Introduction to the Special Section on Computational Modeling and Understanding of Emotions in Conflictual Social Interactions

ROSSANA DAMIANO and VIVIANA PATTI, Dipartimento di Informatica University of Turin
CHLOÉ CLAVEL, LTCI, Telecom Paris, Institut Polytechnique de Paris
PAOLO ROSSO, Universitat Politècnica de València

CCS Concepts: • **Human-centered computing** → **Social media**; • **Computing methodologies** → **Intelligent agents**; **Natural language processing**;

Additional Key Words and Phrases: Affective computing, socio-affective behavior, social interactions, hate speech detection

ACM Reference format:
Rossana Damiano, Viviana Patti, Chloé Clavel, and Paolo Rosso. 2020. Introduction to the Special Section on Computational Modeling and Understanding of Emotions in Conflictual Social Interactions. *ACM Trans. Internet Technol.* 20, 2, Article 8 (May 2020), 5 pages.
https://doi.org/10.1145/3392334

1 INTRODUCTION

In today's media and social media, the expression of social, cultural, and political opinions often features a strong affective component, especially when it occurs in highly polarized contexts (discussions on political elections, migrants, civil rights, etc.). Interactions of this type can easily degenerate from fruitful discussions to conflicts, characterized by negative manifestations of opinions such as hate speech. Hate speech, recognized as an extreme, yet typical, expression of opinion, is increasingly intertwined with the spread of defamatory, false stories [6, 9, 17]. At the same time, interpersonal conflict has emerged as a major cause of failure and discrimination in different social contexts, ranging from institutionalized organizations such as workplaces and schools to personal relationships [20].

Detecting and monitoring conflicts is relevant because conflicts may exacerbate the inequalities latent in our societies, thus contributing to exclusion of specific groups of people, such as young

The editorial work of C. Clavel for this special issue was partially supported by a grant overseen by the French National Research Agency (ANR17-MAOI) and by the European project H2020 ANIMATAS (MSCA-ITN-ETN 7659552). The editorial work of V. Patti was partially funded by Progetto di Ateneo/CSP 2016 (*Immigrants, Hate and Prejudice in Social Media*, S1618_L2_BOSC_01). P. Rosso was partially funded by Spanish MICINN under the research project MISMIS-FAKEnHATE on MISinformation and MIScommunication in social media: FAKE news and HATE speech (PGC2018-096212-B-C31).
Authors' addresses: R. Damiano and V. Patti, Dipartimento di Informatica University of Turin, Corso Svizzera 185, Turin, Italy, 10149; emails: rossana.damiano@unito.it, patti@di.unito.it; C. Clavel, LTCI, Telecom Paris, Institut Polytechnique de Paris, 46 rue Barrault, Paris, France, 75013; email: chloe.clavel@telecom-paristech.fr; P. Rosso, Universitat Politècnica de València, Camino de Vera, s/n - 46022, València, Spain.

1533-5399/2020/05-ART8
https://doi.org/10.1145/3392334

people (bullying and cyberbullying), women (misogyny), and immigrants (hate speech). In this sense, the availability of computational models of emotions in conflictual social interactions can have practical effects in contrasting discrimination and achieving the objectives of equality and cohesion of future societies. At the interpersonal level, conflicts can negatively affect the quality of life and the emotional well-being of the people involved. Understanding the affective dynamics of conflicts is important also for modeling human behavior in social settings that involve artificial agents, as well as for designing socially aware artificial systems.

Current approaches for monitoring and circumscribing the spread of conflict-related phenomena mostly rely on standard affective models that do not account for emotions as complex cognitive, social, and cultural constructs behind linguistic behavior. Moral emotions, for example, possess a potential for advancing sentiment analysis in social media, especially since they provide insights on the motivations behind hate speech [4, 11].

Based on the background sketched previously, the goal of this special issue is twofold:

- On the modeling side, including finer-grained accounts of emotions in computational models of interpersonal and social interactions, with the goal of monitoring and dealing with conflicts in social media and human-to-human and human-to-agent interactions.
- On the processing side, leveraging the recent advances in machine learning and reasoning techniques to design more effective computational models of interpersonal and social conflict.

Last but not least, the development of novel approaches requires the availability of shared resources, such as specialized datasets, annotation models, benchmarking tools, and linguistic resources, to inspire and validate models and methods within an experimental paradigm [1].

2 THE SPECIAL SECTION ON COMPUTATIONAL MODELING AND UNDERSTANDING OF EMOTIONS IN CONFLICTUAL SOCIAL INTERACTIONS

The articles included in this special section address the role of emotions in the emergence and manifestation of conflicts in different contexts, ranging from interpersonal relations to social media, and provide tools for studying, detecting, and classifying conflicts with novel resources and methods.

In "Sub-Population Models of Couples' Conflicts: Automatically Detecting Interpersonal Conflict Through the Ambulatory Monitoring of Couples' Physiological Signals, Audio Samples, and Linguistic Indices" [10], the authors address the detection of conflict in romantic relations through the use of acoustic, linguistic, and physiological indices obtained through wearable devices, coupled with self-reported data. For each sub-population, feed-forward neural networks were trained for conflict classification using both hierarchical and adaptive methods. Sub-population models outperformed the model trained on the entire population, and slightly better results were obtained with adaptive training. Different models turned out to be suitable for different subpopulations, whereas the self-reported features collected through questionnaires were more effective by themselves than all separate signal-based indices; the situation was reversed when signal indices were gathered. However, the classification of conflict was shown to not be uniform across sub-populations, with populations with higher variability in attachment and satisfaction parameters being easier to deal with for the classification task. The results reported by the authors open a new perspective on the use of machine learning techniques for health applications, as they highlight the need for interdisciplinary models and methods to tackle domains that are characterized by intrinsic complexity, such as interpersonal relations.

In "A Multilingual Evaluation for Online Hate Speech Detection" [5], the authors address the task of automatically detecting hate speech in social networks by addressing the multilinguality

challenge. One of the main contributions of the article consists of identifying a deep learning architecture suitable for the task in a cross-lingual environment, as well as selecting the best features (e.g., word and emoji embeddings, social network specific features, or emotion and hate lexicons [2]) that could be exploited to develop better hate speech detection systems. The proposed architectures have been evaluated on freely available datasets in English, German, and Italian, and the impact of various parameters on hate speech detection in different languages is discussed. The multilinguality issue is one of the challenges for the abusive language detection task in past years. In fact, most popular social media, such as Twitter, Facebook, and Instagram, go multilingual, fostering their users to interact in their primary language. Therefore, there is a considerable urgency to develop a robust approach for abusive language detection in a multilingual environment [12], as well as for supporting better compliance to government demands for counteracting the hate speech phenomenon (e.g., see the recently issued EU Commission's *Code of Conduct on Countering Illegal Hate Speech Online* [7]). Even if most studies so far have focused on English, the availability of abusive language datasets developed for different languages in past 2 years, including Italian, [3, 8, 16], Spanish [1, 8], and German [19], constitutes an excellent condition for progress on this front.

The article "Mandola: Monitoring, Detecting, Visualizing, and Reporting the Spread and Penetration of Online Hate-Related Speech Using Machine Learning and Big Data Approaches" [13] describes the development of a big data processing framework for detecting, monitoring, and visualizing hate speech online in social media, by collecting real-time data on the fly from different sources, including Twitter. The architecture of the proposed framework includes a data processing pipeline where the output of an automatic hate speech detection engine is exploited to feed a visualization dashboard, which allows users to explore the outcome of the hate speech monitoring process in the form of different map-based visualizations. The hate speech detection model is based on a novel ensemble-based classification algorithm relying on traditional machine learning and deep learning techniques. The model was successfully evaluated against different state-of-the-art approaches for hate speech detection based on available benchmark datasets annotated for hate and offensiveness. A use case related to a real hate event in the United States is presented to show the functionality of the MANDOLA framework and visualization dashboard in action. The article tackles a very important problem by combining, in a novel and interesting way, natural language processing and machine learning techniques for hate speech detection with big data analysis and visualization techniques. Mandola's data collection engine is built to be compliant with new General Data Protection Regulations, and processes information in real time and on the fly without storing any user-specific information (apart from the statistical output and metadata resulted from the processing). This is an aspect to be appreciated considering the new challenges provided by regulations for social media monitoring platforms.

The article "Detecting Misogyny and Xenophobia in Spanish Tweets: Create Appropriate Language Resources for Hate Speech Detection in Spanish" [14] addresses the detection of hate speech in Spanish tweets for the partially superimposed domains of misogyny and xenophobia, which represent important social problems in today's society for which effective tools are strongly needed. The authors compare different, yet standard, machine learning approaches for hate speech detection, namely a lexicon-based approach, a supervised machine learning approach, and a deep learning approach. The main conclusion is that the combination of lexicon-based and supervised machine learning approaches yields the best results. The effectiveness of deep learning techniques remains open, as in this study it turned out to be less effective than other simpler approaches due to the limited size of the dataset. As the approaches tested in this article originally were developed for the English language, their use for a different language provides a reference method for their adaptation to other languages, including the creation of a lexicon for the target language from

available resources. Despite the portability of the proposed approach across languages, the authors warn about the need to account for the language-specific phenomenology of the hate speech. To do so, they report some significant features of abusive language in Spanish, which shows a larger variety of expressions than English, partly determined by the way opposite polarities can be combined in Spanish to form complex expressions.

The work presented in "Emo2Vec: Learn Emotional Embeddings to Identify Sentimentally Similar Words Models" [18] contributes to the building of more effective computational models of emotions and sentiment for the task of social network analysis. The studied social networks consist of hotel and movie reviews, news headlines, and microblogs in English and Chinese. The authors tackle the challenging issue of adding sentiment and emotional information to pre-trained word embeddings. The emotional information is built from a dimensional and categorical psychological model of emotions—Plutchik's wheel model [15]—and takes the form of a lexicon that is encoded in a vector and added to pre-trained representation. The vector computes the position of the words in Plutchik's wheel model. The pre-trained word embeddings are modified according to their proximity to emotional words from the provided lexicon. Evaluation of the proposed representation is carried out by comparing performance of the emotional embeddings with conventional embeddings for two tasks: sentiment analysis (polarity classification) and multicategory emotion classification. Results show that the emotional embeddings increase the performance of the classification systems on the social network datasets.

3 CONCLUDING REMARKS

Emotions and sentiment can provide useful conceptual and practical tools for detecting and monitoring conflicts in social interactions of different types, as investigated by the articles in this special issue, potentially contributing to revising and refining theoretical models of conflict from a data-driven perspective.

The occurrence of conflicts at the global level, in the past decade, has offered several opportunities for collecting data on conflicts for several controversial topics through social media and sensors. Unfortunately, environmental factors of different types—of which the present COVID-19 epidemic is an example, with its accompaniment of controversies—are likely to renew this opportunity in the immediate future as well. In this sense, initiatives aimed at gathering further data will be crucial in the future to improve our computational models of emotions in conflictual interactions and thus release useful tools for dealing with conflicts at the global level.

ACKNOWLEDGMENTS

We would like to thank Editor in Chief Liu Ling and the editorial staff of ACM and ACM TOIT for their support in difficult times, and all authors for their valuable contributions.

REFERENCES

[1] Valerio Basile, Cristina Bosco, Elisabetta Fersini, Debora Nozza, Viviana Patti, Francisco Manuel Rangel Pardo, Paolo Rosso, and Manuela Sanguinetti. 2019. SemEval-2019 Task 5: Multilingual detection of hate speech against immigrants and women in Twitter. In *Proceedings of the 13th International Workshop on Semantic Evaluation.* 54–63. DOI: https://doi.org/10.18653/v1/S19-2007

[2] Elisa Bassignana, Valerio Basile, and Viviana Patti. 2018. Hurtlex: A multilingual lexicon of words to hurt. In *Proceedings of the 5th Italian Conference on Computational Linguistics (CLiC-it'18).* 6. http://ceur-ws.org/Vol-2253/paper49.pdf

[3] Cristina Bosco, Felice Dell'Orletta, Fabio Poletto, Manuela Sanguinetti, and Maurizio Tesconi. 2018. Overview of the EVALITA 2018 hate speech detection task. In *Proceedings of the 6th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA'18) co-located with the 5th Italian Conference on Computational Linguistics (CLiC-it'18).* 9. http://ceur-ws.org/Vol-2263/paper010.pdf

[4] William J. Brady, Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel. 2017. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences* 114, 28 (2017), 7313–7318. DOI: https://doi.org/10.1073/pnas.1618923114

[5] Michele Corazza, Stefano Menini, Elena Cabrio, Sara Tonelli, and Serena Villata. 2020. A multilingual evaluation for online hate speech detection. *ACM Transactions on Internet Technology: Special Section on Emotions in Conflictual Social Interaction* 20, 2 (March 2020), Article 10, 22 pages. DOI: https://doi.org/10.1145/3377323

[6] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. In *Proceedings of the 11th International Conference on Web and Social Media (ICWSM'17)*. 512–515. https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15665/14843

[7] EU Commission. 2016. The EU Code of Conduct on Countering Illegal Hate Speech Online. Retrieved April 18, 2020 from https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online_en#theeucodeofconduct.

[8] Elisabetta Fersini, Paolo Rosso, and Maria Anzovino. 2018. Overview of the task on automatic misogyny identification at IberEval 2018. In *Proceedings of the 3rd Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval'18) co-located with the 34th Conference of the Spanish Society for Natural Language Processing (SEPLN'18)*. 214–228. http://ceur-ws.org/Vol-2150/overview-AMI.pdf.

[9] Paula Fortuna and Sérgio Nunes. 2018. A survey on automatic detection of hate speech in text. *ACM Computing Surveys* 51, 4 (July 2018), Article 85, 30 pages. DOI: https://doi.org/10.1145/3232676

[10] Krit Gupta, Aditya Gujral, Theodora Chaspari, Adela C. Timmons, Sohyun Han, Yehsong Kim, Sarah Barrett, Stassja Sichko, and Gayla Margolin. 2020. Sub-population specific models of couples' conflict. *ACM Transactions on Internet Technology: Special Section on Emotions in Conflictual Social Interaction* 20, 2 (March 2020), Article 9, 20 pages. DOI: https://doi.org/10.1145/3372045

[11] Jonathan Haidt2003. The moral emotions. In *Handbook of Affective Sciences*, R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (Eds.). Oxford University Press, Oxford, UK, 852–870.

[12] Endang Wahyu Pamungkas and Viviana Patti. 2019. Cross-domain and cross-lingual abusive language detection: A hybrid approach with deep learning and a multilingual lexicon. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. 363–370. DOI: https://doi.org/10.18653/v1/P19-2051

[13] Demetris Paschalides, Dimosthenis Stephanidis, Andreas Andreou, Kalia Orphanou, George Pallis, Marios D. Dikaiakos, and Evangelos Markatos. 2020. MANDOLA: A big-data processing and visualization platform for monitoring and detecting online hate speech. *ACM Transactions on Internet Technology: Special Section on Emotions in Conflictual Social Interaction* 20, 2 (March 2020, Article 11, 21 pages. DOI: https://doi.org/10.1145/3371276

[14] Flor-Miriam Plaza-Del-Arco, M. Dolores Molina-González, L. Alfonso Ureña López, and M. Teresa Martín-Valdivia. 2020. Detecting misogyny and xenophobia in Spanish tweets using language technologies. *ACM Transactions on Internet Technology: Special Section on Emotions in Conflictual Social Interaction* 20, 2 (March 2020), Article 12, 19 pages. DOI: https://doi.org/10.1145/3369869

[15] Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist* 89, 4 (2001), 344–350.

[16] Manuela Sanguinetti, Fabio Poletto, Cristina Bosco, Viviana Patti, and Marco Stranisci. 2018. An Italian Twitter corpus of hate speech against immigrants. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC'18)*. 8. https://www.aclweb.org/anthology/L18-1443.

[17] Anna Schmidt and Michael Wiegand. 2017. A survey on hate speech detection using natural language processing. In *Proceedings of the 5th International Workshop on Natural Language Processing for Social Media*. 1–10. DOI: https://doi.org/10.18653/v1/W17-1101

[18] Shuo Wang and Xiaofeng Meng. 2020. Emo2Vec: Learning emotional embeddings via multi-emotion category. *ACM Transactions on Internet Technology: Special Section on Emotions in Conflictual Social Interaction* 20, 2 (March 2020), 1–17.

[19] Michael Wiegand, Melanie Siegel, and Josef Ruppenhofer. 2018. Overview of the GermEval 2018 shared task on the identification of offensive language. In *Proceedings of GermEval 2018, the 14th Conference on Natural Language Processing (KONVENS'18)*. 1–10.

[20] W. Wilmot and J. Hocker. 2013. *Interpersonal Conflict* (9th ed.). McGraw-Hill, New York, NY.