УДК 004.4/.5+159.99
doi:10.31799/1684-8853-2022-5-12-22
EDN: CXBRCS

Научные обзоры
Surveys

Аналитический обзор методов автоматического распознавания вовлеченности пользователя в виртуальную коммуникацию

А. А. Двойникова[a], *младший научный сотрудник, orcid.org/0000-0001-8047-6639*
И. А. Кагиров[a], *научный сотрудник, orcid.org/0000-0003-1196-1117*
А. А. Карпов[a], *доктор техн. наук, профессор, orcid.org/0000-0003-3424-652X, karpov@iias.spb.su*
[a]*Санкт-Петербургский Федеральный исследовательский центр РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ*

Введение: *решение автоматическими средствами задачи распознавания и оценивания степени вовлеченности пользователя в процесс человеко-машинного взаимодействия или телекоммуникации является актуальным в области компьютерного распознавания состояний человека. Это необходимо для проектирования приложений дистанционного обучения, бизнеса и развлечений.* ***Цель:*** *провести сравнительный анализ существующего информационного обеспечения и методов в области автоматического распознавания и оценивания вовлеченности пользователя в процесс человеко-машинного взаимодействия или виртуальной коммуникации и обосновать методологию построения корпуса данных, основанного на идее многомодальной коммуникации.* ***Результаты:*** *выполненный анализ работ показал, что в большей части известных корпусов (data sets) отсутствуют данные по онлайн-коммуникации в естественных условиях и не всегда учитываются разные модальности взаимодействия в системе «человек — машина — человек». Для многоуровневой классификации, определяющей степень вовлеченности, важны, помимо видеомодальности, также акустическая и текстовая информация. Перспективным оказывается учет «языка тела» пользователя: мимика, движения тела, рук и головы. Кроме того, для правильной оценки вовлеченности в корпусе данных должна содержаться метаинформация по психоэмоциональным состояниям коммуникантов. Наиболее перспективным подходом для автоматического оценивания степени вовлеченности следует признать нейросетевой.* ***Практическая значимость:*** *на основании полученных аналитических выводов планируется разработка оригинальной программной системы для автоматического распознавания вовлеченности и формирование набора данных для обучения ее вероятностных моделей. Также представленный обзор позволяет сформулировать основные требования к подобным системам и является вкладом в решение задачи автоматического распознавания психоэмоциональных состояний человека.* ***Обсуждение:*** *анализ работ позволяет сделать вывод, что понятие вовлеченности в контексте распознавания эмоций отличается от распространенного в психологии. Вовлеченность пользователя (коммуниканта) в информационно-коммуникационной сфере — проявление различной степени эмоциональной, когнитивной и поведенческой составляющих психической активности человека в процессе взаимодействия с собеседником или компьютерной системой, имеющее динамический характер.*

Ключевые слова *— вовлеченность пользователя, информационное обеспечение, автоматические системы распознавания вовлеченности, многомодальность, искусственные нейронные сети.*

Для цитирования: Двойникова А. А., Кагиров И. А., Карпов А. А. Аналитический обзор методов автоматического распознавания вовлеченности пользователя в виртуальную коммуникацию. *Информационно-управляющие системы*, 2022, № 5, с. 12–22. doi:10.31799/1684-8853-2022-5-12-22, EDN: CXBRCS
For citation: Dvoynikova A. A., Kagirov I. A., Karpov A. A. Analytical review of methods for automatic detection of user engagement in virtual communication. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 5, pp. 12–22 (In Russian). doi:10.31799/1684-8853-2022-5-12-22, EDN: CXBRCS

Введение

В современном мире средства онлайн-коммуникации уже давно и прочно являются одним из основных способов межличностного взаимодействия. В период пандемии COVID (с 2020 г.) актуальность дистанционных технологий только возросла. Не в последнюю очередь это касается учебных процессов, организации рабочих совещаний и телеконференций [1]. Другой существенной составляющей онлайн-коммуникации является повседневное общение и развлечение (прежде всего, массовые многопользовательские игры, зачастую требующие от участников обсуждения командных действий).

Перенос многих форм социального взаимодействия в дистанционный формат порождает ряд новых проблем, одной из которых является низкая вовлеченность (Engagement) одного или нескольких участников в коммуникативный процесс (в широком смысле этого слова). Отсутствие живого контакта между участниками онлайн-мероприятий (эффект присутствия), во-первых, может способствовать рассеянности внимания собеседников и, во-вторых, лишает коммуникантов возможности прямо оценивать уровень заинтересованности собеседника.

Эта проблема актуальна, прежде всего, для онлайн-обучения. Если вовлеченность учащихся

в учебный процесс и раньше оказывалась в центре внимания научного сообщества, то сегодня разработку методов и средств объективного измерения степени вовлеченности в образовательный процесс можно признать необходимостью [2, 3].

Вовлеченности как синониму осознанной активной внимательности посвящен целый ряд психологических исследований [4, 5], результаты которых находят практическое применение во многих областях человеческой деятельности. Например, людям, работающим в сфере управления, иногда критически важно удерживать свое внимание на рабочих процессах (мониторинг, визуальный контроль и принятие решений на основе видеоданных). В чрезвычайных ситуациях оператору необходимо быстро принять правильное решение, а этого можно достичь только при вовлеченности (в другой терминологии — активном включении) оператора в рабочий процесс. Поэтому психологи совместно с техническими специалистами разрабатывают различные тренажеры для тренировки внимания [4, 5]. Системы управления и предоставления информации должны не просто передавать визуальные данные, но и оценивать, в том числе, эмоциональное состояние людей, облегчая процесс принятия решений [6].

Главной целью настоящей статьи является сравнительный аналитический обзор основных методов распознавания вовлеченности участников виртуальной коммуникации, в последнее время появившихся в контексте исследований по машинному обучению. Под виртуальной коммуникацией здесь подразумевается как общение между собеседниками посредством телеконференций, так и взаимодействие пользователя с системой (просмотр видео, онлайн-игры и пр.). Другой важной задачей оказывается выполнение аналитического обзора существующего информационного обеспечения, созданного для распознавания вовлеченности участников.

Термин «вовлеченность» был введен в научный оборот в работе [7], и, начиная с конца XX в., он активно используется в научном сообществе, в первую очередь среди специалистов по психологии, педагогике и социологии. Тем не менее во всех этих дисциплинах вовлеченность трактуется по-разному, и на сегодня не существует консенсуса относительно его научного содержания [8]. Можно констатировать, что понятие вовлеченности прочно вошло в систему научного знания, однако эксплицитное раскрытие этого термина является предметом научных дискуссий [9]. Вовлеченность является сложным психологическим и социальным феноменом; различные исследователи понимают этот термин с разных позиций [10]. Необходимо подчеркнуть при этом, что разнообразие мнений о вовлеченности напрямую проистекает из сложной природы этого феномена.

В работе [11] выделяется три основных компонента социальной установки человека: когнитивный (познавательный) [12], аффективный (эмоциональный) [13, 14] и поведенческий [15, 16]. В зависимости от того, в каком компоненте применяется термин «вовлеченность», его определение может несколько отличаться [17].

Также существуют работы, в которых проводятся исследования корреляции всех трех аспектов вовлеченности: познавательного, эмоционального и поведенческого. Так, в статье [18] указывается, что наиболее глубокая погруженность в работу отмечается возросшей когнитивной составляющей, что подразумевает у индивида процесс осмысления задачи. Исследование [19] демонстрирует взаимосвязь между когнитивной вовлеченностью и обучаемостью студентов. В работе [13] экспериментальным путем доказывается связь между эмоциональной и поведенческой вовлеченностью.

В контексте автоматического распознавания вовлеченности следует учитывать тот факт, что автоматическими методами можно анализировать только внешнее проявление, а не истинную вовлеченность человека. Вовлеченность в области автоматического анализа — это проявление различной степени эмоциональной, когнитивной и поведенческой составляющих человека в процессе взаимодействия с собеседником или компьютерной системой, имеющее динамический характер. По различным уровням интенсивности ряд авторов выделяет вовлеченность бинарную (вовлечен/не вовлечен) [20, 21], трехуровневую (не вовлечен, вовлечен формально, сильно вовлечен) [22], четырехуровневую (очень низкую, низкую, высокую и очень высокую) [23–25] и даже десятиуровневую [26]. Помимо категориального разделения степени вовлеченности, также активно применяется и непрерывная оценка вовлеченности [27, 28].

Обзор информационного обеспечения для распознавания вовлеченности

Для того чтобы разработать систему автоматического распознавания вовлеченности, необходимо построить модель с учетом реальных данных. Все существующие корпусы, содержащие разметку данных по вовлеченности, можно разделить на две группы по принципу взаимодействия участников. К первой группе относятся корпусы, при записи которых участники коммуницировали друг с другом, такие как NoXi [27], MEDICA [28], MHHRI [29], RECOLA [20], Emotion Miner Data Corpus (EMDC) и Sümer [2]. Другая группа пред-

ставляет собой корпусы, содержащие сценарии, в которых участники взаимодействуют с компьютерной системой, например просматривают обучающие видео или играют в онлайн-игры. К таким корпусам относятся Engagement Recognition dataset [21], EngageWild [23], DAiSEE [24], корпуса авторов Kamath [22], Whitehill [25], Psaltis [29]. В табл. 1 содержится сравнительный обзор корпусов, применяемых для анализа вовлеченности.

Как видно из таблицы, существует несколько представительных корпусов для автоматического распознавания вовлеченности. Стоит отметить, что большинство рассмотренных корпусов включают в себя только видеоданные. Это обусловлено тем, что сценарии для записи подразумевают только просмотр участниками видеофрагментов или участие в онлайн-играх. В таких корпусах представлена вовлеченность участников во взаимодействие с компьютерной системой. В многомодальных корпусах содержатся данные по вовлеченности участников в процесс коммуникации друг с другом. Проявление вовлеченности в диалоге с собеседником зависит не только от собственного интереса к теме разговора, но и от черт личности собеседника, его коммуникативных навыков, а также от проявления эмоционального состояния обоих коммуникантов.

Наибольшая вариативность информантов среди одномодальных корпусов представлена в корпусах Sümer [2] (128 чел.) и DAiSEE [24] (112 чел.). При этом корпус EMDC имеет наибольший объем видеоданных (140 ч) по сравнению со всеми рассмотренными корпусами. Корпус NoXi [27] содержит в себе данные большого количества участников (87 чел.), которые общались между собой на различные темы на семи языках. Исходя из этого можно сказать, что NoXi является наиболее репрезентативным корпусом для обучения многомодальных систем автоматического распознавания вовлеченности.

Еще одним немаловажным фактором при анализе данных для распознавания вовлеченности является учет эффекта Хоторна (эффект наблюдателя) [30] при записи участников эксперимента. Эффект Хоторна представляет собой реакцию — изменение поведения человека в ответ на осознание того, что за ним наблюдают со стороны. Во всех рассмотренных корпусах участники экспериментов знали о предстоящей записи, поэтому могли вести себя не вполне естественно.

■ ***Таблица 1.*** Сравнительная характеристика корпусов для автоматического анализа вовлеченности участников
■ ***Table 1.*** Comparative characteristics of corpora collected for automatic analysis of communicant engagement

Корпус	Модальность				Объем, ч	Разметка данных
	В	А	Т	Ф		
NoXi (https://noxi.aria-agent.eu/) [27]	+	+	+	–	25,3	Вовлеченность (непрерывная), эмоции, жесты, мимика
MEDICA [28]	+	+	+	–	1,1	Вовлеченность (непрерывная), внимательность, стресс, эмоции, нерешительность
MHHRI [26]	+	+	–	+	4,2	Вовлеченность (10 уровней), характеристики личности
RECOLA [20]	+	+	–	+	3,8	Вовлеченность (бинарная), согласие, доминирование, производительность, взаимопонимание
Emotion Miner Data Corpus (EMDC) (https://www.prweb.com/releases/2018/03/prweb15339730.htm)	+	+	–	–	140	Разговор (вовлеченность), эмоции, психические состояния и поведение, личность и ситуация
Корпус Sümer [2]	+	–	–	–	47	Вовлеченность (6 уровней)
Корпус Whitehill [25]	+	–	–	–	22,6	Вовлеченность (4 уровня)
DAiSEE (https://www.iith.ac.in/~daisee-dataset/) [24]	+	–	–	–	25,2	Вовлеченность (4 уровня), скука, замешательство, разочарование
Engagement Recognition dataset [21]	+	–	–	–	20	Вовлеченность (3 уровня), эмоции
Корпус Psaltis [29]	+	–	–	–	12	Вовлеченность (бинарная)
EngageWild (https://www.sites.google.com/view/emotiw2020) [23]	+	–	–	–	8,5	Вовлеченность (4 уровня)
Корпус Kamath (https://github.com/e-drishti/wacv2016) [22]	+	–	–	–	4408 кадров	Вовлеченность (3 уровня)

Примечание: В — видео; А — аудио; Т — текст; Ф — физиологические сигналы.

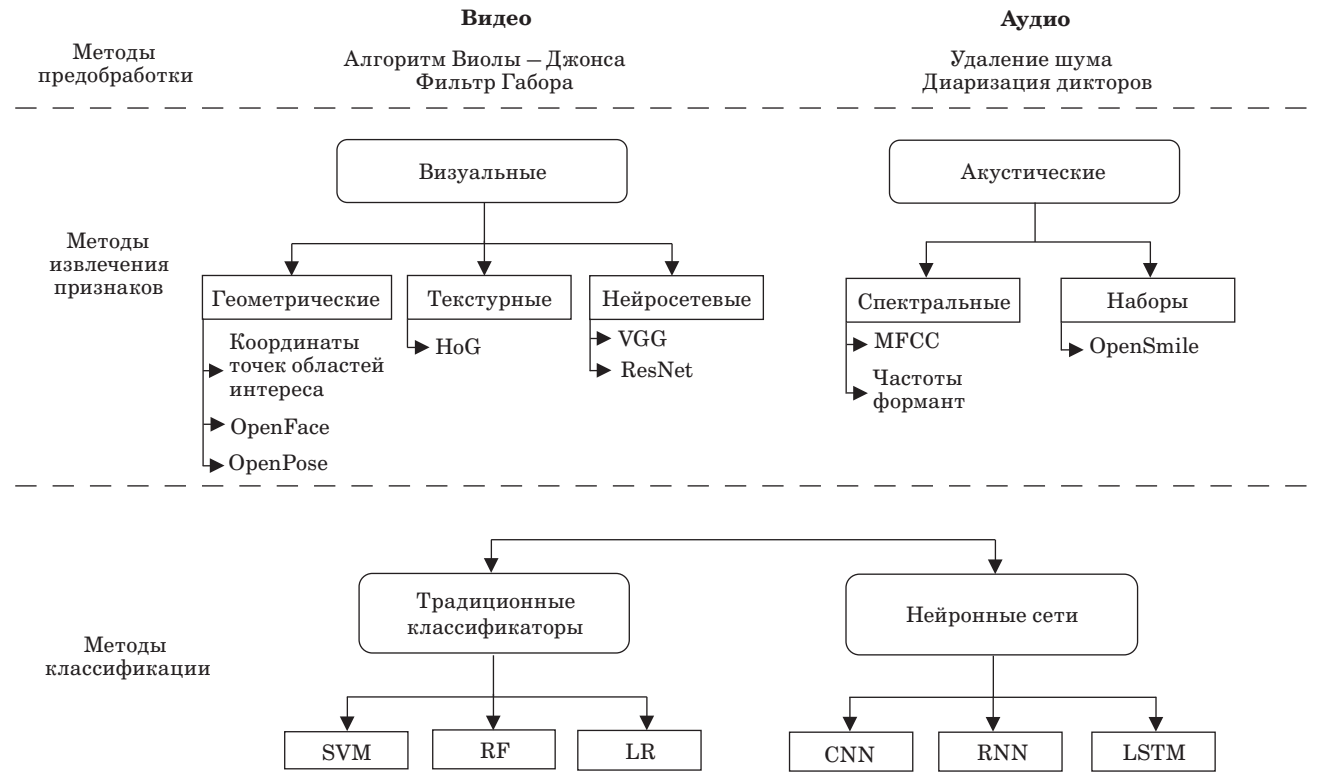
В корпусах DAiSEE и Kamath [22] использовалась веб-камера, которая позволяла приблизить условия эксперимента к естественным, потому что веб-камера незаметна и является довольно привычным предметом современной обстановки.

Обзор методов для распознавания вовлеченности

Большинство автоматических систем распознавания вовлеченности участников виртуальной коммуникации или взаимодействия с компьютерной системой основываются на анализе репрезентативных данных. Наличие текстовой информации в высказывании диктора может символизировать об однозначной вовлеченности диктора. Визуальные и акустические характеристики человека могут передавать как вовлеченность человека в виртуальную коммуникацию, так и ее отсутствие. Поэтому наиболее репрезентативными модальностями являются видео- и аудиоданные. При анализе видеоданных необходимо в первую очередь детектировать область лица, а также жесты рук, головы, позы тела и пр. Затем проводится извлечение визуальных и акустических признаков, которые в последующем подаются на вход классификатора, и на выходе даются вероятностные предсказания степени вовлеченности участников. Алгоритм распознавания вовлеченности с применяемыми методами предобработки данных, извлечения признаков и классификации представлен на рисунке.

Методы предобработки. Для выделения репрезентативной информации из видеоданных необходимо детектировать графические области интереса, которые представляют собой границы областей лица, тела, рук и пр. Алгоритм Виолы — Джонса наиболее часто используется для детектирования области лица в задаче распознавания вовлеченности [22]. К изображениям детектированных областей интереса также может применяться нормализация, например с помощью фильтра Габора, который преобразует изображения в оттенки серого [25]. Данные методы позволяют выделить релевантную информацию из изображений для анализа мимики лица и жестов.

Методы извлечения признаков. Все визуальные признаки, используемые для анализа вовлеченности, можно разделить на три группы: геометрические, текстурные и нейросетевые. Геометрические признаки используются для определения области положения губ [31], зрачков [31], позы тела [32], глаз [23, 33], головы [23], [32, 33]. Благодаря этому можно определить, откры-



■ Систематизация методов распознавания вовлеченности пользователей в виртуальную коммуникацию
■ Systematization of methods for recognition of user engagement into virtual communication

ты или закрыты глаза у собеседника, смотрит ли диктор в монитор или отвлекается, по положению губ можно сказать о том, произносит ли речь диктор, или анализировать эмоции (например, улыбка показывает радость). Также по жестам тела можно определить степень вовлеченности диктора, например, если рука подпирает голову, то это говорит о низкой заинтересованности. OpenFace и OpenPose — программные инструментарии, позволяющие находить координаты лица и частей тела соответственно. Помимо этого, OpenFace также выделяет единицы действия лица, например, закрыты глаза или открыты. OpenFace и OpenPose применялся в работах [2, 34–38] для распознавания вовлеченности собеседников.

Наиболее часто используемым методом извлечения текстурных признаков для распознавания вовлеченности [21, 22] является гистограмма ориентированных градиентов (Histogram of Oriented Gradients, HoG). HoG позволяет выделять контурную информацию объектов, удаляя при этом фоновый шум. Некоторые исследователи используют предобученные нейронные сети, такие как VGG [32, 38], ResNet (Residual Neural Network) [31], для извлечения визуальных признаков. Однако нейросетевые признаки не несут в себе физической информации об объектах, но зачастую позволяют достигать высокого результата классификации уровня вовлеченности.

Для распознавания вовлеченности в основном используют спектральные акустические признаки, такие как мел-частотные кепстральные коэффициенты (Mel-Frequency Cepstral Coefficients, MFCC), частоты формант и пр. [26, 28]. Данные характеристики отражают физическую составляющую речевого сигнала и зависят от особенностей работы артикуляционных органов каждого человека. Также одним из используемых наборов признаков для распознавания вовлеченности [31] является OpenSmile, который включает в себя 65 низкоуровневых дескрипторов.

Методы классификации. Существует множество методов классификации вовлеченности, их можно разделить на две группы: традиционные детерминированные методы классификации и искусственные нейронные сети (см. рисунок).

Эффективным традиционным методом распознавания вовлеченности является метод опорных векторов (Support Vector Machine, SVM) [22, 25, 26, 29]. Авторы работы [26] применяли SVM для бинарной классификации вовлеченности на основе анализа различных типов данных: видео, аудио и физиологических сигналов. Многие исследователи использовали SVM для классификации только видеоданных по кадрам [22, 25, 29]. В работе [22] автоматическая система распознавала три степени вовлеченности, используя SVM с обучением на нескольких ядрах. Другой метод традиционных классификаторов, который используется для распознавания вовлеченности, — случайный лес (Random Forest, RF) [2].

При наличии большого объема данных наиболее эффективными методами машинной классификации являются нейронные сети. Для анализа видеоданных можно выделить две основные группы нейронных сетей: прямого распространения, например сверточные (Convolutional Neural Network, CNN), и с обратными связями, например рекуррентные (Recurrent Neural Network, RNN). В работе [21] для бинарной классификации вовлеченности авторы используют архитектуру нейросети VGG-B, предварительно обученную на эмоциональном корпусе данных. Авторы статьи [35] для распознавания вовлеченности предлагают подход интегрального распознавания (End-to-End), основанный на использовании архитектур нейронных сетей ResNet и Temporal Convolutional Network (TCN).

Рекуррентные нейронные сети также являются эффективными в задачах распознавания вовлеченности участников виртуальной коммуникации. В работе [24] использовалась архитектура нейросети InceptionNet V3 [39] для покадрового анализа видеоданных и нейронная сеть Long-Term Recurrent Convolutional Network (LRCN) для классификации всего видео целиком. Авторы статьи [36] применяют нейронную сеть с управляемыми рекуррентными нейронами (Gated Recurrent Unit, GRU). Также для анализа вовлеченности используются нейронные сети с долгой краткосрочной памятью (Long Short-Term Memory, LSTM) [23, 34]. В статье [37] используются несколько LSTM-сетей: для анализа диктора, собеседника и обоих коммуникантов.

Для повышения уровня точности распознавания вовлеченности применяется объединение модальностей. В работе [32] анализируются видео- и аудиоданные с помощью байесовской сети (Bayesian Network, BN), а также RNN, а в работе [31] — с помощью логистической регрессии (Logistic Regression, LR).

Анализ исследований по распознаванию вовлеченности

В настоящее время существует ряд коммерческих систем для анализа вовлеченности персонала компаний в их трудовую деятельность. Наиболее известными являются системы TalentTech (https://talenttech.ru/engagement/), Happy Job (https://happy-job.ru/), Gallup (https://www.gallup.com/workplace/229424/employee-

engagement.aspx). Однако данные системы анализируют вовлеченность сотрудников компании при помощи методологии, основанной на опросниках [40–42]. Коммерческих систем, основанных на анализе аудиовизуальной информации проявления вовлеченности, на данный момент не существует. Именно поэтому далее в настоящей статье рассматриваются некоммерческие автоматические системы, работающие с аудио и визуальными данными.

Характеристики систем для автоматического распознавания вовлеченности пользователя приведены в табл. 2. Для оценки эффективности применяются коэффициент корреляции Пирсона (Pearson's Correlation Coefficient, PCC), среднеквадратическая ошибка (Mean Squared Error, MSE), а также точность (Precision) и F-мера (F-score).

Как показывает обзор существующих работ, в настоящее время чаще всего система проводит анализ видеоданных, а не аудио и текста. Это обусловлено тем, что видеоданные являются более репрезентативными для анализа вовлеченности и в ситуациях, когда взаимодействие человека происходит не с собеседником, а, например, с информационной системой. В таких случаях заинтересованность проявляется только в мимике и жестах. Проявление речевой активности, как правило, говорит о том, что человек вовлечен, поэтому при бинарной классификации вовлеченности нет смысла анализировать акустические и смысловые признаки. Однако при многоуровневой классификации, показывающей интенсивность вовлеченности, анализ аудио и текстовой модальностей может помочь повысить точность распознавания.

Исследований, направленных на анализ физиологических сигналов человека для решения данной задачи, довольно мало. Это связано с тем, что сбор таких данных является достаточно трудоемким процессом, к тому же носимые участниками эксперимента датчики могут мешать естественному поведению человека. Также анализ аудио и текстовой модальностей в области распознавания вовлеченности является недостаточно изученным. Актуальность данной задачи проявляется в ситуациях, когда анализ видеоданных становится невозможным, например, при неработающей камере при виртуальной коммуникации.

Использование нейронных сетей для распознавания вовлеченности показывает более высокую точность по сравнению с применением традиционных методов классификации (см. табл. 2). Одной из причин эффективности нейронных сетей является возможность использовать предварительно обученные модели на других корпусах с большим объемом данных с переносом обучения (Transfer Learning). Некоторые существующие экспериментальные исследования показывают

■ ***Таблица 2.*** Сравнение исследований по автоматическому распознаванию вовлеченности
■ ***Table 2.*** Comparison of studies on automatic engagement recognition

Метод (работа)	Корпус	Модальность	Методы классификации	Показатель оценки	Значение показателя
[38]	DAiSEE	В	RNN	Precision	0,39
[24]			InceptionNet		0,47
			LRCN		0,58
[35]			ResNet, TCN		0,64
[36]	EngageWild		GRU	MSE	0,07
[33]			TCN		0,08
[34]			LSTM		0,06
[23]				PCC	0,25
[37]	NoXi			Средняя F-score	0,96
[32]		А+В	RNN	PCC	0,60
			BN		0,74
[26]	MHHRI	А	SVM	F-score	0,65
		В			0,58
		Ф			0,54
		А+В			0,56
		А+Ф			0,59
		В+Ф			0,52

эффективность использования предобученных моделей на эмоциональных корпусах для анализа вовлеченности. Это связано со значительной корреляцией между эмоциональным состоянием человека и вовлеченностью как в коммуникациях, так в образовательных и игровых [21, 43–45] взаимодействиях между людьми.

Исходя из экспериментальных исследований [24, 35, 38], можно сделать вывод, что с помощью архитектуры ResNet получается более эффективное распознавание вовлеченности, чем с RNN. На визуальном корпусе DAiSEE авторы работы [35] достигают точности 64 %, используя нейронную сеть ResNet. LSTM [34] позволяет уменьшить MSE относительно GRU [36] и TCN [33] до 0,06 на корпусе EngageWild.

Экспериментальные исследования показывают, что для повышения точности распознавания вовлеченности, помимо анализа мимики, эффективно анализировать поворот головы, жесты и направление взгляда [29, 32]. Также, как правило, объединение нескольких модальностей: видео, аудио, текста, физиологических сигналов — позволяет улучшить точность распознавания вовлеченности. Однако в работе [26] эксперименты показали противоположный результат: одномодальные подходы превзошли многомодальные. Авторы предполагают, что это связано со способом объединения модальностей на уровне принятия решения. Вероятно, объединение на уровне признаков позволит повысить точность одномодального распознавания.

Заключение

Как показывает обзор существующих корпусов, для анализа вовлеченности практически все данные собираются в лабораторных (контролируемых) условиях. В том случае, когда авторы корпусов рассматривают сценарии коммуникации между людьми, общение происходит только между двумя участниками. Ни один из описанных выше корпусов не включает в себя коммуникацию между людьми в группе. К тому же на сегодня все больше коммуникативных актов происходит в онлайн-режиме. Также стоит отметить, что большинство многомодальных корпусов англоязычные, меньшая часть — на французском и немецком языках.

Аналитический обзор существующих методов анализа вовлеченности позволяет сформулировать следующие основные требования к разрабатываемой нами программной системе автоматического распознавания вовлеченности:

1. Многомодальный анализ данных вербальных и невербальных сигналов проявлений вовлеченности коммуникантов.

2. Распознавание вовлеченности с высокой точностью (не менее 70 %).

3. Учет эмоционального состояния коммуникантов.

4. Поддержка офлайн- и онлайн-режимов.

5. Возможность интеграции в существующие системы телеконференций.

6. Использование нейросетевых подходов.

Основываясь на этих выводах, авторы статьи планируют собрать собственный корпус данных, в котором должны быть удовлетворены следующие требования:

1. Многомодальность: корпус должен включать в себя видео, аудио и текстовые данные.

2. Включение сценариев коммуникации группы от двух и более людей.

3. Запись данных в естественных условиях с использованием современных средств телекоммуникации (например, Zoom, Signal, Яндекс, Телемост и т. п.).

4. Вербальная коммуникация дикторов между собой на русском языке.

5. Минимизация эффекта Хоторна при записи данных.

6. Разметка данных по меткам вовлеченности и психоэмоциональных состояний информантов.

Таким образом, мы планируем собрать новый русскоязычный многомодальный корпус, содержащий записи коммуникаций людей в группе. На основе этих данных будет разработана программная система для распознавания вовлеченности с использованием многомодального анализа информации.

Финансовая поддержка

Настоящий обзор выполнен в рамках ведущей научной школы РФ (грант № НШ-17.2022.1.6), а также частично в рамках бюджетной темы (№ FFZF-2022-0005).

Литература

1. **Pregowska A., Masztalerz K., Garlińska M., Osial M.** A worldwide journey through distance education — from the post office to virtual, augmented and mixed realities, and education during the COVID-19 pandemic. *Education Sciences*, 2021, vol. 11, no. 3, pp. 1–26. doi:10.3390/educsci11030118
2. **Sümer Ö., Goldberg P., d'Mello S., Gerjets P., Trautwein U., Kasneci E.** Multimodal engagement analysis from facial videos in the classroom. *IEEE Transactions on Affective Computing*, 2021, 16 p. doi:10.1109/TAFFC.2021.3127692

3. **Nkomo L., Daniel B.** Sentiment analysis of student engagement with lecture recording. *TechTrends*, 2021, vol. 65, no. 2, pp. 213–224. doi.org:10.1007/s11528-020-00563-8
4. **Дозорцев В. М., Назин В. А.** Компьютерные тренажеры как инструмент моделирования операторской деятельности в психологическом эксперименте. *Актуальные проблемы психологии труда, инженерной психологии и эргономики*: тр. Института психологии РАН, 2013, вып. 5, с. 81–103.
5. **Соколов В. Н., Коротеев Г. Л.** Принципы и технологии построения адаптивных обучающих сред. *Актуальные проблемы психологии труда, инженерной психологии и эргономики*: тр. Института психологии РАН, 2013, вып. 5, с. 57–81.
6. **Двойникова А. А., Карпов А. А.** Аналитический обзор подходов к распознаванию тональности русскоязычных текстовых данных. *Информационно-управляющие системы*, 2020, № 4, с. 20–30. doi:10.31799/1684-8853-2020-4-20-30
7. **Kahn W. A.** Psychological conditions of personal engagement and disengagement at work. *The Academy of Management Journal*, 1990, vol. 33, no. 4, pp. 692–724. doi:10.2307/256287
8. **Kelders S. M., van Zyl L. E., Ludden G.** The concept and components of engagement in different domains applied to eHealth: A systematic scoping review. *Frontiers in Psychology*, 2020, vol. 11, Article 926. doi:10.3389/fpsyg.2020.00926
9. **De Vreede T., Andel S., de Vreede G.-J., Spector P. E., Singh V., Padmanabhan B.** What is engagement and how do we measure it? Toward a domain independent definition and scale. *Proc. of the 52nd Hawaii Intern. Conf. on System Sciences (HICSS 2019)*, 2019, pp. 1–10. doi:10.24251/HICSS.2019.092
10. **Calder B. J., Malthouse E. C., Schaedel U.** An experimental study of the relationship between online engagement and advertising effectiveness. *Journal of Interactive Marketing*, 2009, vol. 23, no. 4, pp. 321–331. doi:10.1016/j.intmar.2009.07.002
11. **Smith M.** An approach to the study of the social act. *Psychological Review*, 1942, vol. 49, no. 5, pp. 422–440. doi:10.1037/h0062907
12. **Posner M. I.** Orienting of attention. *Quarterly Journal of Experimental Psychology*, 1980, vol. 32, no. 1, pp. 3–25. doi:10.1080/00335558008248231
13. **Li Y., Lerner R. M.** Interrelations of behavioral, emotional, and cognitive school engagement in high school students. *Journal of Youth and Adolescence*, 2013, vol. 42, no. 1, pp. 20–32. doi:10.1007/s10964-012-9857-5
14. **Truss C., Soane E., Edwards C., Wisdom K., Croll A., Burnett J.** *Working Life: Employee Attitudes and Engagement 2006*. Chartered Inst. of Personnel and Development, 2006. 54 p.
15. **Fredricks J. A., McColskey W.** The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. *Handbook of research on student engagement*/ S. L. Christenson et al. (eds.). Springer Science+Business Media, 2012. Pp. 763–782. doi:10.1007/978-1-4614-2018-7_37
16. **Coates H.** The value of student engagement for higher education quality assurance. *Quality in Higher Education*, 2005, vol. 11, no. 1, pp. 25–36. doi:10.1080/13538320500074915
17. **Greene J. A., Plumley R. D., Urban C. J., Bernacki M. L., Gates K. M., Hogan K. A., Demetriou C., Panter A. T.** Modeling temporal selfregulatory processing in a higher education biology course. *Learning and Instruction*, 2021, vol. 72, pp. 101201. doi:10.1016/j.learninstruc.2019.04.002
18. **Boekaerts M.** Engagement as an inherent aspect of the learning process. *Learning and Instruction*, 2016, vol. 43, pp. 76–83. doi:10.1016/j.learninstruc.2016.02.001
19. **Miller B. W.** Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist*, 2015, vol. 50, no. 1, pp. 31–42. doi:10.1080/00461520.2015.1004068
20. **Ringeval F., Sonderegger A., Sauer J., Lalanne D.** Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. *Proc. of IEEE Intern. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, 2013, pp. 1–8. doi:10.1109/FG.2013.6553805
21. **Mohamad Nezami O., Dras M., Hamey L., Richards D., Wan S., Paris C.** Automatic recognition of student engagement using deep learning and facial expression. *Joint European Conf. on Machine Learning and Knowledge Discovery in Databases*, 2019, pp. 273–289. doi:10.1007/978-3-030-46133-1_17
22. **Kamath A., Biswas A., Balasubramanian V.** A crowdsourced approach to student engagement recognition in e-learning environments. *Proc. of the 2016 IEEE Winter Conf. on Applications of Computer Vision (WACV)*, 2016, pp. 1–9. doi:10.1109/WACV.2016.7477618
23. **Kaur A., Mustafa A., Mehta L., Dhall A.** Prediction and localization of student engagement in the wild. *Proc. of 2018 Conf. on Digital Image Computing: Techniques and Applications (DICTA)*, 2018, pp. 1–8. doi:10.1109/DICTA.2018.8615851
24. **Gupta A., D'Cunha A., Awasthi K., Balasubramanian V.** DAiSEE: Towards user engagement recognition in the wild. *Journal of Latex Class Files*, 2015, vol. 14, no. 8, 12 p. doi:10.48550/arXiv.1609.01885
25. **Whitehill J., Serpell Z., Lin Y.-Ch., Foster A., Movellan J. R.** The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, 2014, vol. 5, no. 1, pp. 86–98. doi:10.1109/TAFFC.2014.2316163
26. **Celiktutan O., Skordos E., Gunes H.** Multimodal human-human-robot interactions (MHHRI) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing*, 2017, vol. 10, no. 4, pp. 484–497. doi:10.1109/TAFFC.2017.2737019

27. **Cafaro A., Wagner J., Baur T., Dermouche S., Torres M. T., Pelachaud C., André E., Valstar M.** The NoXi database: multimodal recordings of mediated novice-expert interactions. *Proc. of the 19th ACM Intern. Conf. on Multimodal Interaction*, 2017, pp. 350–359. doi:10.1145/3136755.3136780
28. **Guhan P., Agarwal M., Awasthi N., Reeves G., Manocha D., Bera A.** ABC-Net: Semi-supervised multimodal GAN-based engagement detection using an affective, behavioral and cognitive model. *arXiv preprint arXiv:2011.08690*, 2020. doi:10.48550/arXiv.2011.08690
29. **Psaltis A., Apostolakis K. C., Dimitropoulos K., Daras P.** Multimodal student engagement recognition in prosocial games. *IEEE Transactions on Games*, 2018, vol. 10, no. 3, pp. 292–303. doi:10.1109/TCIAIG.2017.2743341
30. **Mayo E.** Hawthorne and the western electric company. *The Social Problems of an Industrial Civilization*. Routledge & Kegan Paul, London, 1949. Pp. 161–182.
31. **Fedotov D., Perepelkina O., Kazimirova E., Konstantinova M., Minker W.** Multimodal approach to engagement and disengagement detection with highly imbalanced in-the-wild data. *Proc. of the Workshop on Modeling Cognitive Processes from Multimodal Data (MCPMD'18)*, 2018, pp. 1–9. doi:10.1145/3279810.3279842
32. **Heimerl A., Baur T., André E.** A Transparent framework towards the context-sensitive recognition of conversational engagement. *Proc. of the 11th Intern. Workshop on Modelling and Reasoning in Context*, 2020, pp. 7–16.
33. **Thomas C., Nair N., Jayagopi D. B.** Predicting engagement intensity in the wild using temporal convolutional network. *Proc. of the 20th ACM Intern. Conf. on Multimodal Interaction*, 2018, pp. 604–610. doi:10.1145/3242969.3264984
34. **Yang J., Wang K., Peng X., Qiao Y.** Deep recurrent multi-instance learning with spatio-temporal features for engagement intensity prediction. *Proc. of the 20th ACM Intern. Conf. on Multimodal Iinteraction*, 2018, pp. 594–598. doi:10.1145/3242969.3264981
35. **Abedi A., Khan S.** Affect-driven engagement measurement from videos. *arXiv preprint arXiv:2106.10882*, 2021. doi:10.48550/arXiv.2106.10882
36. **Niu X., Han H., Zeng J., Sun X., Shan Sh., Huang Y., Yang S., Chen X.** Automatic engagement prediction with GAP feature. *Proc. of the 20th ACM Intern. Conf. on Multimodal Interaction (ICMI'18)*, 2018, pp. 599–603. doi:10.1145/3242969.3264982
37. **Dermouche S., Pelachaud C.** Engagement modeling in dyadic interaction. *Proc. of the 2019 Intern. Conf. on Multimodal Interaction (ICMI'19)*, 2019, pp. 440–445. doi:10.1145/3340555.3353765
38. **Dresvyanskiy D., Minker W., Karpov A.** Deep learning based engagement recognition in highly imbalanced data. *Speech and Computer. SPECOM 2021. Lecture Notes in Computer Science*. Springer, Cham, 2021. Vol 12997. Pp. 166–178. doi:10.1007/978-3-030-87802-3_16
39. **Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z.** Rethinking the inception architecture for computer vision. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.
40. **O'Brien H., Toms E.** The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 2010, vol. 61, no. 1, pp. 50–69. doi:10.1002/asi.21229
41. **Fuller K. A., Karunaratne N. S., Naidu S., Exintaris B., Short J. L., Wolcott M. D., Singleton S., White P. J.** Development of a self-report instrument for measuring in-class student engagement reveals that pretending to engage is a significant unrecognized problem. *PLOS ONE*, 2018, vol. 13, no. 10, pp. e0205828. doi:10.1371/journal.pone.0205828
42. **Koto I., Harneli M., Winarni E.** Primary school teacher strategy to promote student engagement in science lessons. *Proc. of Intern. Conf. of Primary Education Research Pivotal Literature and Research UNNES 2018 (IC PEOPLE UNNES 2018)*, 2018, pp. 122–127. doi:10.2991/icpeopleunnes-18.2019.25
43. **Garn A. C., Dasigner T., Simonton A., Simonton K. L.** Predicting changes in student engagement in university physical education: Application of control-value theory of achievement emotions. *Psychology of Sport and Exercise*, 2017, no. 29, pp. 93–102. doi:10.1016/j.psychsport.2016.12.005
44. **D'Errico F., Paciello M., Cerniglia L.** When emotions enhance students' engagement in e-learning processes. *Journal of e-Learning and Knowledge Society*, 2016, vol. 12, no. 4, pp. 173676.
45. **Верховяк О. В., Карпов А. А.** Автоматический анализ эмоционально окрашенной речи. *Голосовой портрет ребенка с типичным и атипичным развитием*/ Е. Е. Лякco, О. В. Фролова (ред.). СПб., Издательско-полиграфическая ассоциация вузов, 2020. С. 149–198.

UDC 004.4/.5+159.99
doi:10.31799/1684-8853-2022-5-12-22
EDN: CXBRCS

Analytical review of methods for automatic detection of user engagement in virtual communication

A. A. Dvoynikova[a], Junior Researcher, orcid.org/0000-0001-8047-6639
I. A. Kagirov[a], Research Fellow, orcid.org/0000-0003-1196-1117
A. A. Karpov[a], Dr. Sc., Tech., Professor, Principal Researcher, orcid.org/0000-0003-3424-652X, karpov@iias.spb.su
[a]St. Petersburg Federal Research Center of the RAS, 39, 14th Line, 199178, Saint-Petersburg, Russian Federation

Introduction: The solution of the task of the recognition and assessment of user engagement in the acts of human-machine interaction or telecommunication, achieved through the use of automatic means, is highly important in computer recognition of human psycho-emotional states. This is necessary for e-learning, business and entertainment applications design. **Purpose:** To conduct a comparative analysis of the current information support in the field of automatic recognition and assessment of user involvement in human-machine interaction or virtual communication, as well as to establish a methodology for building a data body based on the idea of multimodal communication. **Results:** The conducted analysis of research papers has shown that in most existing databases there is a substantial lack of data for natural online communication. Moreover, not all databases contain different modalities in "human – machine – human" communication system. Text and audio modalities turn out to be important for a multilevel engagement classification task, aimed at the determination of engagement intensity. It is also promising to take into account "body language" features, such as facial expressions, movements of the body and the head, gestures. For the correct assessment of involvement, an engagement database must contain metadata on the psycho-emotional states of communicants. Neural network-based approaches to the automatic detection of user engagement show the best performance. **Practical relevance:** Based on the obtained analytical conclusions, the authors of the paper are going to elaborate an original software system for automatic recognition of user engagement, and to collect a data set for machine learning purposes. The presented review formulates basic requirements for such systems and contributes to the solution of the problem of automatic recognition of psycho-emotional states. **Discussion:** The survey leads to the conclusion that the notion of engagement as understood in studies on automatic emotion recognition differs from that used in psychology. User (or communicant) engagement in terms of info- and communicative sphere implies the manifestation of a person's mental activity level (emotional, cognitive, and behavioral components) changing dynamically while interacting with another person or computer system.

Keywords – user engagement, information support, automatic emotion recognition systems, multimodality, artificial neural networks.

For citation: Dvoynikova A. A., Kagirov I. A., Karpov A. A. Analytical review of methods for automatic detection of user engagement in virtual communication. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 5, pp. 12–22 (In Russian). doi:10.31799/1684-8853-2022-5-12-22, EDN: CXBRCS

Financial support

This survey was carried out in the framework of the Council for Grants of the President of Russia for Leading scientific schools (grant No. NSH-17.2022.1.6), as well as due as part of Russian state research (No. FFZF-2022-0005).

References

1. Pregowska A., Masztalerz K., Garlińska M., Osial M. A worldwide journey through distance education – from the post office to virtual, augmented and mixed realities, and education during the COVID-19 pandemic. *Education Sciences*, 2021, vol. 11, no. 3, pp. 118. doi:10.3390/educsci11030118
2. Sümer Ö., Goldberg P., d'Mello S., Gerjets P., Trautwein U., Kasneci E. Multimodal engagement analysis from facial videos in the classroom. *IEEE Transactions on Affective Computing*, 2021, 16 p. doi:10.1109/TAFFC.2021.3127692
3. Nkomo L., Daniel B. Sentiment analysis of student engagement with lecture recording. *TechTrends*, 2021, vol. 65, no. 2, pp. 213–224. doi.org:10.1007/s11528-020-00563-8
4. Dozortsev V. M., Nazin V. A. Computer simulators as a tool for modeling operator activity in a psychological experiment. *Tr. Instituta psihologii RAN "Aktual'nye problemy psihologii truda, inzhenernoj psihologii i ergonomiki"* [Proc. of the Institute of Psychology of RAS "Actual Problems of Occupational Psychology, Engineering Psychology and Ergonomics"], 2013, iss. 5, pp. 81–103 (In Russian).
5. Sokolov V. N., Koroteev G. L. Principles and technologies for building adaptive learning environments. *Tr. Instituta psihologii RAN "Aktual'nye problemy psihologii truda, inzhenernoj psihologii i ergonomiki"* [Proc. of the Institute of Psychology of RAS "Actual Problems of Occupational Psychology, Engineering Psychology and Ergonomics"], 2013, iss. 5, pp. 57–81 (In Russian).
6. Dvoynikova A. A., Karpov A. A. Analytical review of approaches to Russian text sentiment recognition *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 4, pp. 20–30 (In Russian). doi:10.31799/1684-8853-2020-4-20-30
7. Kahn W. A. Psychological conditions of personal engagement and disengagement at work. *The Academy of Management Journal*, 1990, vol. 33, no. 4, pp. 692–724. doi:10.2307/256287
8. Kelders S. M., van Zyl L. E., Ludden G. The concept and components of engagement in different domains applied to eHealth: A systematic scoping review. *Frontiers in Psychology*, 2020, vol. 11, Article 926. doi:10.3389/fpsyg.2020.00926
9. De Vreede T., Andel S., de Vreede G.-J., Spector P. E., Singh V., Padmanabhan B. What is engagement and how do we measure it? Toward a domain independent definition and scale. *Proc. of the 52nd Hawaii Intern. Conf. on System Sciences (HICSS 2019)*, 2019, pp. 1–10. doi:10.24251/HICSS.2019.092
10. Calder B. J., Malthouse E. C., Schaedel U. An experimental study of the relationship between online engagement and advertising effectiveness. *Journal of Interactive Marketing*, 2009, vol. 23, no. 4, pp. 321–331. doi:10.1016/j.intmar.2009.07.002
11. Smith M. An approach to the study of the social act. *Psychological Review*, 1942, vol. 49, no. 5, pp. 422–440. doi:10.1037/h0062907
12. Posner M. I. Orienting of attention. *Quarterly Journal of Experimental Psychology*, 1980, vol. 32, no. 1, pp. 3–25. doi:10.1080/00335558008248231
13. Li Y., Lerner R. M. Interrelations of behavioral, emotional, and cognitive school engagement in high school students. *Journal of Youth and Adolescence*, 2013, vol. 42, no. 1, pp. 20–32. doi:10.1007/s10964-012-9857-5
14. Truss C., Soane E., Edwards C., Wisdom K., Croll A., Burnett J. *Working Life: Employee Attitudes and Engagement 2006*. Chartered Inst. of Personnel and Development, 2006. 54 p.
15. Fredricks J. A., McColskey W. *The measurement of student engagement: A comparative analysis of various methods and student self-report instruments*. In: *Handbook of research on student engagement*. S. L. Christenson et al. (eds.). Springer Science+Business Media, 2012. Pp. 763–782. doi:10.1007/978-1-4614-2018-7_37
16. Coates H. The value of student engagement for higher education quality assurance. *Quality in Higher Education*, 2005, vol. 11, no. 1, pp. 25–36. doi:10.1080/13538320500074915

17. Greene J. A., Plumley R. D., Urban C. J., Bernacki M. L., Gates K. M., Hogan K. A., Demetriou C., Panter A. T. Modeling temporal selfregulatory processing in a higher education biology course. *Learning and Instruction*, 2021, vol. 72, pp. 101201. doi:10.1016/j.learninstruc.2019.04.002
18. Boekaerts M. Engagement as an inherent aspect of the learning process. *Learning and Instruction*, 2016, vol. 43, pp. 76–83. doi:10.1016/j.learninstruc.2016.02.001
19. Miller B. W. Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist*, 2015, vol. 50, no. 1, pp. 31–42. doi:10.1080/00461520.2015.1004068
20. Ringeval F., Sonderegger A., Sauer J., Lalanne D. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. *Proc. of IEEE Intern. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, 2013, pp. 1–8. doi:10.1109/FG.2013.6553805
21. Mohamad Nezami O., Dras M., Hamey L., Richards D., Wan S., Paris C. Automatic recognition of student engagement using deep learning and facial expression. *Joint European Conf. on Machine Learning and Knowledge Discovery in Databases*, 2019, pp. 273–289. doi:10.1007/978-3-030-46133-1_17
22. Kamath A., Biswas A., Balasubramanian V. A crowdsourced approach to student engagement recognition in e-learning environments. *Proc. of the 2016 IEEE Winter Conf. on Applications of Computer Vision (WACV)*, 2016, pp. 1–9. doi:10.1109/WACV.2016.7477618
23. Kaur A., Mustafa A., Mehta L., Dhall A. Prediction and localization of student engagement in the wild. *Proc. of 2018 Conf. on Digital Image Computing: Techniques and Applications (DICTA)*, 2018, pp. 1–8. doi:10.1109/DICTA.2018.8615851
24. Gupta A., D'Cunha A., Awasthi K., Balasubramanian V. DAiSEE: Towards user engagement recognition in the wild. *Journal of Latex Class Files*, 2015, vol. 14, no. 8, 12 p. doi:10.48550/arXiv.1609.01885
25. Whitehill J., Serpell Z., Lin Y.-Ch., Foster A., Movellan J. R. The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, 2014, vol. 5, no. 1, pp. 86–98. doi:10.1109/TAFFC.2014.2316163
26. Celiktutan O., Skordos E., Gunes H. Multimodal human-human-robot interactions (MHHRI) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing*, 2017, vol. 10, no. 4, pp. 484–497. doi:10.1109/TAFFC.2017.2737019
27. Cafaro A., Wagner J., Baur T., Dermouche S., Torres M. T., Pelachaud C., André E., Valstar M. The NoXi database: multimodal recordings of mediated novice-expert interactions. *Proc. of the 19th ACM Intern. Conf. on Multimodal Interaction*, 2017, pp. 350–359. doi:10.1145/3136755.3136780
28. Guhan P., Agarwal M., Awasthi N., Reeves G., Manocha D., Bera A. ABC-Net: Semi-supervised multimodal GAN-based engagement detection using an affective, behavioral and cognitive model. *arXiv preprint arXiv:2011.08690*, 2020. doi:10.48550/arXiv.2011.08690
29. Psaltis A., Apostolakis K. C., Dimitropoulos K., Daras P. Multimodal student engagement recognition in prosocial games. *IEEE Transactions on Games*, 2018, vol. 10, no. 3, pp. 292–303. doi:10.1109/TCIAIG.2017.2743341
30. Mayo E. Hawthorne and the western electric company. *The Social Problems of an Industrial Civilization*. Routledge & Kegan Paul, London, 1949. Pp. 161–182.
31. Fedotov D., Perepelkina O., Kazimirova E., Konstantinova M., Minker W. Multimodal approach to engagement and disengagement detection with highly imbalanced in-the-wild data. *Proc. of the Workshop on Modeling Cognitive Processes from Multimodal Data (MCPMD'18)*, 2018, pp. 1–9. doi:10.1145/3279810.3279842
32. Heimerl A., Baur T., André E. A Transparent framework towards the context-sensitive recognition of conversational engagement. *Proc. of the 11th Intern. Workshop on Modelling and Reasoning in Context*, 2020, pp. 7–16.
33. Thomas C., Nair N., Jayagopi D. B. Predicting engagement intensity in the wild using temporal convolutional network. *Proc. of the 20th ACM Intern. Conf. on Multimodal Interaction*, 2018, pp. 604–610. doi:10.1145/3242969.3264984
34. Yang J., Wang K., Peng X., Qiao Y. Deep recurrent multi-instance learning with spatio-temporal features for engagement intensity prediction. *Proc. of the 20th ACM Intern. Conf. on Multimodal Iinteraction*, 2018, pp. 594–598. doi:10.1145/3242969.3264981
35. Abedi A., Khan S. Affect-driven engagement measurement from videos. *arXiv preprint arXiv:2106.10882*, 2021. doi:10.48550/arXiv.2106.10882
36. Niu X., Han H., Zeng J., Sun X., Shan Sh., Huang Y., Yang S., Chen X. Automatic engagement prediction with GAP feature. *Proc. of the 20th ACM Intern. Conf. on Multimodal Interaction (ICMI'18)*, 2018, pp. 599–603. doi:10.1145/3242969.3264982
37. Dermouche S., Pelachaud C. Engagement modeling in dyadic interaction. *Proc. of the 2019 Intern. Conf. on Multimodal Interaction (ICMI'19)*, 2019, pp. 440–445. doi:10.1145/3340555.3353765
38. Dresvyanskiy D., Minker W., Karpov A. *Deep learning based engagement recognition in highly imbalanced data*. In: *Speech and Computer. SPECOM 2021. Lecture Notes in Computer Science*. Springer, Cham, 2021. Vol 12997. Pp. 166–178. doi:10.1007/978-3-030-87802-3_16
39. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.
40. O'Brien H., Toms E. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 2010, vol. 61, no. 1, pp. 50–69. doi:10.1002/asi.21229
41. Fuller K. A., Karunaratne N. S., Naidu S., Exintaris B., Short J. L., Wolcott M. D., Singleton S., White P. J. Development of a self-report instrument for measuring in-class student engagement reveals that pretending to engage is a significant unrecognized problem. *PLOS ONE*, 2018, vol. 13, no. 10, pp. e0205828. doi:10.1371/journal.pone.0205828
42. Koto I., Harneli M., Winarni E. Primary school teacher strategy to promote student engagement in science lessons. *Proc. of Intern. Conf. of Primary Education Research Pivotal Literature and Research UNNES 2018 (IC PEOPLE UNNES 2018)*, 2018, pp. 122–127. doi:10.2991/icpeopleunnes-18.2019.25
43. Garn A. C., Dasigner T., Simonton A., Simonton K. L. Predicting changes in student engagement in university physical education: Application of control-value theory of achievement emotions. *Psychology of Sport and Exercise*, 2017, no. 29, pp. 93–102. doi:10.1016/j.psychsport.2016.12.005
44. D'Errico F., Paciello M., Cerniglia L. When emotions enhance students' engagement in e-learning processes. *Journal of e-Learning and Knowledge Society*, 2016, vol. 12, no. 4, pp. 173676.
45. Verkholyak O. V., Karpov A. A. *Automatic analysis of emotionally charged speech*. In: *Golosovoj portret rebenka s tipichnym i atipichnym razvitiem* [Voice portrait of typical and atypical children]. E. E. Lyakso, O. V. Frolova (Eds.). Saint-Petersburg, Izdatelsko-poligraficheskaya associaciya vuzov Publ., 2020. Pp. 149–198 (In Russian).