

ИССЛЕДОВАНИЕ ОБЛАСТИ ПРИМЕНЕНИЯ И МЕТОДОВ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ЧЕЛОВЕЧЕСКИХ ЭМОЦИЙ

Аннотация. В статье представлен обзор области применения и методов автоматического распознавания человеческих эмоций. Приведен краткий список речевых баз данных и их характеристик. Намечен план для дальнейшей работы на основе этого исследования.

Ключевые слова: распознавание эмоций, методы ИАД, вейвлеты, распознавание образов.

Введение

Предметом данного исследования являются область применения и методы автоматического распознавание эмоций человека. В первую очередь стоит определиться с самим понятием слова эмоция.

Существует несколько различных определений понятия «эмоция», в том числе нижеследующие: 1. По толковому словарю Д.Н. Ушакова, эмоция описывает душевное переживание, чувство, часто сопровождаемое какими-либо инстинктивными выразительными движениями. 2. По Медицинской энциклопедии, эмоция – это реакция, проявляющаяся в виде субъективно окрашенного переживания индивида, и отражающая при этом значимость для него того или иного воздействующего раздражителя, а также результата собственного действия (т.е. удовлетворение или неудовлетворение).

Первое определение – более общее. Оно позволяет понять, что такое эмоция с точки зрения человека, как объекта проявления и восприятия эмоции. Однако дополнение в скобках уже уточняет, что распознать эмоцию можно и машинным способом. Второе определение – ближе к тому, как с задачей распознавания эмоций работает компьютер. Он считывает реакцию с получаемых им данных – будь то изображение человека, запись его голоса или текста речи – и соотносит со шкалой удовлетворенности (положительный спектр эмоций) или неудовлетворенности (отрицательный спектр эмоций).

До недавнего времени способностью распознавать эмоции обладал только человек, но в настоящее время компьютеры, в том числе программы искусственного интеллекта, активно учатся распознавать человеческие эмоции. Зачем это может быть необходимо?

Области применения распознавания эмоций

Медицина. Анализ эмоционального состояния пациента часто очень важен для врача, потому что не всегда человек может сам правильно выразить свои чувства, а иногда не умеет их выражать в силу возраста или течения заболевания. В этом случае круглосуточный мониторинг именно за эмоциональным фоном может дать ощутимый результат. Он позволит вовремя отследить изменение настроения пациента и привлечь к нему внимание врача или дежурной медсестры.

В этом же ключе, программа по распознаванию эмоций может стать незаменимым помощником и в личном профилактическом мониторинге, указывая использующему ее человеку, в какие моменты он испытывал те или иные эмоции, позволяя самостоятельно контролировать как сами эмоции, так и их внешнее проявление. Что является актуальным в современном мире, где у детей и подростков встречаются проблемы с

проявлением и контролем своего эмоционального состояния, особенно в период так называемого «переходного возраста».

Также некоторые психические заболевания (такие как болезнь Альцгеймера или Паркинсона) можно выявить на ранней стадии именно по изменению эмоциональной реакции на окружение. Человек не всегда может заметить эти изменения, а также может списать их на проявление усталости, нервного напряжения, перепадов настроения. В этом случае компьютер сможет вовремя подсказать один из возможных диагнозов, потому что он будет просто анализировать данные, а не искать причину их возникновения.

Криминалистика. Считывание эмоционального состояния отдельных людей и целых групп поможет выявлять потенциальную опасность в общественных местах и при большом скоплении людей. Программы по распознаванию лиц уже проходят успешную проверку в реальных условиях, поэтому вполне возможно, что и использование более анонимного в этом случае выявления эмоций, сможет работать для анализа больших групп людей.

Возможность эмоционального мониторинга социальных сетей, позволяющая выявлять всплески негативных эмоций в тех или иных обсуждениях и предупреждать о возможных сетевых преступлениях. Такая программа будет реагировать на частоту использования эмоционально окрашенных слов, а также на использование популярных в последнее время «смайлов», «эмоджи» и прочих сетевых визуальных проявлений эмоционального состояния пользователей. Так же можно научить компьютер выявлять вспышки гнева в сообщениях, возможные угрозы и призывы к правонарушениям или мошенничеству.

Реклама. Эмоциональный маркетинг должен уметь подстраиваться под эмоциональную реакцию человека. В идеале программа будет показывать только те рекламные ролики, которые будут соответствовать

предпочтениям конкретного человека, фильтруя те, которые вызывают у него негативный эмоциональный отзыв и вычеркивая их из возможного списка предложений. Возможен также вариант, что, если человек может быть заинтересован тем или иным товаром, компьютер будет выбирать для него тот вариант рекламы, который будет интересен именно для него. Это позволит более точно определять покупательские ниши, особенно в сочетании с ключевыми запросами самих пользователей.

Эмоциональный маркетинг может помочь нейромаркетингу, облегчая проведение исследований после запуска рекламы. Считывание эмоций с аудио или видеозаписи пользователей, позволит сэкономить время и ресурсы на опросах, охватит аудиторию, которая не участвует в опросах, а также более точно считывает эмоции в случае, когда сами пользователи не могут их выразить – например, в случае рекламы детских товаров.

Игровая индустрия. Здесь для распознавания эмоций довольно широкая возможность применения. Например, подстройка уровня сложности игры в зависимости от эмоционального состояния игрока с возможностью настройки – чтобы, когда пугаются или нервничают новички, игра упрощалась, а если возрастает эмоциональная напряженность опытного игрока, на его пути начинают появляться все более сложные препятствия, которые исчезают, когда он успокаивается и/или устает от игры. Еще один вариант – эмоциональная анимация персонажей в зависимости от состояния игроков, что может найти отклик для разнообразных онлайн игр, которые основаны на взаимодействии нескольких пользователей.

Не менее важным аспектом игровой индустрии будет и анализ медицинских показателей для предотвращения возможных случаев неустойчивой психо - эмоциональной составляющей игрока. Когда он будет переставать нормально реагировать на происходящее в игре,

программа будет сама выключать игру, а в случае необходимости посылать вызов о медицинской помощи.

Робототехника. С возможностью распознавать человеческие эмоции, а в дальнейшем и моделировать необходимое настроение, роботы еще больше приблизятся к тем андроидам, которые пока заполняют рассказы фантастов и предстают перед зрителями на экранах кинотеатров. Сейчас, когда машины умеют делать многие виды работ быстрее и лучше людей, а кибернетика позволяет создавать материалы почти неотличимые от биологических тканей, робототехники решили, что пришло время дополнить свои разработки в части искусственного интеллекта эмоциональной составляющей, чтобы еще больше приблизить роботов к человеку. Способность машины только понимать эмоции, позволит намного упростить общение человек-машина, а когда роботы научатся также и проявлять нужные в конкретных ситуациях эмоции, будет сделан существенный шаг вперед к тому будущему, которое нам показывают фантасты.

Образование. Искусственный интеллект уже начинает внедряться в систему мониторинга в образовательных организациях. Распознавание эмоций помогает, во-первых, выявить асоциальное поведение учащихся, распознать отклонение в поведении детей и предупредить возможные конфликтные ситуации. Во-вторых, оно может способствовать анализу образовательного процесса, выявлять успеваемость и возможные трудности в усвоении учебного материала. В-третьих, мониторинг эмоционального состояния аудитории во время занятия может подсказать как преподавателю, так и самим учащимся, момент, когда необходимо успокоиться или переключиться на другой вид деятельности, чтобы процесс обучения проходил более эффективно.

HR. В некоторых компаниях уже начали применять анализ эмоционального состояния сотрудников с целью перераспределения задач

в случае выявления усталости или перегруженности отделов, или работников. Также при выявлении недовольства отдельных работников поставленной задачей, можно скорректировать их мотивацию тем или иным способом, чтобы увеличить производительность.

Также часть компаний стали использовать эмоциональный анализ кандидатов при собеседовании при устройстве на работу. Во-первых, для определения вопроса, при ответе на которые претендент мог соврать, что очень важно при приеме сотрудников в такие компании, как охранные агентства, банки или страховые компании, что позволяет снизить риск мошенничества. Во-вторых, программы могут анализировать соответствие людей той или иной должности в случае приема кандидата на должность, связанную с общением с людьми или повышенной ответственностью. Система анализирует такие параметры, как стрессоустойчивость, умение контролировать свои эмоции, отсутствие или наличие в речи некоторых ключевых слов.

Законодательный аспект распознавания эмоций

Стоит сказать несколько слов о законодательном аспекте распознавания эмоций, потому что этот вопрос часто возникает при упоминании предмета этого исследования.

Пожалуй, единственной проблемой при использовании эмоциональных технологий – это приватность и защита персональных данных. Поскольку эмоции довольно часто проявляются на уровне, не контролируемом сознанием, и являются откликом на окружающие проявления, то эмоциональные данные считаются довольно личными.

Их анализ подпадает, в частности, под действие европейского регламента по защите персональных данных (GDPR), действующего с 2018 года. Текст регламента можно найти в свободном доступе на английском [1] и русском [2] языках. В частности, данные, которые можно

использовать для алгоритмов машинного обучения должны быть: деперсонализованны, т.е. отделены от идентификации людей (биометрических данных); группового формата, т.е. можно анализировать несколько людей, но не единичные личности; при индивидуальном анализе при ведении аудио или видео съемке, человек должен знать об этом и быть согласен, иначе это может считать правонарушением и повлечь за собой ответственность.

Методы

Над проблемой распознавания эмоций работают ученые разных областей знаний – психологии, биологии, информатики.

Информатики в своих работах по распознаванию эмоций используют набор дискретных эмоциональных состояний (гнев, раздражение, страх, радость, печаль, удивление и нейтральное состояние) [3], потому что при классификации и кластеризации такой подход обеспечивает большую точность результата, чем в случаях, когда используется многовариативная классификация.

Определять эмоции можно по нескольким видам сигналов, передаваемых человеком. Чаще всего для этого используют аудио (речь) и видео (изображение лица) данные. Но также для распознавания эмоций может использоваться сам текст речи, жесты, поза, походка.

Для решения задач классификации и кластеризации при распознавании эмоций используют несколько разных математических подходов и методов. В целом можно их разделить на два раздела – методы интеллектуального анализа данных (ИАД), которые включают в себя различные методы классификации на основе искусственных нейронных сетей, и методы математического анализа данных, в частности на основе проведения вейвлет-разложения входного сигнала с последующим анализом.

Больше всего исследований проведено в области использования методов ИАД. Создано несколько приложений для компьютеров и смартфонов, позволяющих оценивать эмоции с той или иной долей вероятности. Существует несколько обзоров некоторых из существующих на данный момент программ (например, обзор в магистерской диссертации Кузнецовой А.С. [4]).

К методам ИАД относятся, например, метод главных компонент, метод k-средних и искусственные нейронные сети – в том числе сверточные нейронные сети и нейронные сети с методом глубокого обучения. Сравнение эффективности этих методов для задач распознавание эмоций можно прочитать, например, в статье [5]. Там же указано, что наиболее эффективным из перечисленных методов ИАД является сверточная нейросеть, хотя процент погрешности определения эмоций не сильно изменяется при смене метода в случае ИАД. Минусом этих методов является отсутствие точного математического описания процесса распознавания эмоций и возможность переобучения нейросетей. Но эти методы, как уже говорилось ранее, сейчас пробуют применяться на практике, потому что требуют минимальных исследовательских затрат после создания запрограммированной нейронной сети.

Другим способом обработки сигналов является вейвлет-преобразование (ВП). Его преимуществом является использование специальных математических функций – вейвлетов – для проведения эмоционального анализа. В целом, ВП переводит цифровой сигнал из частотной области в частотно-временную, позволяя выявить изменение эмоционального фона по времени.

На данный момент существует несколько исследований, в которых авторы пытаются установить связь картин вейвлет-разложений как аудио-, так и видеосигналов с эмоциями дикторов. В частности, это работа

Брехидина А.И. [6], который создал свою выборку данных и провел ее анализ.

Базы данных. Существует несколько официальных баз данных для эмоциональных исследований по аудио- или видеосигналам. Достаточно большой их список представлен в статье о Женевском сете акустических параметров [7]. Среди них рассмотрены следующие базы данных:

1. FAU AIBO [8]. Эта база данных была основой для первого международного исследования эмоциональных состояний. Она включает в себя записи общения детей с домашним роботом-питомцем Sony Aibo. Записи проводились в двух разных школах, и в них участвовало в общей сложности 51 ребенок 10-13 лет. Записи были поделены на эмоциональные высказывания, выражающие 5 различных эмоций – гнев, возбуждение, нейтральное состояние, радость и расслабленность. Эти эмоции для исследования обычно делят на две группы – негативные (гнев, возбуждение) и позитивные (все остальные).

2. TUM-AVIC [9]. База данных содержит англоязычные эмоциональные записи голоса 21 человека. Вся база разбита на проявление 3 видов эмоций по отношению к продолжению диалога с собеседником – скука (субъект не заинтересован в продолжении разговора), нейтральное состояние (когда невозможно определить желание о продолжении разговора) и заинтересованность в продолжении диалога. Эмоции заложены в наборы ключевых фраз говорящих. Дикторы для данной базы говорили на английском, хотя для большинства из них родным языком был немецкий.

3. EMO-DB [10]. Эта Берлинская эмоциональная база данных является наиболее популярной среди исследователей. Она содержит высказывания 10 дикторов, каждый из которых произносит по-немецки по 10 эмоционально-нейтральных предложений (например, «Der Lappen liegt auf dem Eisschrank» - «Ткань лежит на холодильнике») с 7

эмоциональными состояниями: гнев, скука, отвращение, страх, радость, грусть и нейтральное состояние.

Берлинская база данных наиболее изучена в эмоциональных исследованиях и около 30% ее отмечено как недостаточно естественное звучание или плохое определение эмоциональной составляющей. Это может быть связано с тем, что запись производилась профессиональными дикторами, что вызывало некоторую неестественность записываемых эмоций. Несмотря на это, данная база остается общедоступной и продолжает использоваться исследователями различных стран мира.

Результаты

Был проведен анализ литературы по областям практического применения распознавания эмоций, по методам, используемым для этого, по базам данных, представляемым для проведения эмоционального анализа речи и изображения.

Следующим шагом исследования было решено проведение анализа эмоционального состояния дикторов из Аудиовизуальной базы данных эмоционально записанных речи и песен Райерсона (RAVDESS) [11]. Эта база данных содержит записи 24-х профессиональных актеров, озвучивающих два высказывания с нейтральным североамериканским акцентом. Запись речи включает в себя спокойные, счастливые, грустные, злые, испуганные, удивленные и недовольные эмоции. В базе данных содержатся и аудио-, и видеоданные.

В качестве первого метода анализа было решено использовать вейвлет-анализ аудио-сигнала. Во-первых, полученные данные можно будет сравнить с некоторыми ранее проведенными исследованиями. Во-вторых, применение вейвлет-анализа для задачи распознавания эмоций встречается реже, чем использование искусственных нейронных сетей,

поэтому при получении определенных результатов возможно проведение сравнения использования этих двух методов.

СПИСОК ЛИТЕРАТУРЫ

1. Текст Регламента по защите персональных данных (GDPR). URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679> (Дата обращения: 16.02.2020).
2. Перевод текста регламента по защите персональных данных (GDPR) на русский язык. URL: <https://gdpr-text.com/ru/> (Дата обращения: 16.02.2020).
3. Давыдов А.Г., Киселев В.В., Кочетков Д.С. Классификация эмоционального состояния диктора по голосу: проблемы и решения // Труды международной конференции «Диалог 2011». М.: РГТУ. 2011. С. 178–185.
4. Кузнецова А.С. Нейросетевые алгоритмы распознавания эмоционального состояния человека по изображениям его лица: дипл. работа (диссертация магистра информатики и вычислительной техники). Томский политехнический университет. Томск. 2019. 108 с.
5. Полякова А.С., Сидоров М.Ю., Семенкин Е.С. Комбинирование подходов кластеризации и классификации для задачи распознавания эмоций по речи // Вестник СибГАУ. Том 17. №2. 2016. С. 335-342.
6. Бредихин А.И. РАСПОЗНАВАНИЕ ЭМОЦИЙ В ЧЕЛОВЕЧЕСКОЙ РЕЧИ С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТОВ // Научное сообщество студентов XXI столетия. ТЕХНИЧЕСКИЕ НАУКИ: сб. ст. по мат. LXII междунар. студ. науч.-практ. конф. № 2(61). URL: [https://sibac.info/archive/technic/2\(61\).pdf](https://sibac.info/archive/technic/2(61).pdf) (дата обращения: 17.02.2020)
7. F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. Andre, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan et al., “The geneva

- minimalistic acoustic parameter set (gemaps) for voice research and affective computing,” *IEEE Transactions on Affective Computing*. vol. 7. no. 2. 2016. pp. 190–202.
8. S. Steidl, *Automatic Classification of Emotion-Related User States in Spontaneous Children’s Speech*. Berlin, Germany: Logos Verlag. 2009.
 9. B. Schuller, R. Muller, F. Eyben, J. Gast, B. Hornler, M. Wollmer, G. Rigoll, A. Hothker, and H. Konosu, “Being bored? Recognizing natural interest by extensive audiovisual integration for Real-life application,” *Image Vis. Comput., Special Issue Vis. Multimodal Anal. Human Spontaneous Behavior*. vol. 27. no. 12. Nov. 2009. pp. 1760–1774.
 10. Burkhardt F, Paeschke A, Rolfes M, Sendlmeier WF, Weiss B. A database of German emotional speech. In: *European conference on speech and language processing (EUROSPEECH)*. Lisbon, Portugal. 2005. Pp. 1517–1520.
 11. Livingstone SR, Russo FA. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE* 13(5): e0196391. 2018. <https://doi.org/10.1371/journal.pone.0196391> (Дата обращения 16.02.2020).