

Восстановление фонетического состава распознанного слова с помощью лексической онтологии

В.С. Мошкин^{1,2}, А.И. Армер¹, Н.А. Крашенинникова²

¹Ульяновский государственный технический университет, ул. Северный Венец 32, Ульяновск, Россия, 432027

²Ульяновский государственный университет, ул. Льва Толстого 42, Ульяновск, Россия, 432970

Аннотация. Улучшение качества распознавания речи в условиях шума можно реализовать, адаптировав алгоритм восстановления распознанного слова к особенностям использования и особенностям предметной области. В статье описан подход к восстановлению фонетической композиции распознанного слова с использованием лексических онтологий. Лексическая онтология содержит лексические отношения между членами предметной области и фонетическим составом каждого из этих слов в терминах таблицы SAMPA + для русского языка.

1. Введение

Распознавание слитной речи представляет собой сложный итерационный процесс, который основан на последовательном членении общего звукового сигнала на слова, а в дальнейшем слова – на фонемы. При этом существует большое количество факторов (одним из которых является наличие шумов) значительно осложняющих распознавание и снижающих точность распознавания. В связи с этим, множество алгоритмов после предварительной обработки речевого сигнала в качестве результата выдают матрицу, состоящую из набора фонем с соответствующими вероятностями их детектирования в распознаваемом слове.

С другой стороны, задача распознавания и преобразования в текст слитной русской речи – актуальная проблема, которую необходимо решить во многих предметных областях (ПрО): медицина, судебное дело, автоматизированное нахождение экстремистских материалов и многих других.

При этом каждой ПрО соответствует особый набор наиболее часто употребляемых терминов. Благодаря априорному определению базового набора терминов анализируемой ПрО, есть возможность повысить вероятность определения верного порядка встречаемости тех или иных фонетических сочетаний в процессе распознавания речи.

Одним из возможных направлений адаптации алгоритмов распознавания речи к соответствующим ПрО является определение и применение лексических онтологий ПрО.

2. Алгоритм пофонемного распознавания слов русской речи

Для первоначального определения набора фонем-кандидатов распознаваемого речевого сигнала, из которых затем будет сформированы речевые выражения исходя из особенностей

Про с учетом результата анализа предметной лексической онтологии, используется алгоритм фонемного распознавания слов русской речи.

Для фонемного распознавания слов русской речи из неограниченного словаря используется следующий алгоритм: с помощью метода, описанного в [1], слово разбивается на составляющие речевые единицы, таким образом, что становятся известны стартовые и конечные отсчёты речевого сигнала для каждой из них. Далее, речевой сигнал каждой речевой единицы последовательно преобразуется в автокорреляционный портрет следующим образом: $s(i)$ - i -й отсчет оцифрованного речевого сигнала; $s(i+k)$ - отсчет, отстоящий от $s(i)$ на k отсчетов. Степень зависимости этих отсчетов выражается выборочным коэффициентом корреляции:

$$R_s(k) = R[s(i), s(i+k)] = \frac{\text{cov}[s(i), s(i+k)]}{\sqrt{\frac{1}{N} \sum_{i=1}^N s^2(i) - m_{s(i)}^2} \sqrt{\frac{1}{N} \sum_{i=1}^N s^2(i+k) - m_{s(i+k)}^2}},$$

$$\text{cov}[s(i), s(i+k)] = \frac{1}{N} \sum_{i=1}^N s(i)s(i+k) - \left[\frac{1}{N} \sum_{i=1}^N s(i) \right] \left[\frac{1}{N} \sum_{i=1}^N s(i+k) \right], \quad (1)$$

где N - число отсчетов отрезка, на котором ищется зависимость; ; $\text{cov}[s(i), s(i+k)]$ — выборочная ковариация $s(i)$ и $s(i+k)$ при $i = 1..N$; $m_{s(i)}$ - выборочное среднее $s(i)$ при $i = 1..N$; $m_{s(i+k)}$ - выборочное среднее $s(i+k)$ при $i = 1..N$. Функция, определенная выборочными коэффициентами корреляции по (1), является автокорреляционной функцией сигнала. С помощью её вычисления осуществляется преобразование сигнала $s(i)$ $i = 1..M$ (M — число отсчетов в сигнале речевой единицы) в двумерное изображение. Для этого $s(i)$ разбивается на отрезки по $N < M$ отсчетов, далее, из значений, определяемых соотношением (1) составляются строки изображения:

$$s(i_m^j), s(i_m^j + k) \quad \begin{matrix} k=1..N \\ j=1, N, 2N, \dots \end{matrix}$$

$$X(j, k) = R \quad (2)$$

Полученное по формуле (2) двумерное изображение $X(i, k)$, где i - номер строки, k — номер столбца, является автокорреляционным портретом (АКП) речевой единицы $s(i)$ с размерами $N \times \frac{M}{N}$.

Точно с таким же значением параметра N в АКП преобразованы эталонные речевые единицы, подготовленные из продиктованных примеров фонетического алфавита SAMPA+. Это даёт возможность определить степень похожести АКП речевой единицы с АКП каждого из эталонов. В результате речевая единица считается соответствующей тому эталону, с которым определена наибольшая похожесть. Похожесть АКП определяется путём нахождения Евклидова расстояния между строками АКП с динамическим определением положения каждой строки [2, 3, 4]. При этом заведомо непохожими считаются АКП, у которых разность количества строк превышает определённое пороговое значение.

3. Модель лексической онтологии

Онтология представляет собой систему, состоящую из множества понятий, их определений и аксиом, необходимых для ограничения интерпретации и использования понятий [5].

OWL (Ontology Web Language) – язык описания онтологий, позволяющий описывать классы и отношения между ними. В основе языка – представление действительности в модели данных «объект – свойство». OWL является переформулировкой дескрипционной логики с использованием синтаксиса XML.

Онтология Про представляет собой совокупность RDF-троек: subject – predicate – object (субъект — отношение — объект). В рамках решения поставленной задачи использовалась OWL-онтология [6, 7].

Особый тип онтологий - лексические (или лингвистические). Их отличительное свойство - использование в одном ресурсе (лексикализованных) понятий (слов) вместе с их языковыми

свойствами. Основным источником понятий в онтологиях данного типа являются значения языковых единиц. Их также отличает набор отношений, обычно свойственный языковым элементам: синонимия, гипонимия, меронимия и ряд других [8].

Для решения задачи восстановления фонетического состава распознанного слова в структуру лексической онтологии были включены элементы фонетических алфавитов, устанавливающих соответствие букв и звуков. Наиболее распространёнными фонетическими алфавитами являются международный фонетический алфавит (МФА) и X-SAMPA (а также расширение SAMPA+, содержащее в том числе транскрипции русского языка) [9].

Формальную модель используемой лексической онтологии ПрО можно представить следующим образом:

$$O = \langle A, C^{A_i}, P^{IPA}, P^{SAMPA+}, R^{A_i} \rangle, i = \overline{1, m},$$

где m – количество охватываемых онтологией предметных областей; $A = \{A_1, A_2, K, A_i\}$ – множество предметных областей, охватываемых онтологией; $C^{A_i} = \{C_1^{A_i}, C_2^{A_i}, K, C_n^{A_i}\}$ – множество терминов в рамках i -ой ПрО; P^{IPA} – множество фонем русского языка в соответствии с МФА; P^{SAMPA+} – множество фонем русского языка в соответствии с фонетическим алфавитом SAMPA+. Данный алфавит включает в себя 89 фонем.

R^{A_i} – множество отношений онтологии в рамках i -ой ПрО следующего вида:

$$R^{A_i} = \{R_C^{A_i}, R_{C-P}^{A_i}, R_P^{A_i}, R_{P_1P_2}^{A_i}\},$$

где $R_C^{A_i}$ – множество отношений, формирующих иерархию терминов онтологии в рамках i -ой ПрО; $R_P^{A_i}$ – множество отношений объектов (Object Properties) и типов данных (Data Type Properties), определяющих связь между элементами множеств P^{IPA} , P^{SAMPA+} , а также соответствующие свойства этих объектов. $R_P^{A_i} = \{hasIPA, hasSampaPlus, hasExample...\}$.

$R_{C-P}^{A_i}$ – множество отношений, определяющих связь между объектами множеств C^{A_i} и P^{IPA} , P^{SAMPA+} в рамках i -ой ПрО. Данное свойство определяет, входит ли фонема в фонетическое представление соответствующего термина.

$R_{P_1P_2}^{A_i}$ – множество отношений, определяющих значение вероятности следования фонемы P_2 за фонемой P_1 в рамках терминов i -ой ПрО. Данное значение зависит как от частоты встречаемости термина в текстах i -ой ПрО, так и от порядка следования фонем в фонетическом представлении конкретного термина:

$$R_{P_1P_2}^{A_i} = \mu_C^{A_i} \times \rho_{j,k}, \tag{3}$$

где $\mu_C^{A_i}$ – степень принадлежности термина предметной области A_i ; $\rho_{j,k}$ – вероятность следования к фонеме за j в рамках терминологии предметной области A_i .

Получение значений данных отношений производится путем проведения статистического анализа крупных корпусов текстов по заданной ПрО и фонетическом разборе каждого выделенного термина. Задача извлечения терминологии решается посредством применения семантических алгоритмов: тезаурусного и вложенных связей [10].

4. Разработка онтологии

В рамках данного исследования была разработана лингвистическая онтология предметной области «Сердечно-сосудистые заболевания» для дальнейшего распознавания слитной речи в форме надиктованных медицинских диагнозов по данной ПрО.

Онтология имеет 4 уровня иерархии, а также содержит 743 термина ПрО, большая часть из которых была извлечена из текстов автоматически посредством применения алгоритма вложенных связей [10].

На рис.1 представлен фрагмент разработанной онтологии, содержащий описание всех фонем, используемых при произношении извлеченной терминологии.

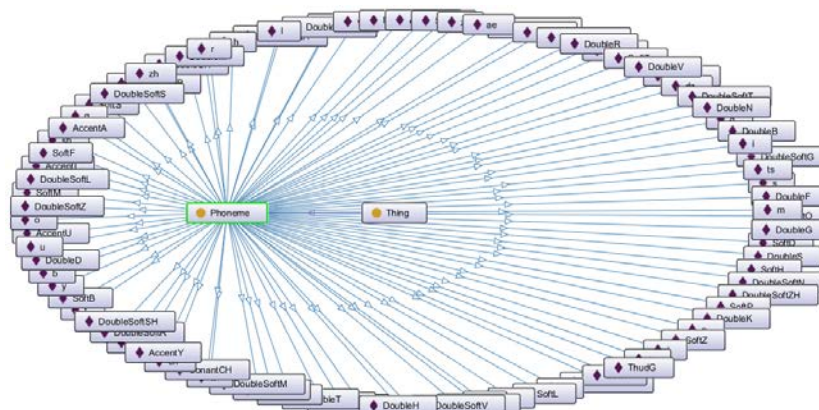


Рисунок 1. Фрагмент лексической онтологии. Фонемы.

Также, в соответствии с выражением (3), были рассчитаны значения вероятностей следования фонем, встречаемых в извлеченных терминах рассматриваемой ПрО. Данные значения позволят ограничить набор выбираемых фонем в процессе восстановления фонетического состава распознанных слов, на которые поделен речевой сигнал.

5. Заключение

Использование лексической онтологии предполагает возможность определения базового набора терминов анализируемой ПрО и, вследствие чего, повышение вероятности определения верного порядка встречаемости тех или иных фонетических сочетаний в процессе распознавания речи.

В рамках развития данного исследования планируется проведения ряда экспериментов по восстановлению фонетического состава распознанных слов с применением разработанной модели лексической онтологии предметной области «Сердечно-сосудистые заболевания» с целью подтверждения эффективности подхода.

6. Благодарности

Работа выполнена при финансовой поддержке РФФИ. Проекты № 16-48-732046 и №16-48-730305.

7. Литература

- [1] Андреев, И.А. Подход к решению задачи членения слитной речи на речевые единицы / И.А. Андреев, А.И. Армер, Н.А. Крашенинникова, В.С. Мошкин // Информационные технологии и нанотехнологии (ИТНТ-2017): сборник трудов III международной конференции и молодежной школы. – Самарский национальный исследовательский университет имени академика С.П. Королева, 2017. – С. 473-476.
- [2] Pienado, A. Speech recognition over digital channels: robustness and standards / A. Pienado, J.C. Segura // John Wiley & Sons Ltd., 2006. – P. 257.
- [3] Keshet, J. Automatic Speech and Speaker Recognition: Large Margin and Kernel Methods / J. Keshet, S. Bengio // John Wiley & Sons Ltd., 2009. – P. 253.
- [4] Gelbart, D. Ensemble Feature Selection for Multi-Stream Automatic Speech Recognition / D. Gelbart // Technical Report No. UCB/ECS-2008-160. – University of California at Berkeley, 2008. – P. 117.
- [5] Соловьев, В.Д. Онтологии и тезаурусы. Учебное пособие / В.Д. Соловьев, Б.В. Добров, В.В. Иванов, Н.В. Лукашевич. – Казань, Москва, 2006.
- [6] Fellbaum, C. WordNet: an Electronic Lexical Database / C. Fellbaum. – MIT Press, Cambridge, MA, 1998.
- [7] Мошкин, В.С. Методики построения нечетких онтологий сложных предметных областей / В.С. Мошкин, Н.Г. Ярушкина // Открытые семантические технологии проектирования

- интеллектуальных систем (OSTIS-2015): материалы V Междунар. научн.техн. конф. (Минск, 19-21 февраля 2015 г.). – Минск: БГУИР, 2015. – С. 401–406.
- [8] Хорошевский, В.Ф. Пространства знаний в сети Интернет и SemanticWeb / В.Ф. Хорошевский // Искусственный интеллект и принятие решений, 2008. – № 1.
- [9] Галунов, В.И. Современные проблемы в области распознавания речи / В.И. Галунов, А.Н. Соловьев // Информационные технологии и вычислительные системы. Вып. 2. – М.: «Рохос», 2004. – С.41-45.
- [10] Андреев, И.А. Оценка терминологичности лексических единиц на основе онтологии предметной области / И.А. Андреев, В.А. Башаев, В.В. Клейн, В.С. Мошкин, Н.Г. Ярушкина // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2015): материалы V Междунар. научн.техн. конф. (Минск, 19-21 февраля 2015 г.). – Минск: БГУИР, 2015. – С. 395-400.

The phonetic composition of the recognized speech recovery using lexical ontology

V.S.Moshkin^{1,2}, A.I. Armer¹, N.A.Krasheninnikova²

¹Ulyanovsk State Technical University, Severny Venetz str. 32, Ulyanovsk, Russia, 432027

²Ulyanovsk State University, Leo Tolstoy str. 42, Ulyanovsk, Russia, 432970

Abstract. Improving the quality of speech recognition in conditions of noisiness can be realized by adapting the algorithm for reconstructing the recognized word to the features of usage and the features of the subject area. The article describes an approach to reconstructing the phonetic composition of a recognized word by using lexical ontologies. The lexical ontology contains lexical relations between terms of the subject domain and the phonetic composition of each of these words in terms of the SAMPA + table for the Russian language.

Keywords: speech recognition, lexical ontology, phonetic composition, SAMPA+.