Н., Бродин А.Г. Многоуровневая система распознавания речи. <http://websound.ru/articles/theory/speech.htm>.
2. Разработка интеллектуальных систем. <http://www.super-profi.com/>.