Р. К. Потапова

ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ ГИБРИДНЫХ СИСТЕМ РАСПОЗНАВАНИЯ РЕЧИ (ПРИМЕНИТЕЛЬНО К ОБУЧАЮЩИМ СИСТЕМАМ)

Автоматическое распознавание речи – междисциплинарная область, в которой работают специалисты различного профиля: инженеры, программисты, лингвисты, математики и многие другие. Подходы к решению этой задачи обусловлены как требованиями к разрабатываемому продукту, так и компетенцией разработчиков, включая лингвистов [1; 2].

Широко применяемыми методами распознавания речи являются скрытые модели Маркова, которые доказали свою эффективность и гибкость при реализации систем распознавания речи. Обычно используются три базовых алгоритма, основанные на скрытых марковских моделях: алгоритм прямого распространения (применяется для распознавания изолированных слов); алгоритм Витерби, применяемый для распознавания слитной речи; алгоритм прямого – обратного распространения, используемый для тренинга моделей, в частности, интонационного и звукового строя различных языков.

Сложность алгоритма поиска зависит от природы пространства, в котором производится поиск, и, следовательно, от ограничений, накладываемых на лингвистические модели. При поиске используются следующие алгоритмы: общий поиск на графе; поиск в глубину; поиск в ширину; эвристический лучевой поиск; иерархический поиск. В задачах, связанных с распознаванием и соотнесением образов, чаще всего используется иерархический вид поиска как позволяющий находить минимальные различия, и следовательно, подбирать максимальные соответствия для различных языковых уровней.

Общая схема работы системы распознавания речи выглядит следующим образом (см. рис. 1).

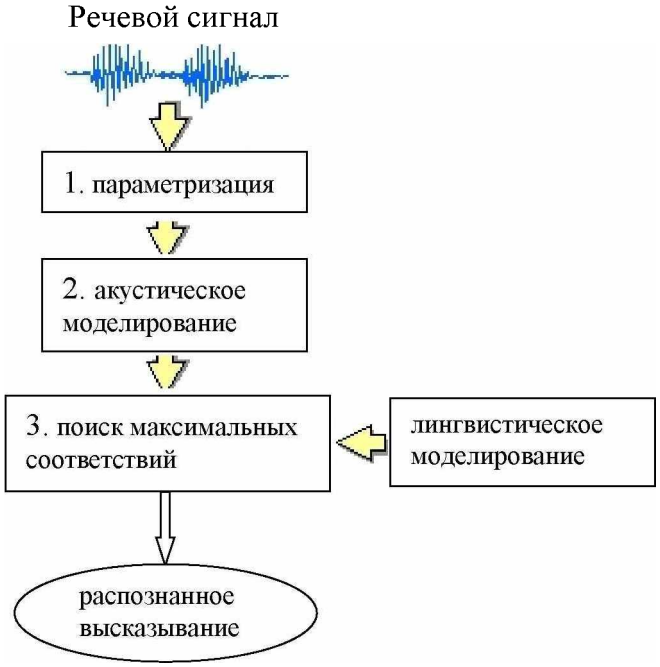

Рис. 1. Принципиальная упрощенная схема работы системы распознавания устной речи

В настоящее время активно разрабатываются алгоритмы распознавания речи, основанные на применении нейронных сетей. Теория нейронных сетей в целом строится на гипотезе, что сложные операции вычисления, выполняемые в рамках одного высокоэффективного модуля системы, можно заменить на ряд параллельно выполняемых простейших операций, производимых в отдельных компонентах системы. Все нейронные сети работают по примерно одинаковым принципам: они состоят из узлов (нейронов), куда поступают данные либо от входов сети, либо от других узлов сети. Входящие сигналы суммируются и посылаются другим элементам по взвешенным связям. Весовой коэффициент ослабляет или усиливает сигнал, идущий по данной связи от одного элемента к другому. Таким образом, нейронная сеть представляет собой соединенные друг с другом простейшие процессоры (элементы, нейроны, узлы), выполняющие элементарные преобразования с входящими сигналами с учетом весовых коэффициентов (коэффициентов, характеризующих связь между элементами), и формирующих выходной сигнал, который поступает на другие элементы, связанные с ним. Каждая связь имеет свой весовой коэффициент. Если он положительный, то входящий сигнал усиливается, в противном случае – подавляется.

Преимущество такого подхода состоит в том, что количество разных функций, по которым вычисляются активность и суммарный вход, ограничено, а диапазон задач, которые могут решать нейронные сети, достаточно велик. Настройка сети на решение новой задачи состоит только в изменении архитектуры сети. Под архитектурой понимается топология сети, структура связей, правила распространения сигналов в сети, правила комбинирования входящих сигналов, правила вычисления активности, правила обучения, т.е. изменения весовых коэффициентов. Как показал ряд исследований, нейронные сети могут быть достаточно эффективны для обработки речи и, в частности, для распознавания речи. Среди задач, с которыми наиболее эффективно справляются нейронные сети, в первую очередь следует назвать классификацию и распознавание образов. Распознавание речи есть приложение теории распознавания акустических и перцептивных слуховых образов [3; 4; 5; 6; 2]. Определенный образ, который может представляться в качестве сегмента речевого сигнала или набора акустических параметров, характеризующих этот сегмент, должен быть отнесен к классу, которому он принадлежит (в качестве класса в рамках различных подходов может быть взята отдельная реализация фонемы, дифон или трифон, слог, или другой элемент, на основе которого выделяются характеризующие параметры класса), что есть задача классификации. Следовательно, алгоритмы распознавания речи, основанные на применении нейронных сетей, могут быть достаточно эффективны. Алгоритмы распознавания речи требуют выполнения огромного объема сложных вычислений. Нейронная сеть, моделируя работу нервной системы человека, сводит этот массив сложнейших вычислений к набору параллельно выполняющихся простейших операций, что ускоряет обработку входящей информации и принятие решения.

Еще одно преимущество использования нейронных сетей для распознавания речи заключается в их эффективности при решении задач классификации и кластеризации. Под классификацией имеется в виду прямое обу-

чение сети – относить ли тот или иной образ к определеному классу в соответствии с инструкциями. То есть выходной образец дает всей сети информацию о том, к какому классу следует научиться относить образец. Если же таких инструкций нет при обучении без управления, то сеть должна научиться группировать образцы самостоятельно. В этом случае принято считать, что сеть должна научиться проводить кластеризацию образов самостоятельно. Преимущество нейронных сетей заключается в том, что они могут сегментировать образцы на группы даже в том случае, если невозможно разделить образцы на классы с помощью линейной зависимости. Задача линейной классификации образцов легко поддается алгоритмизации. Однако при обработке устной речи, к сожалению, такого идеального варианта не встречается. Приходится иметь дело с нелинейными зависимостями. А если к этому добавить проблему динамического изменения значений элементов векторов признаков и размера векторов признаков во времени, то алгоритмизация из сложной превращается в практически невозможную.

В качестве наиболее практичного подхода к распознаванию слитной речи в настоящее время принято использовать **гибридные системы**, представляющие интеграцию нейронных сетей в рамках подхода, основанного на марковских статистических моделях.

Гибридные сети имеют ряд преимуществ при разработке систем распознавания речи. Во-первых, повышается точность моделей, так как нейронные сети формируют более точные акустические непараметрические модели, так как предположения о форме дистрибуции в таких моделях отсутствуют (а это один из основных источников ошибок при акустическом моделировании нейронными сетями). Во-вторых, гибридные сети отличаются повышенной чувствительностью к контексту. Цепи Маркова исходят из предположения о том, что участки обрабатываемого сигнала не зависят друг от друга, и каждый фрейм обрабатывается изолированно. Для того, чтобы иметь возможность использовать информацию о коартикуляции, интерференции и т.д. (т.е. о взаимовлиянии соседних сегментов речи друг на друга), необходимо искусственно расширять границы фрейма и включать в обрабатываемый сегмент соседние. Нейронные сети, с другой стороны, способны естественным образом обрабатывать сегменты любого размера (лишь бы они был одинаковые). Поэтому нейронные сети более чувствительны к фонетическому контексту (окружению). В-третьих, нейронные сети лучше разделяют образцы по группам, ибо классификация и кластеризация входных данных – естественное свойство нейронных сетей. В-четвертых, нейронные сети имеют преимущество в том случае, если данных для тренировки недостаточно, так как им необходимо меньше параметров для разделения образцов на классы [7].

На эффективность и точность работы распознающей и обучающей системы влияют следующие факторы [8; 9; 10]: размер словаря и наличие в нем схожих слов (число ошибок, как правило, сильно увеличивается с увеличением словаря). Наличие слов, которые легко спутать, также представляют определенную проблему; манера произнесения: система может быть нацелена на распознавание изолированно произнесенных слов, пословной речи, когда каждое слово в предложении отделяется паузой, или слитной (естественной) речи. Последний случай наиболее сложный, но он же – необ-

ходимое условие, если мы хотим построить обучающую систему, занимающуюся не только дриллингом отдельных сегментов речи в изолированном окружении, но именно обучением произношению как на сегментном, так и на супрасегментном уровнях в естественной речи; темпоральная вариативность: проблема вариативности в темпе произнесения и в определении момента начала артикуляции очередного сегмента (на темп влияет время артикуляции ударных слогов и время физических пауз); акустическая вариативность – проблема различий в произношении (акцент, индивидуальные особенности речи, громкость, акустические внешние условия и т.д.).

Проблема темпоральной вариативности решается с помощью алгоритмов динамического программирования, чаще всего используются DTW и алгоритм Витерби. Вариативность, которую приходится учитывать при разработке систем автоматического распознавания речи, вызвана междикторской, внутридикторской, ситуативной и т.д. вариативностью речи и вариативностью сигнала, в том числе искажениями сигнала [1].

В настоящее время уровень систем автоматического распознавания речи таков, что точное определение места артикуляции того или иного звука имеет первостепенное значение, поэтому распознавание иноязычной речи не настолько надежно, как распознавание речи говорящего на родном языке диктора.

Для моделирования иностранного произношения в обчающих системах предлагаются следующие подходы: использование репрезентативных массивов данных для обучения системы; применение независящих от диктора моделей по алгоритму Витта и Янга; проведение лингвистического моделирования; расширение словаря за счет вариантов произнесения после обработки правил произношения в родном языке диктора и выявления закономерностей интерференции; использование звукового словаря специфических иноязычных акцентов; классификация иноязычных акцентов на основе их акустических характеристик.

Специальная целенаправленная фонетическая практика позволяет избавиться от возможных произносительных ошибок, увеличить стабильность реализации отдельных звуков в определенных контекстах.

Таким образом, для разработки обучающих фонетических систем наиболее продвинутого вида оптимальным способом является метод повышения устойчивости к иноязычному акценту. При его применении система способна распознать интерферированную речь, и в то же время, используя алгоритм Витта, выделить ошибку реализации речевого сегмента, после чего возможно применение блока формирования правил коррекции произнесения [8; 9; 1].

Таким образом, для повышения эффективности работы, например, лингвистической обучающей системы (фонетический аспект) представляется целесообразным рассмотреть вопрос использования не одного, а комплекса алгоритмов для реализации базового метода распознавания речи. Инкорпорирование элементов нейронных сетей, а также гибридного метода для распознавания и обучения иноязычной речи представляется более перспективным, нежели использование традиционных методов. Для моделирования иноязычного акцента оптимальным является расширение словаря за счет вариантов произнесения после обработки общих правил произношения на родном языке обучающегося и выявления закономерностей интерязыковой интерференции.

ЛИТЕРАТУРА

1. *Потапова, Р. К.* Приоритетные направления современной прикладной лингвистики / Р. К. Потапова // Конверсия в машиностроении. – 2004. – № 3–4.– С. 92–100; 97–103.
2. *Потапова, Р. К.* Новые информационные технологии и лингвистика / Р. К. Потапова. – 10 изд. – М., 2014.
3. *Потапова, Р. К.* Звучащая речь как объект исследования в фундаментальной и прикладной лингвистике / Р. К. Потапова, В. В. Потапов // Ежегодник «Акустика речи и прикладная лингвистика» : сб. тр. РАО и МГЛУ. – М., 2002. – С. 6–28.
4. *Потапова, Р. К.* Речь: коммуникация, информация, кибернетика / Р. К. Потапова. – 4 изд. доп. – М., 2010.
5. *Потапова, Р. К.* Основы речевой акустики / Р. К. Потапова, В. Г. Михайлов. – М. : Рема, 2012.
6. *Потапова, Р. К.* Речевое управление роботом / Р. К. Потапова. – 2 изд., доп., перераб. – М., 2012.
7. *Потапова, Р. К.* Междисциплинарность в исследовании речевой полиинформативности / Р. К. Потапова, В. В. Потапова, Н. Н. Лебедева, Т. В. Агибалова. – М. : Языки славянской культуры, 2015.
8. *Потапова, Р. К.* Некоторые подходы к реализации речевых компонентов в компьютерных лингвистических обучающих системах / Р. К. Потапова, М. Ю. Ордин // Акустика речи. Медицинская и биологическая акустика : матер. 13 сессии Рос. акустического об-ва. – М., 2003. – Т. 3. – С. 145–149.
9. *Potapova, R. K.* Articulation Models in Educational Software with Embedded ASR Components / R. K. Potapova, M. Yu. Ordin // Proceedings of Intern. conf. “SPECOM’ 2003”. – Moscow, 2003. – P. 360–364.
10. *Potapova, R. K.* Modern CALL systems with elements of acoustic feedback / R. K. Potapova // Proceedings of Intern conf. “SPECOM’ 2003”. – Moscow, 2003. – P. 53–60.