Ирина Александровна Пономарева — ООО „ЦРТ“, Санкт-Петербург; научный сотрудник; E-mail: ponomareva@speechpro.com

Наталья Александровна Томашенко — аспирант; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кафедра речевых информационных систем; ООО „ЦРТ“, Санкт-Петербург; младший научный сотрудник; E-mail: tomashenko-n@speechpro.com

Рекомендована кафедрой речевых информационных систем

Поступила в редакцию 22.10.13 г.

УДК 81'322.6

П. Г. Чистиков, О. Г. Хомицевич, С. В. Рыбин

СТАТИСТИЧЕСКИЕ МЕТОДЫ АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ МЕСТ И ДЛИТЕЛЬНОСТИ ПАУЗ В СИСТЕМАХ СИНТЕЗА РЕЧИ

Рассмотрены статистические методы определения местоположения и длительности пауз в системе синтеза речи. Применение таких методов позволяет добиться лучших результатов по сравнению с использованием алгоритмов, основанных на правилах.

Ключевые слова: *пауза, синтез речи, статистические модели.*

Введение. Корректная просодическая разметка в системах синтеза речи необходима для естественного звучания синтезированной речи. Обычно достаточно длинные предложения разбиваются на отдельные фрагменты, которые разделяются паузами. Такие паузы делают речь более понятной и естественной, разрешая неоднозначные трактовки смысла предложений.

Многие системы синтеза речи при определении мест пауз опираются только на знаки препинания. Однако большие участки текста, расположенные между этими знаками, могут звучать монотонно и осложнять восприятие речи, что делает актуальной задачу определения мест пауз на подобных участках. При синтезе русской речи дополнительно возникает другая проблема — знаки пунктуации традиционно используются для обособления различных вводных конструкций, таких как „может быть“, „конечно“ и т.д., которые не выделяются паузами в устной речи.

Кроме того, системы синтеза речи должны не только определять места пауз, но и их продолжительность как внутри предложений, так и между ними. Самым простым решением этой задачи является задание различных констант, регламентирующих длительность пауз. Но так как длительность естественных (в речи человека) пауз является очень вариативной величиной, необходим специальный метод, позволяющий вычислять длительность пауз в зависимости от контекста и структуры предложения [1].

Использование пауз в естественной речи зависит от ряда факторов. Наиболее значимым из них является синтаксическая структура предложения: паузы зачастую располагаются между синтаксически связными компонентами [2, 3]. Однако длина предложения, семантика определенных слов и другие особенности также имеют значение [4]. В системах синтеза речи эти факторы могут быть учтены путем задания правил, определяющих, после какого слова в предложении должна стоять пауза [5, 6], или путем обучения статистических моделей на большом речевом корпусе, на основе которых будут вычисляться вероятности наличия пауз после того или иного слова [7, 8].

В системе синтеза русской речи компании ООО „ЦРТ“ для определения мест пауз применяются алгоритмы, основанные на правилах, а длительность пауз определяется на основе определенных констант [9]. Такой подход достаточно хорош, однако невозможно учесть все случаи, встречающиеся в различных текстах, помимо того, разработка подобных правил для новых (для системы) языков требует большого количества времени. Преимущество методов машинного обучения — простота применения, при наличии размеченного речевого корпуса достаточного объема. Ожидается, что статистические модели будут более детально эмитировать поведение человека, нежели правила, основанные на знаниях экспертов. В настоящей работе исследуются пути, позволяющие улучшить алгоритм, основанный на правилах, путем использования статистических методов и просодического анализа.

Классификаторы CART и RF. Классификатор CART [10] применяется для определения местоположения и длины пауз: определяется длительность паузы между словами (там, где она равна нулю или меньше заданного порога, пауза отсутствует). Также этот классификатор использовался только для определения длины пауз: в этом случае предсказывается длительность паузы только между теми словами, где она была поставлена на предыдущих этапах обработки текста. Классификатор RF [11] применялся только для определения местоположения пауз.

CART — рекурсивный метод разбиения набора данных на основе минимизации функции:

$$G(C_1, C_2) = \frac{D(C_1)T(C_1) + D(C_2)T(C_2)}{T(C_1) + T(C_2)}, \qquad (1)$$

где

$$D(C) = \frac{\sum_{i=1}^{|C|}\sum_{j=i}^{|C|} d(U_i, V_j)}{T(C)},$$

$$T(C) = \frac{1}{2}|C|\left(|C|-1\right).$$

$|C|$ — размер кластера C, $d(U,V)$ — расстояние между векторами признаков U и V, критерием останова служит минимальное число элементов в кластере (в настоящей работе — три).

RF выполняет классификацию данных на основе множества признаков путем создания иерархии („деревьев“) запросов на основе предсказанных значений признаков в каждой точке. Лист каждого дерева содержит информацию обо всех наблюдениях характеризуемой величины, признаки которой лежат в одной области значений. В разработанной экспериментальной системе применяется „лес решений“, содержащий 100 деревьев, где каждое дерево построено на 60 % случайно выбранных данных, это снижает чувствительность алгоритма к шуму в обучающих данных.

Описание эксперимента. Использовался размеченный звуковой корпус, состоящий из записей девяти дикторов (4 мужчин и 5 женщин). Записанный материал — русскоязычная художественная литература и новостные статьи. Итоговая база данных содержит более 50 часов речи, включая 38 000 пауз, для выполнения тестов она была разделена на тестовую и обучающую выборки в отношении 85 к 15 % соответственно.

При обучении модели определения местоположения учитывались только паузы внутри предложений исходя из того, что границы предложений известны (текст с высокой точностью разделяется на предложения процедурой нормализации текста, встроенной в систему синтеза речи). Для определения длительности модель обучалась на основе данных о паузах как внутри предложений, так и между ними.

Для решения задачи классификации использовались

— пунктуация: знак препинания после текущего слова, после двух предыдущих и после двух последующих слов;

— число слов и слогов: слов и слогов в предложении, слов и слогов от предыдущей паузы до текущего слова и от текущего слова до конца предложения;

— грамматические признаки: часть речи, падеж, признак „является ли слово собственным существительным“ (имена, названия и т.д.) — данная информация получается с помощью морфологического словаря, входящего в состав системы синтеза речи;

— признаки согласования: согласуется ли грамматическая форма текущего слова с формой двух последующих слов;

— регистр первой буквы слова: является ли она в двух предыдущих, в текущем или двух последующих словах заглавной или нет.

Как при обучении, так и при тестировании, для снижения числа ошибок вычисления грамматических признаков необходима процедура разрешения неоднозначности для слов-омонимов и омографов (замОк — зАмок). В настоящей статье используется предложенный в работе [12] подход, точность работы которого составляет 96 %.

Определение местоположения пауз. В табл. 1 сравниваются результаты использования метода автоматического определения мест пауз (CART и RF) с результатами работы базового, основанного на правилах, подхода, встроенного в систему синтеза русской речи, использованную при выполнении настоящей работы. Тестовая выборка содержит 47 819 пар слов внутри предложения, между которыми возможна пауза, и 6186 пауз (аналогичные показатели для обучающей выборки равняются 264 336 и 32 630 соответственно).

Таблица 1

Результаты определения местоположения пауз

№	Параметр	Базовый алгоритм	CART	RF
1	Правильных границ между словами	43 254 (90 %)	44 358 (93 %)	44 865 (94 %)
2	Правильных пауз	5042 (82 %)	5176 (84 %)	4695 (76 %)
3	Ложных срабатываний	3421 (55 %)	2451 (40 %)	1463 (24 %)
4	Ложных отклонений	1144 (18 %)	1010 (16 %)	1491 (24 %)
5	Эффективность, %	82	84	76
6	Точность, %	60	68	76
7	F-мера, %	69	75	76

В табл. 1 в строке № 1 указано число корректно идентифицированных „пауз“ или „не пауз“ среди всех пар слов тестовой выборки; в строке № 2 указаны корректно определенные паузы. Эта мера использовалась в работе [7] и приведена здесь для сравнения, в равной степени как и показатели ложных срабатываний (FA), ложных отклонений (FR) и F-меры (F-score). Результаты работы обоих классификаторов превосходят результаты работы базовой версии алгоритма: значение F-меры больше, а показатели ложных срабатываний и ложных отклонений более сбалансированы. Результаты работы CART и RF различаются незначительно, однако, что немаловажно, уровень FA/FR при применении RF может быть настроен произвольным образом.

Результаты работы сравнимы с результатами, описанными в литературе. Так, например, для английского языка количество „правильных границ между словами“ составляет 91,1 %, а F-мера равна 71,9 % [7]. В работе [8] данный показатель составляет 74,4 %.

Определение длительности пауз. При проведении экспериментов по определению длительности пауз классификатор изначально был обучен для работы со всеми типами пауз. Затем было решено определять длительность пауз между предложениями и внутри них отдельно. Для представления результатов работы системы использовалась мера NRMSD (нормализованное среднее квадратичное отклонение, Normalized Root-Mean-Square Deviation), схожая с мерой, применяемой в работе [1]; чем это значение ближе к нулю, тем результат лучше.

В табл. 2 приведены результаты применения обобщенной (создана для определения всех пауз в наборе данных) и специализированных моделей (две различные модели для определения пауз внутри и между предложениями).

Таблица 2

Результаты определения длительности пауз

Модель	Паузы внутри предложения	Паузы между предложениями
Обобщенная	0,25	0,23
Специализированные	0,19	0,16

Проанализировав таблицу, можно сделать вывод, что специализированные модели позволяют более точно описать параметры пауз как между предложениями, так и внутри них.

Интеграция с системой синтеза речи. Как было сказано выше, базовая версия алгоритма определения местоположения пауз основана на правилах, а длительность пауз задается соответствующими константами. Паузы делятся на четыре типа в зависимости от их длины: один тип пауз между предложениями и три — внутри предложений. Такой подход является достаточно статичным для ритмики предложений. Как следствие, было решено использовать алгоритм, основанный на статистических методах, в первую очередь были внедрены модели, на основе которых отдельно определялась длительность пауз внутри предложений и между ними. Затем был интегрирован статистический подход для определения мест пауз. В итоге была получена система, работа которой включает следующие этапы:

1) расстановка пауз согласно знакам препинания (за исключением мест, где знаки препинания не предполагают пауз);

2) разбиение паузами длинных цепочек слов без знаков препинания на основе статистической модели (CART или RF);

3) определение длительности полученных пауз на основе статистики.

Сравнив подходы с использованием классификаторов CART и RF, можно отметить следующее. Очевидным преимуществом использования CART является маленький размер модели, что является важным показателем при реализации системы синтеза речи. Однако, как показано в табл. 1, RF дает лучшие результаты при определении местоположения пауз. Более того, не все ошибки одинаково критичны: в некоторых случаях пауза недопустима, в то время как в других имеет право быть. В работе CART возникают более серьезные ошибки по сравнению с RF, хотя это может быть выявлено только на основе экспертных оценок. В основном ошибки CART заключаются в паузах внутри синтаксически связанных цепочек: после предлогов, союзов и других служебных слов, использующихся для связи последовательности слов; между модификатором (прилагательное, наречие и т.д.) и существительным или глаголом, к которому он относится. Такого рода ошибки практически отсутствуют при использовании классификатора RF. Помимо того, модель с использованием RF является более гибкой, поскольку она может быть настроена с целью увеличения или уменьшения количества пауз в синтезируемой речи, что может быть полезно для практических приложений системы синтеза речи. Например, увеличение количества пауз снижает темп речи.

Сравнение результатов работы полученной экспериментальной системы с базовой — сложная задача, она требует применения MOS-оценки (Mean Opinion Score), что является целью будущих исследований. Однако для определения предпочтений слушателей был проведен тест. Было выбрано 25 предложений, содержащих большие последовательности слов, не разделенные знаками препинания. Результаты оценивались 18 русскоязычными экспертами: 10 предпочли предложенную систему, 4 не смогли определить, что лучше, 4 предпочли базовую.

Заключение. В работе представлен подход к определению местоположения и длительности пауз в системе синтеза речи на основе статистических моделей. Результаты экспериментов показали, что модели, построенные на основе алгоритмов классификации данных CART и RF, в сравнении с основанным на правилах подходом дают лучшие результаты.

Модель CART допускает более критические, по сравнению с RF ошибки, в основном заключающиеся в лишних паузах между синтаксически связанными участками текста. Таким образом, в системе синтеза речи предпочтительно использовать модель RF. Такая экспериментальная система получила положительную оценку слушателей-экспертов.

Был проведен ряд экспериментов по определению длительности пауз на основе алгоритма CART: обнаружено, что он работает лучше в случае применения различных моделей для предсказаний длительности пауз внутри предложений и между ними. На основе экспертных оценок был сделан вывод, что это решение позволяет повысить естественность синтезированной речи.

Работа выполнена при государственной финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01).

СПИСОК ЛИТЕРАТУРЫ

1. *Parlikar A., Black A. W.* Modeling Pause-Duration for Style-Specific Speech Synthesis // Proc. of Interspeech. Portland, OR, USA, 2012. P. 446—449.
2. *Bachenko J., Fitzpatrick E.* A computational grammar of discourse-neutral prosodic phrasing in English // Computational linguistics. 1990. Vol. 16 (3). P. 155—170.
3. *Tepperman J., Nava E.* Where should pitch accents and phrase breaks go? A syntax tree transducer solution // Proc. of Interspeech. Florence, Italy, 2011. P. 1353—1356.
4. *Zellner B.* Pauses and the temporal structure of speech // Fundamentals of speech synthesis and speech recognition / Ed. by *E. Keller*. Chichester: John Wiley, 1994. P. 41—62.
5. *Abney S.* Parsing by chunks // Principle-Based Parsing Computation and Psycholinguistics. 1991. Vol. 44. P. 257—278.
6. *Atterer M.* Assigning Prosodic Structure for Speech Synthesis: A Rule-based Approach // Proc. of Speech Prosody. Aix-en-Provence, 2002. P. 147—150.
7. *Black A. W., Taylor P.* Assigning phrase breaks from part-of-speech sequences // Computer Speech & Language. 1998. Vol. 12, N 2. P. 99—117.
8. *Busser B., Daelemans W., Bosch A. V. D.* Predicting phrase breaks with memory-based learning // Proc. of 4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis. 2001. P. 29—34.
9. *Хомицевич О. Г., Соломенник М. В.* Автоматическая расстановка пауз в системе синтеза русской речи по тексту // Компьютерная лингвистика и интеллектуальные технологии: Матер. Междунар. конф. „Диалог“. М.: Изд-во РГГУ, 2010. Вып. 9 (16). С. 531—537.
10. *Loh W.-Y.* Classification and Regression Tree Methods // Encyclopedia of Statistics in Quality and Reliability. Wiley, 2008. P. 315—323.
11. *Breiman L. Cutler A.* Random Forests [Электронный ресурс]: <http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm>.
12. *Хомицевич О. Г., Рыбин С. В., Аничкин И. М.* Использование лингвистического анализа для нормализации текста и снятия омонимии в системе синтеза русской речи // Изв. вузов. Приборостроение. 2013. Т. 56, № 2. С. 42—46.

Сведения об авторах

Павел Геннадьевич Чистиков — ООО „ЦРТ“, Санкт-Петербург; научный сотрудник; E-mail: chistikov@speechpro.com

Ольга Гурьевна Хомицевич — PhD; ООО „ЦРТ“, Санкт-Петербург; старший научный сотрудник; E-mail: khomitsevich@speechpro.com

Сергей Витальевич Рыбин — канд. физ.-мат. наук, доцент; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кафедра речевых информационных систем; E-mail: rybin@speechpro.com

Рекомендована кафедрой речевых информационных систем

Поступила в редакцию 22.10.13 г.