Emocijų atpažinimas taikant dirbtinius neuroninius tinklus

Alina Belova

VU Matematikos ir informatikos fakultetas,
Naugarduko g. 24, Vilnius
alina.belova@mif.stud.vu.lt

Santrauka. Šiame straipsnyje pristatoma dirbtinių neuroninių tinklų, skirtų emocijų atpažinimui dialogų tekste, apžvalga ir lyginamoji analizė. Dialogų tekstas analizuojamas taikant įprastinius rekurentinius neuroninius tinklus, tačiau didžiausias dėmesys skiriamas DialogueGCN taikymui. DialogueGCN pagalba modeliuojamas pokalbio emocinis kontekstas, turintis įtakos emocijų klasifikavimui. Straipsnyje pateikiami dviejų duomenų aibių (IEMOCAP, MELD) emocijų atpažinimo rezultatai, kurie tik dalinai patvirtina DialogueGCN taikymo pranašumus. Išdėstomi galimi rezultatų paaiškinimai bei tolimesnių darbų alternatyvos.

Raktiniai žodžiai: emocijų atpažinimas, teksto analizė, dirbtinis neuroninis tinklas, grafo neuroninis tinklas, rekurentinis neuroninis tinklas.

1 Įvadas

Emocijų atpažinimas, taikant dirbtinius neuroninius tinklus (DNT), yra viena iš greitai besivystančių šiuolaikinių kompiuterių sistemų sričių. Taikymai apima daugelį kompiuterių sistemų: sveikatos apsaugos, edukacinės, protingų namų [1, 2, 3]. Galima išskirti nemažai sunkumų, susijusių su emocijų klasifikavimu DNT pagalba: psichologinių teorijų nevienodumas, įvesties duomenų kanalų įvairovė, emocijų anotavimas ir t.t. [4, 5, 6].

Šio darbo tikslas – atlikti lyginamąją žmogaus emocijų, perteikiamų rašytinio teksto pavidalu, klasifikatorių analizę. Pagrindinis tyrimo dėmesys skirtas DNT pagrindu veikiantiems modeliams, kurie klasifikuoja emocijas į kelias klases. Siekiant ištirti įvesties duomenų įtaką klasifikavimo tikslumui, čia pristatomi dviejų, nepriklausomai sukurtų duomenų aibių analizės rezultatai. Darbo metu ištirti DNT modeliai, kurių pagalba analizuojamas pokalbio emocinis kontekstas (DialogueRNN[1], DialogueGCN[2]), ir DNT architektūros,

[1] angl. *Dialogue Recurrent Neural Network*
[2] angl. *Dialogue Graph Convolutional Network*

DOI: https://doi.org/10.15388/LMITT.2021.1

neatliekančios pokalbio konteksto modeliavimo (klasikiniai natūralios kalbos apdorojimo modeliai). Dėl straipsnio ribotumo rekurentinių neuroninių tinklų (angl. *recurrent neural network*, RNN) analizė nėra akcentuojama, pagrindinis dėmesys skiriamas grafo neuroninio tinklo (angl. *graph neural network*, GNN) analizei ir taikymui.

2 Emocijų atpažinimo metodai

Tyrimo vykdymui buvo naudojamos IEMOCAP (angl. *Interactive Emotional Dyadic Motion Capture*) ir MELD (angl. *Multimodal EmotionLines Dataset*) duomenų aibės: analizuojami teksto įrašai anglų kalba [7, 8, 9]. Tarp aibių galima išskirti šiuos skirtumus, turinčius įtakos emocijų atpažinimui:

IEMOCAP. Aibę sudaro 6 emocijų klasės (žr. 3.2 poskyris) ir 151 dialogas. Dialogai – dirbtinai sukurti, vieno pokalbio dalyvių skaičius – 2, vieno dialogo vidutinis replikų skaičius – 50.

MELD. Aibę sudaro 7 emocijų klasės (žr. 3.2 poskyris) ir 1433 pokalbių įrašai. Pokalbiai – iš serialo „Draugai“, vieno pokalbio dalyvių skaičius – iki 9 asmenų, vieno pokalbio vidutinis replikų skaičius – 10.

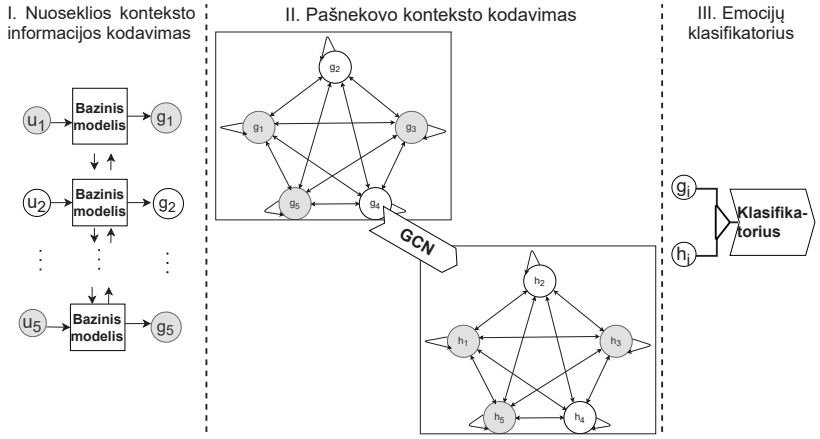
Skaičiavimams atlikti buvo naudojami viešai prieinami replikų bruožai, GloVe[3] sakinių vektoriai su dimensija 300, ir DialogueGCN4[4] kodas, po nežymių originalaus kodo atnaujinimų ir modifikacijų [10].

DialogueGCN modelį sudaro RNN ir GNN architektūros: RNN pagalba modeliuojama nuosekli pokalbio informacija, GNN pagalba modeliuojamas pokalbio emocinis kontekstas (žr. 1 pav.). GNN pagrindą sudaro grafo struktūra (viršūnės – pokalbio replikos, briaunos – sąryšiai tarp replikų), kuri sąlygoja erdvinės informacijos tarp grafo viršūnių išsilaikymą, net ir po konvoliucijos operacijos atlikimo. Grafo konvoliucijos (GCN, II-a DialogueGCN modelio dalis) pasekoje kiekvienas mazgas įgauna vis daugiau informacijos iš vis platesnio kaimyninių mazgų spektro – taip vykdomas emocinio konteksto modeliavimas.

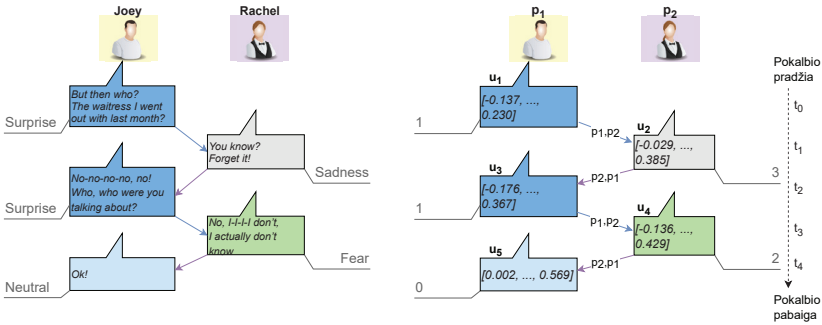
Pokalbio analizės DialogueGCN pagalba pavyzdžiui paimkime dviejų asmenų pokalbį iš MELD duomenų aibės ir sukonstruokime pokalbio grafą (žr. 2 pav.). Pradžioje, 2a pav. pokalbio replikos patalpinamos į grafo viršūnes. 2b pav. vaizduojamas detaliau apdorotas dialogo grafas: su $u_1 \ldots u_5$ replikų bruožais ir sąryšiais tarp viršūnių. Galutinį pokalbio grafą analizuoja 1 pav.

[3] https://nlp.stanford.edu/projects/glove/
[4] https://github.com/declare-lab/conv-emotion

vaizduojamas DialogueGCN modelis, vykdantis tolimesnį replikų emocijų atpažinimą.



1 pav. Trijų dalių DialogueGCN modelis, analizuojantis dviejų asmenų $u_1 \ldots u_5$ replikas (pilki ir balti rutuliukai). I – bazinis modelis, taikomas nuoseklios pokalbio informacijos modeliavimui (gaunami replikų bruožai $g_1 \ldots g_5$). II – GNN, taikomas pokalbio emocinio konteksto modeliavimui (gaunami replikų bruožai $h_1 \ldots h_5$). III – emocijų klasifikatorius, sujungiantis *gi* ir *hi* bruožus bei priskiriantis kiekvienai pokalbio replikai tam tikrą emocijų klasę. Pokalbio duomenis galima apdoroti su GNN (taikant I, II, III) ir be GNN (taikant I, III). Schema nubraižyta vadovaujantis [10] darbu.



2 pav. Dviejų asmenų (*Joey* ir *Rachel*) pokalbio grafas: replikoms priskirtos emocijų klasės (*surprise*, „1“; *sadness*, „3“; *fear*, „2“; *neutral*, „0“). (a) – replikos aprašomos teksto pavidalu; emocijos nėra užkoduotos. (b) – replikas ($u_1 \ldots u_5$) nusako skaitiniai GloVe vektoriai su dimensija 300 (skaitomumo sumetimais iš 300 skaičių rodomi tik pirmas ir paskutinis); emocijos nusakytos skaičiais (0 . . . 3); briaunoms tarp viršūnių yra priskirti simboliniai žymėjimai (p_1, p_2 ir p_2, p_1); grafo dešinėje įvesta pokalbio laiko ašis ($t_0 \ldots t_4$).

Šio tyrimo metu buvo naudojami 6 DNT modeliai:

- Modeliai be GNN – naudojami 3 baziniai rekurentiniai tinklai (rekurentinis vartų vienetas[5] (angl. *gated recurrent unit*, GRU), trumpalaikės ir ilgalaikės atminties tinklas[6] (angl. *long short-term memory*, LSTM) ir DialogueRNN[7]) be II-os DialogueGCN dalies [11, 12, 13].
- Modeliai su GNN – naudojami 3 baziniai rekurentiniai tinklai (GRU, LSTM, DialogueRNN) su II-a DialogueGCN dalimi, t.y. taikomos visos DialogueGCN modelio dalys.

3 Rezultatai

Skaičiavimams atlikti buvo naudojamas asmeninis nešiojamas kompiuteris su sistemos specifikacija: *RAM atmintis* – 15,5 GB; *CPU* – Intel® Core™ i5-8300H CPU @ 2,30 GHz ×8; *VGA* – Intel® UHD Graphics 630 (CFL GT2); *OS* – Ubuntu 18.04.5 LTS; *OS architektūra* – 64-bit. DNT skaičiavimai buvo vykdomi *Python* aplinkoje, CPU pagalba. Darbo metu buvo atlikta 120 DNT skaičiavimų paleidimų: po pusę su IEMOCAP ir MELD duomenų aibėmis.

3.1 Bendra rezultatų apžvalga

1, 2 lentelėse pateikiami geriausi emocijų klasifikavimo rezultatai, gauti taikant 6 DNT modelius (žr. 2 skyrius). Lentelėse rezultatai yra vertinami pagal bendrą tikslumą ir F_1 matą. Nedidelės tikslumo ir F_1 mato reikšmės yra daug geresnės už atsitiktines, turint galvoje, kad sprendžiamas klasifikavimo uždavinys su 6 (IEMOCAP, žr. 1 lentelė) ir 7 (MELD, žr. 2 lentelė) klasėmis. Be to, išryškėja porą tendencijų:

- GNN tinklo taikymas pagerina IEMOCAP duomenų klasifikavimo tikslumą. MELD duomenų analizė, atlikta su GNN pagalba, duoda žemesnius rezultatus nei be GNN taikymo.
- LSTM tinklas duoda geriausius rezultatus dviems atvejais: analizuojant IEMOCAP su GNN taikymu ir analizuojant MELD be GNN taikymo.

3 pav., 4 pav. vaizduojamos tikslumo ir F_1 dispersijos, nustatytos klasifikuojant IEMOCAP ir MELD duomenis su GRU ir LSTM baziniais modeliais (su GNN ir be GNN taikymo). Iš grafikų galima daryti tokius apibendrinimus:

[5] https://pytorch.org/docs/stable/generated/torch.nn.GRU.html
[6] https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html
[7] https://github.com/declare-lab/conv-emotion/tree/master/DialogueRNN

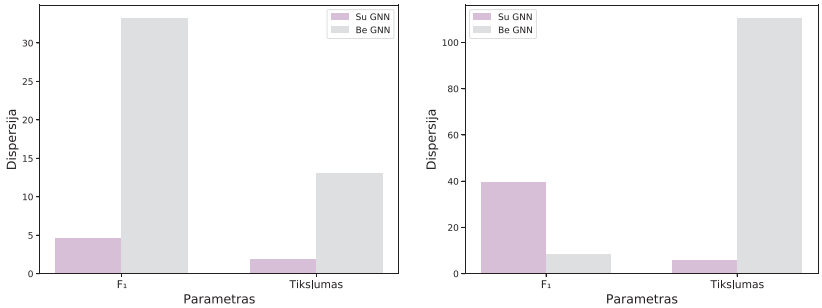
- GNN tinklo taikymas stabilizuoja bazinių tinklų (GRU, žr. 3 pav., ir LSTM, žr. 4 pav.) skaičiavimų tikslumą, nepriklausomai nuo analizuojamos duomenų aibės.
- MELD duomenų analizės atveju, GNN taikymas padidina GRU ir LSTM skaičiavimų *F*1 dispersijas, lyginant su šių modelių taikymu be GNN.

1 lentelė. Geriausių IEMOCAP rezultatų F_1 ir tikslumas: mėlyna spalva pažymėtos didžiausios reikšmės. Skaičiavimams naudojami 6 modeliai: GRU su GNN, LSTM su GNN, DialogueRNN su GNN, GRU be GNN, LSTM be GNN, DialogueRNN be GNN.

	GRU		**LSTM**		**DialogueRNN**	
	*F*1	*Tikslumas*	*F*1	*Tikslumas*	*F*1	*Tikslumas*
Su GNN	59,73	57,42	*61,49*	*61,00*	54,96	55,58
Be GNN	39,53	40,23	40,02	42,01	*41,11*	*42,54*

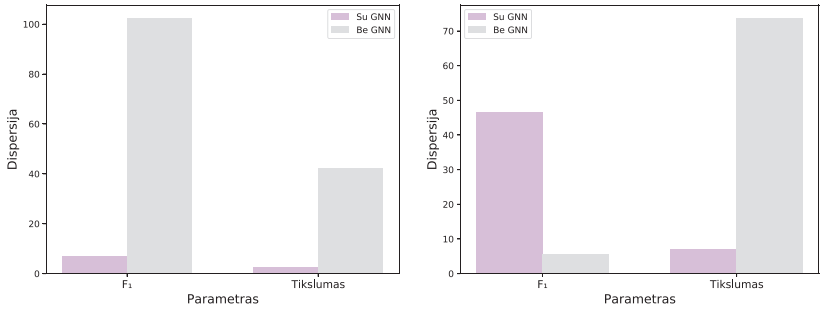
2 lentelė. Geriausių MELD rezultatų *F*1 ir tikslumas: mėlyna spalva pažymėtos didžiausios reikšmės. Skaičiavimams naudojami 6 modeliai: GRU su GNN, LSTM su GNN, DialogueRNN su GNN, GRU be GNN, LSTM be GNN, DialogueRNN be GNN.

	GRU		**LSTM**		**DialogueRNN**	
	*F*1	*Tikslumas*	*F*1	*Tikslumas*	*F*1	*Tikslumas*
Su GNN	*52,01*	54,75	51,74	55,02	51,02	*55,1*
Be GNN	75,1	73,3	*76,14*	*75,8*	69,61	59,63



(a) IEMOCAP (b) MELD

3 pav. Klasifikavimo rezultatų, gautų su GRU baziniu modeliu, tikslumo ir F_1 dispersijos: su GNN (alyvų spalva), be GNN (pilka spalva). Duomenų aibės: IEMOCAP (a), MELD (b).



(a) IEMOCAP (b) MELD

4 pav. Klasifikavimo rezultatų, gautų su LSTM baziniu modeliu, tikslumo ir F_1 dispersijos: su GNN (alyvų spalva), be GNN (pilka spalva). Duomenų aibės: IEMOCAP (a), MELD (b).

3.2 Pavienių emocijų klasių atpažinimo analizė

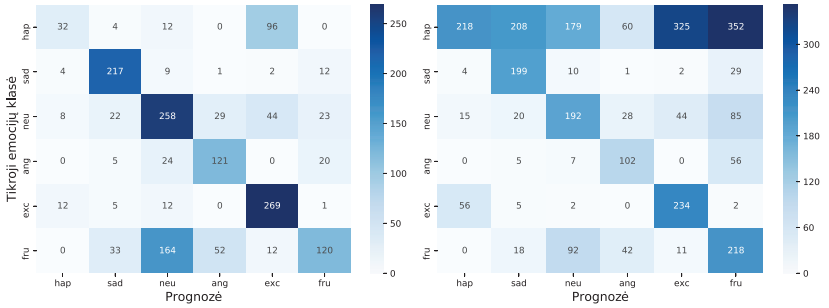
Čia yra pristatomi emocijų klasifikavimo rezultatai, gauti taikant LSTM bazinį modelį. Naudojami sekantys emocijų žymėjimai:

IEMOCAP. Emocijų klasės: „hap“ – laimė (angl. *happiness*), „sad“ – liūdesys (angl. *sadness*), „neu“ – neutrali (angl. *neutral*), „ang“ – pyktis (angl. *anger*), „exc“ – jaudulys (angl. *excitement*), „fru“ – frustracija (angl. *frustration*).

MELD. Emocijų klasės: „neu“ – neutrali (angl. *neutral*), „sur“ – nuostaba (angl. *surprise*), „fea“ – baimė (angl. *fear*), „sad“ – liūdesys (angl. *sadness*), „joy“ – džiaugsmas (angl. *joy*), „dis“ – pasibjaurėjimas (angl. *disgust*), „ang“ – pyktis (angl. *anger*).

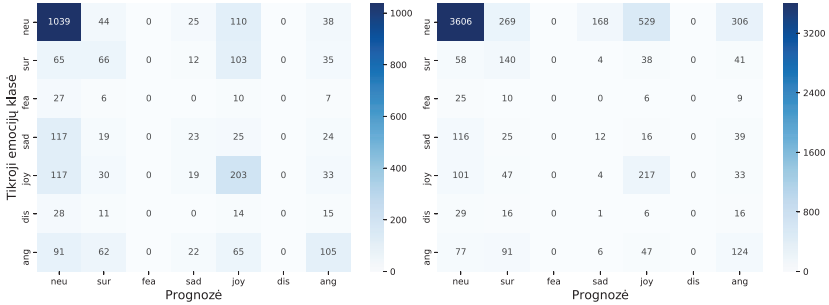
Iš 5 pav. vaizduojamų IEMOCAP rezultatų galima matyti didelį klasifikavimo pagerėjimą, taikant GNN (žr. 5a pav.). 5b pav. sumaišymo matricoje teisingai nustatytų emocijų įstrižainė yra blyškesnių spalvų, lyginant su pirma matricos eilute (t.y. teisingai aptiktų atvejų skaičius yra mažesnis). Be to, 5b pav. matome, kad gana didelis „hap“ emocijų skaičius buvo klaidingai priskirtas kitoms klasėms („sad“, „neu“, „ang“, „exc“, „fru“), lyginant su 5a pav. rezultatų tvarkingumu.

6 pav. vaizduojami MELD klasifikavimo rezultatai: juose negalime įžiūrėti tvarkingai apibrėžtos sumaišymo matricos įstrižainės. Tačiau galima pastebėti GNN taikymo pranašumus: tarp neutralios ir liūdesio klasių yra mažiau klaidų bei mažiau neutralių emocijų yra priskiriama džiaugsmo klasei. Be to, matome, kad kelių emocijų klasės („fea“, „sad“, „dis“) nebuvo teisingai aptiktos nei su GNN, nei be GNN.



(a) LSTM su GNN (b) LSTM be GNN

5 pav. IEMOCAP klasifikavimo rezultatai su LSTM, kai GNN yra taikomas (a) ir GNN nėra taikomas (b).



(a) LSTM su GNN (b) LSTM be GNN

6 pav. MELD klasifikavimo rezultatai su LSTM, kai GNN yra taikomas (a) ir GNN nėra taikomas (b).

4 Išvados

Šio tyrimo metu buvo atlikta lyginamoji emocijų, išreiškiamų teksto pavidalu, atpažinimo analizė: analizei taikomi RNN grįsti metodai (GRU, LSTM, DialogueRNN) bei GNN grįstas modelis DialogueGCN). Sudėtingos GNN struktūros taikymas pasiteisino tik dalinai, ilgų pokalbių su dviems pašnekovais (IEMOCAP duomenų aibė) analizės atveju. Taikant GNN trumpų dialogų analizei (MELD duomenų aibė) buvo gauti žemesni rezultatai, lyginant su paprastesnių modelių (GRU, LSTM) taikymu.

Be to, GNN taikymas, nepriklausomai nuo duomenų aibės, didina kai

kurių emocijų klasifikavimo tikslumą (pvz., liūdesys, neutralumas) ir duoda stabilesnio tikslumo rezultatus. Taigi, emocijų, perteikiamų teksto pavidalu, atpažinimas reikalauja tolimesnių tyrimų vykdymo: modifikuojant GNN tinklų architektūrą bei taikant RNN modelių patobulinimus.

Literatūra

[1] Daily, S. B. et al. (2017). Chapter 9 - Affective Computing: Historical Foundations, Current Applications, and Future Trends. In: Emotions and Affect in Human Factors and Human-Computer Interaction. Ed. by Jeon, M. San Diego: Academic Press, 213–231.

[2] Pieroni, M. et al. (2015). Affective Internet of Things: Mimicking human-like personality in designing smart-objects. In: 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT). 400–405.

[3] Kaklauskas, A., Lill, I., Puust, R., & Ubarte, I. (2020). 9 - Affective Internet of Things. In: Start-Up Creation (Second Edition). Ed. by Pacheco-Torgal, F. et al. Second Edition. Woodhead Publishing Series in Civil and Structural Engineering. Woodhead Publishing, 203–233.

[4] Pantic, M., & Rothkrantz, L. (2003). Toward an affect-sensitive multimodal human-computer interaction. Proceedings of the IEEE, 91(9), 1370–1390.

[5] Kołakowska, A. et al. (2013). Emotion recognition and its application in software engineering. 2013 6th International Conference on Human System Interactions, HSI 2013, 532–539.

[6] Poria, S., Majumder, N., Mihalcea, R., & Hovy, E. (2019). Emotion Recognition in Conversation: Research Challenges, Datasets, and Recent Advances. IEEE Access, 7, 100943–100953.

[7] Busso, C. et al. (2008). IEMOCAP: interactive emotional dyadic motion capture database. Language Resources and Evaluation, 42(4), 335–359.

[8] Zahiri, S., & Choi, J. D. (2018). Emotion Detection on TV Show Transcripts with Sequence-based Convolutional Neural Networks. In: Proceedings of the AAAI Workshop on Affective Content Analysis. AFFCON'18. New Orleans, LA, 44–51.

[9] Poria, S. et al. (2019). MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 527–536.

[10] Ghosal, D. et al. (2019). DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In: EMNLP-IJCNLP. Hong Kong, China: Association for Computational Linguistics, 154–164.

[11] Cho, K. et al. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Lin- guistics.

[12] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.

[13] Majumder, N. et al. (2019). DialogueRNN: An Attentive RNN for Emotion Detection in Con- versations. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 6818–6825.