MINISTERUL EDUCAŢIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA
Universitatea Tehnică a Moldovei
Facultatea Calculatoare, Informatică şi Microelectronică
Departamentul Ingineria Software și Automatică

Admis la susţinere
Șef departament:
I. Fiodorov, conf. univ. dr.
„_____”____2022

Analiza predicției personalității utilizând date din rețele sociale

Teză de masterat

Masteranda:_____Sclifos Corina,
gr. TI-201M

Conducător:_____Duca Ludmila,
asistent universitar

Consultant:_____Cojocaru Svetlana,
asistent universitar

Chișinău, 2022

ADNOTARE

Teza de master are la bază următoarea structură de redare:

- introducere;
- analiza domeniului de studiu;
- metodologia;
- arhitectura sistemului;
- colectarea şi procesarea datelor;
- concluzii;
- bibliografie.

Cuvintele cheie: Neural Network Language Model (NNLM), Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), dataset predicţie, Big 5, model neuronal.

Scopul acestei teze de masterat constă în:

- analiza metodelor de predicţie a personalităţii;
- analiza procesării limbajului natural;
- analiza modelului de limbaj de reţea neuronală;
- analiza tipurilor de personalitate şi specificul lor;
- implimentarea unui model.

Personalitatea caracterizează în mod distinct un individ și influențează profund comportamentele acestuia. Rețelele de socializare oferă comunității virtuale o oportunitate fără precedent de a genera conținut și de a împărtăși aspecte din viața lor care reflectă adesea personalitatea lor. Interesul pentru utilizarea învățării profunde pentru a deduce trăsături din amprente digitale a crescut recent; cu toate acestea, au fost prezentate lucrări foarte limitate care explorează informațiile despre sentimente transmise. Prin urmare, prezentul studiu a folosit o abordare computațională pentru a clasifica personalitatea din rețelele sociale prin măsurarea percepțiilor publice care stau la baza factorilor care cuprind trăsăturile.

În cercetarea raportată în această teză, se planifică dezvoltarea unui sistem de detectare a personalității bazat pe sentimente pentru a deduce trăsături din texte scurte bazate pe dimensiunile personalității „Big 5”. Am exploatat spiritul Neural Network Language Model (NNLM) folosind un model unificat care combină o rețea neuronală recurentă numită Long Short-Term Memory (LSTM) cu o rețea neuronală convoluțională (CNN).

ANNOTATION

The master's thesis has the following rendering structure:

- introduction;
- study analysis;
- methodology;
- system architecture;
- data collection and processing;
- conclusions;
- bibliography.

Keywords: Neural Network Language Model (NNLM), Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), dataseţ prediction, Big 5, neural model.

The purpose of this master's thesis is to:

- analysis of personality prediction methods;
- analysis of natural language processing;
- analysis of the neural network language model;
- analysis of personality types and their specificity;
- implementation of a neuronal model.

Personality distinctly characterizes an individual and profoundly influences behaviors. Social networks offer the virtual community or unprecedented opportunities to generate content and share aspects of their lives that often reflect their personality. The interest in using deep learning to deduce features from the fingerprint has recently increased; however, very limited papers have been presented that explore information about the feeling conveyed. Therefore, the present study used a computational approach to classify personality in social networks by measuring public perceptions that underlie traits.

In the research reported in this thesi şit is planned to develop a personality detection system based on feelings to deduce features from short texts based on the dimensions of the "Big 5" personality. We exploited the spirit of the Neural Network Language Model (NNLM) using a unified model that combines or recurring neural network called Long Short-Term Memory (LSTM) with a convolutional neural network (CNN).

CUPRINS

INTRODUCERE

În ultimii ani, creşterea şi colectarea informaţiei a început într-un ritm accelerat odată cu apariţia reţelelor sociale, în special sub formă de tipuri de date textuale. Conform raportului Social Media Trend publicat în [1], există 3,8 miliarde de utilizatori activi de social media din lume începând cu ianuarie 2020, cu o creştere estimată de 9,2% a utilizatorilor în fiecare an. Adesea, oamenii folosesc reţelele sociale pentru a se exprima asupra anumitor probleme legate de viaţa lor şi de fiinţele familiale, psihologie, probleme financiare, interacţiunea cu societatea şi mediul, precum şi politica. În unele cazuri, aceste expresii pot fi folosite pentru a caracteriza comportamentul individual şi personalitatea. Studiile anterioare (de exemplu [2]) demonstrează că există o corelaţie puternică între personalităţile utilizatorilor şi comportamentul lor online pe reţelele de socializare. Exemple de aplicaţii care pot profita de informaţiile despre personalitatea utilizatorilor sunt sistemele de recrutare, sistemele de consiliere personală, marketing online, sistemele de recomandări personale şi sistemele de notare a creditelor bancare [3].

Datorită ambiguităţilor inerente ale limbajelor naturale, dezvoltarea unui model eficient de predicţie a personalităţii bazat pe mesajul pe care utilizatorul îl partajează pe social media poate fi o sarcină extrem de dificilă. Având în vedere aceste ambiguităţi, s-au făcut multe progrese în domeniul procesării limbajului natural (NLP). Până în prezent NLP a permis computerelor să înţeleagă cuvinte sau propoziţii scrise în limbajul uman [4]. Concepte lingvistice precum partea de vorbire (substantive, verbe, adjective) şi structurile gramaticale sunt de obicei utilizate în NLP.

Predicţia automată a personalităţii a devenit un subiect larg discutat. Referinţele menţionate mai sus arată că personalitatea poate fi definită ca un model de influenţă sau personalitate folosit pentru a caracteriza indivizii unici. Predicţia personalităţii existente a exploatat algoritmul de învăţare profundă şi învăţare automată pentru a îmbunătăţi acurateţea clasificării. Cu toate acestea, această abordare are limitări pentru a extrage caracteristici contextuale din propoziţie datorită limitării algoritmului de calcul şi a problemei vocabularului în afara utilizării setului de date predefinit.

În acestă lucrare, procesarea limbajului natural, extragerea textului, gruparea vor fi folosite pentru a face analiza personalităţii din datele reţelelor sociale.

BIBLIOGRAFIE

1. Violino B. Social media trends. Association for Computing Machinery. Commun ACM.
2. Riccardi G. Alam F Stepanov EA. Personality traits recognition on social network—Facebook. AAAI Workshop—Technical Report.
3. Myint PH. Aung ZMM. Personality prediction based on content of facebook users: a literature re- view. Proceedings - 20th IEEE/ACIS International Conference on Software Engineering, Artifcial Intelligence, Networking and Parallel/Distributed Computing.
4. Kuchy L. Ben-Porat O Hirsch S. Predicting strategic behavior from free text.
5. F. J. Damerau N. Indurkhya. Handbook of natural language processing (2nd ed. 2010).
6. Schutze H. Manning C. Foundations of statistical natural language processing. (1999).
7. I. Zitouni D. Bikel. Multilingual natural language processing applications.
8. Rath S. K. Tripathy A. Agrawal A. Classification of sentiment reviews using n-gram machine learning approach. (2016).
9. Y. Kim. Con-volutional Neural Networks for Sentence Classification 2014.
10. Bengio Y. LeCun Y. The HANDBOOK of BRAIN theory AND NEURAL networks.
11. Collobert R. Natural language processing (almost) from scratch.
12. S. Haykin. NEURAL Networks: A Comprehensive FOUNDATION.
13. D. Shi. A study on neural network language modeling.
14. Danihelka I. Graves A. Kalchbrenner N. Grid long short-term memory. (2015).
15. Shalev A. Mandelbaum A. Word embeddings and their use in sentence classification tasks. (2016).
16. A. Black W. Ling C. Dyer. Two/Too Simple Adaptations of Word2Vec for Syntax Problems.
17. Craswell N. Diaz F. Mitra B. Query expansion with locally- trained word embeddings. (2016).
18. Alaedini Z Dalvi-Esfahani M Niknafs A. Social Media Addiction and Empathy: Moderating impact of personality traits among high school students. Telematics Inform. 2020.
19. Tang Y. Han S Huang H. Knowledge of words: An interpretable approach for personality recognition from social media
20. Cuzzocrea A Howlader P Pal KK. Predicting facebook-users' personality based on status and linguistic features via fexible regression analysis techniques. Proc ACM Symposium Appl Comput