

Применение модели VGG-Face для определения принадлежности человека к музыкальному сообществу

М.Д. Поляк

Санкт-Петербургский государственный университет
аэрокосмического приборостроения
Санкт-Петербург, Россия
markpolyak@gmail.com

Я.О. Сениченкова

Санкт-Петербургский государственный университет
аэрокосмического приборостроения
Санкт-Петербург, Россия
senichenkova02@inbox.ru

Аннотация — Рассматривается применение методов и алгоритмов машинного обучения для определения персональных черт человека, таких как владение профессиональными знаниями и навыками в сфере музыкального исполнительского искусства. Проверяется частная гипотеза о влиянии классического музыкального образования на лицо человека. Полученные результаты позволяют утверждать о потенциальной возможности в автоматическом режиме определять, является ли человек музыкантом симфонического оркестра, на основе одного лишь изображения его или ее лица.

Ключевые слова — классификация, распознавание изображений, VGGFace, машина опорных векторов, метод главных компонент, t-SNE.

I. ВВЕДЕНИЕ

Современные методы распознавания лиц могут быть использованы для решения задач идентификации, как в правоохранительных органах, так и в бизнесе, особенно в сфере обслуживания. Данные технологии могут упростить предоставление индивидуализированных услуг, таких как реклама, персональное обслуживание клиентов (например, персонализированное меню ресторана), а также персонализацию других продуктов, например, игр [1]. Алгоритмы машинного обучения по одной лишь фотографии лица позволяют узнать такую информацию, как наличие психических расстройств [2] (в том числе наличие шизофрении [3] или биполярного аффективного расстройства [4]), политические предпочтения [5], сексуальную ориентацию [6], наличие высшего образования [7].

Возникают вопросы об этичности извлечения персональной информации из фотографий лиц людей. Определение личных характеристик, таких как сексуальная ориентация и политические предпочтения, с помощью физиогномики вызывает сомнения, так как это является неподтвержденной наукой. Однако, определение эмоций, основанное на патогномике, может быть рассмотрено как более научная задача [8]. Патогномика основывается на отражении эмоций на лице, следах переживаний, образе жизни и других биологических процессах. В отличие от классической физиогномики, которая предполагает определение черт характера и поведения по форме лица без научного обоснования и биологической интерпретации, патогномика позволяет сделать выводы на основе биологических процессов.

В данной работе впервые рассматривается гипотеза о том, что музыкальное образование и профессия музыканта симфонического оркестра оставляют биологический след на человеческом лице. С точки

зрения авторов этой статьи задача определения профессии человека по фотографии его или ее лица относится к задаче патогномики и, следовательно, не должна приводить к возникновению вопроса о «лженаучности». Профессия музыканта симфонического оркестра была выбрана для данного исследования специально не только потому, что весьма затруднительно представить какие-либо обвинения в расизме по признаку наличия или отсутствия музыкального образования, но еще и потому, что фотографии музыкантов находятся в открытом доступе и их достаточно легко собрать.

Научная новизна исследования заключается в применении классических методов машинного обучения к новой области – определению неявных черт и особенностей человека, таких как его профессия. Впервые, продемонстрирована возможность определять род деятельности человека по его лицу на безобидном примере определения профессии музыканта. Предлагаемый подход может иметь большую ценность в сфере маркетинга и продаж, и в то же время представлять серьезную угрозу конфиденциальности. Для проверки поставленной гипотезы собран и размечен новый набор данных, большое внимание уделено анализу качества собранного датасета.

II. МОДЕЛЬ

Изображение, размером 224x224 поступает на вход предобученной сети VGGFace [9]. Далее изображение преобразуется в числовой вектор, называемый эмбедингом, содержащий информативные признаки. Затем эмбединг передается в классификатор, реализованный с помощью машины опорных векторов [10].

III. НАБОР ДАННЫХ

Для обучения и тестирования классификатора были собраны фотографии, соответствующие классам «Музыкант» и «Немузыкант». Итоговая версия датасета содержит 1318 изображений.

Для проверки того, что построенный классификатор действительно научился отличать лица музыкантов симфонического оркестра от не имеющих отношения к исполнению симфонической музыки людей был проведен углубленный анализ собранного набора данных. Первым шагом была осуществлена визуализация подготовленного набора данных путем преобразования эмбедингов в двумерные вектора с использованием двух алгоритмов уменьшения размерности: метода главных компонент (PCA) и алгоритма t-SNE.

Из рис. 1 можно сделать вывод, что два класса не являются линейно-разделимыми в пространстве признаков из двух главных компонент, однако присутствует ярко выраженный кластер изображений, соответствующий классу «Музыкант». Две первые компоненты описывают 15,7% и 7,1% дисперсии данных соответственно, что в сумме дает менее 22,8% объясненной дисперсии.

Использование алгоритма t-SNE не дает каких-либо явно выраженных кластеров (рис. 2). Возможная интерпретация заключается в том, что, либо эмбединги, сгенерированные с помощью VGGFace, являются набором случайных чисел и не несут никакого смысла, либо набор данных достаточно широк, разнообразен и оба класса более-менее полно представлены. С точки зрения авторов данной работы последнее более вероятно, т.к. при сборе данных брались изображения из различных источников.

IV. ОБУЧЕНИЕ МОДЕЛИ

По итогам обучения классификатора были получены следующие значения метрик качества на тестовой выборке из 184 изображений: accuracy 0,70; F1-score 0,734; ROC AUC score 0,699. Матрица ошибок для тестовой выборки из 184 изображений приведена в таблице I. Полученные значения метрик оценки качества показывают, что классификатор смог найти зависимости в данных.

V. ЗАКЛЮЧЕНИЕ

Полученные результаты показывают, что алгоритмы машинного обучения позволяют с точностью порядка 70% по одной фотографии лица определить, является ли профессией человека выступление в составе симфонического оркестра. Несмотря на наличие небольших смещений в обучающей выборке классификатор на основе машины опорных векторов оказался способен найти различия между фотографиями музыкантов симфонического оркестра и фотографиями немусыкантов. Данное исследование показывает, что на изображениях лиц существуют незаметные для человеческого взгляда особенности, которые с легкостью могут быть найдены алгоритмами машинного обучения.

Полученные результаты свидетельствуют о том, что технологии распознавания лиц представляют реальную угрозу конфиденциальности, поскольку в некоторых случаях профессия человека может быть определена по фотографии его или ее лица без предварительного на то согласия. Данное исследование может иметь практическое применение в сфере маркетинга для формирования персональных рекомендаций (рекламы) товаров и услуг, например, в магазинах, торговых центрах и т.д., за счет обработки изображений лиц посетителей, получаемых с существующих камер видеонаблюдения.

Таблица I. МАТРИЦА ОШИБОК ДЛЯ ТЕСТОВОЙ ВЫБОРКИ

True class	Predicted class	
	Not musician	Musician
Not musician	53	38
Musician	17	76



Рис. 1. Визуализация эмбедингов: метод главных компонент

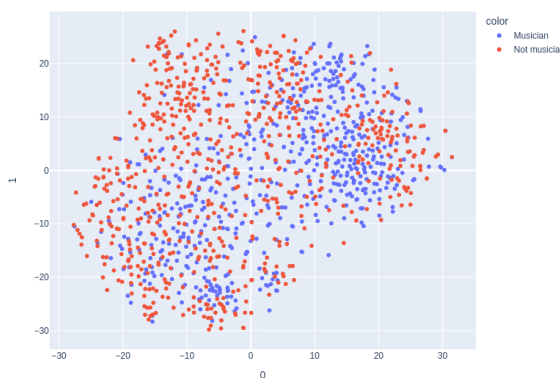


Рис. 2. Визуализация эмбедингов изображений: метод t-SNE

ЛИТЕРАТУРА

- [1] Blom, P.M. Towards personalised gaming via facial expression recognition / P.M. Blom, S. Bakkes, C. Tien Tan, S. Whiteson, D. Roijers, R. Valenti, T. Gevers // Proceedings of the tenth annual AAAI conference on artificial intelligence and interactive digital entertainment (AIIDE 2014). – 2014. – P. 30-36.
- [2] An, H. Mental health detection from speech signal: A convolution neural networks approach / H. An, X. Lu, D. Shi, J. Yuan, R. Li, T. Pan // 2019 International Joint Conference on Information, Media and Engineering (IJCIME). – 2019. – P. 436-439.
- [3] Kucharska-Pietura, K. Perception of facial and vocal affect by people with schizophrenia in early and late stages of illness / K. Kucharska-Pietura, A.S. David, M. Masiak, M.L. Phillips // The British Journal of Psychiatry. – 2005. – Vol. 187(6). – P. 523-528.
- [4] Harmer, C.J. Enhanced recognition of disgust in bipolar illness / C.J. Harmer, L. Grayson, G.M. Goodwin // Biological psychiatry. – 2002. – Vol. 51(4). – P. 298-304.
- [5] Kosinski, M. Facial recognition technology can expose political orientation from naturalistic facial images // Scientific reports. – 2021. – Vol. 11(1). – P. 100.
- [6] Wang, Y. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images / Y. Wang, M. Kosinski // Journal of Personality and Social Psychology. – 2018. – Vol. 114(2). – P. 246-257.
- [7] Zulkashev, R. Automatic analysis of face images for college degree verification / R. Zulkashev, M. Polyak // IEEE proceedings of ITNT 2023 : The IXth International Conference on information technology and nanotechnology. – 2023. – P. 10139216.
- [8] Bendel, O. The uncanny return of physiognomy // 2018 AAAI Spring Symposium Series. – 2018.
- [9] Parkhi, O. Deep face recognition / O. Parkhi, A. Vedaldi, A. Zisserman // BMVC 2015-Proceedings of the British Machine Vision Conference 2015. – 2015.
- [10] Chapelle, O. Support vector machines for histogram-based image classification / O. Chapelle, P. Haffner, V.N. Vapnik // IEEE transactions on Neural Networks. – 1999. – Vol. 10(5). – P. 1055-1064.