УДК 519.87

Вестник СибГАУ
2014. № 5(57). С. 51–56

ЭФФЕКТИВНАЯ ПРОЦЕДУРА АУТЕНТИФИКАЦИИ СТУДЕНТА ПО РЕЧИ В ДИСТАНЦИОННОМ ОБРАЗОВАНИИ

К. Ю. Брестер[1], С. Р. Вишневская[1], О. Э. Семенкина[1], М. Ю. Сидоров[2]

[1] Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева
Российская Федерация, 660014, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31
E-mail: christina.bre@yandex.ru, vishni@ngs.ru, semenkina.olga@mail.ru
[2]Ульмский университет
Германия, 89081, г. Ульм, аллея им. Альберта Эйнштейна, 43
E-mail: maxim.sidorov@uni-ulm.de

В настоящее время на базе практически каждого университета студентам предоставляется возможность получения заочного образования, кроме того, ведутся курсы дистанционного обучения. Из-за широкого спектра преимуществ в последние годы дистанционное образование привлекает все больше и больше людей, что вызывает необходимость создания официального стандарта, включающего ряд требований, которым должна соответствовать дистанционная форма обучения. Так, например, необходимость верификации личности студентов включена во многие зарубежные образовательные стандарты в качестве обязательной процедуры. В случае если преподаватели лишены визуального контакта со своими учениками, появляется необходимость в разработке эффективной технологии для проверки личности студента в дистанционном режиме. Предложена процедура аутентификации студента по речи, основанная на использовании акустических характеристик, извлеченных из речевых сигналов. В настоящее время пока остается открытым вопрос выбора надежной и эффективной классификационной модели, поскольку невозможно в онлайн-режиме исследовать различные классификаторы для определения наиболее эффективного, сохраняя при этом высокую производительность системы при взаимодействии с пользователем. Поэтому, чтобы повысить надежность предлагаемого подхода, были разработаны алгоритмические схемы, основанные на коллективном принятии решений с целью учета предсказаний различных классификаторов. Для исследования эффективности данной процедуры использовались базы данных, содержащие звукозаписи на немецком, английском и японском языках. Согласно полученным результатам применение описанного подхода позволяет получить высокую точность распознавания личности говорящего по речи (до 100 % для некоторых баз данных). Разработанные алгоритмические схемы обеспечивают гарантированный уровень эффективности и являются надежной альтернативой произвольному выбору классификационной модели.

Ключевые слова: дистанционное обучение, аутентификация студентов по речи, классификатор, коллективное принятие решений.

Vestnik SibGAU
2014, No. 5(57), P. 51–56

EFFECTIVE SPEECH-BASED STUDENT AUTHENTICATION PROCEDURE IN DISTANCE LEARNING

C. Y. Brester[1], S. R. Vishnevskaya[1], O. E. Semenkina[1], M. Y. Sidorov[2]

[1] Siberian State Aerospace University named after academician M. F. Reshetnev
31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660014, Russian Federation
E-mail: christina.bre@yandex.ru, vishni@ngs.ru, semenkina.olga@mail.ru
[2] Ulm University
43, Albert-Einstein-Allee, Ulm, 89081, Germany
E-mail: maxim.sidorov@uni-ulm.de

Nowadays it is almost impossible to find a university that does not provide its students with online courses or correspondence education. Due to various advantages, distance learning has attracted more and more people in recent years. As a result, some of the requirements that this educational format has to satisfy have been included in legislation systems. The necessity to authenticate students remotely is presented as a compulsory procedure in many official documents. When teachers are deprived of face-to-face contact with their students, there is a need to find an appropriate way to verify their personality distantly. In this paper we propose the speech-based student authentication procedure which operates with some acoustic characteristics extracted from voice signals. However, there is one crucial question related to the classification model providing high performance. It is almost impossible for the online systems to vary

classifiers and determine the most effective one while interacting with a user. Therefore, to increase the reliability of our proposal we elaborated some classification schemes based on collective decision making to take into account predictions of different classifiers. To prove the effectiveness of this approach, we used a number of multi-lingual corpora (German, English, Japanese). According to the results obtained, a high level of speaker recognition was achieved (up to 100 % of F-score values). The developed algorithmic schemes provide a guaranteed level of effectiveness and might be used as a reliable alternative to the occasional choice of a classification model.

Keywords: distance learning, speech-based student authentication, classifier, collective decision making.

Введение. Все больше и больше людей осознают преимущества дистанционного обучения и заочного образования. Во-первых, многие университеты по всему миру предлагают онлайн-курсы, которые доступны для всех студентов. Это прекрасная возможность одновременного изучения нескольких дисциплин. Во-вторых, нет никакой необходимости каждый день добираться до и от места обучения, что означает экономию транспортных расходов. Кроме того, по сравнению с традиционной системой образования, дистанционное обучение является гораздо более гибким, так как студенты имеют возможность составлять собственное расписание и осваивать материал в индивидуальном темпе. Одним из главных достоинств этого учебного процесса является его низкая стоимость.

Постоянно растущая популярность дистанционного образования приводит к необходимости разработки стандартов, содержащих требования, которым должен соответствовать данный формат обучения. Можно привести широкий спектр различных требований [1], однако в данной статье рассматривается лишь один из ключевых моментов, а именно, вопрос аутентификации личности дистанционного студента.

Самыми распространенными предложениями для дистанционной верификации личности обучающегося являются:

– использование биометрических данных;

– запрос персональных данных (ответ на контрольные вопросы);

– проверка с помощью клавиатурного почерка.

Однако первый подход требует высокотехнологичных устройств (например, для сканирования отпечатков пальцев), что увеличивает стоимость дистанционного обучения в значительной степени, в то время как низкая стоимость является его существенным преимуществом. Кроме того, ни одна из этих процедур не может гарантировать полное отсутствие людей, находящихся поблизости, которые могут дать подсказки во время дистанционного экзамена.

Принимая во внимание все эти детали, был разработан альтернативный подход для аутентификации дистанционных студентов в автоматическом режиме. В ходе учебного процесса студенты участвуют в диалоге «ученик–учитель» или «студент–система», что определяет возможность накопления устных ответов учащихся (в формате звукозаписей). В свою очередь, акустические характеристики голоса являются отличительными признаками одного человека от другого. Следовательно, на промежуточном или финальном экзамене система может сравнить текущий речевой сигнал с накопленными ранее голосовыми записями. Данный подход эффективен прежде всего в случае, когда учащиеся должны ответить на вопрос быстро в режиме реального времени: у них есть только несколько секунд, чтобы обдумать свой ответ, и нет времени на то, чтобы понять объяснения других людей.

Очевидным преимуществом аутентификации студентов по речи является то, что обучающимся не требуются дополнительные устройства (за исключением микрофонов, которые обычно и так встроены в компьютер).

Разрабатываемый подход. Разрабатываемый подход состоит из нескольких этапов. Во-первых, необходимо извлечь акустические характеристики из набора имеющихся звукозаписей. В рамках конференции «INTERSPEECH 2009» учеными был предложен набор акустических характеристик, используемых для описания любого речевого сигнала. Данная совокупность признаков включает максимальное, минимальное, среднее значения или среднеквадратическое отклонение числовых характеристик, описывающих речевой сигнал, его высоту, вибрации, интенсивность и т. п. Суммарное количество признаков – 384. Для извлечения из голосовой записи описанного набора признаков используются программные системы OpenSMILE [2] и Praat [3].

Все извлеченные атрибуты или наиболее информативные из них [4; 5] должны быть привлечены в процесс обучения классификационных моделей, вид которых и способ построения может существенно различаться [6–8].

На заключительном этапе речевой сигнал, подлежащий анализу, конвертируется в вектор признаков (также с использованием OpenSMILE, Praat), который затем подается в качестве входных данных на уже обученный классификатор.

На втором этапе необходимо выбрать классификационную модель. Однако невозможно знать заранее, какая из них окажется наиболее эффективной в конкретном случае. Поэтому для повышения надежности процедуры распознавания были разработаны технологии принятия решений коллективом классификаторов. В данной работе мы исследуем эффективность трех алгоритмических схем, которые позволяют учитывать предсказания различных моделей для принятия окончательного решения [9].

Схема 1. Для каждого тестового примера необходимо определить k ближайших соседей из набора данных для обучения. Предсказание модели, которая правильно классифицирует эти примеры (k ближайших соседей), используется в качестве окончательного решения. Если несколько моделей демонстрируют равную эффективность, необходимо выбрать одну из них случайным образом.

Схема 2. Для каждого тестового примера модели голосуют за разные классы в соответствии с их собственными прогнозами. Окончательное решение

определяется как коллективный выбор, основанный на правиле большинства.

Схема 3. Объединяем схемы 1 и 2 следующим образом:

– выполняем процедуру голосования, как это описано в схеме 2;

– если несколько классов имеют максимальное количество голосов, применяем схему 1.

Важно, что во всех этих схемах нет ограничений на количество классификаторов. Но, безусловно, целесообразно включать в коллектив модели, демонстрирующие высокую эффективность. Поэтому перед применением описанных схем был исследован набор стандартных классификаторов с целью выявления наиболее эффективных моделей.

Результаты исследования разработанного подхода. На практике для оценки результатов работы классификатора нередко используется матрица неточностей (*англ.* confusion matrix) [10], столбцы которой соответствуют экспертным решениям (истинное значение класса), а строки – предсказаниям классификатора (см. рисунок). Размерность матрицы $N \times N$, где N – число различных классов в выборке.

Матрица неточностей демонстрирует работу алгоритма и позволяет оценить его эффективность для каждого класса, содержащегося в выборке. Для этого вводятся специальные метрики *полнота* и *точность*, определяемые следующим образом. Пусть построена матрица неточностей $A = (a_{ij})$, тогда полнотой в пределах класса l (*англ.* recall) назовем величину, равную доле экземпляров данного класса, найденных классификатором, относительно всех примеров данного класса в тестовой выборке:

$$\text{recall}_l = \frac{a_{ll}}{\sum_i a_{il}}. \qquad (1)$$

Точностью в пределах класса l (*англ.* precision) назовем величину, равную доле примеров в тестовой выборке, действительно принадлежащих классу l, относительно всех экземпляров, которые были отнесены к данному классу:

$$\text{precision}_l = \frac{a_{ll}}{\sum_j a_{lj}}. \qquad (2)$$

Чем выше точность и полнота, тем качество работы классификатора лучше. Однако при решении практических задач редко удается добиться максимальных значений обеих метрик одновременно. Поэтому часто используют такой показатель, как F-score, объединяющий в себе информацию и о точности, и о полноте классификатора:

$$\text{F-score} = 2 \cdot \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}, \qquad (3)$$

где $\text{Recall} = \sum_l \text{recoll}$, $\text{Precision} = \sum_l \text{precision}$.

Исследование эффективности предложенного подхода проводилось с использованием баз данных Berlin [11], SAVEE [12], VAM [13] и UUDB [14], содержащих характеристики голосовых записей на немецком, английском, немецком и японском языках соответственно (табл. 1).

В первом эксперименте разрабатываемый подход был исследован с привлечением следующих классификаторов [15]:

– полносвязный персептрон (MLP) с одним скрытым слоем; для обучения использовался алгоритм обратного распространения ошибки;

– машины опорных векторов (SVM), для обучения которых применялся метод последовательной минимальной оптимизации Дж. Платта;

– логистическая регрессия (Logit);

– наивный байесовский классификатор (Naive Bayes);

– деревья решений, для построения которых использовался алгоритм J48 (модификация метода C4.5);

– ансамбль деревьев решений (Random Forest);

– бэггинг (Bagging);

– аддитивная логистическая регрессия (LogitBoost);

– алгоритм генерирования правил 1R (One Rule).

Для сравнения эффективности работы классификаторов была использована процедура кроссвалидации: каждая выборка случайным образом делились на 6 стратифицированных частей. По полученным матрицам неточностей для всех баз данных были вычислены метрики F-score (значения представлены в табл. 2).

		Действительные значения			
		Класс_1	Класс_2	…	Класс_N
Предсказанные значения	Класс_1	a_{11}	a_{12}	…	a_{1N}
	Класс_2	a_{21}	a_{22}	…	a_{2N}
	…	…	…	…	…
	Класс_N	a_{1N}	a_{2N}	…	a_{NN}

Общий вид матрицы неточностей

Таблица 1

Описание используемых баз данных

Название базы данных	Язык	Объем базы данных	Число говорящих
Berlin	Немецкий	535	10
SAVEE (Surrey Audio-Visual Expressed Emotion)	Английский	480	4
VAM (Vera am Mittag)	Немецкий	947	14
UUDB	Японский	4836	47

Таблица 2

Значения метрики F-score для стандартных классификаторов и алгоритмических схем, основанных на коллективном принятии решений, %

	Berlin	SAVEE	UUDB	VAM
MLP	90,01	**100,00**	49,89	**76,71**
SVM	90,04	99,90	66,47	75,17
Logit	87,84	99,17	**83,26**	75,05
Naive Bayes	61,28	97,95	41,07	44,45
J48	51,17	95,02	53,93	29,81
Random Forest	50,00	98,75	61,38	32,17
Bagging	61,36	95,26	68,93	48,01
LogitBoost	67,41	98,35	67,24	50,69
OneR	19,49	72,66	15,61	5,22
Схема 1	88,60	99,79	82,40	68,18
Схема 2	90,64	**100,00**	82,64	71,80
Схема 3	**90,70**	**100,00**	83,16	71,49

В целом для всех баз данных удалось достичь высокой точности распознавания. Так, например, все тестовые звукозаписи из набора SAVEE были классифицированы безошибочно полносвязным персептроном. А для базы данных Berlin наибольшие значения метрики F-score, полученные с помощью машин опорных векторов и нейронной сети, превысили 90 %.

Однако можно заметить, что не существует модели, демонстрирующей наибольшую эффективность для всех представленных наборов звукозаписей. Значения метрики F-score существенно меняются при выборе нового классификатора. Модель, позволяющая получить наилучшие результаты на одной базе данных, может быть худшим классификатором на другом наборе голосовых записей. К примеру, полносвязный персептрон демонстрирует наибольшую эффективность на базе данных SAVEE (100 %), в то время как для UUDB значения метрики F-score, полученные с помощью данной модели, существенно ниже, чем результаты других классификаторов (49,89 %).

Анализ полученных результатов показал, что для представленных баз данных персептрон (MLP), машины опорных векторов (SVM) и логистическая регрессия (Logit) являются наиболее эффективными моделями, поэтому было решено включить именно их в коллектив классификаторов для исследования предложенных алгоритмических схем.

В ходе тестирования описанных подходов было выявлено, что схема 2 и схема 3 демонстрируют наибольшую эффективность на задаче распознавания говорящего по сравнению со схемой 1 и почти всегда большую, чем отдельные классификаторы (за исключением одного случая с базой VAM, заслуживающего отдельного рассмотрения).

Для базы данных Berlin значения метрики F-score, полученные в рамках схем 2 и 3 выше, чем наилучшее значение той же метрики, найденное с помощью стандартной модели (машины опорных векторов). Применение схем 2 и 3 к набору звукозаписей SAVEE также позволяет получить наивысшую точность распознавания. Для баз данных UUDB и VAM алгоритмические схемы, основанные на коллективном принятии решений, демонстрируют результаты, сравнимые с наилучшими значениями F-score стандартных классификаторов (значительно превышают средний уровень F-score).

Таким образом, для задачи распознавания личности говорящего предложенные схемы коллективного принятия решений (в частности, схема 2 и схема 3) являются надежной альтернативой случайному выбору классификатора.

Заключение. В статье описана процедура аутентификации дистанционного студента по устной речи. Вопрос верификации личности обучающегося является одним из ключевых аспектов повышения качества дистанционного образования, поэтому отражен в зарубежных стандартах в качестве обязательного требования.

Для исследования эффективности предложенного подхода были использованы наборы голосовых звукозаписей на разных языках. Анализ полученных результатов показал, что акустические характеристики

голоса являются довольно индивидуальными, поскольку применение классификаторов, обученных на признаках, извлеченных из рассматриваемых звукозаписей, позволяет распознать говорящего с высокой точностью (до 100 % для некоторых баз данных).

В ходе исследования было показано, что не существует модели, демонстрирующей одинаковую эффективность для всех рассматриваемых баз данных, поэтому были предложены алгоритмические схемы, основанные на принятии решений коллективом классификаторов. В свою очередь, применение данных подходов позволяет избежать выбора определенной модели и обеспечить при этом достаточно высокую точность распознавания.

Библиографические ссылки

1. Higher Education Opportunity Act (Public Law 110-315). USA. Aug. 14. 2008.

2. Eyben F., Wöllmer M., Schuller B. Opensmile: the munich versatile and fast opensource audio feature extractor // Proceedings of the International Conference on Multimedia. 2010. ACM. P. 1459–1462.

3. Boersma P. Praat, a system for doing phonetics by computer // Glot international. 2002. 5(9/10). P. 341–345.

4. Self-adaptive multi-objective genetic algorithms for feature selection // Proceedings of International Conference on Engineering and Applied Sciences Optimization (OPT-i'14) / C. Brester [et al.]. 2014. P. 1838–1846.

5. Brester Ch., Sidorov M., Semenkin E. Acoustic Emotion Recognition: Two Ways of Features Selection Based on Self-Adaptive Multi-Objective Genetic Algorithm // Proceedings of the International Conference on Informatics in Control, Automation and Robotics (ICINCO). 2014. P. 851–855.

6. Хритоненко Д. И., Семенкин Е. С. Distributed self-configuring evolutionary algorithms for artificial neural networks design // Вестник СибГАУ. 2013. № 4 (50). С. 112–116.

7. Становов В. В., Семенкин Е. С. Самонастраивающийся эволюционный алгоритм проектирования баз нечетких правил для задачи классификации // Системы управления и информационные технологии. 2014. № 3 (57). С. 30–35.

8. Akhmedova Sh., Semenkin E. Co-Operation of Biology Related Algorithms Meta-Heuristic in ANN-Based Classifiers Design // Proceedings of the World Congress on Computational Intelligence (WCCI'14). 2014.

9. Попов Е. А., Семенкина М. Е., Липинский Л. В. Принятие решений коллективом интеллектуальных информационных технологий // Вестник СибГАУ. 2012. № 5 (45). С. 95–99.

10. Goutte C., Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation // ECIR'05 Proceedings of the 27th European conference on Advances in Information Retrieval Research. 2005. P. 345–359.

11. A database of german emotional speech / F. Burkhardt [et al.] // In Interspeech. 2005. P. 1517–1520.

12. Haq S., Jackson P. Machine Audition: Principles, Algorithms and Systems, chapter Multimodal Emotion Recognition // IGI Global, Hershey PA. 2010. P. 398–423.

13. Grimm M., Kroschel K., Narayanan S. The vera am mittag german audio-visual emotional speech database // In Multimedia and Expo : IEEE International Conference on, IEEE. 2008. P. 865–868.

14. Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics / H. Mori [et al.] // Speech Communication. 2011. 53.

15. The WEKA Data Mining Software: An Update, SIGKDD Explorations / M. Hall [et al.]. 2009. Vol. 11, Iss. 1.

References

1. Higher Education Opportunity Act (Public Law 110-315), Aug. 14, 2008, USA.

2. Eyben F., Wöllmer M., and Schuller B. Opensmile: the munich versatile and fast opensource audio feature extractor. *Proceedings of the international conference on Multimedia*, 2010. ACM, P. 1459–1462.

3. Boersma P. Praat, a system for doing phonetics by computer. *Glot international,* 2002, no. 5(9/10), p. 341–345.

4. Brester C., Semenkin E., Sidorov M., Minker W. Self-adaptive multi-objective genetic algorithms for feature selection. *Proceedings of International Conference on Engineering and Applied Sciences Optimization (OPT-i'14)*, 2014, p. 1838–1846.

5. Sidorov M., Brester Ch., Minker W., Semenkin E. Speech-Based Emotion Recognition: Feature Selection by Self-Adaptive Multi-Criteria Genetic Algorithm. *LREC, 2014*, p. 3481–3485.

6. Khritonenko D. I., Semenkin E. S. Distributed self-configuring evolutionary algorithms for artificial neural networks design. *Vestnik SibGAU*. 2013, no. 4 (50), p. 112–116.

7. Stanovov V. V., Semenkin E. S. [Self-adjusted evolutionary algorithm for fuzzy rules design in classification problems]. *Sistemy upravleniya i informatsionnye tekhnologii*. 2014, no. 3 (57), p. 30–35 (In Russ).

8. Akhmedova Sh., Semenkin E. Co-Operation of Biology Related Algorithms Meta-Heuristic in ANN-Based Classifiers Design. *Proceedings of the World Congress on Computational Intelligence (WCCI'14)*, 2014.

9. Popov E. A., Semenkina M. E., Lipinskiy L. V. [Decision making with intelligent information technology ensemble]. *Vestnik SibGAU*. 2012, no. 5 (45), p. 95–99 (In Russ).

10. Goutte C., Gaussier E. A probabilistic interpretation of precision, recall and *F*-score, with implication for evaluation. *ECIR'05 Proceedings of the 27th European conference on Advances in Information Retrieval Research*, 2005, p. 345–359.

11. Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W. F., and Weiss B. A database of german emotional speech. *In Interspeech,* 2005, p. 1517–1520.

12. Haq S., Jackson P. Machine Audition: Principles, Algorithms and Systems, chapter Multimodal Emotion Recognition. *IGI Global, Hershey PA*. Aug. 2010, p. 398–423.

13. Grimm M., Kroschel K., and Narayanan S. The vera am mittag german audio-visual emotional speech database. *In Multimedia and Expo, IEEE International Conference on*, IEEE, 2008, p. 865–868.

14. Mori H., Satake T., Nakamura M., and Kasuya H. Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics. *Speech Communication*, 53, 2011.

15. Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H. The WEKA Data Mining Software: An Update, SIGKDD Explorations, 2009, Vol. 11, Iss. 1.