DOI 10.37882/2223-2966.2023.12.36

АНАЛИЗ СУЩЕСТВУЮЩИХ МЕТОДОВ И МОДЕЛЕЙ РАСПОЗНАВАНИЯ РЕЧИ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

ANALYSIS OF EXISTING METHODS AND MODELS OF SPEECH RECOGNITION BASED ON NEURAL NETWORKS

Fu Wenwei

Summary. The purpose of this work is to analyze existing methods and models of speech recognition based on neural networks. The features and characteristics of the most effective neural network models used for speech recognition are studied. The positive and negative sides of these models are highlighted. The models were compared and the most promising ones were identified. In conclusion, the paper notes the high prospects of using neural networks and, in particular, the convolution model of a neural network for speech recognition.

Keywords: speech, model, neural network, CNN neural network, LSTM neural network, Kohonen neural network.

Фу Вэньвэй
Аспирант,
Санкт-Петербургский государственный университет
312833703@qq.com

Аннотация. Цель данной работы заключается в проведении анализа существующих методов и моделей распознавания речи на основе нейронных сетей. Изучены особенности и характерные черты наиболее эффективных моделей нейронных сетей, применяемых с целью распознавания речи. Выделены положительные и отрицательные стороны данных моделей. Проведено сравнение моделей и выделены наиболее перспективные из них. В заключение работы отмечается высокая перспективность применения нейронных сетей и, в частности, конволюционной модели нейронной сети для распознавания речи.

Ключевые слова: речь, модель, нейронная сеть, CNN нейронная сеть, LSTM нейронная сеть, нейронная сеть Кохонена.

Введение

Современная научная революция и продолжающееся технологическое развитие оказывают существенное влияние на все современные сферы жизни мировых государств. Разработка естественных для населения технологий работы с компьютерными системами выступает одной из ключевых задач современного научного общества.

Одной из важнейших является технология, связанная с речевым вводом данных, который представляется наиболее приятным и удобным методом для каждого пользователя. Первоначально разработка методик распознавания речи начиналась с освоения способов выделения информативных данных, которыми можно было бы описать изучаемый звук. После решения данного вопроса внимание было обращено к нахождению решения проблемного вопроса, связанного с осуществлением классификации полученных сигналов с помощью системы информативных признаков [1].

В настоящее время высокой степенью актуальности обладают задачи, связанные с созданием технологий, применяемых для распознавания речи, на основе искусственного интеллекта. Использование речевых интерфейсов представляется наиболее удобным для того, чтобы осуществлять управление компьютером и любыми автоматизированными комплексами, нежели чем применять для этих целей стандартизированные графические интерфейсы. С помощью речи конечный пользователь может одновременно решать несколько задач, которые никоим образом не связаны с устройствами, применяемыми для ввода данных в компьютерную систему, потому что его руки будут полностью свободны и могут быть использованы для выполнения других действий. Помимо этого, автоматизированные системы, предназначенные для распознавания речи, можно с успехом применять для проведения автоматизированного стенографирования, перевода текста, автоматизированных справочных центрах и т.п.

Конечно, в настоящее время разработано достаточно большое число готовых решений на основе нейронных сетей, которые используются для распознавания речи, однако каждое из них обладает собственным набором недостатков. В частности, таковыми могут быть низкая точность распознавания речи и высокая степень зависимости работы нейронной сети от доступа к иным системам. Все это делает актуальным задачу анализа существующих методов и моделей распознавания речи на основе нейронных сетей, который, несомненно, поможет выбрать наиболее подходящий вариант для пользователя или же послужит базой для дальнейшего развития и совершенствования подобных систем.

Материалы и методы

В процессе проведения исследования использовались такие методы, как анализ и синтез имеющегося ма-

териала, обобщение имеющихся сведений по проблемной тематике, индукция.

Результаты

В процессе распознавания голоса нейронные сети чаще всего применяются на второй стадии расчетов, когда производится вычисление так называемых локальных метрик [2]. Для статистических распознавателей, осуществляющих процесс непрерывного наблюдения, подобного рода метрики выступают в качестве монотонных функций правдоподобия векторных признаков. Используемые речевые распознаватели, имеющие дискретное наблюдение, первоначально осуществляют процесс квантования векторов и определяют для каждого из них уникальные символы из имеющейся таблицы кодов. После этого, на основе данных символов с применением специализированных таблиц, в которых содержится вероятности появления символов для всех уникальных векторов, происходит вычисление локальных метрик. Подобного рода вычисления можно с успехом проводить с помощью однослойных перцептронов [3], внешний вид которых показан на рисунке 1.

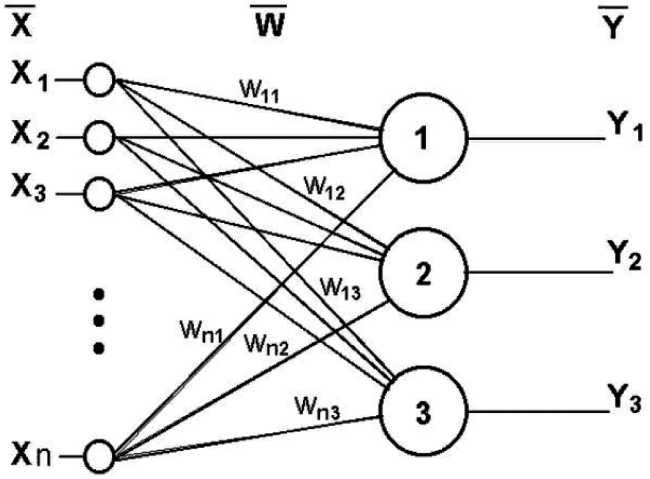

Рис. 1. Однослойный перцептрон [3]

Число узлов таких перцептронов полностью коррелирует с количеством примененных уникальных символов. Количество входов перцептрона совпадает с количеством возможных символов.

Ключевые преимущества использования подобного рода нейронной сети заключаются в простоте аппаратно-программной модели, а также ее реализации, и достаточно высокой скорости обучения. Среди ключевого недостатка можно выделить возможность решения простых задач распознавания голоса, что является не совсем применимым для современных сложных решений.

Процесс квантования векторов можно реализовать с применением нейронной сети Кохонена (см. рисунок 2). Сети подобного вида организованы в виде двумерного узлового массива, в котором для каждого из вероятных символов имеется свой собственный узел. Узлы производят расчет евклидова расстояния между поступающим на вход сетевым вектором и эталонным значением, который представляется весами узла, по результатам вычисления которого производится выбор узла, который обладает минимальным евклидовым расстоянием. Вес подобной сети может определяться либо на основе алгоритма Кохонена либо любых его модификаций, либо любым из существующих стандартных алгоритмов векторного квантования, в которых метрикой выступает евклидово расстояние [4].

Среди основных преимуществ нейронных сетей Кохонена, применяемых для распознавания речи, можно выделить высокую устойчивость к данным, имеющим высокую шумовую составляющую, наличие возможности упрощения внутренней структуру и легкую обучаемость. Среди ключевых недостатков можно выделить определенность количества кластеров и эвристичность применяемого алгоритма обучения.

Применение многослойных нейросетей (см. рисунок 3) для распознавания речи также является актуальным для задач понижения размерности векторных признаков, которые препроцессор получает в на первом этапе процесса распознавания. Число выходов подобного рода нейронных сетей совпадает с количеством входов. Помимо этого, они имеют в своей структуре один или несколько слоев, в которых размещены скрытые узлы. В процессе обучения многослойной нейронной сети вес будет подбираться таким, чтобы сеть имела возможность интерпретировать на выходе все возможные варианты входного вектора с использованием слоя, в котором расположены скрытые узлы. По завершению процесса обучения выходы таких узлов можно применять как векторы входа, которые обладают меньшей размерностью, с целью проведения последующей обработки и процесса распознавания речи [5].

Среди ключевых преимуществ применения многослойных нейронных сетей можно выделить возможность потокового обучения, поддержку функции распараллеливания и отсутствие использования каждого из входных значений для достаточно больших выборок. Однако, несмотря на это, данные сети обладают определенными недостатками, ключевыми из которых являются достаточно медленная сходимость, достаточно часто возникающие застревания в местных минимумах и, в ряде случаев, отсутствие возможности переобучения сети.

За последние годы наибольший интерес представляет применение динамических нейросетевых классификаторов, которые были разработаны конкретно для задач, связанных с распознаванием речи. В них

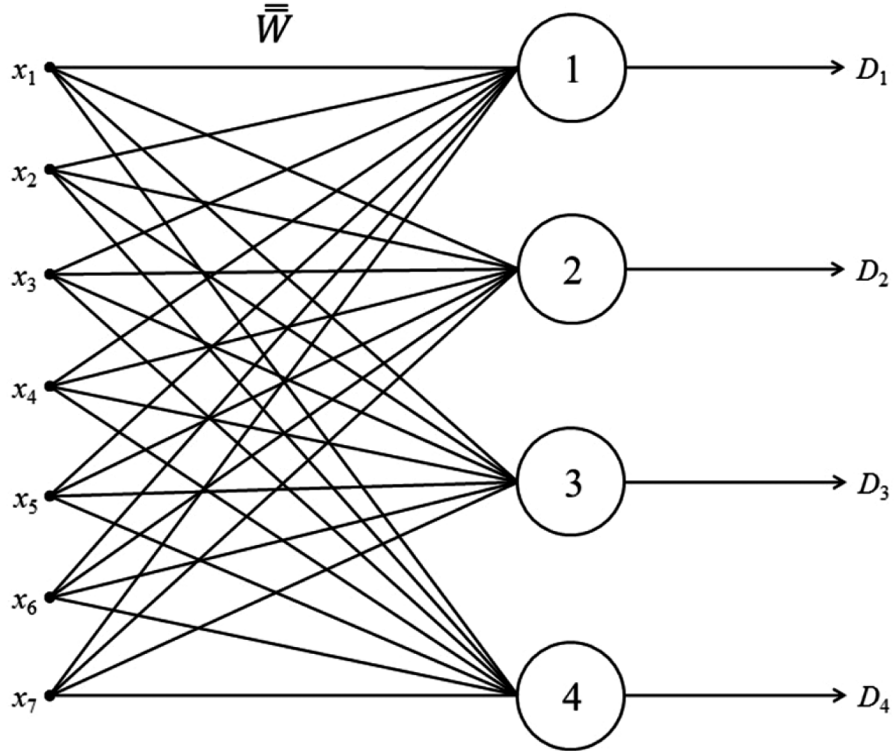

Рис. 2. Нейронная сеть Кохонена [4]

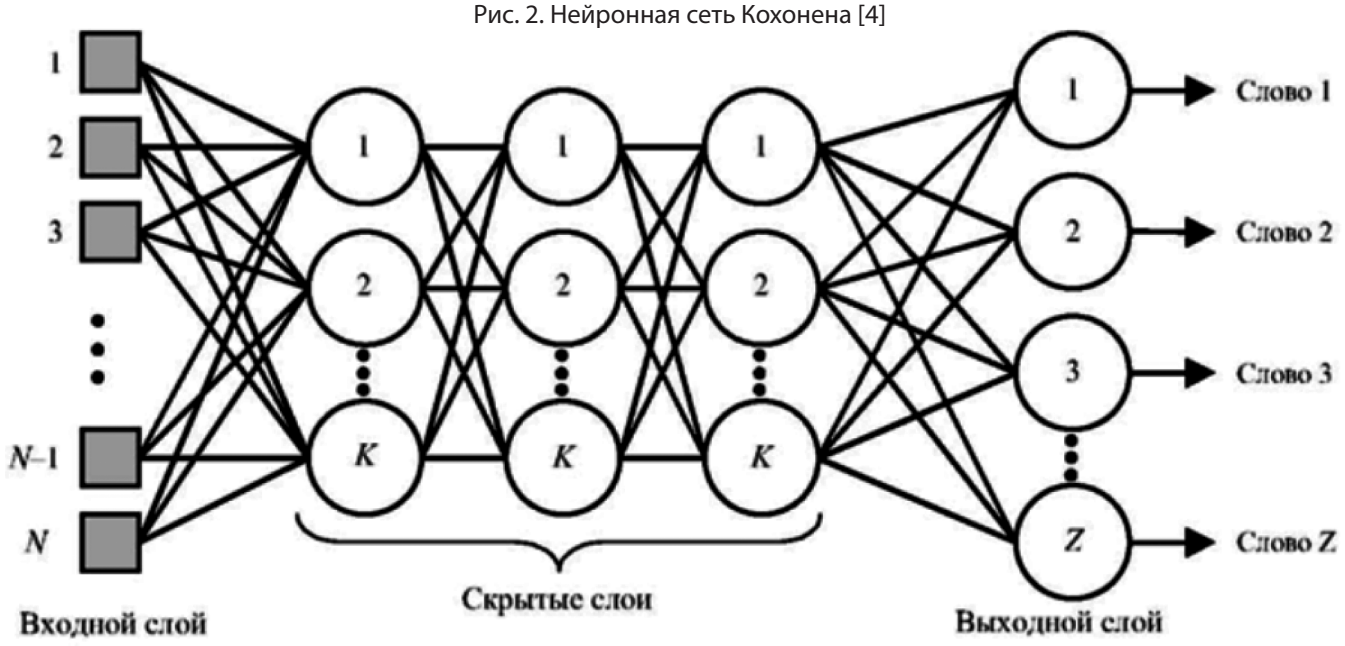

Рис. 3. Структура многослойной нейронной сети для распознавания речи [6]

входят небольшие задержки времени и узлы, которые отвечают за процесс временного интегрирования. В настоящее время они получили название рекуррентных связей. Подобного рода классификаторы являются достаточно устойчивыми к незначительным временным сдвигам, возникающим в выборках при обучении и контроле, вследствие чего для получения высокого результата их работы не нужна идеальная сегментация речевой информации. Применение динамических нейронных сетей для задач, связанных с распознаванием речи, дает возможность избавиться от недостатков, которые были приведены для вышеперечисленных нейронных сетей, что подтверждается проведёнными экспериментами [7].

Одним из примеров подобной нейронной сети, которая может использоваться для распознавания речи, является нейронная сеть с временными задержками.

По своей сути архитектура данной нейронной сети основана на многослойном перцептроне, однако каждый из его узлов сдержит временные задержки (см. рисунок 4). Данное нововведение делает данную нейронную сеть инвариантной к непродолжительным временным смещениям.

На следующем рисунке показана архитектура трехслоной нейронной сети с временными задержками, которая может применяться для распознавания не более трех различных фонем. Процесс обработки нейросетью входных векторов аудиоданных схожа с прохождением окон временных задержек над образами узлов нижнего уровня. Такая простая структура нейросети с временными задержками является наиболее подходящей для стандартной СБИС-реализации с подгружаемыми весами снаружи [8].

В последнее время большая часть работ в сфере искусственного интеллекта и машинного обучения посвящена сверточным или конволюционным нейронным сетям (CNN) [8-10], которые чаще всего используются для обработки различных изображений и иных типов информации, которую можно преобразовать в изображение. Высокая степень популярности конволюционных нейронных сетей, несмотря на их достаточно высокую сложность, дороговизну обучения и неэффективность процесса обучения без учителя, обусловлена следующими важнейшими преимуществами:

— достаточно высокая точность при решении любых задач, связанных с обработкой различных изображений;
— высокая скорость работы с большими объемами входных данных;
— наличие инвариантности к смещению, за счет чего в достаточно сильной степени облегчается процесс обучения данных нейронных сетей.

Конволюционные нейронные сети могут также применяться для распознавания речь, веди любой аудиосигнал можно разложить на спектр, который в дальнейшем может быть обработан нейросетью. В качестве примера на рисунке 6 приведен пример архитектуры сверточной нейронной сети, которая может быть использована для распознавания голоса.

К числу популярных нейронных сетей, применяемых для распознавания голоса, относится нейронная сеть с «длительной кратковременной памятью» (LSTM) [11]. Нейроны в данной сети обладают существенно наиболее сложной структурой и включают в себя блок памяти, состоящий из гейтов входа, забывания и выхода, который имеет собственные состояния вектора. Для каждого из внутренних векторов характерна одинаковая размерность, которая полностью эквивалентна количеству нейронов в нейросети. На следующем рисунке показана архитектура нейросети и одного из ее слоев.

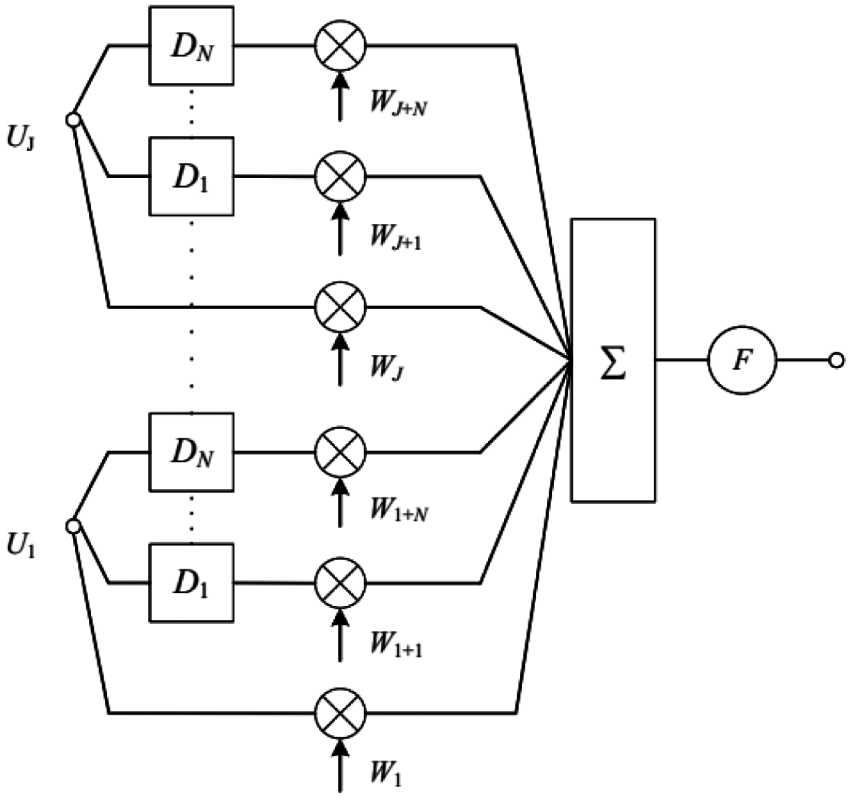

Рис. 4. Схема узла нейронной сети с задержками [8]

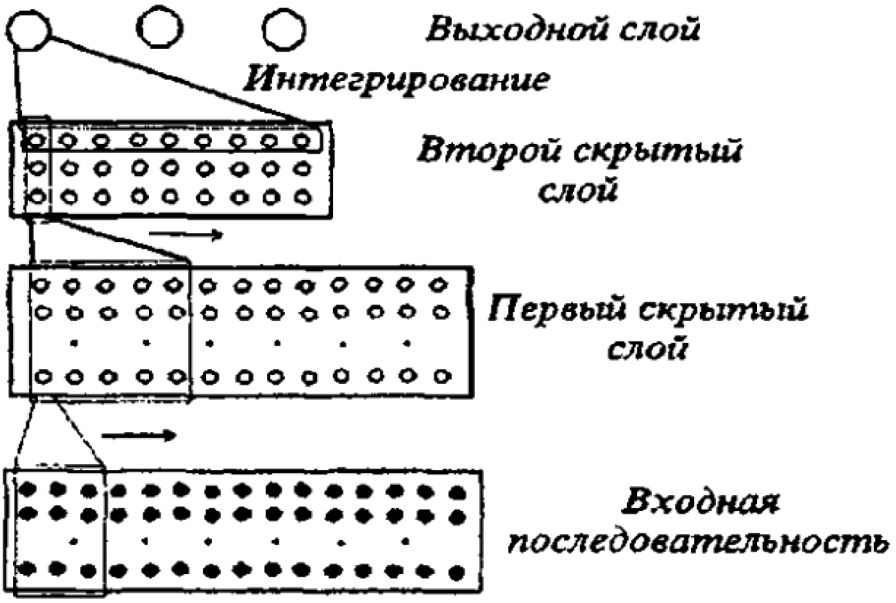

Рис. 5. Архитектура нейронной сети с временными задержками [9]

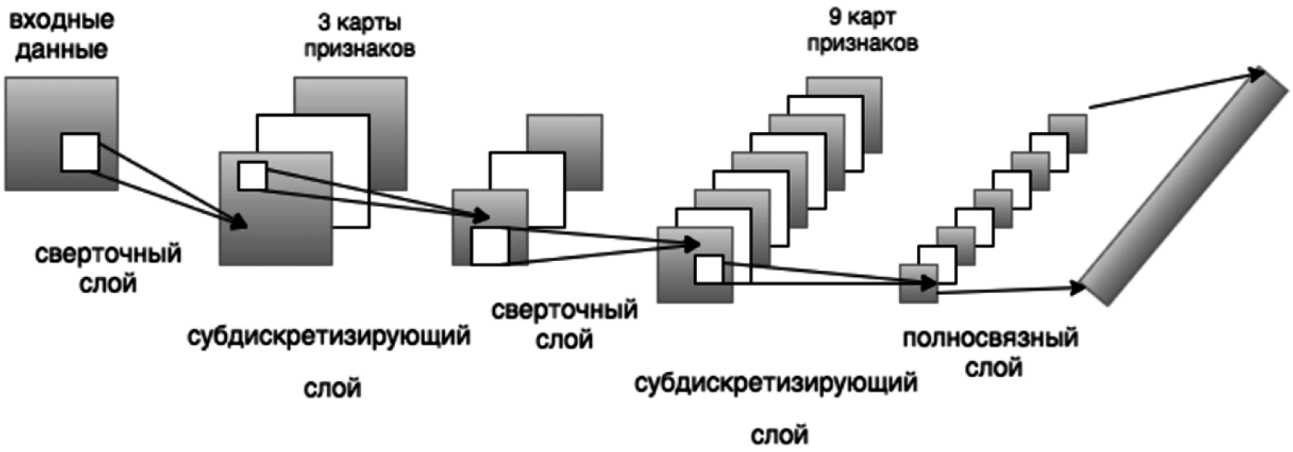

Рис. 6. Сверточные нейронные сети для распознавания голоса [10]

Модель LSTM произвела своего рода революцию в области обработки естественных языков. Она может применяться для предсказания следующего слова на основе предыдущих слов, что может быть использовано, например, для автоматического перевода с одного языка на другой, в системах распознавания речи и в чатботах.

Заключение

Таким образом, в результате проведенного анализа можно сделать вывод о том, что все нейросетевые методы, которые применяются для распознавания речи, дают возможность существенно увеличить скорость распознания речи посредством распараллеливания производимых вычислений. Наибольшей перспективностью обладают конволюционные нейронные сети, которые лишены недостатков статических нейронных сетей. Несмотря на это, у данных сетей имеются свои недостатки, вследствие чего это делает актуальным задачу анализа существующих методов и моделей распознавания речи на основе нейронных сетей, который, несомненно, поможет выбрать наиболее подходящий вариант для пользователя или же послужит базой для дальнейшего развития и совершенствования подобных систем.

ЛИТЕРАТУРА

1. Романюк А.Г. Использование глубокого обучения нейросети для распознавания голосовых команд пользователя / А.Г. Романюк, А.Н. Смирнов, В.М. Антонова // Журнал радиотехники. — 2019. — № 11. — С. 10–20.
2. Тампель И.Б. Автоматическое распознавание речи: уч. пос. / И.Б. Тампель, А.А. Карпов. — СПб.: Лань, 2017. — 152 с.
3. Гафаров Ф.М. Искусственные нейронные сети и их приложение: уч. пос./ Ф.М. Гафаров, А.Ф. Галимянов. — Казань: Издательство Казанского государственного университета, 2018. — 121 с.

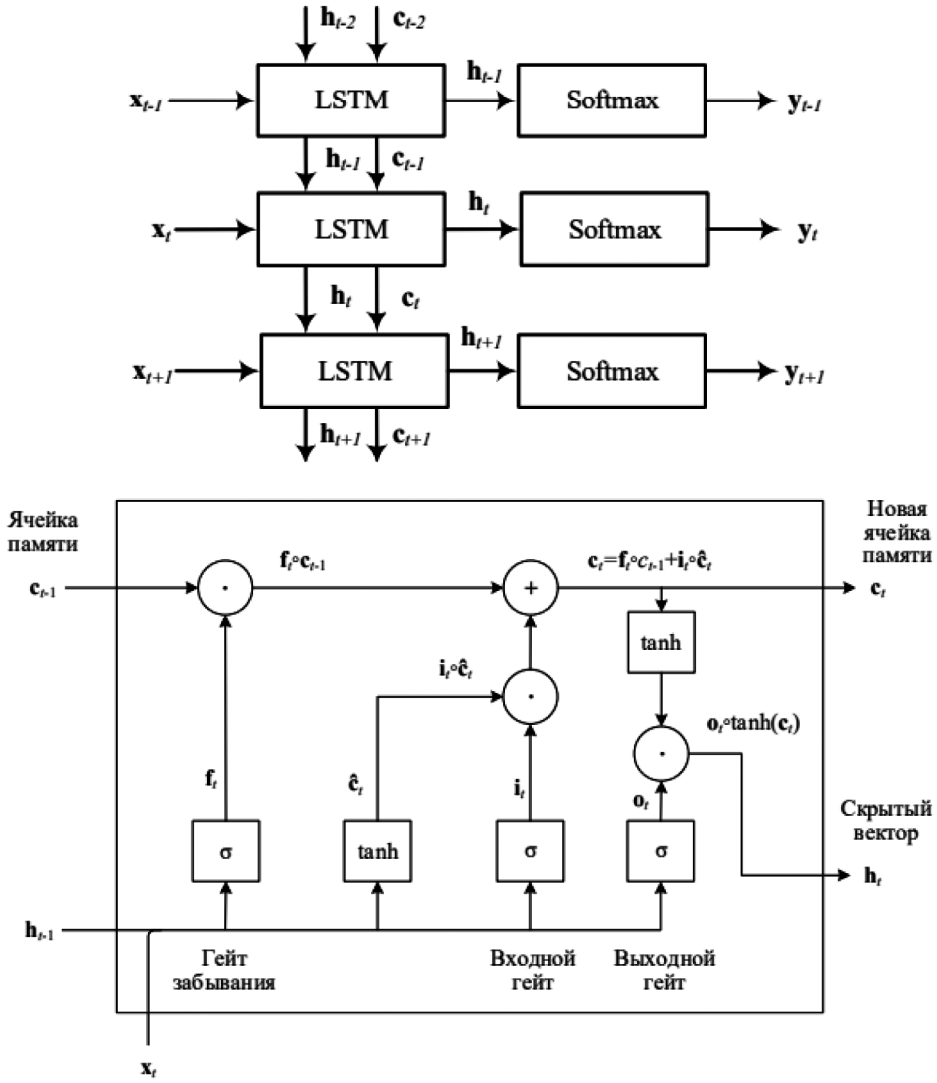

Рис. 7. Архитектура сети и слоя LSTM [8]

4. Кан К.А. Нейронные сети. Эволюция / К.А. Кан. — М.: Литрес, 2018. — 288 с.
5. Вакуленко С.А. Практический курс по нейронным сетям / С.А. Вакуленко, А.А. Жихарева. — СПб: Университет ИТМО, 2018. — 71 с.
6. Данков Н.И. Исследование возможностей нейросетевых технологий в области идентификации голоса / Н.И. Данков // Экономика и качество систем связи. — 2018. — № 3 (9). — С. 47–52.
7. Назаров М.Н. Нейронные сети с динамическими коэффициентами и перестраиваемыми связями на основе интегрированного обратного распространения / М.Н. Назаров // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. — 2018. — № 28 (2). — С. 260–274.
8. Кипяткова И.С. Методы и модели автоматического распознавания речи: уч. пос. / И.С. Кипяткова, А.А. Карпов, С.В. Кулешов, А.А, Зайцева. — СПб ФИЦ РАН, 2021. — 116 с.
9. Гапочкин А.В. Нейронные сети в системах распознавания речи / А.В. Гапочкин // Science Time. — 2014. — № 1. — С. 29–36.
10. Скрипачев В.О. Особенности работы сверточных нейронных сетей / В.О. Скрипачев, М.В. Гуйда, Н.В. Гуйда, А.О. Жуков // International Journal of Open Information Technologies. — 2022. –№ 12 (10). — С. 53–61.
11. Гусенков А.М. Генерация поисковых запросов на основе нейронных сетей / А.М. Гусенков, А.Р. Ситтикова // Научный сервис в сети Интернет: труды XXII Всероссийской научной конференции. Москва, 21–25 сентября 2020 г. — М.: ИПМ им. М.В. Келдыша, 2020. — С. 210–228.