



ТЕХНИЧЕСКИ УНИВЕРСИТЕТ – СОФИЯ
Факултет по Компютърни системи и технологии
Катедра „Компютърни системи“

маг. инж. Ралица Николаева Райнова

**Методи и алгоритми за извличане на информация
за емоционално състояние**

А В Т О Р Е Ф Е Р А Т

на дисертация за придобиване на образователна и научна степен
"ДОКТОР"

област 5. Технически науки
професионално направление 5.3. Комуникационна и компютърна техника
научна специалност „Системи с изкуствен интелект“

Научни ръководители:

проф. д-р инж. Милена Кирилова Лазарова–Мицева
доц. д-р инж. Аделина Пламенова Алексиева-Петрова

СОФИЯ, 2021 г.

Дисертационният труд е обсъден и насочен за защита от Катедрения съвет на катедра „Компютърни системи“ към Факултет по Компютърни системи и технологии на ТУ-София на редовно заседание, проведено на 08.06.2021г.

Публичната защита на дисертационния труд ще се състои на 21.09.2021 г. от 15.00 часа в Конферентната зала на БИЦ на Технически университет – София на открито заседание на научното жури, определено със заповед № ОЖ-5.3-16/ 15.06.2021 г. на Ректора на ТУ-София в състав:

1. доц. д-р Аделина Пламенова Алексиева–Петрова – председател
2. доц. д-р Мария Ангелова Ангелова–Димитрова – научен секретар
3. проф. дтн Стойчо Димитров Стойчев
4. проф. д-р Атанас Велков Атанасов
5. доц. д-р Александър Димов Димов

Рецензенти:

1. проф. дтн Стойчо Димитров Стойчев
2. доц. д-р Мария Ангелова Ангелова–Димитрова

Материалите по защитата са на разположение на интересуващите се в канцеларията на Факултет по компютърни системи и технологии на ТУ-София, блок № 1, кабинет № 1443А.

Дисертантът е задочен докторант към катедра „Компютърни системи“ на Факултет по Компютърни системи и технологии. Изследванията по дисертационната разработка са направени от автора, като някои от тях са подкрепени от научноизследователски проекти.

Автор: маг. инж. Ралица Николаева Райнова

Заглавие: Методи и алгоритми за извличане на информация за емоционално състояние

Тираж: 30 броя

Отпечатано в ИПК на Технически университет – София

I. ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД

Актуалност на проблема

През последните години се наблюдава повишаване на нивата на хора със симптоми на тревожност, депресия, психологически стрес и други депресивни и тревожни разстройства. Обичайните признаци и симптоми, с продължителност от поне две седмици, включват: потиснато настроение, безпокойство или раздразнителност, чувство за вина, безполезност, апатия, намалено самочувствие, песимизъм. Всички тези признаци и симптоми намират отражение в емоционалното състояние на човек. Една от задачите на борбата срещу депресивните разстройства е тяхното ранно диагностициране. Автоматизираното откриване на белези за депресивно състояние и тревожни разстройства е сложна задача, за решаването на която могат да се използват методи и средства за извличане на информация от разнородни входни модалности.

Цел на дисертационния труд, основни задачи и методи за изследване

Целта на научните изследвания в дисертационния труд е: да се изследват възможностите за ранно диагностициране на депресивни разстройства чрез неинвазивни методи, базирани на извличане на информация за емоционално състояние на човека и да се предложи агент-базиран подход за тяхното разпознаване чрез мултимодални данни.

За постигане на поставената цел са определени следните задачи:

1. Да се проучат съществуващите методи и алгоритми за извличане на информация за емоционално състояние.
2. Да се предложи подход за извличане на информация за емоционално състояние с използването на множество модалности.
3. Да се проектира мултиагентна архитектура за извличане на информация за емоционално състояние с използване на мултимодални данни.
4. Да се оцени експериментално възможността за използване на проектираната мултиагентна архитектура при извличането на информация за емоционално състояние, чрез валидиране на предложения подход за множество модалности.

Научна новост

Изследвани са методи и алгоритми за извличане на характеристики и класификация на емоции с използване на вербални, визуални и физиологични данни, които да повишат точността при откриването на депресивни състояния.

Предложени са абстракции и са описани основните модели на мултимодална мултиагентна система за извличане и обработка на информация за емоционално състояние.

Предложен е модел на агент за комбиниране на резултатите от различни агенти за откриване на емоции, който позволява балансиране на данните при обединяването им според типа модалност, пол, възраст и етническа група;

Практическа приложимост

Проектирана е мултимодална мултиагентна система за извличане и обработка на информация за емоционално състояние. На базата на експериментални изследвания е оценена точността на класификация на избраните методи и алгоритми за класификация на

емоции при използване на единствена модалност и е доказана необходимостта от използване на предложената мултимодална мултиагентна система за извличане и обработка на информация за емоционално състояние, особено по отношение на емоциите с негативна валентност, които се проявяват при депресивни състояния.

Апробация

Резултатите от дисертационния труд са разглеждани, обсъждани и публикувани в:

1. Осма международна научна конференция „Computer Science '2018“, Кавала, Гърция, 13–15 септември 2018.
2. Международна научна конференция „Автоматика и информатика '2019“, София, България, 3–5 октомври 2019.
3. Девета международна научна конференция „Computer Science '2020“, Велинград, България
4. XXVIII национална конференция с международно участие „TELECOM '2020“, София, България, 29–30 октомври 2020.
5. Списание “Computer and Communications Engineering”, ТУ–София.
6. Списание „Автоматика и информатика“, САИ.

Публикации

Основни постижения и резултати от дисертационния труд са публикувани в 7 научни статии, от които 1 самостоятелна и шест като водещ съавтор. Статиите са публикувани в български и международни реферирани и индексирани издания.

Структура и обем на дисертационния труд

Дисертационният труд е в обем от 183 страници, като включва увод, 4 глави за решаване на формулираните основни задачи, списък на основните приноси, списък на публикациите по дисертацията и използвана литература.

Цитирани са общо 225 литературни източници, като 205 са на латиница и 2 на кирилица, а останалите са Интернет адреси. Работата включва общо 77 фигури и 29 таблици. Номерата на фигурите и таблиците в автореферата съответстват на тези в дисертационния труд.

II. СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД

ГЛАВА 1. Същност на автоматизираното извличане на информация за емоционално състояние

Белезите, по който може да бъде разпознато емоционалното състояние най-често са свързани с наблюдение на човешкото поведение. Но не винаги истинското емоционално състояние може да се определи коректно единствено чрез мимиките на лицето или жестове на тялото. За да се определят с по-голяма сигурност емоциите, които човек преживява, са необходими други източници на информация: промяна на тона по време на неговото словесно изразяване, повишаване на телесната температура, пулс и други. Ако към използваните мултимодални данни се добавят и коефициенти за характеристики, като темперамент, години, пол, етническа група и се проследяват в определен времеви диапазон може да се повиши точността на определяне на емоционалното състояние близо до реалното за даден човек, което би допринесло за откриване на признаците и симптомите на депресивни разстройства.

1.1 Представяне на емоции

От гледна точка на представянето на емоциите и използването на това представяне за компютърно базиран анализ и разпознаване на емоционално състояние са възможни два основни подхода: представяне с категории и представяне в многомерно пространство.

Към първия вид представяне са разгледани са някои от най-значимите изследвания от началото на XIX век, сред които дискретния модел на емоциите и системата за кодиране на лицеве изразения, разработена под ръководството на Пол Екман, които определят седем основни категории: щастие, тъга, изненада, страх, гняв, отвращение и презрение. Диференциалната теория за емоциите на Карол Изард и група учени, която включва десет различни емоции. Мултидисциплинарните изследвания на Едуард Пунцет и Палау Геа - които от своя страна разкриват повече от триста човешки емоции.

Към втория подход за представяне са разгледани, набор от измерения описващи афекта като взаимно свързани по систематичен начин в пространствен модел, познат като кръгов модел на Ръсел, триизмерния модел на емоционалното състояние - удоволствие, възбуда и господство (Pleasure - Arousal - Dominance), разработен от Алберт Мехрабян и Джеймс Ръсел и хибридният модел на Робърт Плутчик.

1.2 Видове модалности за разпознаване на емоции

При автоматичното разпознаване на емоциите, за да се опростят, емоциите се третираят като многостранно явление, състоящо се от: субективни чувства (семантичният канал на гласа и текста); поведенческа реакция (тяло и жестове); изразителна реакция (глас и лице); физиологична реакция (електроенцефалограма, електродермална реакция, телесна температура и др.);

От гледна точка на сензорните технологии за отдалечено регистриране и разпознаване на емоции, използваните модалности могат да се класифицират като: вербални данни; визуални данни; физиологични данни.

Разгледани са накратко технологиите които използват вербални (звукови) данни, визуални данни и физиологични данни за разпознаване на емоции, както и използваните нива на абстракция и основни характеристики.

1.3 Системи за разпознаване на емоции

Независимо от вида на входните данни откриването на емоции следва аналогични стъпки във всяка система, които могат да се обобщят, както следва (фиг. 1-11):

- предварителна обработка на входните данни,
- идентифициране на ключовите точки за откриване на характеристики,
- класификация и определяне на емоцията.



Фиг. 1-11 Блок диаграма на система за разпознаване на емоции

Системите за разпознаване на емоции, се класифицират:

- Според вида на предлаганите подходи на: системи базирани на знания, системи базирани на статистически методи и хибридни системи;
- Според това дали използват еднородни или нееднородни входни величини на: едно модални и мултимодални системи;
- Според използваните методи и техники на: използващи техники за обработка на сигнали, използващи алгоритми за машинно обучение и системи с изкуствен интелект.

Съществуващите системи за разпознаване на емоции са базирани предимно на едно-модални входни данни и много малка част от тях са имплементирани с използването на интелигентни агенти (Фиг. 1-13).



Фиг. 1-13 Процентно съотношение на общото количество публикации за разпознаване на емоции и тези с мултимодални и мултиагентни технологии

Използването на множество модалности може да повиши точността за разпознаване на емоции, а имплементация с интелигентни агенти би увеличила бързодействието и ефективността на системата за разпознаване на емоции.

1.4 Изводи към първа глава

В резултат на направените проучвания в областта на методите и алгоритмите за извличане на информация за емоционално състояние, представени в първа глава на дисертационния труд могат да бъдат направени следните по-важни изводи:

- Човешките емоции са важни индикатори за психичното здраве на хората, особено тези като тъга, мъка, страх или тревога, които са първоначалните депресивни промени на настроението и се проявяват с понижен емоционален контакт, склонност към песимизъм, намаляване или дори изчезване на способността за изпитване на радост, удоволствие, за смях или веселба.
- Съществуват различни начини за представяне на човешките емоции, като най-често използваните са: представяне на емоции чрез категории и представяне на емоции в многомерно пространство.
- Автоматизираното откриване на белези за депресивно състояние и тревожни разстройства е сложна задача.
- За регистриране и разпознаване на човешки емоции могат да се използват различни модалности: вербални, визуални и физиологични. За всяка модалност могат да се използват различни характеристики, с използването на които могат да бъдат приложени разнообразни методи за разпознаване на емоционално състояние.
- Съществуващите системи за разпознаване на емоции са базирани предимно на едно-модални входни данни и много малка част от тях са имплементирани с използването на интелигентни агенти.

ГЛАВА 2. Методи за извличане на информация за емоционално състояние

Извличането на информация от данни е процес на откриване на характеристики, които са съществени за набора данни и използването им за класификация и разпознаване на свойства на обекта, за който се отнасят данните. Един и същи обект може да бъде описан с различни източници данни и да има свое отражение в тях. Всеки източник на данни описва обекта с определено множество характеристики, които могат да бъдат открити и описани на базата на различните данни. Извличането на информация от данни се разделя на три отделни етапа, за всеки от които се прилагат специфични методи за обработване на данните (фиг. 1-11).

[1] Предварителна обработка – методите в тази група се съсредоточават върху нормализация на данните и прилагане на различни филтри, които да намалят шума и подобрят качеството на сигнала с цел да подпомогнат прилагането на методите от следващите етапи на обработка;

[2] Извличане на характеристики – методите в тази група определят кои са най-подходящите характеристики, с които да се опишат свойствата на конкретен обект, за който се отнасят данните;

[3] Класификация – под класификация се разбира процедура за подреждането на характеристики за изучаваните обекти (поредица от вектори) по групи (кълъстери) въз основа на определени признаци и критерии така, че обектите от една и съща група да са подобни помежду си и да се различават от останалите. Тези критерии обикновено са разстояние или сходство на описанието на характеристиките.

2.1 Предварителна обработка

Предварителната обработка се извършва преди да се започне извличането на характеристики и включва процеси като: избор на ниво на абстракция - нормализация, избор на коефициент на сегментация, и филтриране.

Процесът на нормализация се състои в регулиране на сигнала до предварително зададено ниво.

Използваните нива на абстракция са предимно ниско и средно ниво при звуковите и физиологични сигнали. Нивото на абстракция за предварителната обработка на изображенията включва всички нива на абстракция, тъй като филтрирането на изображенията изисква обработка на ниво пиксел, т.е. ниско ниво на абстракция, а за откриване на лицата и техните характеристики се използват средно и високо ниво на абстракция (регион, форма и шаблон).

Филтрирането е основна стъпка на етапа на предварителна обработка, която има за цел максимизиране на съотношението сигнал/шум.

2.2 Извличане на характеристики

Целта на етапа на извличане на характеристики при звукови сигнали е откриването на такова представяне на речевия звуков сигнал, което да запазва частта, характеризираща информацията в него и да е надеждна при регистриране на звукови сигнали в различни условия, включително при промени в околната среда или говорителя за един и същ речев сигнал. Често използвани характеристики за звуковите сигнали са: форма и амплитудна обвивка на звуковата вълната, интензитет, спектрограма, спектрално преобръщане и спектрален центроид - определени от времевата област. Енергийно ниво на честотната лента, честотна лента определени в честотната област. За извличането на характеристики на звуков сигнал се използват различни методи: кепсър коефициенти на честотите на Мел, линейните прогнозни кепстрални коефициенти, логаритмични коефициенти на честотна мощност.

При избор на характеристики, които да се извлекат от визуални данни целта е характеристиките да са устойчиви на различни промени в размера, местоположението или завъртането на изображенията. Повечето методи за разпознаване на лица използват специално разработени функции включващи контрол на ръбовете, биометрична идентификация, 3D модели и други.

Физиологични сигнали, които могат да се използват за разпознаване на емоционално състояние на човек са: сърдечен ритъм, кръвно налягане, електродермална активност, температура на кожата, честота на дишане и качество на съня. Физиологичните сигнали са силно зависими от човека и обкръжаващата го среда. За това при тях се измерва първоначална стойност и се търси промяна във времето. Измерването и проследяването на изменения във физиологични характеристики, като сърдечна честота, галванична реакция на кожата, температура на кожата, артериално кръвно налягане, честота и изменения на пулса се използва в допълнение към резултатите от другите модалности.

2.3 Алгоритми за класификация

Алгоритмите за класификация чрез машинно обучение имат за цел да изградят математически модел по примерни данни, наречени „обучаващи данни“, на базата на който да се прогнозира изменението на данните или да се вземат решения, без да са изрично програмирани за това. Според управлението на алгоритъма за машинно обучение се разграничават три групи методи за обучение: обучение с учител, подсилено обучение и обучение без учител (самообучение). Разгледаните основните алгоритми за класификация,

които намират приложение на етапа на класификация на извлечените характеристики на базата на сигнали от различни модалности за определяне на емоционално състояние на човек са:

- алгоритъм за класификация K means - същността на алгоритъма е определяне на скрити структури в неклаифицирани данни с цел да се намерят в общия набор от данни групи от такива, които имат общи характеристики;
- алгоритъм за класификация K Nearest Neighbor - класът на принадлежност на даден обект се определя от класа, към който принадлежат K най-близки до него обекти от обучаващата извадка;
- алгоритъм за динамично времево изкривяване - ако два обекта имат характеристики разположени в близост един до друг в пространството, то те вероятно принадлежат към един и същи клас;
- скрити модели на Марков - стохастичен подход, който се основава на двукомпонентен статистически модел за класификация: верига на Марков, която произвежда последователност от състояния и набор от изходни разпределения, управляващи начина, по който последователността от състояния се преобразува в последователност от наблюдения;
- метод на главни компоненти - използва се за намаляване на размерността на представянето на данните като се запазва значимата информацията на основния набор от данни;
- модел на Гаусова смес - вероятностен модел, който приема, че всички екземпляри във входното множество данни са генерирани от смес от краен брой Гаусови разпределения с неизвестни параметри;
- метод на опорните вектори - целта на оптимизацията при използване на метод на опорните вектори е да се определи дискриминантната функция, която определя границите при вземане на решения, разделящи класовете чрез увеличаване на разстоянието между тези граници и точки в близост до определените от границите хиперравнини, наречени опорни вектори;
- невронни мрежи с различни архитектури - математически модели, съставени от елементи, наречени изкуствени неврони, които симулират наблюдавани свойства за структурата и функционирането на биологичните нервни системи. В тази част са разгледани различни активиращи функции, архитектури, обучение на НМ. Както и дълбоки НМ и техните архитектури.

2.4 Изводи към втора глава

В резултат на направените проучвания на методите и алгоритмите за извличане на информация за емоционално състояние във втора глава на дисертационния труд могат да бъдат направени следните по-важни изводи:

- Всички използвани подходи за всеки един от етапите на разпознаване на емоции при всяка от разгледаните модалности – визуални, звукови и физиологични данни, имат предимства и недостатъци от гледна точка на изискванията към входните данни и обучаващите множества, време за изпълнение и за обучения, точност и надеждност на получените резултати.
- С използване на представените подходи, методи и алгоритми се постига висока точност на откриване и разпознаване на емоции с използване на единствена модалност (от 92.3% до 97.8%).
- Всеки един от методите за предварителна обработка, извличане на характеристики и класификация има ограничения и недостатъци по отношение на изискванията към използваните входни данни и наличните обучаващи множества.

- За да се постигне автономност и универсалност при решаване на задачата за разпознаване на емоции с използване на различни модалности и технологии за регистриране на входните данни е необходимо комбиниране на различни подходи за всяка една от фазите на разпознаване.

На базата на тези изводи може да се направи заключение, че за да се постигне високо ниво на точност и ефективност следва да се използва система за разпознаване на емоции, базирана на активни обекти, които да се включват динамично в системата и да взаимодействат помежду си и с обкръжаващата ги среда. Системата следва да може да предоставя една и съща услуга по различен начин в зависимост от входните параметри и да предоставя възможност за персонализиране. Всички тези характеристики дефинират необходимостта от моделиране на една мултиагентна система за разпознаване на емоционално състояние.

ГЛАВА 3. Проектиране на мултиагентна архитектура за извличане на информация за емоционално състояние от мултимодални данни

През последните години навлиза нов модел на проектиране на разпределени системи базирани на идеята да могат да се реорганизират и да адаптират функциите си, т.е. “интелигентни” разпределени системи. Тези системи се изграждат върху middleware и предоставят ново ниво на абстракция. За изграждането на такъв вид системи се прилага технология тясно свързана с изкуствения интелект, а именно **интелигентните агенти**. През призмата на тези технологии разпределените програмни системи се превръщат в **мултиагентни системи**.

Ранното диагностициране на депресивните разстройства чрез неинвазивни методи, базирани на извличане и разпознаване на информация за емоционално състояние на човека се основава на комбинирането и изпълнението на различни компоненти и алгоритми за разпознаване чрез мултимодални данни. Проектирането и изграждането на софтуерна рамка, която да решава тези проблеми, трябва да отговаря на основните свойства, като гъвкавост и адаптивност. Ето защо в трета глава на дисертационния труд е представена агент-базирана архитектура на разпределена система за извличане на информация за емоционално състояние от мултимодални данни.

3.1. Интелигентни агенти

За всеки възможен набор от информация, възприемана от околната среда, рационалният агент избира действие, с което да максимизира ефективността на поведението си, въз основа на възприетата информация и на всички знания, вградени в него.

Интelligentните агенти, освен атрибути и методи, притежават вярвания, желания и намерения, които ги свързват със заобикалящата ги среда, и им осигуряват психични състояния, от които зависи тяхното поведение.

3.2. Методологии за моделиране на мултиагентни системи

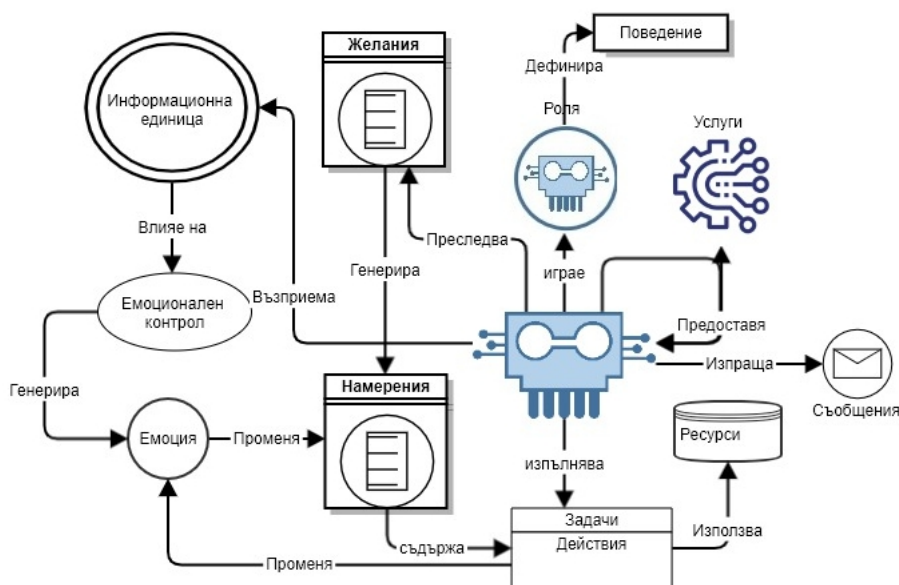
Резюмирана сравнителна таблица на методологиите за проектиране и разработка на мултиагентните системи е представена в таблица 3.2.

Табл. 3-2 Сравнителна таблица на методологиите за моделиране на мултиагентни системи

МЕТОДОЛОГИЯ	МОДЕЛИ / АБСТРАКЦИИ	ЕЗИК ЗА ОПИСАНИЕ	СРЕДА ЗА РАЗРАБОТКА	ЕТАПИ НА РАЗРАБОТКА
Vowel Engineering (AEIO)	Агент, Околна среда, Взаимодействия, Организация	Unicon	MAGMA	не специфицира
MAS - CommonKADS	Агент, Задача, Експерт, Координация, Комуникации, Организация, Дизайн	CML Conceptual Modelling Language	Не специфицира	Концептуализация, Анализ, Дизайн, Имплементация, Валидация
MaSE (Multi-agent systems Software Engineering)	Агент, Цели, Роля, Задачи, Разговори	Java-based graphical development	AgentTool	Създаване на класове на агенти Изграждане на разговори Сглобяване на класове агенти Моделиране на ролите
ZEUS	Агент, Роля, Комуникации, Убеждения, Ресурси, Компетенции	Java-based Graphical development	ZEUS	Идентификация на онтологията, агентите, услугите и знанията Определение, координация и организация на агента Визуализация и верификация.
TROPOS	Актьор, Цели, План, Ресурси, Зависимости, Възможности, Вярвания	UML	Не специфицира	Ранни изисквания, Късни изисквания, Архитектурен дизайн, Подробен дизайн
GAIA	Агент, Услуги, Познати, Роли, Организация, Взаимодействия	спецификации подобни на карти, които са текстови описания на това, което се моделира	Не специфицира	Анализ на високо ниво Дизайн на високо ниво Класически обектно-ориентирани техники за дизайн
MESSAGE/ UML	Агент, Организация, Роля, Ресурси, Задача, Взаимодействие, Протокол на взаимодействие, Цели, Информационен обект, Съобщение	UML	AUML - Extended UML for Agent Modelling	Анализ на високо ниво, Архитектурен дизайн
BDI	Агент, Вярвания, Желания, Намерения	DSML4BDI - domain-specific modeling language	Jason BDI	Външно ниво на абстракция, Вътрешно ниво на абстракция

3.3. Проектиране на система за извличане и обработка на информация за емоционално състояние

Методологията избрана за разработване на предложената в дисертационния труд мултиагентна система за извличане и обработка на информация за емоционално състояние е базирана на множество модели.

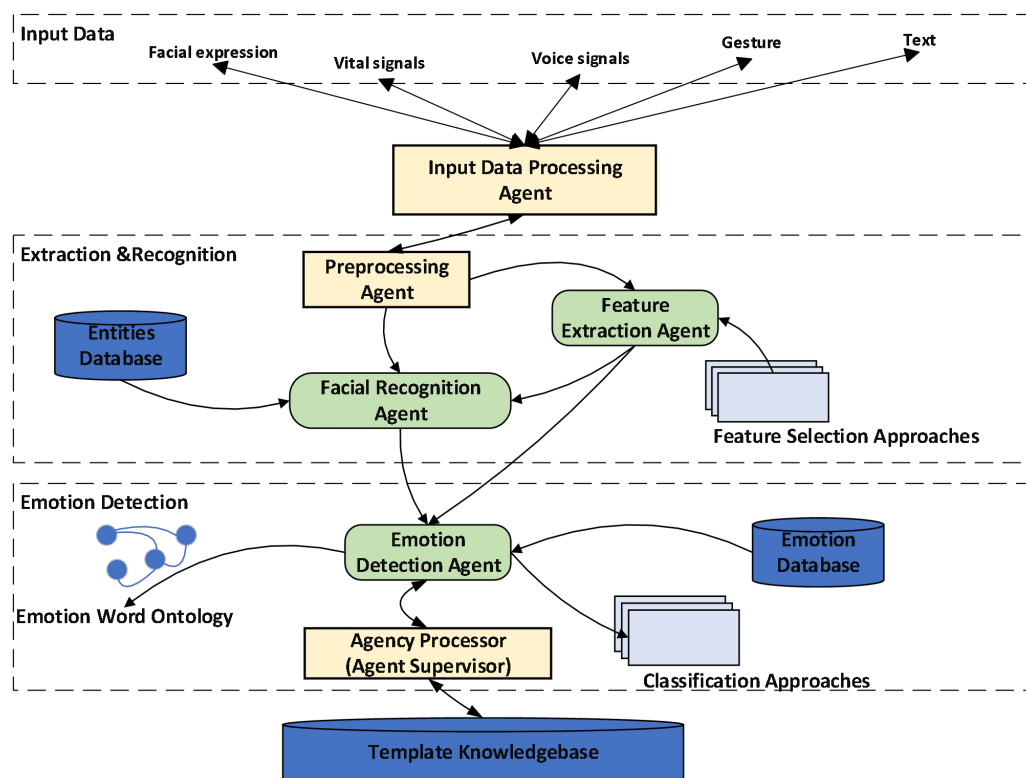


Фиг. 3-9 Основни абстракции на мета модела Агент

Основните абстракции на системата включват:

- Агент - този мета модел центрира вниманието си на основните функционалности на агента и неговия контрол. Описва конкретни агенти и техните състояния, през целия цикъл на живота им;
- Желания - това са целите, към които се стреми агента. Генерират план за действие;
- Намерения (план за действие) - групира конкретните задачи, които е способен да извършва агента. Попълват се при възникването на желания. В контекста на определена роля, са групирани повтарящи се функционалности и свойства на група агенти, например: за агентите от организация Input_data, агент може да играе роля на филтър;
- Роля - временна външна характеристика на агента, която позволява той да изпълнява услуги в зависимост от външни за него условия.
- Поведение - с цел да се добавят възможности за обучение на агента управлението на поведението се извършва с помощта на неговите състояния като тези емоционални състояния могат да бъдат динамично променяни с течение на времето. Поведенията могат да включват набор от процедури, предаване на съобщения между агенти или преходи в машина на състояния;
- Ресурси - използва се за представяне на неавтономни обекти като бази данни или външни програми, използвани от агентите;
- Информационна единица - капсулиран обект информация, активира възприятие, влияе на емоционалния контрол;
- Възприятия - всичко което идва от средата, активира емоционален контрол

Системната инфраструктура за разпознаване на емоции е базирана на многослойна мултимодална архитектура с мобилни интелигентни агенти, където интелигентните агенти изпълняват алгоритми за откриване на емоции с използване на входни данни от различни източници и в различни формати.



Фиг. 3-10 Основни компоненти и комуникации в агент-базирана архитектура на система за извличане и обработка на информация за емоционално състояние от различни източници

Система може да се раздели на три основни модула:

- Обработка на входните данни (Input_data). Този модул има за цел обработката на получените от околната среда входни данни. Входните данни могат да бъдат описани с различни файлове. Модулът за обработка на входните данни трябва да идентифицира получените от обкръжението данни и да ги обработва. Изходът от този модул са филтрирани, нормализирани и етикетирани данни. Етикетите на данните са свързани с типа информация която съдържат.
- Извличане на характеристики и разпознаване (Extraction_and_Recognition). Модулът получава данни от предходния модул и в зависимост от идентификационния етикет избира метод за извличане на характеристики. Обработва получените данни до генериране на характеристичен вектор. Този характеристичен вектор се използва за разпознаване на определен индивид. Ако такъв индивид не съществува, характеристичния вектор се записва в База за индивиди. Изходът от този модул е набор характеристични вектори с етикет за личност.
- Разпознаване на емоции (Emotion_Extraction) - Модулът получава характеристичните вектори от предишния модул и ги подава на подходящ класификатор за разпознаване на емоция. Изходът от този модул са записи за моментното състояние на изследвания индивид.

Системата съдържа бази данни с визуална, звукова и физиологична информация (неутрални индивидуални шаблони) за всеки един от изследваните субекти. Инициализира се със събитието постъпване на информационна единица на входа. Детекторите генерират информационни единици - външни събития, базирани на околната среда. Тези събития се добавят към опашката на събития. Ролята на AgentFactory е идентифицирането, извличането и начална обработка на данните. В зависимост от вида данни се инициализира съответния агент - от организация InputDataPreprocessing.

Агентите са определени с техните свойства: мярка за ефективност, околна среда, изпълнителни механизми, сензори.

Табл. 3-3 Агент за извличане на характеристики от изображения

FaceInputDataPreProcessingAgent	
Ефективност	Правилно откриване на шаблон от изображение; преобразуване във вектор; избиране на най-подходящ модел; автоматично открива лица; правилно изрязване на лице; подравняване; идентифициране или регистриране: време, модалност, идентификационен номер, години, пол, раса
Среда	Частично наблюдавана; сохастична, динамична, мулти агентна
Изпълнители	DeepFace
Сензори	Изображение, съобщение

При постъпване на информационна единица звуков сигнал, AgentFactory инициализира VoiceInputDataPreProcessingAgent, но не преди да е получено съобщение от FaceInputDataPreProcessingAgent, за разпознат обект на изследване

Табл. 3-4 Агент за извличане на характеристики от звуков сигнал

VoiceInputDataPreProcessingAgent	
Ефективност	Правилно откриване на гласова проба; Компенсиране и усилване на високочестотната част на сигнала; Сегментиране и блокиране на кадри; Получаване на характеристичен вектор. време, модалност, идентификационен номер
Среда	Частично наблюдавана; сохастична, динамична, мулти агентна
Изпълнители	Високочестотен филтър; Hamming / Nap прозорец; Бързо преобразуване на Фурие;
Сензори	характеристичния вектор от гласова проба; съобщение

За разлика от звуковия сигнал, наличието на първоначални физиологични данни за всеки изследван обект е задължително, тъй като физиологичните сигнали са силно чувствителни към човека, климата и момента от деня. Целта на агента се състои да открие положителни или отрицателни вариации от предварително измервани физиологични характеристики.

Таблица 3.5Агент за извличане на характеристики от физиологични данни

PhysiologicalInputDataPreProcessingAgent	
Ефективност	Получава вектор стойности от суровия физиологичен сигнал; Нормализира; Сегментира и блокира кадрите; Генерира на характеристичен вектор. време, модалност, идентификационен номер
Среда	Частично наблюдавана; сохастична, динамична, мулти агентна
Изпълнители	BiTalino OpenSignals или друг;
Сензори	PsychoBIT, MuscleBIT; характеристичен вектор; съобщение

Агентът за извличане на емоции се активира при получаване на информационна единица съобщение от агентите от организация InputDataPreprocessing. Целта на този агент е да извлече информация за емоцията на изследвания обект.

Таблица 3.6 Агент за извличане на емоции

EmotionDetectionAgent	
Ефективност	Правилно разпознаване на емоции от изображение; Правилно разпознаване на емоции от глас; Генериране на теглови коефициенти от физиологични сигнали; време, идентификационен номер, години, пол, раса, емоция
Среда	Частично наблюдавана; сохастична, динамична, мулти агентна
Изпълнители	Невронна мрежа или друг класификатор
Сензори	Изображение, характеристичния вектор от гласова проба; характеристични изменения; съобщение

Всеки модул за откриване на емоции получава на изхода си списък с откритите емоции, грешки и процент на кореспонденция. Тези изходи са входовете за AgencyProcessor, чиято задача е да ги комбинира и балансира, според типа модалност, пол, възраст и етническа група.

Ролята на AgencyProcessor е да балансира получените резултати с подходящите тегла в зависимост от входните параметри, пола, възрастта и етническите характеристики.

$$E_t = F[(i_0 w_0 + \varepsilon_0) \dots (i_n w_n + \varepsilon_n)] * p$$

3.4 Изводи към трета глава

По важните резултати и изводи от моделирането на представената в трета глава мултимодална мултиагентна система за извличане и обработка на информация за емоционално състояние са следните:

- Съществуват разнообразни методологии за проектиране на агент-базирани системи, които обхващат различни етапи и аспекти от жизнения цикъл на създаване на този клас системи;
- За проектиране на мултимодална и мултиагентна система за извличане и обработка на информация за емоционално състояние следва да се използва методология, която да позволява нейното разширяване и адаптация;
- Предложената мултимодална архитектура на система за извличане и обработка на информация за емоционално състояние се състои от отделни модули и автономни интелигентни агенти;
- Предложени са абстракции и са описани основните модели за тази система;
- Мултиагентната система позволява да се обособят независими компоненти, които изпълняват паралелно процесите за извличане на емоции от трите модалности. По този начин се получава гъвкава архитектура, която увеличава бързодействието при обработката на информацията.

ГЛАВА 4. Анализ на експериментални резултати

Експерименталната оценка на предложената мултиагентна архитектура е базирана на изследване на производителността на отделните агенти за извличане на характеристики и класификация на визуални, звукови и физиологични данни. За експерименталната оценка на производителността на агентите за извличане на характеристики от визуални, звукови и физиологични данни, както и на агента за извличане на емоции са използвани различни множества с обучаващи и тестови данни.

4.1 Опитна постановка

За извършване на експериментите е използвана опитна постановка, изградена от следните хардуерни и софтуерни компоненти: Мини компютър Nvidia Jetson Nano, който се захранва от четириядрен ARM A57 процесор с работна честота от 1.43GHz и 128-ядрен NVIDIA Maxwell графичен чип, допълнени от 4GB RAM. Общата производителност достига до 472 GFLOPS; Лаптоп Toshiba Satellite - Intel Core I7, 4 GB RAM, 240GB SSD; Лаптоп Raizer 5000 - Intel Core I7-10750H Processor 6core/12, GeForce GTX 1660Ti 6GB GDDR6 VRAM; Wi-Fi Bluetooth модул AC8265 за Jetson Nano - 2.4GHz / 5GHz, 300Mbps / 867Mbps, WiFi protocol: 802.11ac, Bluetooth версия: 4.2, NIC interface: NGFF (M.2) + 2 антени; Камера Raspberry Pi Camera Board V2.1, Интерфейс: CSI (Camera Serial Interface), Sony IMX219, 8MP сензор, фиксиран фокус, разделителна способност: за изображения : до 3280 x 2464 пиксела, за видео: 1080p30, 720p60 и 640x480p60/90; Камера Logitech C920 Pro HD WebCam, Интерфейс: USB Video Device Class (UVC), автоматичен фокус, разделителна способност: 1080p30, 720p30, вграден стерео микрофон; BITalino (r)evolution Starter Kits - HeartBIT, MuscleBIT, PsychoBIT; Външен твърд диск Seagate Expansion 2.5, 1 TB, 5400 rpm, 8 MB, USB 3.0, модел: STEA1000400; Операционни системи: Ubuntu 18.04 LTS 64-bit, върху карта памет SONY SF32U 100 MB/s Micro SD - 32GB; Windows 10 10.0 64-Bit; Софтуерни среди и библиотеки: IDE - PyCharm (Community Edition) 64-BitPython 3.6.9 64-bit; OpenCV 4.1.1.

4.2 Използвани множества данни

За експерименталната оценка на производителността на агентите за извличане на характеристики от визуални, звукови и физиологични данни, както и на агента за извличане на емоции са използвани няколко множества с обучаващи и тестови данни. На табл. 4-2 са обобщени множествата данни използвани за извличане на емоционална информация от звук.

Табл. 4-2 Множества данни използвани за извличане на емоционална информация от звук

Множество данни	RAVDESS	SAVEE	TESS	ОБЩО
Брой файлове	1440	480	2800	4720

За извличане на емоционална информация от изображения са използвани VGG Face Dataset – е наборът от данни, който се състои от изображения за 2 622 личности и FER-2013 dataset, се състои от 34 034 етикетирани изображения.

За извличане на емоционална информация от физиологични данни са използвани Stress Recognition in automobile Drivers, съдържа колекция от мултипараметрични записи от здрави доброволци, направени по време на шофиране, CLAS - База данни за когнитивно натоварване, въздействие и стрес. DEAP - съдържа мултимодални записи (видео и физиологични) на 32-ма здрави участници.

4.3 Експериментални резултати от извличане на информация за емоционално състояние

За оценка на качеството и производителността на разпознаването на емоции с различни класификатори, внедрени в агентите, са използвани следните метрики:

- Точност (accuracy) - ако $\hat{y}_i$ е предсказаната стойност на i -тата извадка, и y_i е съответната реална стойност, тогава частта от правилните прогнози, върху n извадки се определя със следната формула:

$$(4.1) \quad accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i)$$

- Пълнота (recall) - определя способността на класификатора да намери всички положителни проби. Изчислява се с формулата:

$$(4.2) \quad recall = \frac{t_p}{t_p + f_n}$$

където t_p е броят на положителни резултати и f_n броят на отрицателните резултати. Най-добрата възможна стойност е 1, а най-лошата - 0.

- Прецизност (precision) - изчислена средна точност между резултатите и прогнозата. Обобщава кривата на прецизност-пълнота като средно претеглена точност, постигната при всеки праг като увеличението на пълнотата от предишния праг се използва за тегло:

$$(4.3) \quad AP = \sum_n (R_n - R_{n-1}) P_n$$

където: P_n и R_n са прецизност и пълнота при n -тия праг.

- f1-мярка (f1-score) - метриката е известна също като балансиран f1-резултат. Може да се интерпретира като средно претеглена стойност на прецизността и пълнотата, като

f1-мярката достига най-добрата си стойност при 1 и най-лошата при 0. Изчислява се по следната формула:

$$(4.4) \quad F1 = 2 \frac{AP * R}{AP + R}$$

В случая на класове с много етикети това е средната стойност на резултата f1 за всеки клас с претегляне в зависимост от средния параметър.

- Матрица на грешките (Confusion Matrix) (Фиг. 4-3) - оценява точността на класификацията. Представлява таблица с редове и колони, която отчита броя на фалшивите положителни, фалшиви отрицателни, истински положителни и истински отрицателни резултати.
- Време за изпълнение (wall clock time) – определя се чрез измерване на пълното изчислително време за изпълнение на теста: изминалото време между момента, в който процеса започва да се изпълнява и момента, в който е завършен.

За извличане на информация за емоционално състояние чрез VoiceInputDataPreProcessingAgent са проведени експерименти с използване на комбинации от различни класификатори, характеристики и множества данни.

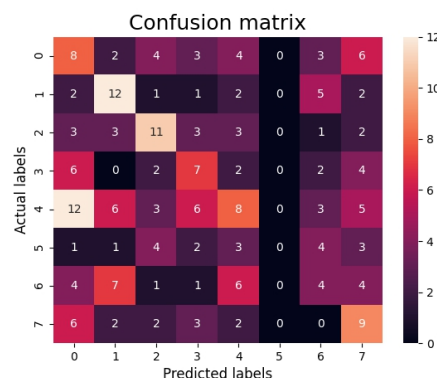
За първоначалните експерименти е използвана е мултимодалната база данни RAVDESS и класификатор Random Forest. Получените експериментални резултати показват, че кепстър коефициентите са напълно достатъчни при използването на този класификатор (табл. 4-3). Резултатите за оценката на качеството на класификация при използване на този класификатор с 40 MFCC характеристики за множеството RAVDESS за различните емоции показват, че най-голяма точност се получава за емоциите „страх“ и „гняв“ (табл. 4-4, фиг. 4-4).

Табл. 4-3 Резултати за точност на класификация на емоции от звукови данни с различни характеристики (RAVDESS - RF)

Модел	Характеристики	Точност %	Train/Test
RF Classifier	LPC -> 17	27.31%	1224, 216
RF Classifier	MFCC -> 40	45.83%	1224, 216
RF Classifier	LPC + MFCC -> 73	45.83%	1224, 216
RF Classifier	LPC+MFCC+MEL ->185	41.67%	1224, 216
RF Classifier	LPC+MFCC+MEL+CROMA -> 197	42.13%	1224, 216

Табл. 4-4 Резултати за качество на класификация за различните емоции от звукови данни за MFCC + RAVDESS + RF

Emotion	Precision	Recall	f1-score	Support
angry	0.52	0.53	0.52	30
calm	0.45	0.96	0.62	25
disgust	0.46	0.65	0.54	26
fearful	0.57	0.70	0.63	23
happy	0.40	0.14	0.21	43
neutral	0.0	0.0	0.0	18
sad	0.47	0.33	0.39	27
surprised	0.33	0.46	0.39	24



Фиг. 4-4 Voice RF Confusion matrix

С използване на класификатор KNN точността на класификацията на емоции от звукови данни е по-добра в сравнение с резултатите при използване на RF класификатора за същите характеристики и множество данни. В сравнение с RF, при използване на KNN тестовата фаза на класификацията е по-бавна. Този класификатор би могъл да доведе до добри резултати само с комбинирането на MFCC и MEL (табл. 4-5).

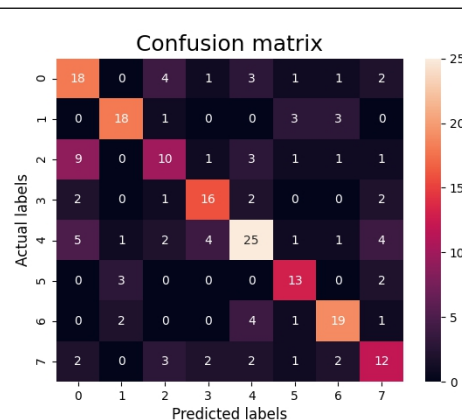
Табл. 4-5 Резултати за точност на класификация на емоции от звукови данни с различни характеристики (RAVDESS + KNN)

Модел	Характеристики	Точност %	Train/Test
KNN Classifier	LPC -> 17	29.17%	1224, 216
KNN Classifier	MFCC -> 40	60.65%	1224, 216
KNN Classifier	LPC + MFCC -> 73	60.65%	1224, 216
KNN Classifier	LPC+MFCC+MEL ->185	61.57%	1224, 216
KNN Classifier	LPC+MFCC+MEL+CROMA -> 197	61.57%	1224, 216

Получените експериментални резултати за качеството на класификация за различни емоции от звукови данни (Табл. 4-6, фиг. 4-5) показват, че най-добре се откриват емоциите „спокойствие“ и „тъга“ с точност съответно от 75% и 70%.

Табл. 4-6 Резултати за качество на класификация за различни емоции от звукови данни за LPC+MFCC+MEL+CROMA, RAVDESS+KNN

Emotion	Precision	Recall	f1-score	Support
angry	0.50	0.60	0.55	30
calm	0.75	0.72	0.73	25
disgust	0.48	0.38	0.43	26
fearful	0.67	0.70	0.68	23
happy	0.64	0.58	0.61	43
neutral	0.62	0.72	0.67	18
sad	0.70	0.70	0.70	27
surprised	0.50	0.50	0.50	24



Фиг. 4-5 Voice KNN Classifier

Експерименталните резултати показват, че с използване на класификатор Gaussian NB точността на класификацията на емоции от звукови данни е по-ниска в сравнение с класификаторите RF и KNN. Получените експериментални резултати, представени в табл. 4-8, показват, че точността на класификация и за трите използвани класификатора се повишава при увеличаване на използваните обучаващи и тестови данни.

Табл. 4-7 Резултати за точност на класификация на емоции от звукови данни с различни характеристики (RAVDESS - Gaussian NB)

Модел	Характеристики	Точност %	Train/Test
Gaussian NB	LPC -> 17	23.15%	1224, 216
Gaussian NB	MFCC -> 40	35.19%	1224, 216
Gaussian NB	LPC + MFCC -> 73	36.11%	1224, 216
Gaussian NB	LPC+MFCC+MEL ->185	32.41%	1224, 216
Gaussian NB	LPC+MFCC+MEL+CROMA -> 197	34.26%	1224, 216

Табл. 4-8 Резултати за точност на класификация на емоции от звукови данни с различни характеристики (RAVDESS + SAVEE + TESS)

Модел	Характеристики	Точност %	Train/Test
RF Classifier	MFCC -> 40	75.53%	4001, 707
KNN Classifier	LPC+MFCC+MEL ->185	84.02%	4001, 707
Gaussian NB	LPC + MFCC -> 73	57.57%	4001, 707

Получените експериментални резултати показват най-висока точност при класификация на емоции от звукови данни с SVM при използване на характеристиките от MFCC-40, които водят до най-добра класификация и при RF. Точността с SVM е 84.15%, при това добавянето на други характеристики не подобрява точността на класификация.

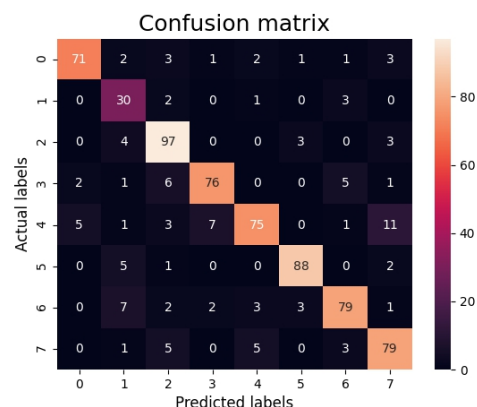
Табл. 4-9 Резултати за качество на класификация за различни емоции от звукови данни за RAVDESS + SAVEE + TESS с SVM

Модел	Характеристики	Точност %	Train/Test
SVM Classifier	LPC -> 17	40.02%	4001, 707
SVM Classifier	MFCC -> 40	84.15%	4001, 707
SVM Classifier	LPC +MFCC -> 57	83.30%	4001, 707
SVM Classifier	LPC +MFCC+MEL -> 185	80.06%	4001, 707
SVM Classifier	LPC +MFCC+MEL+CROMA -> 197	80.20%	4001, 707

Резултатите за качеството на класификация за различни емоции от звукови данни с класификатор SVM (табл. 4-10, фиг. 4-6) показват, че най-голяма точност е постигната за разпознаване на неутрално емоционално състояние – 93 %, достигнатата точност за емоцията „гняв“ е 91 %, а най-ниска точност се получава за спокойно емоционално състояние, което може да се обясни с оскъдните обучаващи данни за тази категория – само 36 гласови файла

Табл. 4-10 Резултати за качество на класификация за различни емоции от звукови данни с SVM

Emotion	Precision	Recall	f1-score	Support
angry	0.91	0.85	0.88	84
calm	0.59	0.83	0.69	36
disgust	0.82	0.91	0.86	107
fearful	0.88	0.84	0.86	91
happy	0.87	0.73	0.79	103
neutral	0.93	0.92	0.92	96
sad	0.86	0.81	0.84	97
surprised	0.79	0.85	0.82	93



Фиг. 4-6 Voice SVM Confusion matrix

Използваният невронно-мрежов класификатор Multy Layered Perceptron MLP се базира на оптимизиране на функцията за регистрация на загуби, използвайки стохастично градиентно спускане. Използвани са 4 001 обучаващи примера и 707 тестови примера от обединените множества данни, формирано от RAVDESS, SAVEE и TESS.

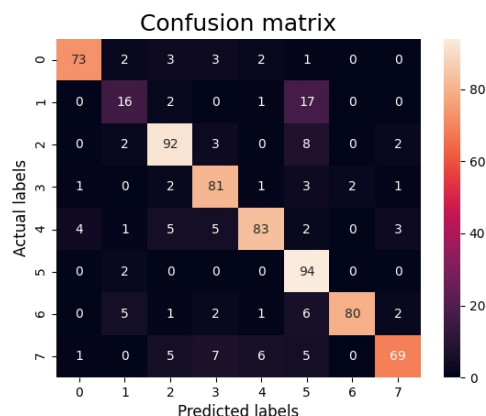
Оценката на качеството на класификация при използване на същия класификатор с 192 характеристики сред които, линейни прогнозни коефициенти и кепстър коефициенти на честотите на мел. Получените резултати показват високи стойности на точност представят емоции като: “тъга” и “гняв”.

Табл. 4-11 Резултати за точност на класификация на емоции от звукови данни с различни характеристики (RAVDESS + SAVEE + TESS) MLP

Модел	Характеристики	Точност %	Train/Test
MLP Classifier	LPC -> 16	62.80%	4001, 707
MLP Classifier	MFCC -> 40	80.62%	4001, 707
MLP Classifier	LPC +MFCC -> 57	80.20%	4001, 707
MLP Classifier	LPC +MFCC+MEL -> 185	80.20%	4001, 707
MLP Classifier	LPC+MFCC+MEL+CROMA ->192	83.17%	4001, 707

Табл. 4-12 Резултати за качество на класификация за различни емоции от звукови данни с MLP Classifier

Emotion	Precision	Recall	f1-score	Support
angry	0.92	0.87	0.90	84
calm	0.57	0.44	0.50	36
disgust	0.84	0.86	0.85	107
fearful	0.80	0.89	0.84	91
happy	0.88	0.81	0.84	103
neutral	0.69	0.98	0.81	96
sad	0.98	0.82	0.89	97
surprised	0.90	0.74	0.81	93



Фиг. 4-7 Voice MLP Confusion matrix

Направените опити за построяване и обучение на 1D CNN от 3 конволюционни и един изходен слой с използване на същите характеристики и множества данни, както в експеримент 5 не води до получаване на по-добра точност на класификация на емоции от звукови данни в сравнение с резултатите от експеримент 5 при използване на класификатор MLP. В същото време обучението на едномерна конволюционна НМ изисква многократно повече време в сравнение с обучението на MLP класификатора.

Конволюционните НМ са особено подходящи за класификация и разпознаване на данни в изображения и звук. Затова за извличане на информация за емоционално състояние чрез FaceInputDataPreProcessingAgent са проведени експерименти с имплементирани класификатори с конволюционни невронни мрежи с различни архитектури.

За извличане на информация за емоционално състояние от изображения първоначално бе използвано множеството данни FER-2013, като наличните данни най-напред са нормализирани. Използвани са 28 709 обучаващи примера и 3 589 тестови примера. Проектирани са конволюционни НМ с различни архитектури, с които е направена оценка на разпознаването на емоции за данните от FER-2013.

От всички проектирани конволюционни НМ най-добри резултати за извличане на емоционална информация от изображения са получени за архитектура със: 4 слоя, 3 конволюционни и един пълно свързан. Времето за обучение на НМ с представената структура с 64 характеристики от обучаващите изображения е 6 часа. Получената точност на класификация на емоции в тестовите изображения е 67.08 %, което е твърде незадоволителен резултат.

Табл. 4-13 Структура на НМ с най-добри получени резултати

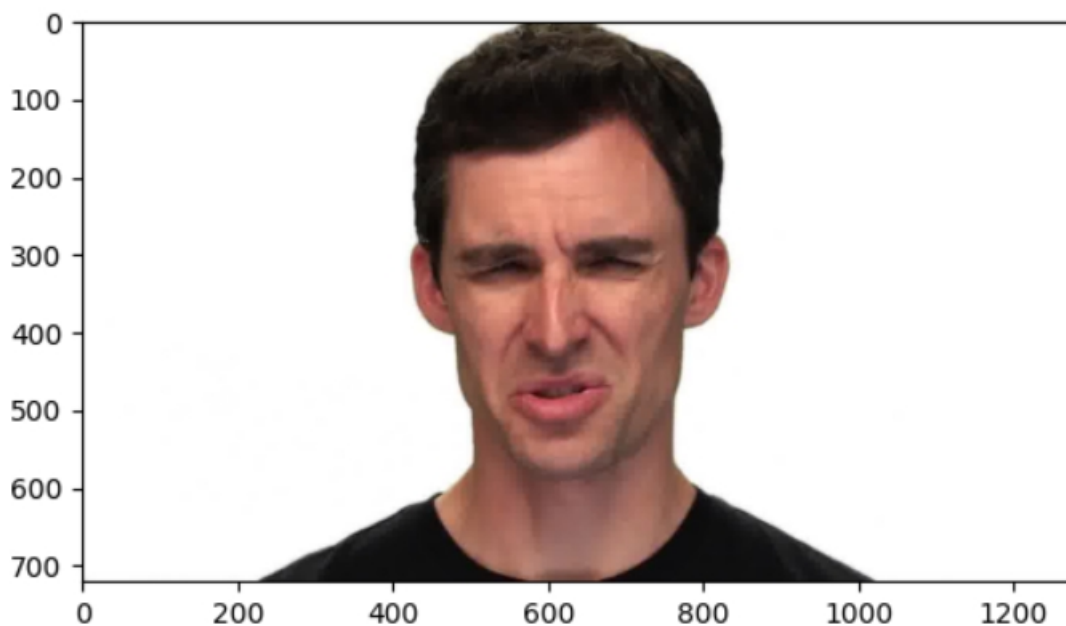
Архитектура на НМ; model: "sequential"		
Слой (тип)	Вид на изходния резултат	Брой параметри
conv2d (conv2D)	(None, 46, 46, 64)	640
conv2d_1 (conv2D)	(None, 44, 44, 64)	36928
conv2d_2 (conv2D)	(None, 42, 42, 32)	18464
max_pooling2d (MaxPooling2D)	(None, 21, 21, 32)	0
dropout (Dropout)	(None, 21, 21, 32)	0
conv2d_3 (Conv2D)	(None, 19, 19, 64)	18496
conv2d_4 (Conv2D)	(None, 17, 17, 64)	36928
conv2d_5 (Conv2D)	(None, 15, 15, 64)	36928
max_pooling2d_1(MaxPooling2D)	(None, 7, 7, 64)	0
dropout_1 (Dropout)	(None, 7, 7, 64)	0
conv2d_6 (Conv2D)	(None, 5, 5, 128)	73856
conv2d_7 (Conv2D)	(None, 3, 3, 128)	147584
max_pooling2d_2 (MaxPooling2D)	(None, 1, 1, 128)	0
flatten (Flatten)	(None, 128)	0
dense (Dense)	(None, 1024)	132096
dropout_2 (Dropout)	(None, 1024)	0

dense_1 (Dense)	(None, 1024)	1049600
dropout_3 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 7)	7175

За разпознаването на емоции от изображения е използвана хибридна рамка за разпознаване на лица DeepFace, публикувана в края на 2020 година. Точността на класификация на използваните модели на дълбоки невронни мрежи: VGG-Face, Google FaceNet, Facebook DeepFace, OpenFace, DeepID, ArcFace и Dlib, които са предварително обучени за класифициране на изображения на лица с огромен набор от данни, достигат и преминават точността на човешко възприемане.

Показаните резултати са от тестови примери от множеството данни RAVDESS (видео данни), където наблюдаваме една изключително висока точност от приблизително 97% на получените резултати за разпознаване на емоции с положителна валентност. За примера от фиг. 4-9 и табл. 4-14 точността е 97.35 за емоция „отвращение“ .

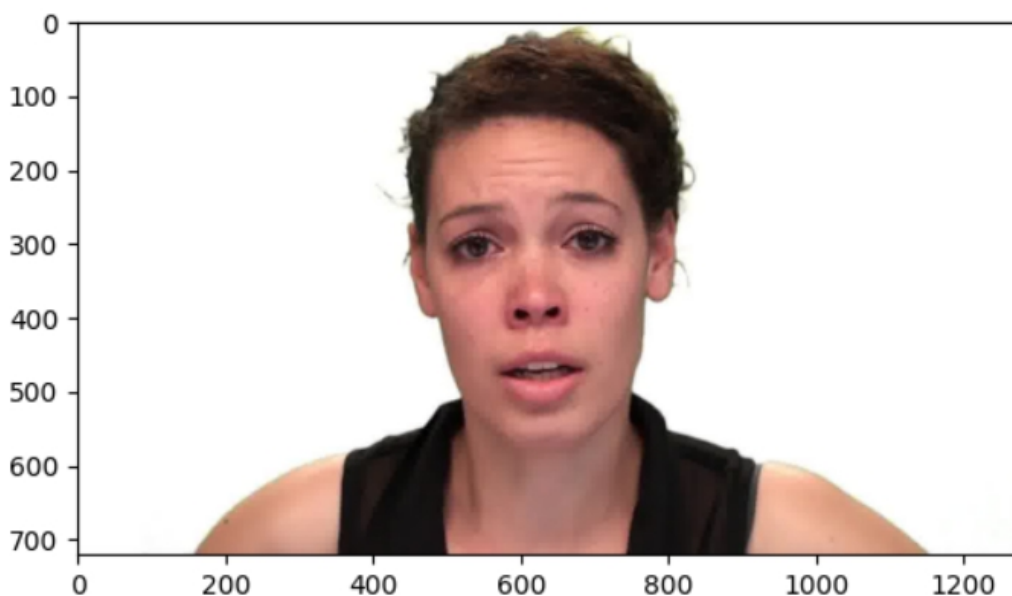
В същото време точността при разпознаване на емоции с отрицателна валентност с използването на тази дълбока НМ, обучена с огромен набор от данни, е по-ниска: за показаните примери точността за емоция „тъга“ е 7.27 за примера от фиг. 4-11 (табл. 4-16)



Фиг. 4-9 Актьор 1 от RAVDESS база данни, аотиран като отвращение

Табл. 4-14 Резултати от класификацията на тестови пример от фиг. 4-9 с дълбока НМ имплементирана с DeepFace

Характеристика	Вероятност
'emotion':	
'angry'	2.0282615531398718
'disgust'	97.35476939139001
'fear'	0.4012733710585083
'happy'	8.330648453843199e-05
'sad'	0.1897003727794138
'surprise'	0.016114891797961807
'neutral'	0.00979655422639003
'dominant_emotion':	'disgust'
'age':	24
'gender'	'Man'
'race':	
'asian'	0.11690548853948712
'indian'	1.1991975829005241
'black'	0.023343769134953618
'white'	58.16018581390381
'middle eastern'	25.403907895088196
'latino hispanic'	15.096458792686462
'dominant_race':	'white'



Фиг. 4-11 Актриса 4 от RAVDESS база данни, аотирана като тъжна

Tabla 4-16 Резултати от класификацията на тестови пример от фиг. 4-11 с дълбока НМ имплементирана с DeepFace

Характеристика	Вероятност
'emotion':	
'angry'	0.7239791564643383
'disgust'	2.1560937568665395e-05
'fear'	19.002722203731537
'happy'	0.14983981382101774
'sad'	7.272630929946899
'surprise'	0.03667088458314538
'neutral'	72.81413674354553
'dominant_emotion':	'neutral'
'age':	31
'gender'	'Woman'
'race':	
'asian'	1.483749132603407
'indian'	0.5813665688037872
'black'	0.05365653778426349
'white'	58.36069583892822
'middle eastern'	7.6199449598789215
'latino hispanic'	31.900587677955627
'dominant_race':	'white'

Използването на НМ изисква дълъг процес на проектиране и избор на хиперпараметри за определяне на подходяща архитектура на НМ. Процесът на обучение на конволюционни и дълбоки НМ изисква модифицирането на много голям брой теглови параметри на модела на НМ с използване на множества с голям брой обучаващи и тестови примери, което води до много дълго време за обучение. Резултатите при използването на предварително обучени модели на дълбоки НМ при разпознаването на емоции с отрицателна валентност от изображения водят до неправилна класификация и незадоволителна точност на разпознаването.

За извличане на информация за емоционално състояние чрез PhysiologicalInputDataPreProcessingAgent са имплементирани следните класификатори: Random Forest, Support Vector Machine, Multi-Layer Perceptron, K-Nearest Neighbors. Използвани са 1332 обучаващи примера и 1200 тестови примера.

Получените резултати от тези експерименти са показани в табл. 4-18. Класификацията на емоциите се базира на характеристики за валентност (удоволствие – недоволство) и възбуда (позитивна – негативна). Експерименталните резултати показват много ниска точност на класификация на емоции с използване на физиологични данни и за двете характеристики валентност и възбуда.

Табл. 4-18 Резултати за точност на класификация на емоции от физиологични данни с различни класификатори

Модел	Характеристики	Точност %
RF Classifier	Valence	43.98%
KNN Classifier	Valence	50.18%
MLP Classifier	Valence	44.01%
SVM Classifier	Valence	36.18 %
RF Classifier	Arousal	71.15 %
KNN Classifier	Arousal	64.80 %
MLP Classifier	Arousal	43.15 %
SVM Classifier	Arousal	35.71 %

Физиологичните данни зависят твърде много от специфичните човешки характеристики и от околната среда. Използването на различни по вид и структура датчици може да доведе до грешки в измерванията и изисква датчиците да бъдат предварително внимателно калибрирани.

За създаването на система за ранно диагностициране на депресивни разстройства не е достатъчно използването на данни само от една модалност. Система от този вид трябва да използва различни входни данни.

За експериментална оценка за извличане на информация за емоционално състояние от мултимодални данни (EmotionDetectionAgent) е използвано множеството данни RAVDESS, което съдържа анотирани звукови данни и видео последователности.

Получените резултати показват, че при разпознаването на емоции с отрицателна валентност в много от случаите се получават различни резултати за класификацията при използване на данни от различните входни модалности, което обосновава необходимостта от използването на AgencyProcessor за комбиниране на резултатите от класификацията. В този случай след комбинацията доминантната емоция е “тъжен”, което съответства на анотацията.



Фиг. 4-16 Актьор 7 от RAVDESS база данни, аотиран като тъжен

Табл. 4-19 Резултати от класификацията на тестови пример от фиг. 4-13 с дълбока НМ имплементирана с DeepFace

Характеристика	Изображение	Звук
'emotion':		
'angry'	76.27641275839957	0.00
'disgust'	4.713425229600099	2.96
'fear'	6.8846129637495235	3.01
'happy'	0.009699108042042543	0.00
'sad'	8.538452063614375	93.72
'surprise'	0.012773089178330652	0.10
'neutral'	3.564634226516989	0.31
'dominant_emotion':	'angry	'sad'
'age':	20	
'gender'	'Man'	
'race':		
'asian'	0.0012520227755885571	
'indian'	0.004218575122649781	
'black'	2.7061983587373106e-05	
'white'	95.30953764915466	
'middle eastern'	3.1040534377098083	
'latino hispanic'	1.5809131786227226	
'dominant_race':	'white'	

Проведените експериментални резултати с мултимодални данни показват, че при разпознаването на емоции с отрицателна валентност в много от случаите се получават различни резултати за класификацията при използване на данни от различните входни модальности, което обосновава необходимостта от използването на EmotionDetectionAgent и AgencyProcessor за комбиниране на резултатите от класификацията.

4.4 Изводи към четвърта глава

По-важните резултати и изводи от проведеното експериментално изследване за извличане на информация за емоционално състояние с използване на различни класификатори, имплементирани в агентите за първоначална обработка на входен сигнал от изображения, звук и физиологични данни са следните:

- Производителността и точността на класификация на всеки един създадените модели са зависими и силно се влияят от използваните характеристики, използваните обучаващи данни и параметрите на модела;
- С използване на единствен входен поток от данни не може да се постигне висока точност на класификация, особено по отношение на откриване на емоциите с негативна валентност, които се проявяват при депресивни състояния;
- Комбинирането на резултатите от различните агенти за откриване на емоции позволява постигане на независимост на извличането на информация за емоционално състояние от наличните входни данни и повишаване на точността на разпознаване;
- За ефективно и ефикасно ранно диагностициране на депресивни разстройства чрез неинвазивни методи, базирани на извличане на информация за емоционално състояние на човека от входен поток с вербални, визуални и физиологични данни може да се приложи подход, базиран на предложената мултимодална мултиагентна системна архитектура за извличане и обработка на информация за емоционално състояние.
- Предложената мултимодална мултиагентна системна архитектура за извличане и обработка на информация за емоционално състояние на човека може да постигне ефективно и ефикасно ранно диагностициране на депресивни разстройства чрез неинвазивни методи, базирани на извличане на информация за емоционално състояние от входен поток с вербални, визуални и физиологични данни.

НАУЧНО-ПРИЛОЖНИ И ПРИЛОЖНИ ПРИНОСИ

Научни приноси:

- Изследвани са методи и алгоритми за извличане на характеристики и класификация на емоции с използване на вербални, визуални и физиологични данни, които да повишат точността при откриването на депресивни състояния.
- Предложени са абстракции и са описани основните модели на мултимодална мултиагентна система за извличане и обработка на информация за емоционално състояние;
- Предложен е модел на агент за комбиниране на резултатите от различни агенти за откриване на емоции, който позволява балансиране на данните при обединяването им според типа модалност, пол, възраст и етническа група;

Научно-приложни приноси:

- Предложени са подходи за първоначална обработка, извличане на характеристики и класификация на входен сигнал от изображения, звук и физиологични данни, както и на агент за извличане на информация за емоционално състояние;
- Предложена е архитектура на мултимодална мултиагентна система за извличане и обработка на информация за емоционално състояние, която позволява независимост на извличането на информация за емоционално състояние от наличните входни данни с вербални, визуални и физиологични данни и повишаване на точността на разпознаване;

Приложни приноси:

- Проектирана е мултимодална мултиагентна система за извличане и обработка на информация за емоционално състояние;
- На базата на експериментални изследвания е оценена точността на класификация на избраните методи и алгоритми за класификация на емоции при използване на единствена модалност и е доказана необходимостта от използване на предложената мултимодална мултиагентна система за извличане и обработка на информация за емоционално състояние, особено по отношение на емоциите с негативна валентност, които се проявяват при депресивни състояния

СПИСЪК НА ПУБЛИКАЦИИТЕ ПО ДИСЕРТАЦИОННИЯ ТРУД

1. **R. Raynova**, M. Lazarova, A. Aleksieva-Petrova, Affective Computing and Emotion Detection Through Image Analyses, Proc. of 8th International Scientific Conference Computer Science'2018, Kavala, Greece, 13–15 Sept. 2018, pp. 19–25, ISBN: 978-619-167-177-9.
2. **R. Raynova**, M. Lazarova, A. Aleksieva-Petrova, Emotion Detection Through Image Analyses, Journal "Computer and Communications Engineering", Vol. 12, No. 1, 2018, pp. 36–45, ISSN 1314-2291.
3. **Р. Райнова**, М. Лазарова, А. Алексиева-Петрова, Мултиагентна мултимодална архитектура за разпознаване на емоции, International Conference "Automatics and Informatics - 2019", 3-5 Oct. 2019, Sofia, Bulgaria, pp. 187–190, ISSN 1313-1850, CD: ISSN 1313-1869.
4. **R. Raynova**, A. Aleksieva-Petrova, Agent Based Emotion Recognition, Journal "Computer and Communications Engineering", Vol. 13, No. 1, 2019, pp. 22–25, ISSN 1314-2291.
5. **Р. Райнова**, М. Лазарова, А. Алексиева-Петрова. Многослойна мултимодална архитектура за управление на процеси при разпознаване на емоции. 2019, кн. 4, стр. 29-36, ISSN 0861-7562 Print, ISSN 2683-1279 Online.
6. **R. Raynova**, Weight Based Model for Emotion Recognition in a Multimodal Agent Based Architecture, Proc. of 9th International Scientific Conference Computer Science'2020, 18–21 Oct. 2020, Velingrad, Bulgaria, pp. 51–57, ISBN: 978-619-167-433-6.
7. **R. Raynova**, A. Aleksieva-Petrova, M. Lazarova, Multi-agent Multimodal Human Emotion Recognition Architecture, Proc. of IEEE 28th National Conference with International Participation TELECOM'2020, 29–30 Oct. 2020, Sofia, Bulgaria, pp. 37–40, ISBN 978-172818717-4, индексирана в Scopus: doi: 10.1109/TELECOM50385.2020.9299534.

SUMMARY

Methods and Algorithms for Emotion Detection and Recognition

Ralitza Raynova

Expressing and perceiving emotions is a vital part of communication between people. The ability to automatically recognize human emotions allows human-machine interactions to resemble the natural way that people communicate as a major enhancement to the human-machine interface. The sensors and technologies that can be used for direct or remote recording of input data for the recognition of emotions can be classified as verbal technologies, visual technologies and physiological technologies. The existing emotion recognition systems are mainly based on unimodal inputs and very few of them are implemented by intelligent agents. Human emotion recognition systems based on input data from different modalities contribute to robust and efficient recognition.

The objective of the dissertation is to explore the possibilities of early diagnosis of depressive disorders through non-invasive methods based on automatic extraction of information of the human emotional state and to propose an agent-based approach for emotion recognition using multimodal input data.

The state-of-the art of the methods and algorithms for pre-processing, feature extraction and classification of human emotions using verbal, visual and physiological data as well as an overview of the existing methodologies for design of multi-agent systems are presented in the thesis. Based on them an architecture of a multi-layered multimodal system with intelligent agents for extraction and processing of information for human emotional state is proposed. The proposed system architecture allows independent information extraction and processing from multimodal input data sources including verbal, visual and physiological data and thus increases the recognition accuracy. Abstractions and main agent models for data processing of verbal, visual and physiological modalities are designed and described. Agents have various objectives, use different modalities, and employ various algorithms for the detection and recognition of emotions. A model of an agent that combines the recognition results of the other intelligent agents according to the modality type and data for the gender, age and ethnic group is also presented.

The proposed system infrastructure allows autonomous operation and utilization of the most suitable recognition algorithms for certain input data as well as generation of neutral models for the object examined. The agent-based system architecture allows efficient and effective simultaneous use of data from different modalities and thus increases the recognition accuracy of the emotional state of a person especially for emotions with negative valence that might reveal depressive disorders.