

Principais linhas de pesquisa 1997-2023

Ivandr  Paraboni

School of Arts, Sciences and Humanities, University of S o Paulo (USP / EACH)
Av. Arlindo Bettio, 1000 - S o Paulo, Brazil
ivandre@usp.br

Abstract. *Este relat rio t cnico resume os principais projetos na linha de pesquisa de processamento de l ngua natural desenvolvidos pelo autor no per odo 1997-2023, incluindo projetos de p s-gradua o e projetos individuais com financiamento de ag ncias p blicas de fomento.*

Key words: Processamento de l ngua natural, Gera o de l ngua natural, pesquisa

1 Resolu o anaf rica

O ingresso na linha de pesquisa em Processamento de L ngua Natural (PLN) ocorreu por ocasi o do projeto de mestrado apresentado em [1] com desdobramentos em [2], tratando do problema de resolu o anaf rica do tipo possessivo com m todos baseados em regras.

Ap s um hiato de v rios anos, esta linha de pesquisa seria brevemente retomada, desta vez com uso de m todos de aprendizado de m quina em [3].

2 Gera o de l ngua natural

Por ocasi o do doutoramento [4], o foco da pesquisa em interpreta o de express es de refer ncia foi direcionado   gera o destas express es ou, mais especificamente, para a tarefa de sele o de conte do destas. Algumas das etapas intermedi rias deste estudo encontram-se em [5–8] que foram posteriormente estendidos em [9–14] j  no  mbito do grupo de pesquisa da EACH-USP.

3 Caracteriza o e atribui o autoral

Uma mudan a significativa de linha de pesquisa ocorreu por volta de 2017, com foco redirecionado (pela primeira vez em quase 20 anos) para a interpreta o de l ngua natural. Foram assim desenvolvidos diversos projetos na  rea de caracteriza o autoral que de modo geral tiravam proveito de m todos incipientes de aprendizado profundo [15–19] e atribui o autoral de dom nio fechado [20], dentre outros.

4 Predição de transtornos de saúde mental

Dentro da mesma linha de caracterização autoral, o projeto SetembroBR obteve maior alcance (inclusive na mídia) ao tratar do problema computacional de detecção de transtornos de saúde mental (depressão e ansiedade) a partir de dados multimodais provenientes de redes sociais. O projeto, ainda em desenvolvimento, apresenta resultados parciais como [21–23].

5 Detecção de posicionamentos

Como um desdobramento da tarefa de construção do cópús SetembroBR, foi desenvolvido em paralelo um segundo cópús, chamado UstanceBR [24, 25], de posicionamentos contrários/favoráveis a tópicos de interesse de discussões em redes sociais. Este cópús, e um recurso anterior chamado BRmoral [26] são a base de uma série de subprojetos dedicados à detecção de posicionamentos em várias modalidades (e.g., a partir de dados textuais e não textuais, de forma supervisionada ou *zero-shot*, dentro outras), como em [27–29].

References

1. Paraboni, I.: Uma arquitetura para a resolução de referências pronominais possessivas no processamento de textos em língua portuguesa. Master’s thesis, PUCRS, Porto Alegre (1997)
2. Paraboni, I., de Lima, V.L.S.: Possessive pronominal anaphor resolution in Portuguese written texts. In: Proceedings of the 17th international conference on Computational linguistics-Volume 2, Association for Computational Linguistics (1998) 1010–1014
3. Cuevas, R.R.M., Paraboni, I.: A machine learning approach to Portuguese pronoun resolution. In: IBERAMIA-2008, Lecture Notes in Artificial Intelligence 5290, Lisboa, Portugal, Springer-Verlag (2008) 262–271
4. Paraboni, I.: Generating references in hierarchical domains: the case of Document Deixis. PhD thesis, University of Brighton (2003)
5. Paraboni, I., van Deemter, K.: Issues for the generation of document deixis. In: Procs. of workshop on Deixis, Demonstration and Deictic Belief in Multimedia Contexts, in association with the 11th European Summers School in Logic, Language and Information (essll99). (1999) 44–48
6. Paraboni, I.: An algorithm for generating document-deictic references. In: Procs. of workshop Coherence in Generated Multimedia, associated with First Int. Conf. on Natural Language Generation (INLG-2000), Mitzpe Ramon. (2000) 27–31
7. Paraboni, I., van Deemter, K.: Generating easy references: the case of document deixis. In: INLG-2002, New York. (2002) 113–119
8. Paraboni, I., Masthoff, J., van Deemter, K.: Overspecified reference in hierarchical domains: measuring the benefits for readers. In: Proceedings of the fourth international natural language generation conference (INLG-2006), Sydney, Australia, Association for Computational Linguistics (2006) 55–62
9. de Lucena, D.J., Paraboni, I., Pereira, D.B.: From semantic properties to surface text: The generation of domain object descriptions. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial* **14**(45) (2010) 48–58

10. Paraboni, I., Galindo, M., Iacovelli, D.: Stars2: a corpus of object descriptions in a visual domain. *Language Resources and Evaluation* **51**(2) (2017) 439–462
11. dos Santos Silva, D., Paraboni, I.: Generating spatial referring expressions in interactive 3D worlds. *Spatial Cognition & Computation* **15**(03) (2015) 186–225
12. Pereira, D.B., Paraboni, I.: Statistical surface realisation of Portuguese referring expressions. In: Gotal-2008, *Lecture Notes in Artificial Intelligence* 5221, Gothenburg, Sweden, Springer-Verlag (2008) 383–392
13. Teixeira, C.V.M., Paraboni, I., da Silva, A.S.R., Yamasaki, A.K.: Generating relational descriptions involving mutual disambiguation. In: *Computational Linguistics and Intelligent Text Processing (CICLing-2014)*, *Lecture Notes in Computer Science* 8403, Kathmandu, Nepal, Springer (2014) 492–502
14. Pereira, D.B., Paraboni, I.: A language modelling tool for statistical NLP. In: 5th Workshop on Information and Human Language Technology (TIL-2007). *Anais do XXVII Congresso da SBC*, Rio de Janeiro, Sociedade Brasileira de Computação (2007) 1679–1688
15. Silva, B.B.C., Paraboni, I.: Learning personality traits from Facebook text. *IEEE Latin America Transactions* **16**(4) (2018) 1256–1262
16. Silva, B.B.C., Paraboni, I.: Personality recognition from Facebook text. In: 13th International Conference on the Computational Processing of Portuguese (PROPOR-2018) LNCS vol. 11122, Canela, Springer-Verlag (2018) 107–114
17. dos Santos, W.R., Ramos, R.M.S., Paraboni, I.: Computational personality recognition from facebook text: psycholinguistic features, words and facets. *New Review of Hypermedia and Multimedia* **25**(4) (2019) 268–287
18. dos Santos, V.G., Paraboni, I., Silva, B.B.C.: Big five personality recognition from multiple text genres. In: *Text, Speech and Dialogue (TSD-2017)* *Lecture Notes in Artificial Intelligence* vol. 10415, Prague, Czech Republic, Springer-Verlag (2017) 29–37
19. da Silva, S.C., Ferreira, T.C., Ramos, R.M.S., Paraboni, I.: Data driven and psycholinguistics motivated approaches to hate speech detection. *Computación y Sistemas* **24**(3) (2020) 1179–1188
20. Custódio, J.E., Paraboni, I.: Stacked authorship attribution of digital texts. *Expert Systems with Applications* **176** (2021) 114866
21. dos Santos, W.R., Funabashi, A.M.M., Paraboni, I.: Searching Brazilian Twitter for signs of mental health issues. In: 12th International Conference on Language Resources and Evaluation (LREC-2020), Marseille, France, ELRA (2020) 6113–6119
22. dos Santos, W.R., de Oliveira, R.L., Paraboni, I.: SetembroBR: a social media corpus for depression and anxiety disorder prediction. *Language Resources and Evaluation* (2023)
23. da Costa, P.B., Pavan, M.C., dos Santos, W.R., da Silva, S.C., Paraboni, I.: BERTabaporu: assessing a genre-specific language model for Portuguese NLP. In: *Recent Advances in Natural Language Processing (RANLP-2023)*, Varna, Bulgaria (2023) 217–223
24. Pavan, M.C., Paraboni, I.: Cross-target stance classification as domain adaptation. In Pichardo Lagunas, O., Martínez-Miranda, J., Martínez Seis, B., eds.: *Advances in Computational Intelligence - MICAI 2022 - Lecture Notes in Artificial Intelligence* vol 13612, Cham, Springer Nature Switzerland (2022) 15–25
25. Pereira, C., Pavan, M., Yoon, S., Ramos, R., Costa, P., Cavalheiro, L., Paraboni, I.: UstanceBR: a multimodal language resource for stance prediction. *arXiv:2312.06374* (2023)

26. Pavan, M.C., dos Santos, V.G., Lan, A.G.J., ao Trevisan Martins, J., dos Santos, W.R., Deutsch, C., da Costa, P.B., Hsieh, F.C., Paraboni, I.: Morality classification in natural language text. *IEEE transactions on Affective Computing* **14**(1) (2023) 857–863
27. dos Santos, W.R., Paraboni, I.: Moral Stance Recognition and Polarity Classification from Twitter and Elicited Text. In: *Recent Advances in Natural Language Processing (RANLP-2019)*, Varna, Bulgaria, INCOMA Ltd. (2019) 1069–1075
28. Pavan, M.C., dos Santos, W.R., Paraboni, I.: Twitter Moral Stance Classification using Long Short-Term Memory Networks. In: *9th Brazilian Conference on Intelligent Systems (BRACIS)*. LNAI 12319, Springer (2020) 636–647
29. Flores, A.M., Pavan, M.C., Paraboni, I.: User profiling and satisfaction inference in public information access services. *Journal of Intelligent Information Systems* **58**(1) (2022) 67–89