

ГЛУБОКИЕ НЕЙРОННЫЕ СЕТИ В ЗАДАЧАХ РАСПОЗНАВАНИЯ РЕЧИ

Авторы: д.ф.ф.м.н. **Онаркулов Максад Каримбердиевич**
(Старший преподаватель кафедры «Прикладная математика и информатика»
Ферганского Государственного Университета).

Юсупов Мирсаид Абдулазиз угли
(Магистрант 2 курса направления «Прикладная математика» Ферганского
Государственного Университета).

Умиржонов Лазизжон Акмалжон угли
(Магистрант 2 курса направления «Прикладная математика» Ферганского
Государственного Университета).

Ключевые слова: Искусственные нейронные сети, глубокие нейронные сети, распознавание речи, гибридные алгоритмы, ИНС/СММ.

Искусственные нейронный сети с временными задержками. В данной статье рассмотрены использование глубоких нейронных сетей в решении практической задачи распознавания речи. Для акустического моделирования также применяются искусственные нейронный сети с временными задержками, которые представляет собой многослойную нейронную сеть прямого распространения, узлы которой модифицированы введением временных задержек. Пример узла с N задержками показан на рисунке 1.

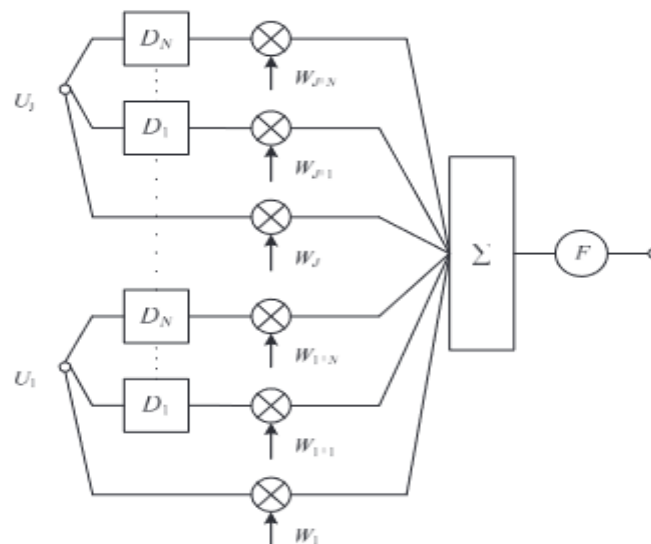


Рисунок 1. Искусственные нейронный сети с временными задержками.

Таким образом, в искусственные нейронный сети встраивается кратковременная память. Введение временной задержки позволяет сделать искусственные нейронный сети инвариантной к временным сдвигам. Использование искусственные нейронный сети с временными задержками позволило получить относительное уменьшение ошибки распознавания слов на 2,6 %. Еще одним типом нейронных сетей являются

рекуррентные искусственные нейронные сети. Наличие обратной связи наделяет искусственные нейронные сети памятью, благодаря чему появляется возможность моделировать динамические процессы. Одним из типов рекуррентных искусственных нейронных сетей, применяемых для акустического моделирования, является сеть LSTM, содержащая специальные элементы, называемые блоками памяти. Блоки памяти содержат ячейки, которые хранят временное состояние сети, а также мультипликативные элементы, называемые гейтами (gates), управляющие потоком информации. Каждый блок памяти содержит входной и выходной гейты, а также гейт забывания.

Гейты определяют, когда входная информация существенна и ее необходимо запомнить, когда следует продолжать запоминать или забыть информацию и когда следует информацию подать на выход. Применение LSTM в гибридной ИНС/СММ модели позволяет снизить ошибку распознавания слов по сравнению с применением глубоких искусственных нейронных сетей.

В последнее время возникают попытки построения так называемых end-to-end систем преобразования речи в текст, использующих только искусственные нейронные сети, без обучения СММ моделей. End-to-end системы состоят из двух подмодулей: кодировщик и декодер.

Кодировщик читает входной сигнал, вычисляет признаки сигнала и преобразует его в промежуточное параметрическое представление.

Декодер преобразует параметрическое представление сигнала в последовательность символов. End-to-end система была построена на базе сверточной нейронной сети и метода нейросетевой темпоральной классификации (Connectionist Temporal Classification; CTC). Разработанный подход тестировался для задачи распознавания фонем, при этом ошибка распознавания фонем составила 18,2%. Сеть LSTM применялась для построения end-to-end системы. Без использования лингвистической информации ошибка распознавания слов составила 27,3%, применение словаря позволило снизить ошибку до 21,9%, с триграммной моделью языка ошибка распознавания слов составила 8,2%. Для распознавания речи был применен метод, использующий преобразователи на основе конечных автоматов. Было показано, что предложенный метод позволяет повысить точность распознавания речи по сравнению со скрытыми марковскими моделями.

Еще одно исследование применения глубоких искусственных нейронных сетей в системе распознавания русской речи представлено, где был предложен метод адаптации к голосу диктора с использованием гибридных контекстно-зависимых ИНС/СММ моделей, основанный на использовании в качестве входных данных для ИНС признаков, вычисленных с помощью модели на гауссовских смесях. Было получено относительное сокращение ошибки распознавания слов на 5-36% в зависимости от адаптационного набора речевых

данных. В нейросетевое моделирование русской речи с использованием графического процессора предложены два подхода к распознаванию спонтанной русской речи: адаптация глубокой нейронной сети с применением метода *i*-векторов идикторозависимые признаки, полученные на слое «узкого горла». Эти методы позволили получить относительное уменьшение количества неправильно распознанных слов на 8,6% и 11,9% соответственно.

Возможность создавать нейросетевые акустические модели есть в наиболее распространенных пакетах современных программных средств для разработки систем автоматического распознавания речи, таких как Kaldi, RWTH ASR (RASR), НТК v3.5.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ:

1. Arun Das, Paul Rad Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey . arXiv:2006.11371, 2020
2. Katharina Weitz, Dominik Schiller, Ruben Schlagowski, Tobias Huber, Elisabeth Andre “Let me explain!”: exploring the potential of virtual agents in explainable AI interaction design . Journal on Multimodal User Interfaces volume 15, pages 87–98 (2021)
3. Alejandro Barredo Arrieta, Natalia Diaz-Rodriguez и др.Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI . arXiv:1910.10045, 2019
4. Абдулазиз угли, Ю. М., Каримбердиевич, О. М., & Махамедин угли, Ё. А. (2022). АЛГОРИТМЫ РАСПОЗНОВАНИЯ РЕЧИ И КЛАССИФИКАЦИЯ МЕТОДОВ РАСПОЗНОВАНИЯ РЕЧИ. CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES, 3(10), 15-19. Retrieved from <https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS/article/view/240>MORE CITATION FORMATS
5. Livingstone S.R., Russo F.A. Surrey Audio-Visual Expressed Emotion (SAVEE) Database. <http://kahlan.eps.surrey.ac.uk/savee/>
6. Kate Dupuis, M. Kathleen Pichora-Fuller Toronto emotional speech set (TESS). University of Toronto, Psychology Department, 2010.
7. 14. Каримов Ш.Т., Хайдарова С.А. Численное решение периодических уравнений с дробно-интегральным оператором вейля в главной части.//Fars Int J Soc Sci Hum 10(12);2022. Publishing centre of Finland. С.152-157.
8. 15. Фармонов Ш., Хайдарова С. Обобщенный метод Бубнова-Галеркина для уравнений с дробно-дифференциальным оператором // Norwegian Journal of Development of the International Science. 2022. №99.С.10-15.