

ИЗВЕШТАЈ О ОЦЕНИ МАСТЕР РАДА

I ПОДАЦИ О КОМИСИЈИ
<ol style="list-style-type: none">1. Датум и орган који је именовео Комисију 09.01.2024. Веће Департмана за математику и информатику Природно-математичког факултета Универзитета у Новом Саду2. Састав Комисије са назнаком имена и презимена сваког члана, звања, назива уже научне области за коју је изабран у звање, датума избора у звање и назив факултета, установе у којој је члан комисије запослен:<ul style="list-style-type: none">• др Душан Јаковетић, ванредни професор, Математичко моделовање, 15.11.2020, Природно-математичког факултета у Новом Саду, председник• др Оскар Марко, научни сарадник, Електроника, телекомуникације и информационе технологије, 22.01.2020, Институт БиоСенс, члан• др Никша Јаковљевић, ванредни професор, Телекомуникације и обрада сигнала, 11.10.2019, Факултет техничких наука у Новом Саду, ментор
II ПОДАЦИ О КАНДИДАТУ
<ol style="list-style-type: none">1. Име, име једног родитеља, презиме: Данијел (Славко) Лазаревић2. Датум рођења, општина, република: 9.8.1996, Добој, Република Српска – Босна и Херцеговина3. Година уписа на основне академске студије, смер/усмерење: 2015. година, смер М4, Дипломирани професор математике
III НАСЛОВ МАСТЕР РАДА
Евалуација система за препознавање емоција у говору заснованих на трансформерима и дугорочно-краткорочној меморији.
IV ПРЕГЛЕД МАСТЕР РАДА
Мастер рад је написан на 49 страна. Садржај рада је организован у 6 поглавља која укључују увод и закључак, са 47 слика и листом коришћене литературе од 41 референце. У уводу је дат опис проблема, његов значај и општи преглед постојећих решења. У наставку је пажња стављена на решења која се заснивају на методама дубоког учења, од којих су изабране две које су деловале перспективно <i>ACRNN (3-dimensional Convolutional Recurrent Neural Network with Attention Model)</i> и <i>GLAM (Global-Aware Multi-Scale Neural Network)</i> . Детаљан опис архитектура <i>ACRNN</i> и <i>GLAM</i> , дат је у трећем поглављу. У четвртном поглављу је дат опис говорних база које су коришћене у експериментима (<i>IEMOCAP</i> и <i>SEAC</i>), као и опис параметара модела који су у овој студији варирани. Након тога у петом поглављу су приказани добијени резултати, праћени одговарајућом дискусијом. На крају се налази закључак у коме су истакнути најважнији резултати до којих се дошло у студији.
V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА МАСТЕР РАДА
Структура мастер рада у потпуности прати формат научних радова. У уводу је објашњена потреба за препознавањем емоција из говорног сигнала, што је и основна тема овог рада. Говорни сигнал је најприроднији вид комуникације између људи и

очекује се да ће постати и између људи и машина. Зарад успешне комуникације неопходно је да се поред онога шта је речено (лингвистичких информација) узме и шири контекст како је то речено и од стране кога (паралингвистичке информације) да би интерпретација била потпуна. Ово је већ нашло примену у аутоматизованим контакт центрима, где у случају детекције негативних емоција конверзација се пребацује на живог оператера. Додатно, у уводу је дат кратак преглед постојећих начина описивања емоција (Екманов дискретни и вишедимензионални модел) и њиховог начина моделовања у последњих двадесетак година. Пошто су последњих пар година доминантни приступи засновани на дубоком учењу акценат је стављен те су представљени у одељку Повезани радови, од којих су два са завидним перформансама изабрана за евалуацију у овој студији.

У одељку Методе описана је детаљно архитектура два изабрана приступа - *ACRNN (3-dimensional Convolutional Recurrent Neural Network with Attention Model)* и *GLAM (Global-Aware Multi-Scale Neural Network)*. Сваки појединачни блок који чини ове моделе детаљно је описан и објашњен, и врло често илустрован одговарајућим сликама улаза и излаза. Због оваквог приступа писању и читалац који нема одговарајуће предзнање у области обраде сигнала и машинског учења може врло лако разумети основне принципе који су коришћени.

У одељку Експерименти дат је опис две јавно доступне базе које су коришћене у самој студији као и циљеви појединачних експеримената. Прва база *IEMOCAP (Interactive Emotional dyadic Motion Capture database in English)* садржи снимке професионалних глумаца који су читали текстове на енглеском с различитим емоцијама (срећа, туга, бес, фрустрација и неутрално). Друга база *SEAC (Serbian Emotional Amateur Cellphone Speech Corpus)* садржи снимке студената и наставника са Факултета техничких наука у Новом Саду који су глумили емоције (срећа, туга, бес, и неутрално) и све је снимано помоћу паметног телефона који се налазио у блиском пољу на растојању 20-40 cm од уста. Експерименти су имали неколико циљева и то: 1) Утврђивање поновљивости публикованих резултата изабраних архитектура на *IEMOCAP* бази; 2) Процена могућности коришћења претходно обучених модела на једном језику (енглеском/српском) за препознавање емоција на другом језику (српском/енглеском) 2.1) без било какве адаптације и 2.2) уз фино подешавање последњих слојева мреже (варирано је од 1 до 3); 3) Утицај додавања нових слојева на *GLAM* архитектури. Пошто су базе биле релативно мале, евалуација је извршена тако што се целокупан скуп за обуку поделио на 10 делова, и у свакој од 10 обука у оквиру једног експеримента је 8 подскупова коришћено за обуку и по један за тестирање и валидацију. На овај начин је практично формирано 10 модела и целокупан скуп је искоришћен за тестирање.

У одељку Резултати су у форми матрица конфузије табеларно представљени добијени резултати које је пратила одговарајућа дискусија. Најбитнија запажања добијена на основу експеримената са две различите архитектуре и базе у различитим сценаријима су издвојени у Закључку.

Кроз практични део рада кандидат је показао да успешно влада програмским језиком *Python* и његовим пакетима за машинско учење *PyTorch*, *TensorFlow* и *SCIKIT-Learn* и да омогући рад старих кодова са новим верзијама пакета, као и да обезбеди окружење са старим верзијама пакета, чиме би се обезбедила поновљивост резултата. При анализи добијених резултата, показао је значајан ниво критичког размишљања и познавања материје да формира одговарајуће закључке.

VI ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА

У случају обуке и тестирања на IEMOCAP бази, ACRNN модел је дао просечна осетљивост од 55%, а у GLAM модела 60%. С аспекта појединачних емоција, у случају ACRNN модела туга и љутња су емоције чија је осетљивост достигла више од 70% при чему су друге две емоције нешто испод 42%, док у случају GLAM модела само емоција која је достигла 70% је срећа, али су остале три емоције између 52% и 65%

У случају обуке и тестирања на SEAC бази, просечна осетљивост ACRNN модела је била свега 44%, а GLAM модела око 56%, што указује на знатно већу сложеност препознавања у српској бази. Једно од потенцијалних објашњења је и већи број говорника, а тиме и њихова варијабилност, али и потенцијална недовољна експресивност глумаца аматера који су учествовали у снимању.

Пошто су начини изражавања емоција зависни од културе очекивано је било да се добију знатно лошији резултати по питању осетљивости уколико је један модел трениран на једном језику, а тестиран на другом. Модели који су тренирани на енглеској бази, а тестирани на српској бази су дали просечну осетљивост од 32% и 45% за случај ACRNN и GLAM модела, а у обрнутом случају 47% и 39%. На основу ових резултата, који су бољи од онога да систем на случај погађа емоцију, можемо закључити да модели успевају издвојити обележја која на исти начин описују емоције у оба језика.

Пошто теорија дубоког учења полази од претпоставке да нижи слојеви неуронске мреже раде издвајање обележја а виши класификацију извршено је фино подешавање последњих слојева и то само последњи, последња два и последња три слоја. У већини случајева су добијени очекивани резултати, односно да с повећањем броја слојева који се фино подешавају добија модел чија осетљивост расте. Изузетак представља фино подешавање на српски GLAM модела обученог на енглеској бази, где је просечна осетљивост пала са 45% на 43%, што је последица начина на који се израчунава просечна осетљивост и где варијације у осетљивости појединачних класа не утичу. Наиме тачност GLAM модела без финог подешавања износи 46%, 52%, 11% и 73%, а са финим подешавањем 3 последња нивоа 33%, 46%, 55% и 37% за неутрално, срећу, тугу и бес, респективно. Види се да је у првом случају бес са 73% је практично замаскирао лоше перформансе које модел има на тугу која је дала 11%. Пошто перформансе нити једног модела са финим подешавањем нису превазишле перформансе модела обученог од нуле на свим подацима, може се закључити да обележја која издвајају нижи нивои и нису тако универзална као што су иницијални резултати указивали.

VII КОНАЧНА ОЦЕНА МАСТЕР РАДА

Мастер рад је урађен у складу са одобреном темом. Сви аспекти наведени у пријави теме су детаљно обрађени и реализовани. Рад је прегледно и квалитетно написан.

VIII ПРЕДЛОГ

На основу укупне оцене, Комисија предлаже да се мастер рад прихвати, а кандидату Данијелу Лазаревићу одобри одбрана.

Нови Сад,

ПОТПИСИ ЧЛАНОВА КОМИСИЈЕ

др Душан Јаковетић, председник

др Оскар Марко, члан

др Никша Јаковљевић, ментор
