

DOI: 10.32347/2412-9933.2020.43.123-

УДК 004.056.5

Чернышев Денис ОлеговичДоктор технических наук, профессор, первый проректор, orcid.org/0000-0002-1946-9242

Киевский национальный университет строительства и архитектуры, Киев

Михайленко Виктор МефодиевичДоктор технических наук, профессор, заведующий кафедрой информационных технологий проектирования и прикладной математики, orcid.org/0000-0002-9573-9873

Киевский национальный университет строительства и архитектуры, Киев

Терейковская Людмила АлексеевнаКандидат технических наук, доцент, доцент кафедры информационных технологий проектирования и прикладной математики, orcid.org/0000-0002-8830-0790

Киевский национальный университет строительства и архитектуры, Киев

РАЗРАБОТКА И ИССЛЕДОВАНИЕ ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ ДЛЯ НЕЙРОСЕТЕВОГО АНАЛИЗА ГОЛОСА СЛУШАТЕЛЕЙ СИСТЕМЫ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

***Аннотация.** Обоснована актуальность задачи внедрения в существующие системы дистанционного обучения инструментальных средств распознавания личности и эмоций слушателей на основании анализа их голоса. Показана перспективность разработки программных средств нейросетевого анализа голоса. Установлено, что в современной научно-прикладной литературе недостаточное внимание уделяется разработке архитектуры указанных средств нейросетевого анализа. В результате проведенных исследований в терминах языка моделирования UML разработано описание архитектуры модуля нейросетевого анализа голоса слушателей системы дистанционного обучения, ориентированного на распознавание личности и эмоций слушателя. Разработаны диаграммы прецедентов, классов и компонентов. Также построена структурная схема модуля распознавания. Особенностью предложенных решений является адаптация архитектуры модуля к использованию нейронной сети для анализа коэффициентов Фурье отфильтрованного голосового сигнала с целью комплексного распознавания личности и эмоций слушателя. Целесообразность использования предложенных архитектурных решений подтверждена с помощью компьютерных экспериментов, направленных на определение эффективности разработанного модуля при его использовании для распознавания эмоций дикторов, записи голосовых сигналов которых представлены в базе данных Toronto emotional speech set. Эксперименты показали, что уже после 100 эпох обучения точность распознавания эмоциональной окраски голосового сигнала для примеров, которые не вошли в учебную выборку, находится в диапазоне значений от 0,94 до 0,95. Таким образом, по достигнутым показателям точности и ресурсоемкости распознавания эмоций разработанный модуль не уступает наиболее известным решениям в данной области. Определено, что направления дальнейших исследований связаны с разработкой модулей нейросетевого анализа таких биометрических параметров, как изображение лица, радужная оболочка глаза и клавиатурный почерк, а также с интеграцией таких модулей в единую систему.*

Ключевые слова: нейронные сети; распознавание эмоций; распознавание личности; дистанционное обучение; защита информации

Введение

В современных условиях одной из наиболее важных стратегических задач в области систем дистанционного обучения (СДО) является повышение доступности качественных образовательных услуг, соответствующих как потребностям каждого отдельного гражданина, так и потребностям всего общества в целом. Решение указанной задачи связано

с повсеместным внедрением в СДО электронных средств обучения, режим использования которых должен коррелироваться с эмоциональным состоянием конкретного слушателя. Это выдвигает на первый план вопрос разработки и внедрения в СДО инструментальных средств распознавания личности и эмоций слушателя на основании анализа биометрических параметров, регистрация которых может быть реализована с помощью

распространенного аппаратного обеспечения [9; 11; 22]. В первую очередь к указанным биометрическим параметрам относятся: лицо человека, радужная оболочка глаза, голос и клавиатурный почерк, для регистрации которых применяется видеочамера, микрофон и клавиатура компьютера. Существенным ограничением на использование средств анализа изображений лица и радужной оболочки глаза является необходимость нивелирования характерных помех, связанных с ракурсом видеосъемки, неравномерностью освещения, разрешением видеочамеры. Недостатки современных средств анализа клавиатурного почерка связаны с несовершенством применяемой теоретической базы, что затрудняет получение достаточно точных результатов, касающихся распознавания личности и эмоций слушателя. В то же время методология распознавания голоса человека достаточно отработана, а характерные помехи, препятствующие качественному распознаванию, возможно нивелировать путем определения эксплуатационных ограничений. Этим объясняется актуальность научно-прикладной проблемы разработки инструментальных средств анализа голоса слушателей системы дистанционного обучения.

Анализ современных исследований и публикаций

Согласно [1; 2], общепринятый алгоритм работы инструментальных средств анализа голоса состоит из следующих этапов:

1. Оцифровка аудиосигнала.
2. Фильтрация шумов.
3. Выделение из сигнала отдельных слов.
4. Обработка оцифрованного сигнала с целью уменьшения объема входных данных системы распознавания.
5. Дополнительная фильтрация спектра.
6. Сжатие спектра с целью учета особенностей восприятия звука человеком и уменьшения количества входных параметров блока распознавания.
7. Сравнение эталонного и испытуемого сигналов с целью определения личности и эмоций диктора.

В настоящее время считается, что реализация первых шести этапов в достаточной степени отработана. Трудности вызывает реализация седьмого этапа. При этом для сравнения эталонного и испытуемого сигналов используются разные методы, основанные на различных математических теориях и адаптированные к различным условиям, определяющим их эффективность при использовании в информационной технологии голосового взаимодействия дистанционного обучения. Характерной особенностью инструментальных средств сравнения эталонного и испытуемого

сигналов является высокая ресурсоемкость. Это объясняется сложностью используемых алгоритмов, математический аппарат которых базируется на байесовском подходе, скрытых марковских процессах, Support Vector Machines, методах динамического программирования, теории нечеткой логики [2 – 4; 12; 15; 17; 20; 21; 25]. Известным путем устранения этого недостатка является использование нейросетевых моделей.

Как свидетельствуют результаты анализа научно-практических работ [1; 5], распространенная технология разработки нейросетевых средств анализа голоса предполагает ассоциацию входных параметров нейронной сети с параметрами, характеризующими фрагмент голосового сигнала, регистрируемый с частотой дискретизации от 8 до 96 кГц. Сложившийся подход к определению указанных параметров базируется на использовании процедуры кратковременного анализа. То есть голосовой сигнал разбивается на временные окна фиксированного размера (квазистационарные фрагменты), длительность которых выбирается в пределах 10–30 мс [18; 22; 23]. Как правило, между фрагментами делают перекрытие, равное половине длины этого фрагмента. Затем к каждому фрагменту применяются алгоритмы извлечения признаков, в качестве которых, как правило, используются так называемые мел-кепстральные коэффициенты (MFCC) [6; 26; 27]. Также на основании данных [16; 19] можно сделать заключение о том, что во многих случаях специфика использования средств распознавания диктора позволяет подавать на вход нейронной сети параметры, характеризующие фиксированный участок голосового сигнала продолжительностью 5 – 15 с, квазистационарные фрагменты которого описываются 20 MFCC. При этом наиболее распространенным видом нейросетевой модели является многослойный перцептрон, представляющий собой глубокую нейронную сеть с прямым распространением сигнала [8; 10]. Структура такой сети показана на рис. 1, а особенности математического обеспечения определяются тем, что в скрытых нейронах используется функция активации ReLU, определяемая выражением (1), а в выходных нейронах используется функция активации Softmax, определяемая выражением (2).

$$y_i = \max(0, x_{s,i}); \quad (1)$$

$$y_i = \exp(x_{s,i}) / \sum_{k=1}^K \left(\exp(x_{s,k}) \right), \quad (2)$$

где y_i – выходной сигнал i -го нейрона выходного слоя; $x_{s,i}$ – суммарный входной сигнал для k -го нейрона выходного слоя; K – количество выходных нейронов.

Таким образом, проведенный анализ позволяет утверждать, что перспективы разработки инструментальных средств анализа голоса связаны с применением в них нейросетевых моделей.

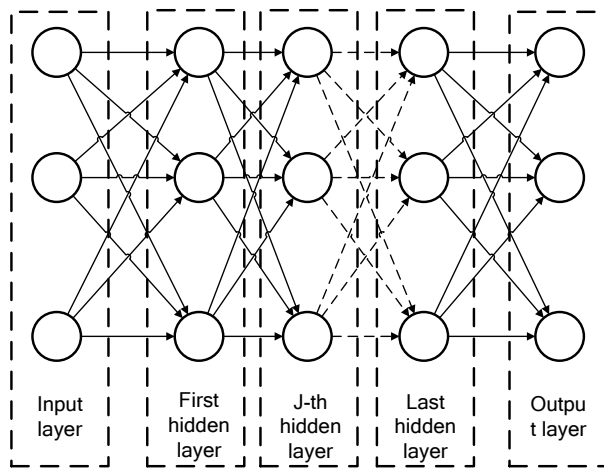


Рисунок 1 – Структурная схема многослойного персептрона

При этом теоретический аспект разработки нейросетевых средств анализа голоса для распознавания личности и эмоций диктора проработан в достаточной степени. В то же время в доступной литературе решение вопросов проектирования и реализации соответствующих программных средств освещено недостаточно полно.

Цель статьи

Основной целью публикации является разработка и исследование программного модуля нейросетевого анализа голоса слушателей системы дистанционного обучения.

Изложение основного материала

В соответствии с результатами [7; 13; 14; 24], первый этап построения программного модуля нейросетевого анализа голоса слушателей СДО ассоциирован с построением соответствующих моделей данных, позволяющих провести анализ предметной области, сформулировать требования и осуществить проектирование структуры программного обеспечения.

Поскольку процесс разработки программного обеспечения не предусматривал возможность распараллеливания, то с позиций упрощения документирования использована каскадная модель жизненного цикла. Для построения моделей применен язык проектирования UML (Unified Modeling Language). Выбор UML объясняется универсальностью, апробированностью и доступностью средств автоматизации процесса проектирования. Для разработки UML-диаграмм использован сервис Rational Rose производства компании IBM.

Построены диаграммы прецедентов, классов и компонентов. Диаграмма прецедентов (вариантов использования), показанная на рис. 2, позволяет представить модуль нейросетевого анализа в виде

множества сущностей или актеров, взаимодействующих с системой путем так называемых вариантов использования. При этом актером (действующим лицом) называется любая сущность, взаимодействующая с системой извне. Это может быть человек, техническое устройство, программа или любая другая система, которая может служить источником воздействия на проектируемую систему. В свою очередь, вариант использования служит для описания сервисов, которые система предоставляет актеру. Таким образом, каждый вариант использования определяет некоторое множество действий, которое реализуется системой при взаимодействии с актером. При этом в диаграмме прецедентов не содержится информация о способе реализации взаимодействия актеров с системой распознавания, что позволяет повысить уровень формализации проектирования. В состав этой диаграммы входят актеры "Administrator", "Monitor" и "Listener".

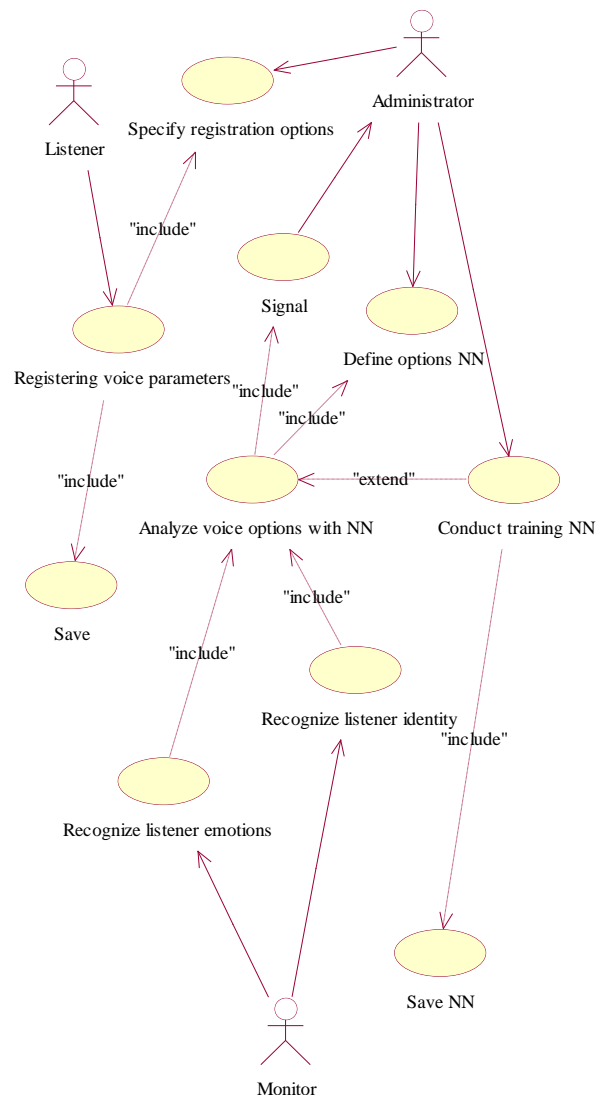


Рисунок 2 – Диаграмма прецедентов модуля нейросетевого анализа голоса слушателя СДО

Актер "Administrator" представляет собой администратора системы и отвечает за настройку параметров модуля нейросетевого анализа голоса слушателя СДО.

Актер "Listener" соотносится со слушателем СДО, который использует голосовой интерфейс и чей голос подлежит нейросетевому анализу.

"Monitor" – актер, который соотносится с блоком мониторинга слушателей СДО.

Разработанная диаграмма прецедентов предусматривает выполнение следующих функциональных задач:

- Register voice parameters – реализация регистрации параметров голоса;
- Specify registration options – определение параметров регистрации голоса;
- Save – сохранение зарегистрированных параметров голоса;
- Recognize listener identity – распознавание личности слушателя;
- Recognize listener emotions - распознавание эмоций слушателя;
- Define options NN – определение архитектурных параметров нейронной сети;
- Conduct training NN – обучение нейронной сети;
- Save NN – сохранение параметров нейронной сети;
- Analyze voice options with NN – анализ зарегистрированных параметров голоса с помощью нейронной сети;

– Signal – сигнализация о результатах нейросетевого анализа, свидетельствующая об итогах распознавания личности и эмоций слушателя.

Разработанная диаграмма прецедентов предусматривает, что между прецедентами используются отношение расширения ("extend") и включения ("include"). Отношение расширения отмечает тот факт, что один из прецедентов может присоединять к своему поведению некоторое дополнительное поведение, характерное для другого варианта использования. Отношение включения между двумя прецедентами указывает, что поведение одного из прецедентов включается как составной компонент в последовательность поведения другого.

Показанная на рис. 3 диаграмма классов определяет типы классов модуля нейросетевого анализа голоса и статические связи, которые существуют между этими классами.

В общем случае диаграмма классов представляет собой граф, вершинами которого являются элементы типа "классификатор", которые связаны между собой различными типами

структурных отношений. Следует заметить, что диаграмма классов может также содержать интерфейсы, объекты и связи между ними. Считается, что диаграмма классов должна отражать статическую структурную модель проектируемой системы. Поэтому диаграмму классов принято считать графическим представлением таких структурных взаимосвязей логической модели системы, которые не зависят от времени. На диаграмме классов предусмотрено отражать основные атрибуты классов и ограничения, которые накладываются на связи между классами. В ее состав входят классы Consumer, Register, Analisators и NeuroNet, предназначенные для определения параметров регистрации и обработки параметров голоса, обучения нейронной сети и ее применения для распознавания личности и эмоций слушателя.

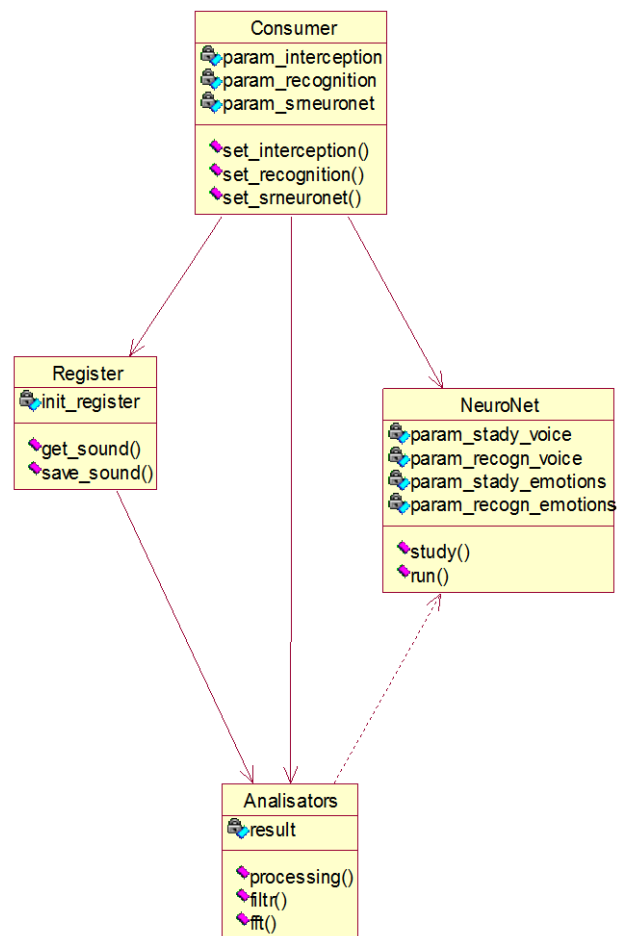


Рисунок 3 – Диаграмма классов модуля нейросетевого анализа голоса слушателя СДО

Показанная на рис. 4 диаграмма компонентов характеризует особенности физического представления модуля нейросетевого анализа и обеспечивает визуализацию структуры программной системы. Также диаграмма компонентов позволяет за счет формирования зависимостей между программными компонентами упростить

компоновку блоков, составляющих архитектуру модуля. Пунктирные стрелки, соединяющие блоки данной диаграммы, показывают отношения взаимозависимости, аналогичные тем, которые имеют место при компиляции исходного кода. Основными графическими элементами диаграммы компонентов являются компоненты, интерфейсы и зависимости между ними.

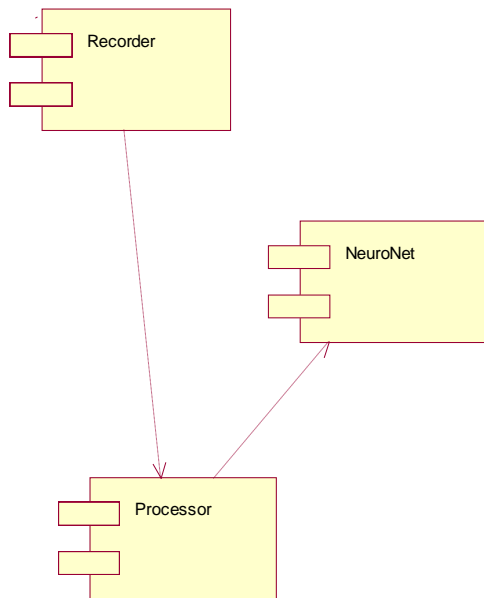


Рисунок 4 – Диаграмма компонентов модуля нейросетевого анализа голоса слушателя СДО

Предполагается, что модуль должен состоять из трех блоков: Recorder, Processor, NeuroNet. Указанные блоки представляют интерфейс получения параметров голоса слушателя, обработку полученных параметров, а также распознавание личности и эмоций слушателя. Также, в дополнение к представленным диаграммам разработана, показанная на рис. 5, структурная схема модуля нейросетевого анализа голоса слушателя СДО. Как показано на рис. 5, модуль состоит из 8 отдельных блоков и 3 хранилищ данных, предназначенных для хранения параметров голоса, обучающей выборки и параметров нейронной сети.

Построение, показанных на рис. 2 – 4 диаграмм и структурной схемы, позволило перейти к разработке программного обеспечения модуля нейросетевого анализа, а также к проведению компьютерных экспериментов, направленных на его верификацию. При этом использованы результаты работ [14; 24; 26], посвященных построению нейросетевых моделей, предназначенных для анализа голосового сигнала с целью распознавания личности и эмоций диктора. Базируясь на этих результатах, в качестве базовой нейросетевой модели использован двухслойный перцептрон, входные параметры которого ассоциированы

с 24 мелкепестральными коэффициентами каждого из квазистационарных фрагментов анализируемого голосового сигнала длительностью 1,2 с. Программное обеспечение разработано с использованием языка программирования Python и библиотек Keras, Numpy и Matplotlib.

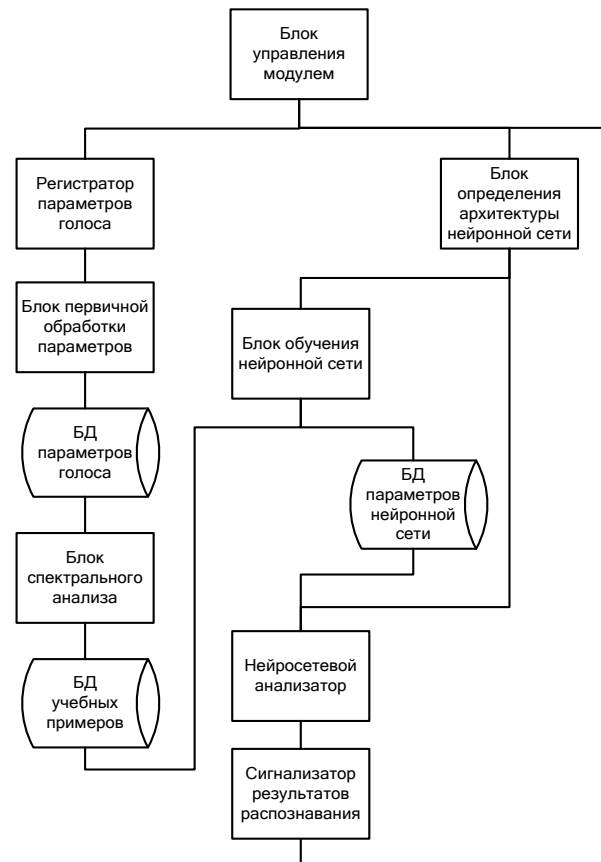


Рисунок 5 – Структурная схема модуля нейросетевого анализа голоса слушателя СДО

Проведены компьютерные эксперименты, направленные на определение эффективности разработанного модуля при его использовании для распознавания эмоций дикторов, записи голосовых сигналов которых представлены в базе данных Toronto emotional speech set. Указанная база данных доступна для скачивания на сайте www.kaggle.com. В ней содержится 2800 записей голосовых сигналов 2 дикторов, в которых отражены 7 базовых эмоций (злость, счастье, удивление, грусть, нейтральность, отвращение и страх). Каждой отдельной эмоции соответствует 200 примеров. Эксперименты показали, что даже после 100 эпох обучения точность распознавания эмоциональной окраски голосового сигнала для примеров, которые не вошли в учебную выборку, находится в диапазоне значений от 0,94 до 0,95. Более точно распознаются эмоции злость, счастье и удивление. Эмоции грусть, нейтральность, отвращение и страх распознаются менее точно. Следует отметить, что в экспериментах использован персональный компьютер со средними

характеристиками (процесор Intel Core i5-9400F (2,9 – 4,1 ГГц), об'єм оперативної пам'яті 16 ГБ, відеокарта nVidia GeForce GTX 1660), навантаженість якого не перевищала 50%. Таким чином, по досягнутим показателям точності і ресурсоемкості розпізнавання емоцій розроблений модуль не поступає найбільш відомим рішенням в даній області, що свідчить про ефективність його застосування для нейросетевого аналізу голосу слухачів системи дистанційного навчання [5; 8; 12].

Разом з тим результати робіт [15; 26] вказують на ефективність розпізнавання особистості і емоцій слухачів СДО на основі комплексного нейросетевого аналізу декількох біометричних параметрів. Крім голосу до таких параметрів в першу чергу відносять зображення обличчя слухача, клавіатурний почерк і радужна оболонка його очей. Тому шляхи подальших досліджень можливо пов'язати з розробкою модулів нейросетевого аналізу згаданих параметрів, а також з інтеграцією таких модулів в єдину систему.

Выводы

Обоснована актуальність удосконалення інструментальних засобів аналізу голосу слухачів системи дистанційного навчання. На основі аналізу відомих публікацій показано перспективність використання засобів нейросетевого аналізу, створення яких ускладнюється існуючою невизначеністю в формалізованому описанні архітектурних рішень. В результаті проведених досліджень в термінах мови моделювання UML розроблена архітектура модуля нейросетевого аналізу голосу слухачів системи дистанційного навчання, орієнтованого на розпізнавання особистості і емоцій слухача. Ефективність використання запропонованих архітектурних рішень підтверджено з допомогою комп'ютерних експериментів. Встановлено, що напрямки подальших досліджень пов'язані з розробкою модулів нейросетевого аналізу таких біометричних параметрів, як зображення обличчя, радужна оболонка очей і клавіатурний почерк, а також з інтеграцією таких модулів в єдину систему.

Список літератури

1. Aitchanov, B., Korchenko, A., Tereykovskiy, I., Tereykovska L. (2017). *Perspectives for using classical neural network models and methods of counteracting attacks on net-work resources of information systems. (2017). News of the national academy of sci-ences of the republic of Kazakhstan. Geology and technical sciences, 5, 425, 202 – 212.*
2. Ajinkya, N. Jadhav, Nagaraj, V. Dharwadkar. (2018). *A Speaker Recognition System Using Gaussian Mixture Model, EM Algorithm and K-Means Clustering. International Journal of Modern Education and Computer Science, 10, 11, 19-28.*
3. Altincay, H. (2003). *Speaker identification by combining multiple classifiers using Dempster–Shafer theory of evidence. Speech Communication, 41, 4, 531 – 547.*
4. Campbell, W., Sturm, D., Reynolds, D. (2006). *Support vector machines using GMM supervectors for speaker verification. IEEE Signal Process. Lett., 13, 5, 308–311.*
5. Ehsan, Variani, Xin, Lei, McDermott, Erik, Moreno, Ignacio Lope & Gonzalez-Dominguez, Javier. (2014). *Deep neural networks for small footprint text-dependent speaker verification. In Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference, pp. 4052–4056.*
6. Ganchev, T., Fakotakis, N., Kokkinakis, G. (2005). *Comparative evaluation of various MFCC implementations on the speaker verification task. 10th International Conference on Speech and Computer. Patras, Greece.*
7. Hu, Z., Tereykovskiy, I., Zorin, Y., Tereykovska, L., Zhibek, A. (2019). *Optimization of Convolutional Neural Network Structure for Biometric Authentication by Face Ge-ometry. In: Hu Z., Petoukhov S., Dychka I., He M. (eds) Advances in Computer Science for Engineering and Education. ICCSEEA 2018. Advances in Intelligent Systems and Computing, 754. Springer, Cham.*
8. Iandola, F.N., Han, S., W. Moskewicz M.W. (2016). *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv:1602.07360v4 [cs.CV], 13. Available at: https://arxiv.org/pdf/1602.07360.pdf.*
9. Ingale, A.B., Chaudhari, D.S. (2012). *Speech emotion recognition. International Jour-nal of Soft Computing and Engineering (IJSCE), 2, 1, 235-238.*
10. Ing-Jr Ding, Chih-Ta Yen, Yen-Ming Hsu. (2013). *Developments of Machine Learning Schemes for Dynamic Time-Wrapping-Based Speech Recognition. Mathematical Prob-lems in Engineering, 56-68.*
11. Juslin, P.N., Laukaa, P. (2003). *Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological bulletin, 129, 5, 770.*
12. Karam, Z., Campbell, W. (2007). *A new kernel for SVM MLLR based speaker recog-nition. In: Proc. Interspeech 2007, Antwerp, Belgium, August 2007, pp.290–293.*
13. Lyon, R.F. (2010). *Machine hearing: An emerging field. IEEE signal processing maga-zine, 27, 5, 131-139.*
14. Makarova, V., Petrushin, V.A. (2002). *RUSLANA: a database of Russian emotional utterances. ICSLP, 2041-2044.*

15. McLaren, Mitchell, Yun, Lei, Scheffer, Nicolas and Ferrer, Luciana. (2014). Application of convolutional neural networks to speaker recognition in noisy conditions. 15th Annual Conference of the International Speech Communication Association, Singapore, September 14-18, pp. 686–690. ISCA.
16. Partila, P., Tovarek, J. (2015). Pattern Recognition Methods and Features Selection for Speech Emotion Recognition System. *Scientific World Journal*, 7.
17. Penagarikano, M., Bordel, G. (2004). Layered Markov models: A New architectural approach to automatic speech recognition. *Machine Learning for Signal Processing XIV - Proceedings of the 2004 IEEE Signal Processing Society Workshop*, pp. 305-314.
18. Ranganathan, H., Chakraborty, S., Panchanathan, S. (2016). Multimodal emotion recognition using deep learning architectures. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 1-9.
19. Satyanand, Singh, Abhay, Kumar, Kolluri, David Raju. (2016). Efficient Modelling Technique based Speaker Recognition under Limited Speech Data. *International Journal of Image, Graphics and Signal Processing*, 8, 11, 41-48.
20. Savchenko, L. V., Savchenko, A.V. (2019). Fuzzy Phonetic Encoding of Speech Signals in Voice Processing Systems. *Journal of Communications Technology and Electronics*, 64, 3, 238-244.
21. Savchenko, V. V., Savchenko, A.V. (2016). Information Theoretic Analysis of Efficiency of the Phonetic Encoding–Decoding Method in Automatic Speech Recognition. *Journal of Communications Technology and Electronics*, 4(61), 430-435.
22. Savchenko, V. V. (2015). The Principle of the Information-Divergence Minimum in the Problem of Spectral Analysis of the Random Time Series Under the Condition of Small Observation Samples. *Radiophysics and Quantum Electronics*, 5(58), 373-379.
23. Savchenko, V. V. (2016). Enhancement of the Noise Immunity of a Voice-Activated Robotics Control System Based on Phonetic Word Decoding Method. *Journal of Communications Technology and Electronics*, 12(61), 1374-1379.
24. Tereikovska, L., Tereikovskiy, I., Aytkhozhaeva, E., Tynymbayev, S., Imanbayev, A. (2017). Encoding of neural network model exit signal, that is devoted for distinction of graphical images in biometric authenticate systems. *News of the national academy of sciences of the republic of Kazakhstan. Geology and technical sciences*, 426, 217–224.
25. Tereikovskiy, I., Parkhomenko, I., Toliupa, S., Tereikovska, L. (2018). Markov model of normal conduct template of computer systems network objects. 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET 2018 – Proceedings. pp. 498 – 501.
26. Tereikovskiy, I., Subach, I., Tereikovskiy, O., Tereikovska, L., Toliupa, S., Nakonech-nyi, V. (2019). Parameter Definition for Multilayer Perceptron Intended for Speaker Identification. 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 2019, pp. 227-231.
27. Vaziri, G., Almasganj, F., Behroozmand, R. (2010). Pathological assessment of patients' speech signals using nonlinear dynamical analysis, *Comput. Biol. Med.*, 40, 1, 54-63.

Стаття надійшла до редакції 02.09.2020

Чернишев Денис Олегович

Доктор технічних наук, професор, перший проректор, orcid.org/0000-0002-1946-9242

Київський національний університет будівництва і архітектури, Київ

Міхайленко Віктор Мефодійович

Доктор технічних наук, професор, завідувач кафедри інформаційних технологій проектування та прикладної математики, orcid.org/0000-0002-9573-9873

Київський національний університет будівництва і архітектури, Київ

Терейковська Людмила Олексіївна

Кандидат технічних наук, доцент інформаційних технологій проектування та прикладної математики, orcid.org/0000-0002-8830-0790

Київський національний університет будівництва і архітектури, Київ

РОЗРОБЛЕННЯ І ДОСЛІДЖЕННЯ ІНСТРУМЕНТАЛЬНИХ ЗАСОБІВ ДЛЯ НЕЙРОМЕРЕЖЕВОГО АНАЛІЗУ ГОЛОСУ СЛУХАЧІВ СИСТЕМИ ДИСТАНЦІЙНОГО НАВЧАННЯ

Анотація. Обґрунтовано актуальність завдання впровадження в наявні системи дистанційного навчання інструментальних засобів розпізнавання особистості і емоцій слухачів на підставі аналізу їх голосу. Показана перспективність розробки програмних засобів нейромережевого аналізу голосу. Встановлено, що в сучасній науково-прикладній літературі недоведена увага приділяється розробці архітектури зазначених засобів нейромережевого аналізу. В результаті проведених досліджень в термінах мови моделювання UML розроблено опис архітектури модуля нейромережевого аналізу голосу слухачів системи дистанційного навчання, орієнтованого на розпізнавання особистості і емоцій слухача. Розроблено діаграми прецедентів, класів і компонентів. Також побудована структурна схема модуля нейромережевого аналізу. Особливістю запропонованих рішень є адаптація архітектури модуля до використання нейронної мережі для аналізу коефіцієнтів Фур'є відфільтрованого голосового сигналу з метою комплексного

розпізнавання особистості і емоцій слухача. Доцільність використання запропонованих архітектурних рішень підтверджена за допомогою комп'ютерних експериментів, спрямованих на визначення ефективності розробленого модуля при його використанні для розпізнавання емоцій дикторів, записи голосових сигналів яких представлені в базі даних *Toronto emotional speech set*. Експерименти показали, що вже після 100 епох навчання точність розпізнавання емоційного забарвлення голосового сигналу для прикладів, які не ввійшли в навчальну вибірку, перебуває в діапазоні значень від 0,94 до 0,95. Отже, за досягнутими показниками точності і ресурсоемності розпізнавання емоцій розроблений модуль не поступається найбільш відомим рішенням в цій сфері. Визначено, що напрями подальших досліджень пов'язані з розробленням модулів нейромережевого аналізу таких біометричних параметрів, як зображення обличчя, райдужна оболонка ока і клавіатурний почерк, а також з інтеграцією таких модулів в єдину систему.

Ключові слова: нейронні мережі; розпізнавання емоцій; розпізнавання особи; дистанційне навчання; захист інформації

Chernyshev Denys

DSc (Eng.), Professor, First vice-rector, orcid.org/0000-0002-1946-9242

Kyiv National University of Construction and Architecture, Kyiv

Mihaylenko Victor

DSc (Eng.), Professor, Head of the Department of Information Technologies of Design and Applied Mathematics, orcid.org/0000-0002-9573-9873

Kyiv National University of Construction and Architecture, Kyiv

Tereikovska Liudmyla

PhD (Eng.), Associate Professor, Associate Professor of the Department of Information Technologies of Design and Applied Mathematics, orcid.org/0000-0002-8830-0790

Kyiv National University of Construction and Architecture, Kyiv

DEVELOPMENT AND RESEARCH OF TOOLS FOR NEURAL NETWORK ANALYSIS OF THE VOICE OF LISTENERS OF THE DISTANCE LEARNING SYSTEM

Abstract. *The urgency of the problem of introducing tools for recognizing the personality and emotions of listeners into the existing distance learning systems based on the analysis of their voice is substantiated. The prospects for the development of software for neural network analysis of voice are shown. It has been established that in the modern scientific and applied literature, insufficient attention is paid to the development of the architecture of these means of neural network analysis. As a result of the research carried out in terms of the UML modeling language, a description of the architecture of the module for neural network analysis of the voice of listeners of the distance learning system, focused on recognizing the personality and emotions of the listener, has been developed. Developed diagrams of use cases, classes and components. The block diagram of the recognition module is also built. A feature of the proposed solutions is the adaptation of the module architecture to the use of a neural network for the analysis of the Fourier coefficients of the filtered voice signal for the purpose of complex recognition of the listener's personality and emotions. The expediency of using the proposed architectural solutions was confirmed with the help of computer experiments aimed at determining the effectiveness of the developed module when using it to recognize the emotions of speakers whose voice recordings are presented in the Toronto emotional speech database. Experiments have shown that after 100 epochs of training, the accuracy of recognizing the emotional coloring of a voice signal for examples that were not included in the training sample is in the range of values from 0.94 to 0.95. Thus, in terms of the achieved indicators of accuracy and resource intensity of emotion recognition, the developed module is not inferior to the most well-known solutions in this area. It has been determined that the directions for further research are related to the development of modules for neural network analysis of such biometric parameters as facial image, iris and keyboard handwriting, as well as with the integration of such modules into a single system.*

Keywords: neural networks; recognition of emotions; personality recognition; distance learning; data protection

Посилання на публікацію

APA Chernyshev, Denys, Mihaylenko, Victor & Tereikovska, Liudmyla. (2020). Development and research of tools for neural network analysis of the voice of listeners of the distance learning system. *Management of Development of Complex Systems*, 43, 123 – 130, [dx.doi.org/10.32347/2412-9933.2020.43.123-130](https://doi.org/10.32347/2412-9933.2020.43.123-130).

ДСТУ Чернышев Д. О. Разработка и исследование инструментальных средств для нейросетевого анализа голоса слушателей системы дистанционного обучения [Текст] / Д. О. Чернышев, В. М. Михайленко, Л. А. Терейковская // *Управление развитием сложных систем*. – 2020. – № 43. – С. 123 – 130, [dx.doi.org/10.32347/2412-9933.2020.43.123-130](https://doi.org/10.32347/2412-9933.2020.43.123-130).