УНИВЕРСИТЕТ ИТМО
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
сентябрь–октябрь 2020 Том 20 № 5 ISSN 2226-1494 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
September–October 2020 Vol. 20 No 5 ISSN 2226-1494 http://ntv.ifmo.ru/en/
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК УНИВЕРСИТЕТ ИТМО ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

УДК 004.932.2 doi: 10.17586/2226-1494-2020-20-5-683-691

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ УСТРАНЕНИЯ ДИСБАЛАНСА КЛАССОВ ЭМОЦИЙ В ВИДЕОДАННЫХ ВЫРАЖЕНИЙ ЛИЦ

Е.В. Рюмина[a,b], А.А. Карпов[a,b]

[a] Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация
[b] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
Адрес для переписки: ryumina_ev@mail.ru

Информация о статье
Поступила в редакцию 20.07.20, принята к печати 30.08.20
Язык статьи — русский

Ссылка для цитирования: Рюмина Е.В., Карпов А.А. Сравнительный анализ методов устранения дисбаланса классов эмоций в видеоданных выражений лиц // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 5. С. 683–691. doi: 10.17586/2226-1494-2020-20-5-683-691

Аннотация

Предмет исследования. Несбалансированность классов в наборах данных негативно влияет на системы машинной классификации, применяемые в таких приложениях искусственного интеллекта как медицинская диагностика заболеваний, обнаружение обмана и управление рисками. Эта проблема в наборах данных выражений лиц также ухудшает эффективность алгоритмов классификации. **Метод.** Рассмотрены основные подходы для уменьшения дисбаланса классов: методы повторной выборки и установление весов классам в зависимости от количества наблюдаемых образцов для каждого класса. Для локализации области лица в потоке кадров использован метод гистограммы направленных градиентов, и применена активная модель формы, которая обнаруживает координаты 68 ключевых ориентиров лица. С помощью координат ключевых ориентиров извлекаются информативные признаки, характеризующие динамику выражений лиц. **Основные результаты.** Результаты исследования показали, что предложенный подход извлечения визуальных признаков повышает точность распознавания эмоций по выражениям лиц. Рассмотренные методы уменьшения дисбаланса классов в наборе данных выражений лиц позволили повысить эффективность машинного классификатора, а также показали, что имеющийся дисбаланс классов в обучающем наборе оказывает значительное влияние на точность. **Практическая значимость.** Предложенный подход извлечения визуальных признаков может быть использован в автоматических системах распознавания эмоций человека по выражениям лиц, а анализ результатов применения методов уменьшения дисбаланса классов данных может быть полезен исследователям в области машинного обучения.

Ключевые слова
дисбаланс классов данных, недостаточная выборка, избыточная выборка, классификация, распознавание выражений лиц, извлечение визуальных признаков, активная модель формы

Благодарности
Исследование выполнено при поддержке Российского научного фонда (проект № 18-11-00145).

doi: 10.17586/2226-1494-2020-20-5-683-691

COMPARATIVE ANALYSIS OF METHODS FOR IMBALANCE ELIMINATION OF EMOTION CLASSES IN VIDEO DATA OF FACIAL EXPRESSIONS

E.V. Ryumina[a,b], A.A. Karpov[a,b]

[a] St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation
[b] ITMO University, Saint Petersburg, 197101, Russian Federation
Corresponding author: ryumina_ev@mail.ru

Article info
Received 20.07.20, accepted 30.08.20
Article in Russian

For citation: Ryumina E.V., Karpov A.A. Comparative analysis of methods for imbalance elimination of emotion classes in video data of facial expressions. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, vol. 20, no. 5, pp. 683–691 (in Russian). doi: 10.17586/2226-1494-2020-20-5-683-691

Abstract

Subject of Research. The imbalance of classes in datasets has a negative impact on machine classification systems used in applications of artificial intelligence, such as: medical diagnostics, fraud detection and risk management. This problem in facial expression datasets also degrades the performance of classification algorithms. **Method.** The paper discusses the main approaches for the class imbalance reduction: resampling methods and setting the weight of classes depending on the number of samples observed for an each class. A histogram of oriented gradients is used for the face area localization in the frame stream, then an active shape model is applied, which detects the coordinates of 68 key facial landmarks. Using the coordinates of key landmarks, informative features are extracted that characterize the dynamics of facial expressions. **Main Results.** The results of the study have shown that the proposed approach to the extraction of visual features exceeds the accuracy of human emotion recognition by facial expressions. The considered methods of the class imbalance reduction in the set of facial expressions have provided the improvement of machine classifier performance and showed that the existing class imbalance in a training set has a significant effect on the accuracy. **Practical Relevance.** The proposed approach to the extraction of visual features can be used in automatic systems for human emotion recognition by facial expressions, and result analysis of applying methods that reduce class imbalance can be useful for researchers in the field of machine learning.

Keywords

data class imbalance, under-sampling, over-sampling, classification, facial expression recognition, visual feature extraction, active shape model

Acknowledgements

This research was supported by the Russian Science Foundation (project No.18-11-00145).

Введение

В задаче распознавания образов при идеальных условиях машинный классификатор обучается на репрезентативном наборе данных, сбалансированном по количеству экземпляров в классах. В реальности практически всегда наблюдается значительный дисбаланс представленных наблюдений для разных классов. Данная проблема оказывает существенное влияние на алгоритмы классификации, применяемые в таких областях как медицинская диагностика заболеваний [1, 2], обнаружение обмана (например, в полиграфах) [3, 4], управление рисками [5, 6] и др. Это, в свою очередь, создает трудности для алгоритмов машинного обучения, так как предсказания обученной модели смещаются в сторону классов, имеющих больше наблюдений. Для решения данной проблемы часто используются следующие методы балансировки классов набора данных:

— недостаточная выборка (under-sampling) для мажоритарных классов [7];
— избыточная выборка (over-sampling) для миноритарных классов [1];
— взвешивание классов (в соответствии с количеством экземпляров) [8].

Несмотря на то, что существуют методы для уменьшения дисбаланса классов, применять их необходимо с осторожностью, так как при использовании методов недостаточной выборки для мажоритарного класса возможно удаление ключевых наблюдений, а при использовании методов избыточной выборки для миноритарного класса — переобучение [9]. По этой причине в работе с несбалансированными мультиклассовыми наборами данных легко потерять точность классификации в одном классе, пытаясь получить ее в другом.

В настоящей работе представлен краткий анализ традиционных методов уменьшения дисбаланса классов, которые исследуются на наборе данных выражений лиц. Предложен подход предварительной обработки видеоданных, заключающийся в нахождение графических областей лиц с последующим определением координат ключевых лицевых ориентиров (точек). Наличие координат позволяет извлечь информативные признаки, а именно: центр тяжести, расстояние от центра тяжести до ключевых лицевых точек и их угловое смещение. Извлеченные признаки подаются на машинный классификатор. Для анализа влияния методов балансирования классов в обучающем наборе вводится искусственный дисбаланс. Осуществлен сравнительный анализ полученных результатов до и после применения методов уменьшения дисбаланса классов. Выводы представлены в последнем разделе статьи.

Краткий аналитический обзор

С проблемой дисбаланса классов сталкиваются многие исследователи и разработчики систем искусственного интеллекта. В основном исследователи используют классические подходы для балансировки классов, к которым относятся подходы, основанные на недостаточной или избыточной выборках. Так, авторы в [4] использовали алгоритмы избыточной выборки Synthetic Minority Over-sampling Technique (SMOTE) [10] и Adaptive Synthetic (ADASYN) [11] для уменьшения дисбаланса классов в базах данных обнаружения обмана в речевых высказываниях. Исследование показало, что алгоритм SMOTE в комбинации с методом опорных векторов позволил повысить эффективность распознавания по показателю невзвешенной средней полноты (Unweighted Average Recall, UAR) на 11 %. В работе [1] авторы применили алгоритм SMOTE для синтетической генерации наблюдений на базе данных, содержащих электрокардиограммы, что позволило уменьшить дисбаланс пяти рассматриваемых классов и повысить эффективность системы. Проблема дисбаланса классов на аудиовизуальном и речевых наборах данных с акцентом на распознавание эмоций диктора по аудиоданным исследована в [12]. Для решения этой проблемы был предложен метод выборочной интерполяции синтетического меньшинства с избыточной выборкой, который является расширением метода SMOTE. Также авторы проводят сравнение с

другими методами избыточной выборки. Результаты исследования показали, что применение методов избыточной выборки позволяет повысить эффективность классификатора.

В исследованиях систем анализа выражений лиц существующие базы данных подразделяются на собранные в лабораторных и в реальных условиях («дикой природы»). Базы данных, собранные в реальных условиях, имеют, как правило, сильный дисбаланс классов, так как эмоции, такие как удивление, отвращение и страх встречаются реже остальных, что сильно затрудняет работу классификатора и смещает предсказания в сторону мажоритарных классов для увеличения точности распознавания эмоций в целом. В базах данных, собранных в лабораторных условиях, имеются другие проблемы, например, наигранность эмоций, что делает применение обученных моделей на таких наборах данных в реальных условиях затруднительным, так как в действительности эмоции проявляются иначе. В работе [7] авторы представили новую базу данных выражений лиц RAF-DB, которая была собрана в реальных условиях с большим дисбалансом классов. Авторы также рассматривают в своей работе традиционные подходы для уменьшения дисбалансов, такие как Random Under-sampling, Random Over-sampling и SMOTE, а также предлагают свой подход Virtual Facial Sample Generation (VFSG), который заключается в предобработке исходных изображений с изменением освещенности и позы головы. Результаты исследования показали, что использование предложенных методов позволяет повысить эффективность классификатора. Для балансировки классов в [13] предложена нейронная сеть The Deep Emo-Transfer Network (DETN). DETN учитывает дисбаланс классов за счет введения весовых параметров, что позволяет повысить точность распознавания выражений лица в несбалансированных наборах данных. Авторы статьи [14] для уменьшения дисбаланса классов в наборах данных выражений лиц использовали метод SMOTE. В работе [15] авторы применяют генеративные состязательные сети (Generative Adversarial Networks, GAN), которые генерируют новые наблюдения изображений лиц, с помощью данного подхода авторы получили увеличение показателя средней точности на 7,38 %.

Исследовательские данные

В данной работе используется аудиовизуальная база данных выражений лиц CREMA-D [16], которая содержит 7 442 файла речевых записей от 91 актера, имеющих различную этническую принадлежность, возрастом от 20 до 74 лет. Актеры с разной интенсивностью имитировали 6 эмоций: гнев, отвращение, счастье, грусть, страх и нейтральность. База данных была оценена 2 443 людьми отдельно для аудио-, видео- и аудиовизуальных данных, точность ручной оценки для рассматриваемых модальностей составила 40,9, 58,2 и 63,6 % соответственно. В табл. 1 представлено распределение данных по классам как в целом по динамическим видеофайлам, так и по статическим кадрам.

Как можно заметить из табл. 1, меньше всего представлено экземпляров видео для класса нейтральность. Тогда как покадровая оценка показала, что видеофайлы имеют разную продолжительность; можно заметить, что меньше всего кадров представлено для трех классов (нейтральность, счастье и страх).

Обзор методов уменьшения дисбаланса классов

Традиционными методами для уменьшения дисбаланса классов в наборе данных являются методы повторной выборки, которые делятся на две категории: недостаточная выборка для мажоритарного класса и избыточная выборка для миноритарного класса.

К первой категории относятся следующие методы.

1. Neighborhood cleaning rule (NCR) [17]. Недостаточная выборка выполняется для наблюдений мажоритарного класса, которые негативно влияют на классификацию миноритарного класса. Для этого все наблюдения классифицируются по правилу трех ближайших соседей, удаляются те наблюдения мажоритарного класса, которые получили верную метку, и те наблюдения, которые являются соседями миноритарного класса и были неверно классифицированы.
2. Tomek Links (TL) [18]. Недостаточная выборка выполняется следующим образом. Для двух наблюдений X_i и X_j, принадлежащих к различным классам, рассчитывается эвклидово расстояние (Euclidean Distance) $dist(X_i, X_j)$. Пара X_i и X_j называется связью Томека, если нет ни одного наблюдения X_n, для которого будет справедлива система уравнений:

$$\begin{cases} dist(X_i, X_n) < dist(X_i, X_j) \\ dist(X_j, X_n) < dist(X_i, X_j) \end{cases},$$

где X_i, X_j и X_n — случайные наблюдения из обучающего набора данных. Так, все наблюдения мажоритарного класса, которые входят в связи Томека, будут удалены из набора данных.
3. One-sided selection (OSS) [19]. На первом этапе в общее множество объединяются все наблюдения мажоритарного класса и выбранные случайным образом n наблюдений из миноритарного класса. Далее по правилу n-ближайших соседей отбираются только те наблюдения, которые получили ошибоч-

Таблица 1. Распределение данных по классам эмоций в базе данных CREMA-D

Количество, шт.	Гнев	Отвращение	Счастье	Грусть	Страх	Нейтральность	Всего
Видео	1 271	1 271	1 271	1 271	1 271	1 087	7 442
Кадр	98 709	108 000	89 392	99 255	95 950	79 219	570 525

ную метку. На втором этапе применяется метод связей Томека.

4. NearMiss (NM) [20]. Сначала находятся евклидовы расстояния между всеми наблюдениями мажоритарного и миноритарного классов. Затем выбираются n наблюдений мажоритарного класса, которые наиболее близки к миноритарному классу. Удаляются те наблюдения из выборки n, которые максимизируют расстояние между наблюдениями двух классов.
5. Random Under-sampling (RUS) [21]. Позволяет случайным образом сбалансировать наблюдения между классами.

Ко второй категории относятся методы.

1. SMOTE [10]. Генерирует синтетические наблюдения следующим образом. Выбирается одно наблюдение из миноритарного класса, по отношению к которому находятся n-ближайших соседей из того же класса, далее из n-ближайших соседей выбирается случайным образом одно наблюдение, находится эвклидово расстояние между выбранными наблюдениями для каждого признака. Полученные расстояния умножаются на случайное число в интервале от (0, 1), затем найденные значения прибавляются к ранее выбранному n-му ближайшему соседу.
2. Borderline-SMOTE (BSMOTE) [22]. Делит наблюдения миноритарного класса на три группы: безопасные (все ближайшие соседи принадлежат к тому же классу, что и класс рассматриваемого наблюдения); опасные (минимум половина ближайших соседей принадлежит к тому же классу, что и класс рассматриваемого наблюдения); шум (все ближайшие соседи принадлежат к классу, отличному от класса рассматриваемого наблюдения). Генерация синтетических наблюдений осуществляется только с опасными наблюдениями идентично методу SMOTE.
3. SMOTE SVM (SMOTE-SVM) [23]. Осуществляет генерацию синтетических наблюдений вокруг тех экземпляров миноритарного класса, которые близки к опорным векторам.
4. Random Over-sampling (ROS) [21]. Позволяет случайным образом сгенерировать синтетические наблюдения миноритарного класса.

Помимо представленных методов широко используется метод взвешивания классов (class weight, CW), который позволяет установить больший вес миноритарным классам, тем самым подсказывая классификатору, на какой класс стоит обратить больше внимания. Вес класса рассчитывается по формуле:

$$\text{Вес}_i = \ln\left(\frac{rN}{n_i}\right),$$

где r — параметр, регулирующий веса (чем больше значение r, тем больше веса похожи на истинные классовые отношения); N — количество наблюдений всех классов; n_i — количество наблюдений в классе i.

Вес устанавливается по рассчитанному значению, если он превышает 1, иначе приравнивается к 1.

Предложенный подход

Аудиовизуальная база данных CREMA-D содержит сырые (необработанные) видеофайлы, и важным этапом является предварительная обработка видеопоследовательностей для дальнейшего распознавания эмоций. Для обнаружения областей лица в потоке кадров использовался метод гистограммы направленных градиентов, а для нахождения координат 68 ключевых ориентиров лица применялась активная модель формы. Для анализа эмоций по выражениям лица наличие координат ключевых ориентиров является недостаточным, поэтому необходимо извлечение векторов информативных признаков, состоящих из расстояний ключевых точек до центра тяжести, которым являются средние значения координат (x, y), а также угловое смещение, учитывающее изменение положения ключевых координат от кадра к кадру. Далее извлеченные векторы признаков, характеризующие динамику лица, подаются на машинный классификатор. В качестве вероятностной модели применяется одномерная сверточная нейронная сеть (Convolutional Neural Network, CNN).

Поскольку дисбаланс классов в рассматриваемой базе данных отличается от естественно встречающегося дисбаланса в реальных условиях, для анализа влияния методов балансирования классов вводится искусственный дисбаланс. Также проведены анализ влияния рассмотренных методов на количественные показатели эффективности классификатора и сравнение полученных результатов предложенного подхода извлечения важных векторов признаков с результатами других исследователей.

Предварительная обработка базы данных

База данных CREMA-D содержит необработанные видеопоследовательности кадров, поэтому необходимо осуществление предварительной обработки (предобработки) видеоданных. Предобработка данных подразумевает детектирование областей лиц, т. е. нахождение их графического местоположения. В качестве детектора областей лиц использовался метод гистограммы направленных градиентов (Histogram of oriented gradients) [24]. Затем на графических местоположениях областей лиц с помощью активной модели формы (Active shape model) [25] обнаруживались 68 ключевых лицевых ориентиров (рис. 1). Оба метода реализованы в библиотеке с открытым исходным кодом Dlib [26]. Для обучающего набора выбирался каждый 3-й кадр видео, для тестового набора — каждый 5-й, это было необходимо, чтобы уменьшить переобучение нейросети схожими векторами признаков. Так как пропорции лиц людей различны, то координаты ключевых ориентиров нормализованы до квадратной области 224 × 224 пикселей. В работе [27] представлен алгоритм для извлечения информативных признаков из координат лицевых ориентиров. Авторы предлагают определить «центр масс» для координат (x, y), затем рассчитать расстояние между «центром масс» и координатами точек, а также найти их угловое смещение. Разработанный алгоритм применен в данной работе, но без учета координат

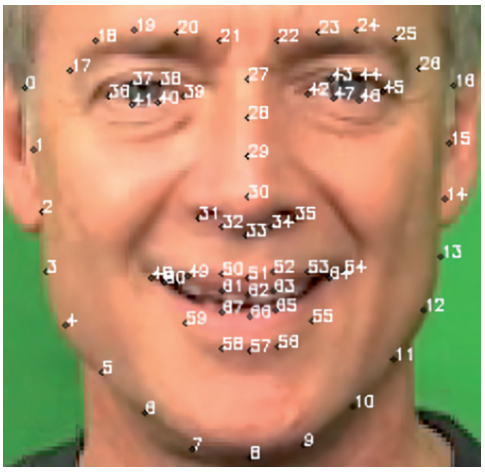

Рис. 1. Пример извлечения 68 ключевых ориентиров лица

ключевых точек с номерами 0–16, а также их расстояния до центра масс и углового смещения, так как они характеризуют форму лица, а не динамику изменения эмоций. Таким образом, для каждого кадра получен вектор размерностью 204 компонента.

При введении искусственного дисбаланса в базу данных CREMA-D, в качестве эталонов для распределения наблюдений по классам рассматривались базы данных AffectNet [28], FER-2013 [29] и RAF-DB [30] (их распределение представлено в табл. 2). Прореживание обучающего набора выполнялось с соблюдением процентного соотношения наблюдений классов в эталонных базах данных.

На рис. 2 представлено распределение наблюдений по классам эмоций в исходных обучающем и тестовом наборах, а также количество наблюдения в прореженных наборах данных, где I — данные прореживаний в соответствии с AffectNet, II — FER-2013 и III — RAF-DB.

Каждый из наборов данных I, II и III подвергался повторной выборке. Параметры для методов повторной выборки устанавливались по умолчанию[1]. Для метода CW параметр r устанавливался в интервале от 0 до 1. Лучший результат получен при параметре r равном 0,75, 0,51 и 0,49 для баз данных I, II и III соответственно. Наблюдения в наборах I, II и III и наблюдения, полученные после применения методов повторной выборки, подавались на вход нейронной сети. Так как извлеченные признаки различны по своим характеристикам (например, градусы и пиксели), то для подачи на классификатор необходимо их нормализовать. Нормализация для обучающего и тестового наборов данных производилась по средним значениям и стандартным отклонениям признаков обучающей выборки.

Эффективность классификатора рассчитывалась с помощью таких показателей как полнота (recall), точность (accuracy) и UAR по следующим формулам:

$$\text{Полнота}_i = \frac{TP_i}{n_i},$$

$$\text{Точность} = \frac{\sum_{i=1}^{K} \text{Полнота}_i \times n_i}{N},$$

$$UAR = \frac{1}{K} \sum_{i=1}^{K} \text{Полнота}_i,$$

где TP_i — верно классифицированные наблюдения в классе i; n_i — количество наблюдений в классе i; K — количество классов.

UAR, в отличие от точности, позволяет объективно оценивать эффективность классификатора для несбалансированных данных, поэтому является важным показателем для оценки работы системы.

[1]Imbalanced-learn API [Электронный ресурс]. https://imbalanced-learn.readthedocs.io/en/stable/api.html#module-imblearn.under_sampling, свободный. Яз. англ. (дата обращения: 05.07.2020).

Таблица 2. Распределение наблюдений по классам эмоций в эталонных базах данных, %

База данных	Гнев	Отвращение	Счастье	Грусть	Страх	Нейтральность
AffectNet	9,30	1,58	49,45	9,52	2,52	27,63
FER-2013	15,64	1,71	28,25	18,91	16,04	19,44
RAF-DB	6,42	6,53	43,46	18,05	2,56	22,99

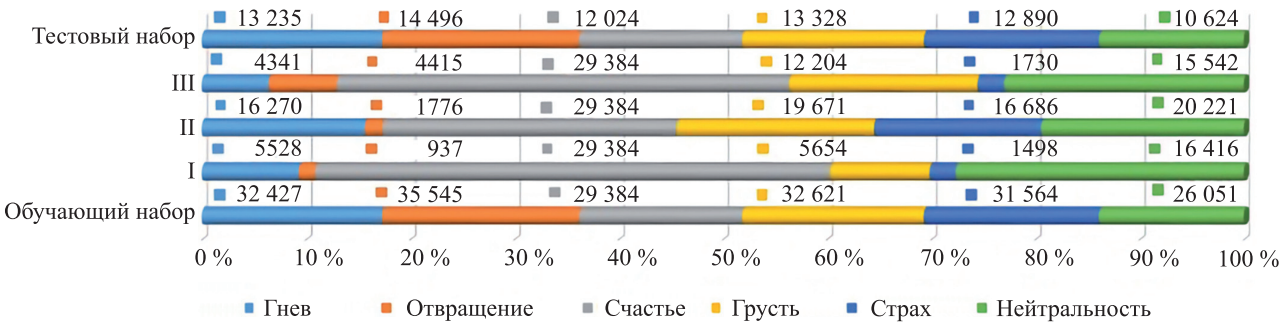

Рис. 2. Распределение наблюдений по классам эмоций в наборах данных

Экспериментальные исследования

Нормализованные векторы использовались в качестве входа одномерной CNN, которая имеет пять сверточных слоев (Conv, выходные фильтры 64, 128 и 256, окно 3, сдвиг 1), два слоя подвыборки (maxpool, окно 3, сдвиг 2), три полносвязных слоя (fc, размерность выходного пространства 256, 128 и 6) с функцией активации ReLU после сверточных и полносвязных слоев, за исключением последнего выходного слоя с функцией активации softmax для определения вероятностей отнесения наблюдения к шести эмоциям. После слоев подвыборки применяется прореживание (dropout) с вероятностью 0,5, после первых двух полносвязных слоев — 0,25. Количество эпох обучения составило 40, применялся оптимизатор Adam со скоростью обучения 0,001 и сокращением веса 0,00005. Для настройки оптимальных параметров использовался поиск по сетке. Архитектура примененной нейросети представлена на рис. 3.

В табл. 3 представлены точность и UAR проведенных экспериментов для различных рассмотренных методов на наборах данных I, II и III. Лучшие результаты по каждому столбцу выделены жирным шрифтом.

Результаты экспериментов показывают, что методы недостаточной выборки (NCR, TL, OSS, NM, RUS) не повышают эффективность классификации в поставленной задаче распознавания эмоций. Одним из значимых недостатков методов повторной выборки является требование дополнительных вычислительных и временных ресурсов. Метод балансировки CW лишен этого недостатка, так как он требует только наличие весовых коэффициентов, однако при использовании данного метода достигается меньший прирост значений точности и UAR. Наибольший прирост эффективности классификации на тестовом наборе продемонстрировал метод SMOTE при обучении на всех прореженных наборах данных. Так, при обучении на наборе данных I получен прирост точности и UAR на 6 % и 5,48 % соответственно. При обучении на наборах данных II и III прирост точности составил 3,05 % и 6 %, для UAR — 2,59 % и 5,58 %. Также стоит отметить, что чем выше дисбаланс классов в прореженных наборах данных (рис. 2), тем достигается больший прирост значений точности и UAR. Лучшее значение точности равно 79,64 % и UAR — 80,31 % получено при обучении на наборе данных II, в котором классы лучше сбалансированы по сравнению с I и III. На рис. 4 представлена матрица спутывания для тестового набора до и после применения метода избыточной выборки SMOTE к обучающему набору (II набор данных).

Как можно заметить из рис. 4 с помощью синтетической выборки для миноритарных классов количество верно классифицированных наблюдений для эмоции отвращение увеличилось на 2 330 (т. е. прирост полноты составил 16,07 % от общего количества наблюдений для эмоции отвращения в тестовом наборе (рис. 2)), гнев – 596 (4,5 %), страх — 100 (0,77 %), для остальных эмоций наблюдается уменьшение верно классифицированных наблюдений в среднем на 232 (2 %).

Рис. 3. Архитектура сверточной нейронной сети

Таблица 3. Экспериментальные результаты для различных методов уменьшения дисбаланса данных

Метод	I, %		II, %		III, %	
	Точность	UAR	Точность	UAR	Точность	UAR
Исходные данные	63,54	65,13	76,59	77,72	68,33	69,23
NCR	54,36	56,04	73,59	74,64	62,30	63,51
TL	62,49	64,00	76,06	77,13	68,44	69,36
OSS	62,02	63,56	76,07	77,16	68,23	69,09
NM	32,93	31,34	26,21	24,41	34,86	33,73
RUS	54,23	54,33	60,04	60,20	59,66	59,64
SMOTE	**69,54**	**70,61**	**79,64**	**80,31**	**74,33**	**74,81**
BSMOTE	67,81	69,10	79,04	79,86	73,02	73,56
SMOTE-SVM	67,45	68,73	79,27	80,19	72,58	73,15
ROS	66,40	67,77	79,27	80,06	72,77	73,32
CW	66,31	67,50	78,19	79,01	71,47	72,15

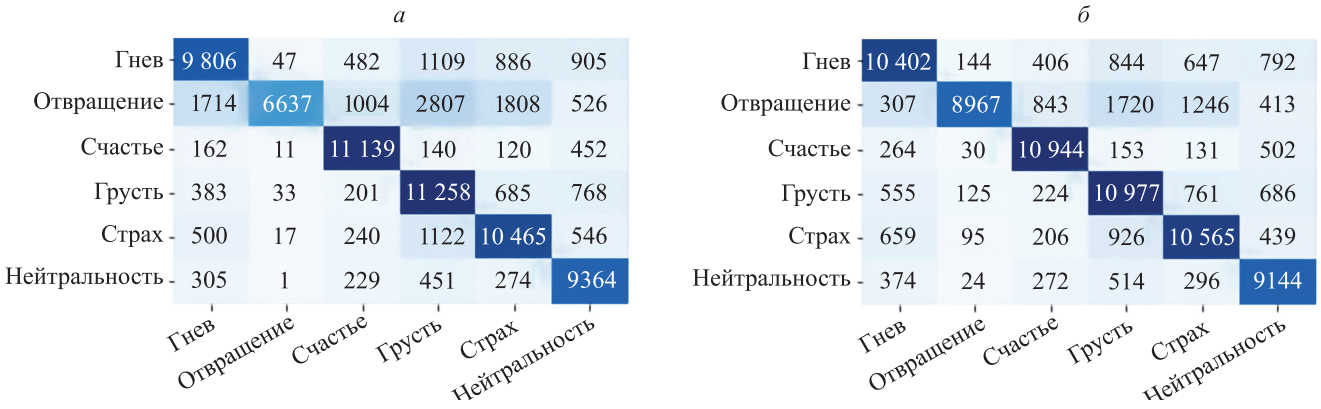

Рис. 4. Матрица спутывания для тестового набора до (*а*) и после (*б*) применения метода избыточной выборки SMOTE к обучающему набору

Сравнение с другими современными результатами

Большинство исследований на аудиовизуальной базе данных CREMA-D проводятся с акцентом на аудиомодальность, что, в свою очередь, приводит к недостаточному анализу визуальной модальности. С помощью рейтинговой системы восприятия видеоданных в работе [16] выполнена оценка распознавания эмоций людьми, специализирующимися на опросах. В работе [31] детектирование областей лиц выполнено при помощи метода ансамбля дерева решений. Затем полученные изображения выравнивались относительно центра глаз, носа и других частей лица, далее обрезались и масштабировались к разрешению 224 × 224. Предобработанные изображения подавались на предобученную нейронную сеть VGG-face. Для нахождения средней точности авторы поделили набор данных на 10 частей и осуществили 10-кратную перекрестную проверку (cross-validation). Для сравнения с работой [31] обучающая и тестовая выборки набора данных CREMA-D были также поделены на 10 частей для осуществления перекрестной проверки. В табл. 4 представлено среднее значение точности распознавания тестовых наборов, полученное в результате перекрестной проверки, а также результаты других исследователей.

Таблица 4. Сравнение полученных результатов с другими исследованиями

Исследование	Средняя точность, %
Cao H. et al. 2014 [16]	58,2
Ghaleb E. et al. 2020 [31]	66,8
Полученный результат	83,0

Как можно заметить из табл. 4 рассмотренный метод значительно превосходит полученные ранее результаты.

Заключение

В работе предложен подход извлечения визуальных признаков с использованием активной модели формы. Для устранения дисбаланса классов рассмотрено использование традиционных методов, таких как недостаточная и избыточная выборки, а также установка большего веса классам, имеющим меньше наблюдений. Совокупность методов позволила повысить показатели эффективности классификатора. Анализ полученных результатов показал, что вычисление информативных признаков из ключевых точек лица позволяет достигать высокой точности классификации эмоций по выражениям лиц. Кроме того, было выявлено, что чем лучше сбалансированы классы в обучающем наборе, тем выше эффективность алгоритмов распознавания выражений. Применение рассмотренных методов недостаточной выборки для балансировки данных не улучшает показатели эффективности в мультиклассовой классификации. Тогда как методы избыточной выборки и взвешивание классов позволяют значительно повысить показатели классификации для нескольких классов, имеющих меньше обучающих наблюдений.

В дальнейшем планируется проведение экспериментов на других базах данных выражений лиц, определение наиболее информативных признаков, вычисленных из координат 68 ключевых ориентиров лица и использование нейронных сетей с длинной краткосрочной памятью, которые позволят изучать динамику изменений векторных признаков во времени.

Литература

1. Pandey S.K., Janghel R.R. Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE // Australasian Physical & Engineering Sciences in Medicine. 2019. V. 42. N 4. P. 1129–1139. doi: 10.1007/s13246-019-00815-9
2. Han W., Huang Z., Li S., Jia Y. Distribution-sensitive unbalanced data oversampling method for medical diagnosis // Journal of Medical Systems. 2019. V. 43. N 2. P. 39. doi: 10.1007/s10916-018-1154-8

References

1. Pandey S.K., Janghel R.R. Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE. *Australasian Physical & Engineering Sciences in Medicine*, 2019, vol. 42, no. 4, pp. 1129–1139. doi: 10.1007/s13246-019-00815-9
2. Han W., Huang Z., Li S., Jia Y. Distribution-sensitive unbalanced data oversampling method for medical diagnosis. *Journal of Medical Systems*, 2019, vol. 43, no. 2, pp. 39. doi: 10.1007/s10916-018-1154-8

3. Ahammad J., Hossain N., Alam M.S. Credit card fraud detection using data pre-processing on imbalanced data — Both oversampling and undersampling // Proc. of the International Conference on Computing Advancements. 2020. doi: 10.1145/3377049.3377113
4. Velichko A., Karpov A. A study of data scarcity problem for automatic detection of deceptive speech utterances // CEUR Workshop Proceedings. 2020. V. 2552. P. 38–46.
5. Sun J., Lang J., Fujita H., Li H. Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates // Information Sciences. 2018. V. 425. P. 76–91. doi: 10.1016/j.ins.2017.10.017
6. Leong C.K. Credit risk scoring with bayesian network models // Computational Economics. 2016. V. 47. N 3. P. 423–446. doi: 10.1007/s10614-015-9505-8
7. Li S., Deng W. Real world expression recognition: A highly imbalanced detection problem // Proc. 9th International Conference on Biometrics (ICB 2016). 2016. P. 7550074. doi: 10.1109/ICB.2016.7550074
8. Kaya H., Karpov A.A. Introducing weighted kernel classifiers for handling imbalanced paralinguistic corpora: Snoring, addressee and cold // Proc. 18th Annual Conference of the International Speech Communication Association (INTERSPEECH 2017). 2017. P. 3527–3531. doi: 10.21437/Interspeech.2017-653
9. Johnson J.M., Khoshgoftaar T.M. Survey on deep learning with class imbalance // Journal of Big Data. 2019. V. 6. N 1. P. 27. doi: 10.1186/s40537-019-0192-5
10. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: synthetic minority over-sampling technique // Journal of Artificial Intelligence Research. 2002. V. 16. P. 321–357. doi: 10.1613/jair.953
11. He H., Bay Y., Garcia E.A., Li S. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning // Proc. of the IEEE International Joint Conference on Neural Networks (IJCNN 2008). 2008. P. 1322–1328. doi: 10.1109/IJCNN.2008.4633969
12. Liu Z.-T., Wu B.-H., Li D.-Y., Xiao P., Mao J.-W. Speech emotion recognition based on selective interpolation synthetic minority over-sampling technique in small sample environment // Sensors. 2020. V. 20. N 8. P. 2297. doi: 10.3390/s20082297
13. Li S., Deng W. Deep emotion transfer network for cross-database facial expression recognition // Proc. 24th International Conference on Pattern Recognition (ICPR 2018). 2018. P. 3092–3099. doi: 10.1109/ICPR.2018.8545284
14. Rashid T.A. Convolutional neural networks based method for improving facial expression recognition // Advances in Intelligent Systems and Computing. 2016. V. 530. P. 73–84. doi: 10.1007/978-3-319-47952-1_6
15. Yi W., Sun Y., He S. Data augmentation using conditional GANs for facial emotion recognition // Proc. of the Progress in Electromagnetics Research Symposium (PIERS-Toyama 2018). 2018. P. 710–714. doi: 10.23919/PIERS.2018.8598226
16. Cao H., Cooper D.G., Keutmann M.K., Gur R.C., Nenkova A., Verma R. CREMA-D: Crowd-sourced emotional multimodal actors dataset // IEEE Transactions on Affective Computing. 2014. V. 5. N 4. P. 377–390. doi: 10.1109/TAFFC.2014.2336244
17. Wilson D.L. Asymptotic properties of nearest neighbor rules using edited data // IEEE Transactions on Systems, Man and Cybernetics. 1972. V. 2. N 3. P. 408–421. doi: 10.1109/TSMC.1972.4309137
18. Tomek I. Two modifications of CNN // IEEE Transactions on Systems, Man and Cybernetics. 1976. V. 6. N 11. P. 769–772. doi: 10.1109/TSMC.1976.4309452
19. Kubat M., Matwin S. Addressing the curse of imbalanced training sets: one-sided selection // Proc. 14th International Conference on Machine Learning. 1997. P. 179–186.
20. Zhang I., Mani I. kNN approach to unbalanced data distributions: a case study involving information extraction // Proc. of Workshop on Learning from Imbalanced Datasets. 2003. P. 42–48.
21. Lemaître G., Nogueira F., Aridas C.K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning // Journal of Machine Learning Research. 2017. V. 18. P. 559–563.
22. Han H., Wang W.-Y., Mao B.-H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning // Lecture Notes in Computer Science. 2005. V. 3644. P. 878–887. doi: 10.1007/11538059_91
23. Nguyen H.M., Cooper E.W., Kamei K. Borderline over-sampling for imbalanced data classification // International Journal of Knowledge Engineering and Soft Data Paradigms (IJKESDP). 2011. V. 3. N 1. P. 4–21. doi: 10.1504/IJKESDP.2011.039875

3. Ahammad J., Hossain N., Alam M.S. Credit card fraud detection using data pre-processing on imbalanced data — Both oversampling and undersampling. *Proc. of the International Conference on Computing Advancements*, 2020. doi: 10.1145/3377049.3377113
4. Velichko A., Karpov A. A study of data scarcity problem for automatic detection of deceptive speech utterances. *CEUR Workshop Proceedings*, 2020, vol. 2552, pp. 38–46.
5. Sun J., Lang J., Fujita H., Li H. Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates. *Information Sciences*, 2018, vol. 425, pp. 76–91. doi: 10.1016/j.ins.2017.10.017
6. Leong C.K. Credit risk scoring with bayesian network models. *Computational Economics*, 2016, vol. 47, no. 3, pp. 423–446. doi: 10.1007/s10614-015-9505-8
7. Li S., Deng W. Real world expression recognition: A highly imbalanced detection problem. *Proc. 9th International Conference on Biometrics (ICB 2016)*, 2016, pp. 7550074. doi: 10.1109/ICB.2016.7550074
8. Kaya H., Karpov A.A. Introducing weighted kernel classifiers for handling imbalanced paralinguistic corpora: Snoring, addressee and cold. *Proc. 18th Annual Conference of the International Speech Communication Association (INTERSPEECH 2017)*, 2017, pp. 3527–3531. doi: 10.21437/Interspeech.2017-653
9. Johnson J.M., Khoshgoftaar T.M. Survey on deep learning with class imbalance. *Journal of Big Data*, 2019, vol. 6, no. 1, pp. 27. doi: 10.1186/s40537-019-0192-5
10. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 2002, vol. 16, pp. 321–357. doi: 10.1613/jair.953
11. He H., Bay Y., Garcia E.A., Li S. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. *Proc. of the IEEE International Joint Conference on Neural Networks (IJCNN 2008)*, 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969
12. Liu Z.-T., Wu B.-H., Li D.-Y., Xiao P., Mao J.-W. Speech emotion recognition based on selective interpolation synthetic minority over-sampling technique in small sample environment. *Sensors*, 2020, vol. 20, no. 8, pp. 2297. doi: 10.3390/s20082297
13. Li S., Deng W. Deep emotion transfer network for cross-database facial expression recognition. *Proc. 24th International Conference on Pattern Recognition (ICPR 2018)*, 2018, pp. 3092–3099. doi: 10.1109/ICPR.2018.8545284
14. Rashid T.A. Convolutional neural networks based method for improving facial expression recognition. *Advances in Intelligent Systems and Computing*, 2016, vol. 530, pp. 73–84. doi: 10.1007/978-3-319-47952-1_6
15. Yi W., Sun Y., He S. Data augmentation using conditional GANs for facial emotion recognition. *Proc. Progress in Electromagnetics Research Symposium (PIERS-Toyama 2018)*, 2018, pp. 710–714. doi: 10.23919/PIERS.2018.8598226
16. Cao H., Cooper D.G., Keutmann M.K., Gur R.C., Nenkova A., Verma R. CREMA-D: Crowd-sourced emotional multimodal actors dataset. *IEEE Transactions on Affective Computing*, 2014, vol. 5, no. 4, pp. 377–390. doi: 10.1109/TAFFC.2014.2336244
17. Wilson D.L. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man and Cybernetics*, 1972, vol. 2, no. 3, pp. 408–421. doi: 10.1109/TSMC.1972.4309137
18. Tomek I. Two modifications of CNN. *IEEE Transactions on Systems, Man and Cybernetics*, 1976, vol. 6, no. 11, pp. 769–772. doi: 10.1109/TSMC.1976.4309452
19. Kubat M., Matwin S. Addressing the curse of imbalanced training sets: one-sided selection. *Proc. 14th International Conference on Machine Learning*, 1997, pp. 179–186.
20. Zhang I., Mani I. kNN approach to unbalanced data distributions: a case study involving information extraction. *Proc. of Workshop on Learning from Imbalanced Datasets*, 2003, pp. 42–48.
21. Lemaître G., Nogueira F., Aridas C.K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 2017, vol. 18, pp. 559–563.
22. Han H., Wang W.-Y., Mao B.-H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. *Lecture Notes in Computer Science*, 2005, vol. 3644, pp. 878–887. doi: 10.1007/11538059_91
23. Nguyen H.M., Cooper E.W., Kamei K. Borderline over-sampling for imbalanced data classification. *International Journal of Knowledge*

24. Déniz O., Bueno G., Salido J., De La Torre F. Face recognition using histograms of oriented gradients // Pattern Recognition Letters. 2011. V. 32. N 12. P. 1598–1603. doi: 10.1016/j.patrec.2011.01.004
25. Cootes T.F., Taylor C.J., Cooper D.H., Graham J. Active shape models-their training and application // Computer Vision and Image Understanding. 1995. V. 61. N 1. P. 38–59. doi: 10.1006/cviu.1995.1004
26. King D.E. Dlib-ml: A machine learning toolkit // Journal of Machine Learning Research. 2009. V. 10. P. 1755–1758.
27. Van Gent P. Emotion Recognition Using Facial Landmarks Python DLib and OpenCV // A tech blog about fun things with Python Embed. Electron. 2016 [Электронный ресурс]. URL: http://www.paulvangent.com/2016/08/05/emotion-recognition-using-facial-landmarks/ (дата обращения: 19.07.2020).
28. Mollahosseini A., Hasani B., Mahoor M.H. AffectNet: A database for facial expression, valence, and arousal computing in the wild // IEEE Transactions on Affective Computing. 2019. V. 10. N 1. P. 18–31. doi: 10.1109/TAFFC.2017.2740923
29. Carrier P. L., Courville A., Goodfellow I.J., Mirza M., Bengio Y. FER-2013 face database: Technical report 1365. Universit de Montral, 2013.
30. Li S., Deng W., Du J. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild // Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017). 2017. P. 2584–2393. doi: 10.1109/CVPR.2017.277
31. Ghaleb E., Popa M., Asteriadis S. Metric learning-based multimodal audio-visual emotion recognition // IEEE Multimedia. 2020. V. 27. N 1. P. 37–48. doi: 10.1109/MMUL.2019.2960219

Engineering and Soft Data Paradigms (IJKESDP), 2011, vol. 3, no. 1, pp. 4–21. doi: 10.1504/IJKESDP.2011.039875
24. Déniz O., Bueno G., Salido J., De La Torre F. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 2011, vol. 32, no. 12, pp. 1598–1603. doi: 10.1016/j.patrec.2011.01.004
25. Cootes T.F., Taylor C.J., Cooper D.H., Graham J. Active shape models-their training and application. *Computer Vision and Image Understanding*, 1995, vol. 61, no. 1, pp. 38–59. doi: 10.1006/cviu.1995.1004
26. King D.E. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 2009, vol. 10, pp. 1755–1758.
27. Van Gent P. Emotion Recognition Using Facial Landmarks Python DLib and OpenCV. *A tech blog about fun things with Python Embed. Electron*, 2016. Available at: http://www.paulvangent.com/2016/08/05/emotion-recognition-using-facial-landmarks/ (accessed: 19.07.2020).
28. Mollahosseini A., Hasani B., Mahoor M.H. AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 2019, vol. 10, no. 1, pp. 18–31. doi: 10.1109/TAFFC.2017.2740923
29. Carrier P. L., Courville A., Goodfellow I.J., Mirza M., Bengio Y. *FER-2013 face database. Technical report 1365*. Universit de Montral, 2013.
30. Li S., Deng W., Du J. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 2017, pp. 2584–2393. doi: 10.1109/CVPR.2017.277
31. Ghaleb E., Popa M., Asteriadis S. Metric Learning-Based Multimodal Audio-Visual Emotion Recognition. *IEEE Multimedia*, 2020, vol. 27, no. 1, pp. 37–48. doi: 10.1109/MMUL.2019.2960219

Авторы

Рюмина Елена Витальевна — программист, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация; студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0002-4135-6949, ryumina_ev@mail.ru
Карпов Алексей Анатольевич — доктор технических наук, доцент, руководитель лаборатории, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация; профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, Scopus ID: 57195330987, ORCID ID: 0000-0003-3424-652X, karpov@iias.spb.su

Authors

Elena V. Ryumina —Software Developer, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation; Student, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0002-4135-6949, ryumina_ev@mail.ru
Alexey A. Karpov — D.Sc., Associate Professor, Laboratory Head, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation; Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, Scopus ID: 57195330987, ORCID ID: 0000-0003-3424-652X, karpov@iias.spb.su