ISSN 1992-6502 (Print) ISSN 2225-2789 (Online)
2019. Т. 23, № 2 (84). С. 121–126 Вестник УГАТУ http://journal.ugatu.ac.ru

УДК 004.934.5

НЕЙРОСЕТЕВОЕ МОДЕЛИРОВАНИЕ ГРАФЕМНО-ФОНЕМНОГО ПРЕОБРАЗОВАНИЯ: АВТОМАТИЧЕСКИЙ МОРФОЛОГИЧЕСКИЙ АНАЛИЗ

М. Р. Бикметова[1], К. Ричмонд[2]

[1] maya.bikmetova@ed-alumni.net, [2] korin@cstr.ed.ac.uk

Исследовательский центр речевых технологий, Эдинбургский университет

Поступила в редакцию 13.04.2019

Аннотация. Предложен альтернативный подход к одному из ключевых этапов автоматического синтеза речи – графемно-фонемному преобразованию. В основе подхода лежит автоматический морфологический анализ слов, поступающих на вход системы. В отличие от классического метода морфологического анализа, базирующегося на принципе конечного автомата, в настоящей статье предлагается решение с использованием нейронных сетей. Эксперименты показывают, что нейросетевой подход значительно эффективнее (доля верных ответов 93,8 %) подхода с использованием конечных автоматов (доля верных ответов 75 %). Предложенный подход позволяет ускорить создание графемно-фонемных моделей, а также снизить трудозатраты на составление и поддержание машиночитаемых словарей произношений.

Ключевые слова: графемно-фонемное преобразование; морфологический анализ; синтез речи; нейронные сети; речевые технологии.

ВВЕДЕНИЕ

Современный мир невозможно представить без технологий искусственного интеллекта. Одно из наиболее быстроразвивающихся направлений – автоматический синтез речи. Его задачей является преобразование текстовых данных в звучащую речь [1]. Архитектура системы селективного синтеза речи представлена на рис. 1. Примерами практического применения синтетической речи являются речевые человеко-машинные интерфейсы и голосовые помощники от компаний Apple [2], Google [3], Amazon [4], Яндекс [5].

Для успешного синтеза речи система должна в первую очередь преобразовывать текстовые данные в последовательность фонем, т.е. производить графемно-фонемное преобразование (ГФП, от англ. grapheme-to-phoneme). Модели ГФП являются неотъемлемой частью любой системы синтеза речи. Большинство таких моделей обучаются на специальных электронных словарях произношений. Однако словари ограничены и трудоемки в создании и поддержке актуальности, поэтому альтернативным подходом к проблеме является использование таких словарей в качестве обучающей выборки для настройки алгоритма, который бы автоматически предсказывал произношение ранее не встречавшихся слов.

Ввиду того, что большинство слов в естественном языке являются производными от сравнительно небольшой группы слов, теоретически нет необходимости хранить в словарях информацию о произношении всех однокоренных слов. Достаточным было бы хранить произношение только основных, непроизводных слов и словообразующих аффиксов, а произношения для производных генерировать автоматически. Например, если в словаре системы имеется произношение слова «*run*», то можно авто-

матически сгенерировать произношение слова *«running»*, поскольку *«running»* = корень *«run»* + суффикс *«-ing»*.

Для практического применения данного подхода необходим надежный способ морфологической сегментации слов на составляющие морфемы.

Таким образом, целью работы является нахождение оптимального способа автоматизации морфологического анализа с последующим применением полученных результатов для улучшения моделей ГФП, применяемых в системах синтеза речи.

Актуальность предлагаемого подхода обусловлена тем, что это значительно облегчит работу по созданию автоматических систем синтеза речи и снизит долю человеческого труда при составлении словарей произношений.

МОРФОЛОГИЧЕСКИЙ АНАЛИЗ КАК АЛЬТЕРНАТИВНЫЙ ПОДХОД К ГРАФЕМНО-ФОНЕМНОМУ ПРЕОБРАЗОВАНИЮ

Одной из первых реализаций ГФП были написанные вручную правила. Позже фонетические правила стали представлять в виде конечных автоматов [6]. Такой подход дает неплохие результаты для языков, в которых письменная форма слова совпадает с его произношением. Однако M. Bisani и H. Ney отмечают, что создание правил произношения слишком трудоемко и требует экспертных знаний в фонетики и фонологии [8]. К тому же, в естественных языках наблюдается множество исключений, которые также приходится учитывать при разработке правил. В результате правила ГФП могут получаться чрезвычайно запутанными и взаимозависимыми. Для такого языка, как английский, задача моделирования ГФП нетривиальна и требует более совершенных методов.

Современные системы синтеза речи обычно опираются на модели ГФП, созданные с помощью алгоритмов машинного обучения на существующих словарях произношений. Традиционно подобные модели используют словесный подход, т.е. происходит замена последовательности графем $g \in G$ на наиболее вероятную последовательность фонем $\phi \in \Phi$, где G – это множество допустимых графем, а Φ – множество допустимых фонем:

$$\phi(g) = arg \max_{\phi' \in \Phi} p(g, \phi').$$

Предлагаемая идея морфемного подхода к задаче ГФП основывается на декомпозиции слов на составляющие морфемы [9] и последующей автоматической генерации произношений для морфем:

$$\phi(m) = arg \max_{\phi' \in \Phi} p(m, \phi'),$$

где $m \in M$ и M – множество допустимых морфем.

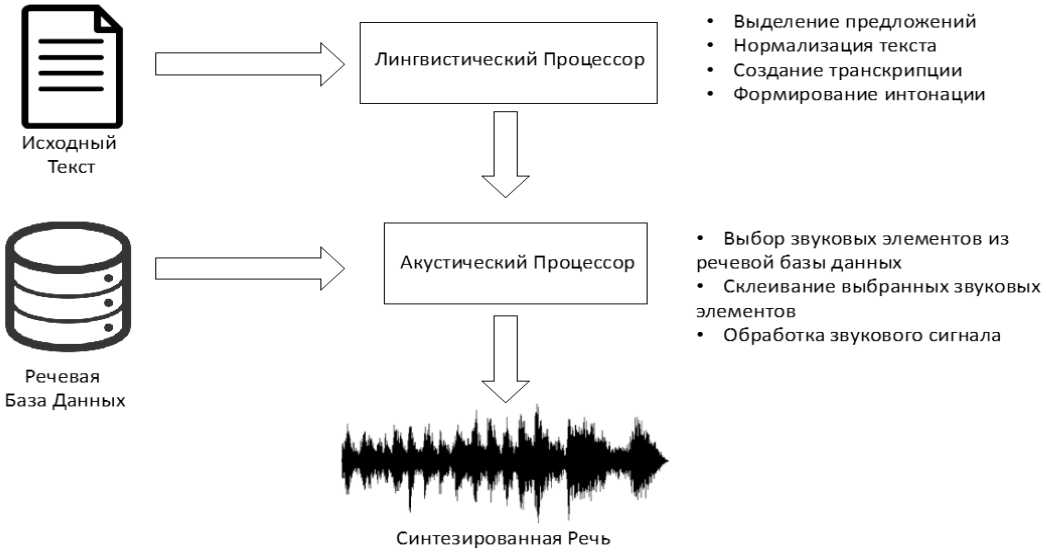

Рис. 1. Схема синтезатора речи

Потенциал морфологического членения в качестве подхода к ГФП подтверждают следующие факты:

1. В [7] предлагается использовать морфологический анализ в тандеме с деривационным функционалом словаря произношений, в качестве способа генерирования произношений для слов, не включенных в словарь.

2. Большинство слов, не включенных в словарь, которые встречаются в процессе синтеза речи, являются производными слов, уже имеющихся в словаре.

В [7] эксперименты по автоматизации морфологического анализа проводились с помощью программы PC-KIMMO [10], основанной на концепте конечных автоматов. Недостатком данного метода является то, что правильный морфологический анализ для слова маловероятен, если корень слова не включен в словарь программы PC-KIMMO. Данный дефект может быть устранен ручным добавлением корней из словаря произношений в словарь программы. Однако поскольку для этого требуется труд эксперта, предполагается, что использование нейронных сетей было бы более гибким подходом. Нейронным сетям не нужны специальные знания в морфологии. Единственное требование для их использования – это достаточное количество размеченных данных для обучения.

В таком случае, нейронные сети могут использоваться для создания морфологического анализа ранее не встречавшихся слов. Такой морфологический анализ затем может служить в качестве основы для генерации последовательности фонем для системы синтеза речи.

Поскольку для отнесения графемы g_k к той или иной морфеме модель должна опираться на контекст, т.е. «помнить», к какой морфеме была отнесена графема g_{k-1}, необходимо использовать нейронную сеть, способную хранить информацию в течение долгих промежутков времени. Таким свойством обладает одна из разновидностей традиционных рекуррентных нейронных сетей – долгая краткосрочная память (LSTM – Long Short-Term Memory) [11]. Подобное свойство нейронной сети обусловлено усложненной архитектурой: LSTM состоит из ячейки памяти и фильтров, контролирующих информационный поток [12].

ОЦЕНКА ЭФФЕКТРИВНОСТИ НЕЙРОСЕТЕВОГО ПОДХОДА

В настоящей работе использована однонаправленная нейронная сеть долгой краткосрочной памяти NMTSmall, реализованная в общедоступной библиотеке OpenNMT [13].

Для обучения нейросетевой модели были использованы машиночитаемые словари произношений английских слов – Unilex [14] и Combilex [15].

На входной слой нейронной сети подается исходная последовательность графем с дополнительным начальным символом <s> и конечным символом <s/>, например <s> u n w i l l e d <s/>. На выходе сеть генерирует ту же последовательность графем, но уже с символами, обозначающими границы морфем: < u n { w i l l } e d >. Так, символ «<» отмечает начало префикса, символ «>» – конец суффикса или окончания, символ «{» – начало основы, символ «}» – конец основы.

Процесс обучения проходит циклически: на вход подается исходная последовательность графем, весовые коэффициенты между нейронами сети выбираются случайным образом, нейросеть генерирует предполагаемые границы морфем. Предсказанная последовательность графем и границ сравнивается с истиной, затем происходит корректировка весов, направленная на уменьшение ошибки. Во время предсказания модель находит наиболее вероятное расположение границ морфем.

Чтобы трансформировать данные словарей в формат, подходящий для подачи на вход нейронной сети, написана программа на языке bash [16], которая добавляет к каждому слову начальные и конечные символы, вставляет пробелы между знаками, удаляет из словаря все слова, несоответствующие кодировке ASCII, а также все слова, помеченные составителями как иностранные. Последнее сделано для того чтобы модель не обучалась на морфологии заимствованных слов, поскольку она может значительно отличаться от английской. Затем,

каждый словарь случайным образом разбивается на обучающую, проверочную и тестовые выборки в соотношении 80/10/10. Размеры полученных выборок представлены в табл. 1.

Таблица 1

Количество слов в выборках

	Слов в выборке		
Словарь	Обучающая	Проверочная	Тестовая
Unilex	95486	11935	11935
Combilex	78000	9750	9750

После 170000 итераций модели NMTSmall с базовыми гиперпараметрами приводят приемлемый морфологический анализ лексем тестовых выборок. В качестве метрики результативности модели используется доля верных ответов (accuracy, A):

$$A = \frac{P}{\text{размер выборки}} \times 100\,\%,$$

где P – количество верных морфологических анализов, сгенерированных моделью.

Результаты отражены в табл. 2. Модель, обученная на словаре Unilex, генерирует морфологически верные анализы слов с долей верных ответов 91,1 %; доля верных ответов модели, обученной на Combilex, составляет 93,8 %. Точность модели FST (Finite State Transducer – конечный автомат), представленной в [7], отражена в нижней строке табл. 2 и составляет лишь 75 %. Можно отметить, что нейросетевая модель оказалась более эффективным методом автоматического морфологического анализа по сравнению с подходом [2], в основу которого положен конечный автомат.

Таблица 2

Доля верных ответов автоматического морфологического анализатора

Модель	*Словарь*	*Доля верных ответов, %*
LSTM	Unilex	91,1
LSTM	Combilex	93,8
FST [7]	Combilex	75

Далее результат морфологического анализа может быть использован для автоматического построения транскрипций для производных слов, подаваемых на вход синтезатора речи. Непосредственное использование модели, автоматически выделяющей морфемы, в качестве графемно-фонемного преобразователя в реальной системе синтеза речи является следующим шагом развития данного вопроса.

ЗАКЛЮЧЕНИЕ

Проведенные эксперименты показали, что нейронные сети долгой краткосрочной памяти справляются с задачей морфологической декомпозиции лучше традиционных конечных автоматов.

Стоит также отметить, что в связи с временными ограничениями настоящей работы, число итераций для обучения модели и начальные параметры нейронной сети были выбраны произвольно. Вероятно, подбор оптимальных параметров модели увеличит точность анализа. Это может рассматриваться как одно из возможных направлений будущей работы.

Направлением для дальнейших работ является исследование корреляции между точностью предсказаний модели и наличием в обучающей выборке частеречной разметки (POS – Part of Speech Tagging). Поскольку морфемный состав слова непосредственно зависит от части речи, предполагается, что предоставление нейросетевой модели информации о части речи обучающего примера позволит значительно улучшить точность автоматического морфологического анализа.

Также стоит рассмотреть решение данной задачи с использованием нейронных сетей с механизмом внимания. Одной из наиболее эффективных архитектур, использующих подобный механизм, является нейросеть Transformer, представленная в 2017 г. исследователями компании Google [17]. Модель Transformer показала прекрасные результаты для различных Sequence-to-Sequence задач, т.е. для задач генерации новых последовательностей на основании входной информации [18]. Вероятно, что данная архитектура также покажет хорошие результаты при работе над задачей автоматического морфологического анализа.

Авторы выражают благодарность д-ру физ.-мат. наук, проф. Р. К. Газизову за высказанные замечания и пожелания по улучшению статьи.

СПИСОК ЛИТЕРАТУРЫ

1. **Рыбин С. В.** Синтез речи: учебное пособие. СПб.: Университет ИТМО, 2014. 92 с. **[** S. V. Rybin, *Speech synthesis,* (in Russian). Spb.: University ITMO, 2014. **]**

2. **AppleSiri** [Электронный ресурс] // URL: http://www.apple.com/ios/siri/ (дата обращения: 15.12.2018).

3. **Google Now** [Электронный ресурс] // URL: http://www.google.com/landing/now/ дата обращения: 15.12.2018).

4. **Amazon Alexa**, [Электронный ресурс] // URL: https://developer.amazon.com/alexa/ (дата обращения: 15.12.2018).

5. **Алиса** [Электронный ресурс] // Яндекс.Ассистент. URL: http://yandex.ru/support/alice/ (дата обращения: 15.12.2018).

6. **Kaplan R. M., Kay M.** Regular models of phonological rule systems // Computational Linguistics. 1994. Vol. 20, no. 3, pp. 331–378. **[** R. M. Kaplan, M. Kay, "Regular models of phonological rule systems", in *Computational Linguistics,* vol. 20, no. 3, pp. 331-378, 1994*.* **]**

7. **Richmond K., Clark R., and Fitt S.** On generating combilex pronunciations via morphological analysis // Proceedings of Interspeech. Pp. 1974–1977, 2010.

8. **Bisani M., Ney H.** Joint-sequence models for grapheme-to-phoneme conversion // Speech communication. 2008. Vol. 50, no. 5, pp. 434–451. **[** M. Bisani, H. Ney, "Joint-sequence models for grapheme-to-phoneme conversion", in *Speech communication*. Vol. 50, no. 5. Pp. 434-451, 2008. **]**

9. **Корочков А. В.** Компьютерное моделирование графемно-фонемного преобразования в английском языке (выбор подхода) // Вестник МГУ. 2003. № 1–2. URL: https://cyberleninka.ru/article/n/kompyuternoe-modelirovanie-grafemno-fonemnogo-preobrazovaniya-v-angliyskom-yazyke-vybor-podhoda (дата обращения: 15.10.2018). **[** A. V. Korochkov, "Computer modeling of grapheme-phonemic transformation in English (choice of approach)", (in Russian), in *Vestnik MGU*, no. 1-2, 2003. Avaliable: https://cyberleninka.ru/article/n/kompyuternoe-modelirovanie-grafemno-fonemnogo-preobrazovaniya-v-angliyskom-yazyke-vybor-podhoda. **]**

10. **Antworth E. L.** PC-KIMMO: a two-level processor for morphological analysis // Occasional Publications in Academic Computing No. 16, Summer Institute of Linguistics, Dallas, Texas. 1990.

11. **Hochreiter S., Schmidhuber J.** Long short-term memory // Neural computation. 1997. Vol. 9, no. 8, pp. 1735–1780.

12. **LSTM:** A search space odyssey / K. Greff et. al. // IEEE transactions on neural networks and learning systems. 2017. Vol. 28, no. 10, pp. 2222–2232. **[** K. Greff et. al., "LSTM: A search space odyssey", in *IEEE transactions on neural networks and learning systems*. Vol. 28, no. 10, pp. 2222-2232, 2017. **]**

13. **Klein G. et al.** Opennmt: Open-source toolkit for neural machine translation // arXiv preprint arXiv:1701.02810. – 2017. **[** G. Klein et. al., "Opennmt: Open-source toolkit for neural machine translation", in *arXiv preprint arXiv:1701.02810*. 2017. **]**

14. **Fitt S.** Documentation and user guide to UNISYN lexicon and post-lexical rules. Center for Speech Technology Research, University of Edinburgh, Tech. Rep. 2000.

15. **Fitt S., Richmond K., Clark R.** Redundancy and productivity in the speech technology lexicon-can we do better? // Ninth International Conference on Spoken Language Processing, Sept. 2006, pp. 165-168. **[** S. Fitt, K. Richmond, R. Clark, "Redundancy and productivity in the speech technology lexicon-can we do better?", in *Ninth International Conference on Spoken Language Processing*, 2006, pp. 165-168. **]**

16. **Ramey C., Fox B.** Bash reference manual. – Network Theory Limited, 2003. P. 204. **[** C. Ramey, B. Fox, *Bash reference manual,* in Network Theory Limited, 2003. **]**

17. **Attention** Is All You Need / A. Vaswani et. al. // Advances in Neural Information Processing Systems. 2017. Pp. 5998–6008. **[** A. Vaswani et. al., "Attention Is All You Need", in *Advances in Neural Information Processing Systems***,** 2017, pp. 5998-6008. **]**

18. **Kaiser L.** (2017). Accelerating Deep Learning Research with the Tensor2Tensor Library. URL: https://ai.googleblog.com/ (дата обращения: 17.07.2018).

ОБ АВТОРАХ

БИКМЕТОВА Майя Рустемовна, магистр в сфере обработки языка и речи (Эдинбургский ун-т, 2018). Вед. специалист отдела прототипов и развития технологий ООО «РН-БашНИПИнефть». Иссл. в обл. синтеза речи.

РИЧМОНД Корин, Ph.D (Эдинбургский ун-т, 2002). Доцент Исследовательского центра речевых технологий, Эдинбургский ун-т. Иссл. в обл. речевых технологий, прикладного машинного обучения.

METADATA

Title: Grapheme-to-phoneme conversion with neural networks: automatic morphological analysis.

Authors: M. R. Bikmetova[1], K. Richmond[2]

Affiliation: Centre for Speech Technology Research (CSTR), University of Edinburgh, Scotland, United Kingdom.

Email: [1] maya.bikmetova@ed-alumni.net, [2] korin@cstr.ed.ac.uk

Language: Russian.

Source: Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), vol. 23, no. 2 (84), pp. 121-126, 2019. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

Abstract: The paper provides an alternative approach to one of the key steps of speech synthesis – grapheme-to-phoneme conversion. The approach is based on automatic morphological analysis of out-of-vocabulary words into their consitutent morphemes. In constrast to a traditional method of morphological analysis based on finite state transducers (FST), we propose here a solution that makes use of neural networks. Experiments show that morphological analysis with the proposed neural network approach is significantly more effective (accuracy 93.8%) than

than morphological analysis with FST (accuracy 75%). The proposed approach allows to speed up the creation of new grapheme-to-phoneme models, as well as making it easier to build and maintain pronunciation dictionaries.

Key words: grapheme-to-phoneme conversion; morphological analysis; speech synthesis; neural networks.

About authors:

BIKMETOVA Maiia Rustemovna, Master of Science in Speech and Language Processing (Univ. Edinburgh, 2018). Lead Specialist of Prototypes and Technology Development Branch in RN-BashNIPINeft LLC.

RICHMOND Korin, Ph.D (Univ. Edinburgh, 2002). Reader in Speech Technology, Centre for Speech Technology Research (CSTR), University of Edinburgh.