УНИВЕРСИТЕТ ИТМО
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
июль–август 2017 Том 17 № 4 ISSN 2226-1494 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
July–August 2017 Vol. 17 No 4 ISSN 2226-1494 http://ntv.ifmo.ru/en

УДК 004.934.5

МЕТОД АВТОМАТИЧЕСКОЙ РАССТАНОВКИ ПАУЗ ДЛЯ КАЗАХСКОГО ЯЗЫКА

А. Калиев[a]

[a] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
Адрес для переписки: kaliyev.arman@yandex.kz

Информация о статье
Поступила в редакцию 19.05.17, принята к печати 07.06.17
doi: 10.17586/2226-1494-2017-17-4-749-752
Язык статьи – русский

Ссылка для цитирования: Калиев А. Метод автоматической расстановки пауз для казахского языка // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 4. С. 749–752. doi: 10.17586/2226-1494-2017-17-4-749-752

Аннотация

Предложен новый метод паузации для систем синтеза интонационной речи, основанный на анализе дистрибутивной семантики в больших текстовых корпусах. Для предсказывания паузы использовался классификатор на основе метода опорных векторов и два речевых корпуса на казахском языке. Предсказание мест паузации проводилось на уровне биграмм, где входными параметрами биграммы служили векторные представления обоих ее лексем и их битовое представление в кластерной модели Брауна. Проведенные исследования показали, что предложенный метод паузации для систем автоматического синтеза казахской речи в повествовательном стиле обеспечивает расстановку пауз с высокой точностью. Экспериментально подтверждена важность использования однородных данных для решения такого рода задач. Предложенный подход может быть использован при создании систем автоматического синтеза речи для множества языков.

Ключевые слова

синтез речи, паузы, кластеризация, текстовый корпус, просодика

Благодарности

Исследование выполнено в рамках научно-исследовательской работы «Разработка и исследование методов и алгоритмов распознавания эмоционального и психофизического состояния человека по многомодальным данным» и поддержано Грантом Правительства Российской Федерации № 616029.

METHOD OF AUTOMATIC PAUSE PLACEMENT FOR KAZAKH LANGUAGE

A. Kaliyev[a]

[a] ITMO University, Saint Petersburg, 197101, Russian Federation
Corresponding author: kaliyev.arman@yandex.kz

Article info
Received 19.05.17, accepted 07.06.17
doi: 10.17586/2226-1494-2017-17-4-749-752
Article in Russian

For citation: Kaliyev A. Method of automatic pause placement for Kazakh language. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2017, vol. 17, no. 4, pp. 749–752 (in Russian). doi: 10.17586/2226-1494-2017-17-4-749-752

Abstract

The paper considers a new pausing method for intonational speech synthesis systems based on the analysis of distributional semantics in large text corpora. The support vector machine and two speech corpora in Kazakh were used for pause prediction. The prediction of pause places was carried out at the level of bigrams, where the input parameters of the bigram were the vector representations of both of its words and their bit string representation in the Brown cluster model. The carried out studies have shown that the proposed pausing method for the automatic speech synthesis systems for the Kazakh language in the narrative style provides high accuracy of pause placement. The importance of homogeneous data usage was confirmed experimentally for solving such problems. Such approach can facilitate the creation of automatic speech synthesis for many languages.

Keywords

speech synthesis, pauses, clustering, text corpora, prosody

Acknowledgements

This work was financially supported by the Government of the Russian Federation, Grant No. 616029.

Для успешного использования систем автоматического синтеза речи необходима высокая естественность речи. Паузы, наряду с интонационным оформлением и ударением, являются одной из важнейших просодических характеристик речи, обеспечивающих ее естественность. Корректная паузация необходима для комфортности восприятия речи, а во многих случаях – и для правильного понимания смысла предложения.

В настоящей работе предлагается способ расстановки пауз для автоматического синтеза речи на основе параметров лексических представлений, полученных из кластерной модели Брауна и др. [1], и векторных представлений слов, полученных алгоритмом Стратоса и др. [2]. Кластеризация Брауна представляет собой форму иерархической кластеризации слов, основанной на распределении скрытых марковских моделей.

Приведем пример стандартной реализации алгоритма кластеризации Брауна.

Входные данные: корпус из n слов $\{w^1,..,w^n\}$, упорядоченных в порядке убывания по частоте появления; количество кластеров m.

Выходные данные: иерархический кластер слов $w^1,..,w^n$.

Шаг 1. Инициализация активных кластеров $C = \{\{w^1\},\dots,\{w^m\}\}$.

Шаг 2. For $i = m+1\ to\ n+m-1$:

Шаг 2a. If $i \le n$: $C = C \cup \{\{w^i\}\}$.

Шаг 2b. Выбираем два кластера c и $\grave{c} \in C$ и объединяем их в один кластер таким образом, чтобы на каждом шаге максимизировать функцию $Quality(C)$.

Функция $Quality(C)$ вычисляет вероятность разделения слов на кластеры C для входного корпуса.

Входные данные: корпус из n слов $\{w^1,..,w^n\}$, позиции слов в корпусе остаются без изменений; активные кластеры C.

Выходные данные: вероятностная оценка разделения слов на активные кластеры C.

$$Quality(C) = p(w^1,..,w^n) = \sum_{i=1}^{n} \log\left(e\left(w_i\middle|C(w_i)\right)q\left(C(w_i)\middle|C(w_{i-1})\right)\right),$$

где для $w_i \in \{w^1,..,w^n\}$ и $c, \grave{c} \in C$,

$e(w_i|c)$ – вероятность принадлежности слова w_i к кластеру c,

$q(c|c')$ – вероятность перехода с кластера $\grave{c}$ в кластер c.

В результате кластеризации каждое слово в кластере получает битовое представление, которое указывает путь до кластера от корня (рисунок).

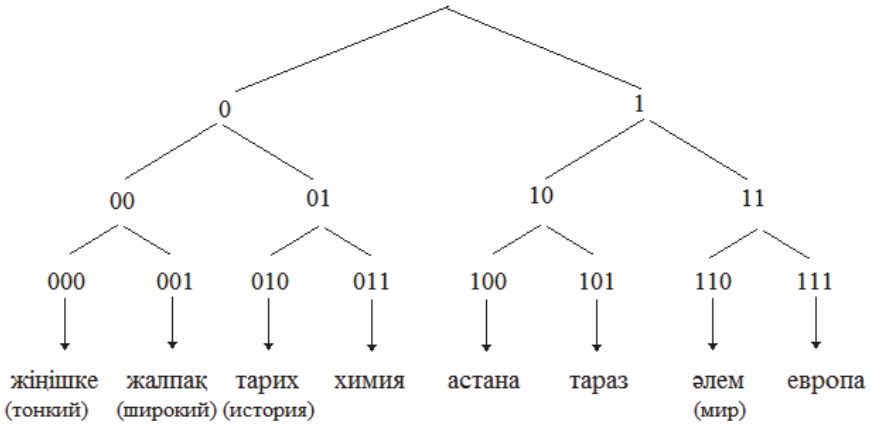

Рисунок. Пример битового представления слов, полученных с помощью алгоритма Брауна

Битовое представление слов применялось ранее для решения многих задач обработки естественных языков, в том числе распознавания именных сущностей [3] и синтаксического анализа зависимостей [4]. Также векторные представления слов получены с помощью алгоритма анализа канонической корреляции слов, вычисленных на основе матрицы совместного появления слов [5]. Особенностью данного алгоритма является то, что он способен эффективно и быстро извлекать векторы небольшой размерности, сохраняя корреляционное отношение между словами.

Алгоритм анализа канонической корреляции.

Входные данные: вектора $x^i \epsilon \mathbb{R}^d$ и $y^i \epsilon \mathbb{R}^g$, полученные с помощью матрицы совместного появления слов.

Выходные данные: матрицы проекции $A \in \mathbb{R}^{d\times p}$ и $B \in \mathbb{R}^{g\times p}$ для $i = 1,\dots,n$

$x^{(i)} \in \mathbb{R}^d \rightarrow \hat{x}^{(i)} = A^T x^{(i)}, \quad \hat{x}^{(i)} \in \mathbb{R}^p,$

$y^{(i)} \in \mathbb{R}^g \rightarrow \hat{y}^{(i)} = B^T y^{(i)}, \quad \hat{y}^{(i)} \in \mathbb{R}^p.$

Шаг 1. Вычисление ковариационных матриц $\hat{C}_{XY} \in \mathbb{R}^{d\times g}$, $\hat{C}_{XX} \in \mathbb{R}^{d\times d}$ и $\hat{C}_{XY} \in \mathbb{R}^{g\times g}$.

$$\left[\hat{C}_{XY}\right]_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_j^{(i)} - \bar{x}_j)(y_k^{(i)} - \bar{y}_k),$$

где

$$\bar{x} = \sum_i \frac{x^{(i)}}{n},\ \bar{y} = \sum_i \frac{y^{(i)}}{n},$$
$$\left[\hat{C}_{XX}\right]_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_j^{(i)} - \bar{x}_j)(x_k^{(i)} - \bar{x}_k),$$
$$\left[\hat{C}_{YY}\right]_{jk} = \frac{1}{n}\sum_{i=1}^{n}(y_j^{(i)} - \bar{y}_j)(y_k^{(i)} - \bar{y}_k).$$

Шаг 2. Применение сингулярного разложения для $\hat{C}_{XX}^{\ -1/2}\hat{C}_{XY}\hat{C}_{YY}^{\ -1/2} \in \mathbb{R}^{d\times g}$:

$$\hat{C}_{XX}^{\ -1/2}\hat{C}_{XY}\hat{C}_{YY}^{\ -1/2} \overset{SVD}{\Longrightarrow} U\Sigma V^T.$$

Пусть $U_p \in \mathbb{R}^{d\times p}$ будут первыми p значениями U, а $V_p \in \mathbb{R}^{g\times p}$ будут первыми p значениями V.

Шаг 3. Определение матриц проекций $A \in \mathbb{R}^{d\times p}$ и $B \in \mathbb{R}^{g\times p}$:

$$A = \hat{C}_{XX}^{\ -1/2}U_p$$
$$B = \hat{C}_{YY}^{\ -1/2}V_p.$$

Предсказания мест паузации проводилось на уровне биграмм. Биграмма – это два слова, которые в текстовом корпусе являются соседними. Входными параметрами биграммы служили векторные представления обоих слов, их битовое представление в кластерной модели Брауна и сами слова. Для классификации биграмм использовался метод опорных векторов [6]. Для получения векторного и битового представлений использовался текстовый корпус казахской Википедии, насчитывающий более 1,7 млн предложений, или 20 млн слов.

Для объективной оценки модели использовалась мера F1 [7]. Мера F1 – это среднее гармоническое значение точности и полноты классификатора. Чем лучше модель, тем ближе значение F1 к 1.

Основной размеченный корпус состоит из записей речи в нейтральном тоне одного диктора (женщины). Было записано 596 предложений, в среднем в каждом предложении 10,4 слов. Общее количество пауз в корпусе 757 или 12,09% по отношению к количеству лексем в корпусе. Учитывались только паузы внутри предложений (синтагм).

Второй корпус состоит из нескольких часов записей казахской речи дикторами разных возрастных групп и разных полов. Общее количество записанных предложений около 7000.

Во время экспериментов 80% корпуса использовалось для обучения и 20% – для тестирования. Для получения адекватной оценки модели производилось 10 экспериментов с одними и теми же параметрами, и в каждом эксперименте предложения для обучения и тестирования классификатора выбирались случайным образом. Конечная оценка F1 является усредненной оценкой по всем 10 экспериментам. Среднее значение F1 получилось равным 0,781 для корпуса с казахской речью одного диктора.

Аналогичные исследования были проведены и на смешенном корпусе, значение F1 = 0,406. Столь низкие результаты объясняются стилистическим различием речи самих дикторов. Так, молодые дикторы, чаще женского пола, имели тенденцию говорить быстро, уменьшая количество пауз в предложениях, тогда как более взрослые дикторы говорили медленнее, чаще вставляя паузы между словами.

П. Чистиков и др. [8] решали похожую задачу для русского языка, используя для этих целей базу из 38000 записанных предложений и морфологический анализатор. Согласно их результатам, была достигнута точность расстановки пауз F1 = 0,76. Из этого следует, что предложенный метод показывает сопоставимую точность предсказывания пауз, используя для этих целей параметры, полученные из больших текстовых корпусов.

Для множества языков, в том числе казахского, для которых морфологические классификаторы еще не разработаны, предложенный подход является вполне приемлемым. В ближайшее время предполагается развить предлагаемый метод для предсказания длин пауз и определения интонационных контуров синтагм.

Литература

1. Brown P.F., Desouza P.V., Mercer R.L. et. al. Class-based n-gram models of natural language // Computational Linguistics. 1992. V. 18. P. 467–479.
2. Stratos K., Kim D., Collins M., Hsu D. A spectral algorithm for learning classbased n-gram models of natural language // Proc. 30th Conf. on Uncertainty in Artificial Intelligence. Quebec, Canada, 2014. P. 762–771.
3. Miller S., Guinness J., Zamanian A. Name tagging with word clusters and discriminative training // Proc. Human Language Technologies and North American Association for Computational Linguistics. 2004. V. 4. P. 337–342.
4. Koo T., Carreras X., Collins M. Simple semi-supervised

References

1. Brown P.F., Desouza P.V., Mercer R.L. et. al. Class-based n-gram models of natural language. *Computational Linguistics*, 1992, vol. 18, pp. 467–479.
2. Stratos K., Kim D., Collins M., Hsu D. A spectral algorithm for learning classbased n-gram models of natural language. *Proc. 30th Conf. on Uncertainty in Artificial Intelligence*. Quebec, Canada, 2014, pp. 762–771.
3. Miller S., Guinness J., Zamanian A. Name tagging with word clusters and discriminative training. *Proc. Human Language Technologies and North American Association for Computational Linguistics*, 2004, vol. 4, pp. 337–342.
4. Koo T., Carreras X., Collins M. Simple semi-supervised

dependency parsing // Proc. 46th Annual Meeting of the Association for Computational Linguistics, ACL-08: HLT. Columbus, USA, 2008. P. 595–603.
5. Lancia F. Word Co-occurrence and Theory of Meaning. 2005. URL: www.soc.ucsb.edu/faculty/mohr/classes/soc4/summer_08/pages/Resources/Readings/TheoryofMeaning.pdf (дата обращения: 25.04.2017).
6. Cortes C., Vapnik V. Support vector networks // Machine Learning. 1995. V. 20. N 3. P. 273–297. doi: 10.1023/A:1022627411411
7. Rijsbergen C.J.V. Information Retrieval. 2nd ed. London: Butterworths, 1979. 152 p.
8. Chistikov P.G., Khomitsevich O.G. Improving prosodic break detection in a Russian TTS system // Lecture Notes in Computer Science. 2013. V. 8113. P. 181–188. doi: 10.1007/978-3-319-01931-4_24

dependency parsing. *Proc. 46th Annual Meeting of the Association for Computational Linguistics, ACL-08: HLT*. Columbus, USA, 2008, pp. 595–603.
5. Lancia F. *Word Co-occurrence and Theory of Meaning*. 2005. Available at: www.soc.ucsb.edu/faculty/mohr/classes/soc4/summer_08/pages/Resources/Readings/TheoryofMeaning.pdf (accessed: 25.04.2017).
6. Cortes C., Vapnik V. Support vector networks. *Machine Learning*, 1995, vol. 20, no. 3, pp. 273–297. doi: 10.1023/A:1022627411411
7. Rijsbergen C.J.V. *Information Retrieval*. 2nd ed. London, Butterworths, 1979, 152 p.
8. Chistikov P.G., Khomitsevich O.G. Improving prosodic break detection in a Russian TTS system. *Lecture Notes in Computer Science*, 2013, vol. 8113, pp. 181–188. doi: 10.1007/978-3-319-01931-4_24

Авторы

Калиев Арман – аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, kaliyev.arman@yandex.kz

Authors

Arman Kaliyev – postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, kaliyev.arman@yandex.kz