

СЕКЦИЯ «ЕСТЕСТВЕННЫЕ И ТОЧНЫЕ НАУКИ: Математика»

УДК 510.63

КЛАСТЕРИЗАЦИЯ ДОКУМЕНТОВ НА ОСНОВЕ ГЕНЕТИЧЕСКОГО АЛГОРИТМА

Бондарев Я. А.

Филиал МГУ имени М. В. Ломоносова в городе Севастополе

Целью работы является адаптация классического генетического алгоритма [1] и его оптимизация для задачи кластеризации текстовых документов. Полученный результат можно применять в задачах интеллектуального поиска, структурирования документооборота.

Классический генетический алгоритм принадлежит классу алгоритмов машинного обучения с подкреплением. Обычно он применяется в задачах оптимизации или в случае наличия среды, с которой можно взаимодействовать. В решаемой задаче средой является некоторый набор текстовых документов, а взаимодействием – отнесение конкретного документа к определенному множеству (кластеру).

В работе приведен возможный способ кодирования исходных данных для последующей обработки генетическим алгоритмом, и все необходимые операторы для реализации генетического алгоритма (селекция, репродукция, мутация, редукция, фитнес-функция).

Также были реализованы надстройки, позволяющие выполнять интеллектуальный поиск в сети Интернет. Под интеллектуальным поиском подразумевается соответствие семантики исходного запроса к поисковому сервису и полученной выдачи.

Например, документы, полученные по запросу «строительный магазин в Севастополе» должны быть действительно страницами различных магазинов, занимающихся реализацией стройматериалов и расположенных в Севастополе. При этом документы-справочники, карты, рекламные предложения будут проигнорированы.

В итоге был построен и реализован плоский нечеткий генетический алгоритм для кластеризации текстовых документов, реализована надстройка для применения результатов исследования на практике.

Список литературы

1. John H. Holland. Genetic Algorithms. Computer programs that "evolve" in ways that resemble natural selection can solve complex problems even their creators do not fully understand. [Электронный ресурс]. URL: https://www.cc.gatech.edu/~turk/bio_sim/articles/genetic_algorithm.pdf.

УДК 004.9:612.821

РАСПОЗНАВАНИЕ ФОНЕМ ПРИ ВНУТРЕННЕМ ПРОГОВАРИВАНИИ НА ОСНОВЕ ДАННЫХ ЭЭГ

Гавриленко Ю.Ю., Саада Д.Ф., Шевченко А.О.

Филиал МГУ имени М.В. Ломоносова в городе Севастополе

МГУ имени М.В. Ломоносова, г. Москва

Речь – стандартный метод общения людей в повседневной жизни. Однако, многие люди имеют проблемы с воспроизведением речи. Это может быть вызвано особенностями нейродегенеративных заболеваний, например, таких как боковой амиотрофический склероз. Также, устная речь не всегда может быть подходящим способом коммуникации в различных военных сценариях. Для решения таких проблем может быть применен интерфейс мозг-компьютер (ИМК) – система, поддерживающая взаимодействие мозга и электронного устройства, в частности, интерфейс безмолвного доступа – система, позволяющая распознавать внутреннее проговаривание.

Известно, что восстановить внутреннее проговаривание по данным биоэлектрической активности мозга возможно, поскольку при внутреннем проговаривании генерируется активность артикуляторов (губ, языка, гортани, нёба), нервных путей между ними и самого головного мозга [1]. Проводившиеся на эту тему исследования касаются распознавания английского, испанского, японского и некоторых других языков – для русского языка таких экспериментов не проводилось, что делает данную работу актуальной.

На сегодняшний день одним из популярных нейроинтерфейсов, которым пользуются люди, страдающие от нейродегенеративных заболеваний, являются интерфейсы, основанные на событийном потенциале P300 – электрофизиологической реакции человека на стимул, возникающей через 300 миллисекунд после его предъявления. Несмотря на свою эффективность, подобные интерфейсы не позволяют достичь высокой скорости генерации речи, необходимой для комфортного повседневного общения, а также требуют от пользователя приобретения навыков использования таких интерфейсов. Интерфейсы безмолвного доступа позволили бы пользователям общаться, не имея дополнительных навыков, поскольку для того, чтобы произнести букву с помощью такого интерфейса достаточно подумать о ней.

Разработка интерфейсов безмолвного доступа находится в экспериментальной стадии. Существуют проблемы, в решении которых исследователи еще не пришли к общему мнению. Одной из таких проблем

является выбор технологии отведения потенциалов, позволяющей достичь оптимального пространственно-временного разрешения и исключающего инвазивность. Другой проблемой является выбор фонологических категорий, распознавание которых будет проведено в эксперименте. Так, исходя из проведенных исследований, известно, что распознавание слов целиком является менее эффективным, чем распознавание фонем [2, 3, 4, 5].

Данная работа посвящена распознаванию фонем русского языка [а], [б], [м], [р], [з], [у], [ф] при внутреннем проговаривании. На первом этапе проводится эксперимент, в котором регистрируется активность головного мозга во время внутреннего проговаривания. Для регистрации биоэлектрической активности головного мозга используется электроэнцефалография, поскольку это неинвазивный метод, являющийся достаточно простым в использовании и позволяющий достичь высокого временного разрешения. Эксперимент проводится в лаборатории психофизиологии факультета психологии МГУ им. М.В.Ломоносова. Далее, собранные данные проходят этап предварительной обработки. К ним применяется полосовой фильтр для отсекающих высоких частот, не имеющих значение в проводимом эксперименте, и алгоритм нормализации, после чего полученные данные разбиваются на эпохи. К обработанным данным применяются методы извлечения признаков, которые наделяют их уникальными характеристиками и преобразовывают к виду, понятному для классификатора. На последнем этапе применяются алгоритмы классификации.

В данной работе была проведена серия экспериментов. Для решения задачи распознавания используются методы машинного обучения. Были разработаны алгоритм извлечения признаков, а также алгоритм классификации фонем на основе полученных признаков. На данном этапе разработка ИМК распознавания внутреннего проговаривания для фонем русского языка, приемлемого по качеству распознавания и удобству использования в повседневной жизни, не достигнута, однако, данное направление активно развивается как с технической, так и с теоретической точки зрения и является перспективным направлением для исследования.

Список литературы

1. Suppes, Patrick & Lu, Zhong-Lin & Han, Bing. (1998). Brain Wave Recognition of Words. Proceedings of the National Academy of Sciences of the United States of America. 94. 14965-9. 10.1073/pnas.94.26.14965. URL: https://www.researchgate.net/publication/13823956_Brain_Wave_Recognition_of_Words.
2. H. Yamaguchi, T. Yamazaki, K. Yamamoto, S. Ueno, A. Yamaguchi, T. Ito, S. Hirose, K. Kamijo, H. Takayanagi, T. Yamanoi et al., "Decoding silent speech in Japanese from single trial EEGs: Preliminary results," Journal of Computer Science & Systems Biology, vol. 2015, 2015. URL: <https://www.omicsonline.org/open-access/decoding-silent-speech-in-japanese-from-single-trial-EEGs-preliminary-results-jcsb-1000202.php?aid=60262>.
3. A.R. Sereshkeh, R. Trott, A. Bricout, T. Chau, Online EEG Classification of Covert Speech for Brain-Computer Interfacing, International journal of neural systems, 27 (2017) 1750033. URL: <https://www.worldscientific.com/doi/abs/10.1142/S0129065717500332>.
4. Brigham, K., Kumar, B.V.: Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy. In: 2010 4th International Conference on Bioinformatics Biomedical Engineering (iCBBE), pp. 1-4 (2010).
5. Kamalakkannan, R., Rajkumar, R.: Imagined speech classification using EEG. Adv. Biomed. Sci. Eng. 1(2), 20-32 (2014). URL: https://www.researchgate.net/publication/309967859_Imagined_Speech_Classification_using_EEG.

УДК 004.942

ПОСТРОЕНИЕ ПОТОКА НА СИМВОЛИЧЕСКОМ ОБРАЗЕ

Кукушкин К.В.

Филиал МГУ имени М.В. Ломоносова в городе Севастополе

Современность устроена таким образом, что сферами производства и управления ежедневно выдвигается ряд реальных задач, которые, разумеется, требуют своего решения. Среди возможных подходов к такому решению следует выделить метод математического моделирования, который подразумевает представление реального объекта в виде некоей целостной системы, являющейся, по сути, математической абстракцией, и дальнейшее применение математического аппарата к смоделированному объекту. Одним из инструментов математического моделирования является компьютерное моделирование, позволяющее с помощью программных средств представить исследуемый объект и произвести с ним необходимые вычисления. В ходе анализа результатов подсчетов и переноса выводов на изначальный исследуемый объект может быть получено решение поставленной задачи.

Цель данной работы – построение и визуализация потока, удовлетворяющего определенным в [4, с. 3] условиям, на графе, представляющем собой символический образ динамической системы [3, с. 13]. В качестве отображения f было выбрано отображение Икеды:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0,6 - 0,9(x \cos \tau - y \sin \tau) \\ 0,9(x \sin \tau + y \cos \tau) \end{pmatrix}, \text{ где}$$