Big Hug: Artificial intelligence for the protection of digital societies

Big Hug: Inteligencia artificial para la protección de la sociedad digital

Arturo Montejo-Ráez[1], María Teresa Martín-Valdivia[1], L. Alfonso Ureña-López[1], Manuel Carlos Díaz-Galiano[1], Miguel Ángel García-Cumbreras[1], Manuel García-Vega[1], Fernando Martínez-Santiago[1], Flor Miriam Plaza-del-Arco[1], Salud María Jiménez-Plaza[1], María Dolores Molina-González[1], Luis-Joaquin García-López[2] and María Belén Díez-Bedmar[3]

[1] *Department of Computer Science, Advanced Studies Center in ICT (CEATIC), Universidad de Jaén, Campus Las Lagunillas, 23071, Jaén, Spain*

[2] *Department of Psychology, Universidad de Jaén, Campus Las Lagunillas, 23071, Jaén, Spain*

[3] *Department of English Studies, Universidad de Jaén, Campus Las Lagunillas, 23071, Jaén, Spain*

Abstract

In this paper, we present the Big Hug Project, which aims to claim protect vulnerable citizens and help them and their families to feel more confident when using social media communication platforms. To this end, it proposes activities for building quality data, research in new algorithms to adapt current solutions to the changing nature of colloquial and informal communication, the evaluation of techniques and methods and the development of demonstrators. This project presents an interdisciplinary approach to early detection of young people at high-risk emotional problems. The involvement of colleagues from the Clinical Psychology and Corpus Linguistics fields, furthermore, provides the project with the necessary interdisciplinary to obtain robust results which may be significant to society.

Keywords

Natural Language Processing, NLP, sentiment analysis, Clinical Psychology, early detection.

1. Introduction

Human language is the main transmission medium involved in social interaction. There are revolutionary Natural Language Processing (NLP) algorithms that can provide means to prevent and predict risky interactions, protecting the most fragile members of our digital societies. Children and adolescents have been identified by the World Health Organization as being at particular risk of psychological distress in these media[1].

Human Language Technologies (HLT) can help us build more confident environments. Thanks to NLP, artificial intelligence solutions are able to model human language and use learned models to extract information and understand the meaning of text flowing through social networks. The combination of deep learning algorithms with linguistic resources and tools, enable the construction of monitoring systems for the early detection of signs of misbehaviours like eating disorders, depression, bullying or suicide tendencies over social media[1, 2].

To this end, the project proposes two years of ac-

[1] https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

SEPLN-PD 2022. Annual Conference of the Spanish Association for Natural Language Processing 2022: Projects and Demonstrations, September 21-23, 2022, A Coruña, Spain

amontejo@ujaen.es (A. Montejo-Ráez); maite@ujaen.es (M. T. Martín-Valdivia); laurena@ujaen.es (L. A. Ureña-López); mcdiaz@ujaen.es (M. C. Díaz-Galiano); magc@ujaen.es (M. García-Cumbreras); mgarcia@ujaen.es (M. García-Vega); dofer@ujaen.es (F. Martínez-Santiago); fmplaza@ujaen.es (F. M. Plaza-del-Arco); sjzafra@ujaen.es (S. M. Jiménez-Plaza); mdmolina@ujaen.es (M. D. Molina-González); ljgarcia@ujaen.es (L. García-López); belendb@ujaen.es (M. B. Díez-Bedmar)
0000-0002-8643-2714 (A. Montejo-Ráez); 0000-0002-2874-0401 (M. T. Martín-Valdivia); 0000-0001-9752-2830 (L. A. Ureña-López); 0000-0001-9298-1376 (M. C. Díaz-Galiano); 0000-0003-1867-9587 (M. García-Cumbreras); 0000-0003-2850-4940 (M. García-Vega); 0000-0002-1480-1752 (F. Martínez-Santiago); 0000-0002-3020-5512 (F. M. Plaza-del-Arco); 0000-0003-3274-8825 (S. M. Jiménez-Plaza); 0000-0002-8348-7154 (M. D. Molina-González); 0000-0003-0446-6740 (L. García-López); 0000-0001-9250-2224 (M. B. Díez-Bedmar)

tivities for building quality data, research in new algorithms to adapt current solutions to the changing nature of colloquial and informal communication, the evaluation of techniques and methods and the development of demonstrators to leverage human-centered solutions that will protect vulnerable citizens and help them and their families to feel more confident when using social media communication platforms. Besides, this project presents an interdisciplinary approach to early detection of young people at high-risk emotional problems. By indicated prevention, scientific community has agreed to name to high-risk individuals who are identified as having some detectable symptoms of emotional disorders but who do not meet criteria or a diagnosis at the current time. The collaboration of colleagues from the Clinical Psychology and Corpus Linguistics fields, furthermore, provides the project with the necessary interdisciplinary approach to obtain robust results which may be significant to society.

Joint efforts of NLP with Corpus Linguistics and Clinical Psychology are sought in this project with a two-fold purpose: a) to analyse the results obtained from the linguistic point of view to fine-tune and complement the NLP findings; and b) to contrast the results with the scientific literature on these disorders in Clinical Psychology.

2. Participants and project funding

The project brings together 3 partners from University of Jaén: SINAI group from Advanced Studies Center in ICT (CEATIC), Department of Psychology and Department of English Studies. This project has been supported by the grant P20_00956 (PAIDI 2020) funded by the Andalusian Regional Government.

3. State of the art

It is estimated 24 million children and young people in the EU suffer from bullying every year, which means that 7 out of 10 suffer some form of harassment or intimidation, whether verbal, physical or through new communication technologies [3]. Navarro-Gómez [4] stated that social networks allow the viral diffusion of degrading contents. Cyberbullying or electronic aggression has already been designated as a serious public health threat and has elicited warnings to the general public from the Centers for Disease Control and Prevention (CDC) [5].

In another study [6], approximately 1 out of 10 people were found to develop some sort of eating disorder, which also caused anxiety, self-harming and a high risk of suicide. May studies have tackled this fact from psychometrics, but better tools for modeling the language used would help [7], even more when eating disorders are rising all around the world. Emotional disorders, like depression and anxiety, affect a quarter of our population during their lifetime [8]. Depression can be studied and identified by monitoring users' posts and activity [1].

In Spain there are 10 suicides a day, twice as many people die by suicide as by traffic accidents, 11 times more than by homicide and 80 times more than by gender violence. A very complete overview on how computers and algorithms can help in preventing or detecting suicide risk is the one recently published by Ji [9]. Recent studies have found that automatic processing of social media communications is an effective way to detect suicidal ideation by applying emotion and sentiment analysis over textual messages [10].

NLP techniques are being applied to the analysis of social media textual data to face new problems like fake-news detection [11], offensive language identification [12], sentiment analysis [13], opinion mining and emotion detection [14]. Social Big Textual Data is challenging, because language varies across time and space, language register is informal, colloquial and full of idioms compared to formal forms of text. Artificial Intelligence has gained a lot of popularity in recent years thanks to advent of Deep Learning techniques [15]. Nevertheless, many of the applications and problems overcome where already attempted with traditional algorithms in machine learning, heuristic approaches or knowledge-based systems. The big difference to previous approaches is that current proposals are data-driven: they are able to learn from large amounts of data and build models to perform different tasks with a level of success never reached by other solutions.

This shift has been especially dramatic for NLP. Linguistic-based methods have been surpassed by end-to-end architectures, where no prior knowledge on language is needed [16], but massive amounts of data are required. During the last two years we have witnessed the birth of amazing models like BERT [17], GPT-2 [18] or Transformer-XL [19], with impressive results in many different tasks. New models seem to learn language linguistic nature from data.

The gross research on NLP is turning towards Transformer based models and exploring how far these architectures are able to learn and perform in human related tasks, being sentiment analysis, emotion detection and hate-speech identification, among them.

There are previous projects in the pursuit of similar goals, like the STOP project [20] or MENHIR [21]. The Big Hug project is not only focused in exploring algorithm and models for early detection of disorders, but also in finding effective ways to transfer these systems to real world applications.

4. Objectives of the project

The main objective is clear: a multidisciplinary project for the research on methods and algorithms to analyse textual streams across time and discover patterns for an early detection of potential harmful situations or behaviours. This global goal can be divided into the following sub-objectives:

1. To identify valid technologies for "listening" the interactions in digital environments.
2. To model different forms of aggressive communication or risky situations.
3. To identify young people at high risk, but by the very first time, via a screening of altogether big data, psychological, linguistic variables.
4. To facilitate the replication of the screening protocol based on a well-defined methodology and analysis plan, if the previous objective is met.
5. To enhancement of our capabilities to feed these artificial intelligences with quality data by means of new techniques and methods to process informal language or colloquial expressions.
6. To adapt human language technologies also to the specific one that is usually used to make apologia of those scenarios.
7. To explore practical solutions which may be integrated in the real world.

5. Conclusion

Dispositions for eating, anxiety and depressive disorders, are multifactorial. Big Hug represents a novel approach for mental disorders, integrating mental health, big data and linguistics measures as predictive measures for early diagnosis.

Research on mental health, for the early diagnosis and treatment of emotional mental health problems in the young is fragmented as researchers have traditionally worked in isolation and few studies examined the same or more than a limited set of risk factors, neglecting novel stratification strategies and development of algorithms. The Big Hug project avoids the problems of fragmentation by co-ordinating and developing joint activities related to early identification in order to coordinate high quality transnational research. The different perspectives and especially the different qualifications of mental-health, applied linguistics and Information and Communication of Technologies (ICT) specialists working in academia could stimulate the discovery of new and creative solutions. Apart from multidisciplinarity, there are relevant transversal aspects in the project.

References

[1] D. E. Losada, F. Crestani, J. Parapar, Overview of erisk 2019 early risk prediction on the internet, in: International Conference of the CLEF for European Languages, Springer, 2019, pp. 340–357.

[2] J. Parapar, P. Martín-Rodilla, D. E. Losada, F. Crestani, eRisk 2021: pathological gambling, self-harm and depression challenges, in: ECIR, Springer, 2021, pp. 650–656.

[3] E. Cross, R. Piggin, T. Douglas, J. Vonkaenel-Flatt, Virtual violence ii: Progress and challenges in the fight against cyberbullying, London: Beatbullying (2012).

[4] N. Navarro-Gómez, El suicidio en jóvenes en españa: cifras y posibles causas. análisis de los últimos datos disponibles, Clínica y Salud 28 (2017) 25–31.

[5] E. Aboujaoude, M. W. Savage, V. Starcevic, W. O. Salame, Cyberbullying: Review of an old problem gone viral, Journal of adolescent health 57 (2015) 10–18.

[6] E. Stice, M. J. Van Ryzin, A prospective test of the temporal sequencing of risk factor emergence in the dual pathway model of eating disorders., Journal of Abnormal Psychology 128 (2019) 119.

[7] T. Wang, M. Brede, A. Ianni, E. Mentzakis, Detecting and characterizing eating-disorder communities on social media, in: Proceedings of the Tenth ACM International conference on web search and data mining, 2017, pp. 91–100.

[8] J. Wang, X. Wu, W. Lai, E. Long, X. Zhang, W. Li, Y. Zhu, C. Chen, X. Zhong, Z. Liu, et al., Prevalence of depression and depressive symptoms among outpatients: a systematic review and meta-analysis, BMJ open 7 (2017) e017173.

[9] S. Ji, S. Pan, X. Li, E. Cambria, G. Long, Z. Huang, Suicidal ideation detection: A review of machine learning methods and appli-

cations, IEEE Transactions on Computational Social Systems 8 (2020) 214–226.

[10] J. J. Glenn, A. L. Nobles, L. E. Barnes, B. A. Teachman, Can text messages identify suicide risk in real time? a within-subjects pilot examination of temporally sensitive markers of suicide risk, Clinical Psychological Science 8 (2020) 704–722.

[11] F. Monti, F. Frasca, D. Eynard, D. Mannion, M. M. Bronstein, Fake news detection on social media using geometric deep learning, arXiv preprint arXiv:1902.06673 (2019).

[12] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, R. Kumar, Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval), arXiv preprint arXiv:1903.08983 (2019).

[13] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Urena-López, A. R. Montejo-Ráez, Sentiment analysis in twitter, Natural Language Engineering 20 (2014) 1–28.

[14] F. M. Plaza-del Arco, M. T. Martín-Valdivia, L. A. Ureña-López, R. Mitkov, Improved emotion recognition in spanish social media through incorporation of lexical knowledge, Future Generation Computer Systems 110 (2020) 1000–1008.

[15] J. Dean, D. Patterson, C. Young, A new golden age in computer architecture: Empowering the machine-learning revolution, IEEE Micro 38 (2018) 21–29.

[16] T. Young, D. Hazarika, S. Poria, E. Cambria, Recent trends in deep learning based natural language processing, ieee Computational intelligenCe magazine 13 (2018) 55–75.

[17] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[18] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI blog 1 (2019) 9.

[19] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, R. Salakhutdinov, Transformer-xl: Attentive language models beyond a fixed-length context, arXiv preprint arXiv:1901.02860 (2019).

[20] D. Ramírez-Cifuentes, A. Freire, R. Baeza-Yates, J. Puntí, P. Medina-Bravo, D. A. Velazquez, J. M. Gonfaus, J. Gonzàlez, et al., Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis, Journal of medical internet research 22 (2020) e17758.

[21] M. Kraus, P. Seldschopf, W. Minker, Towards the Development of a Trustworthy Chatbot for Mental Health Applications, in: MultiMedia Modeling, Springer, 2021, pp. 354–366.