Proceedings of the Informing Science + Information Technology Education Conference

An Official Publication of the Informing Science Institute InformingScience.org

InformingScience.org/Publications

Online July 24-25,2024

EMOTIONAL ANALYSIS IN HEBREW TEXTS: ENHANCING MACHINE LEARNING WITH PSYCHOLOGICAL FEATURE LEXICONS

[ABSTRACT]

Ron Keinan*	Department of Computer Science and Engineering, Jerusalem College of Technology, Jerusalem, Israel	ronke21@gmail.com
Efraim Margalit	Department of Computer Science and Engineering, Jerusalem College of Technology, Jerusalem, Israel	efraimefraim@hotmail.com
Dan Bouhnik	Department of Computer Science and Engineering, Jerusalem College of Technology, Jerusalem, Israel	bouhnik@jct.ac.il

* Corresponding author

ABSTRACT

Aim/Purpose	This paper addresses the challenge of emotional analysis in Hebrew texts, specifically focusing on enhancing machine learning techniques with psychological feature lexicons to improve classification accuracy in identifying depression.
Background	Emotional analysis in Hebrew texts presents unique challenges due to the language's intricate morphology and rich derivation system. This paper seeks to leverage advanced machine learning methods augmented with carefully crafted psychological feature lexicons to address these challenges and improve the identification of depression from online discourse.
Methodology	The study involves scraping and analyzing a dataset consisting of over 350K posts from 25K users on the "Camoni" health-related social network spanning 2010-2021. Various machine learning models, including SVM, Random Forest, Logistic Regression, and Multi-Layer Perceptron, were employed alongside en-

Accepted by Editor Eli Cohen │ Received: March 7, 2024 │ Revised: April 17, 2024 │ Accepted: April 19, 2024.

Cite as: Keinan, R., Margalit, E., & Bouhnik, D. (2024). Emotional analysis in Hebrew texts: Enhancing machine learning with psychological feature lexicons. In M. Jones (Ed.), *Proceedings of InSITE 2024: Informing Science and Information Technology Education Conference*, Article 3. Informing Science Institute. https://doi.org/10.28945/5279

semble methods such as Bagging, Boosting, and Stacking. Features were selected using TF-IDF, incorporating both word and character n-grams (Aisopos et al., 2016; HaCohen-Kerner et al., 2018). Pre-processing steps, including punctuation removal, stop word elimination, and lemmatization, were applied, to handle the challenges in Hebrew as a reach morphological language (Amram et al., 2018; Tsarfaty et al., 2019). Then hyperparameter tuning was conducted to optimize model performance across different languages. Following this, the models were enriched with features extracted from sentiment lexicons conducted by professional psychologists. (Shapira et al., 2021).

Contribution This paper contributes to the field by demonstrating the efficacy of integrating psychological feature lexicons into machine-learning models for emotional analysis in Hebrew texts. Addressing the unique linguistic challenges, it advances the understanding of depression detection in online communities and informs the development of more effective preventive measures and treatments.

Findings Through experimentation, it was discovered that enriching the models with features from sentiment lexicons significantly improved classification accuracy. Among the sentiment lexicons tested, six were identified as particularly enhancing: Negative emojis, positive emojis, neutral emojis, Hostile words, Anxiety words, and No-Trust words. The coverage and the quality of a feature lexicon are and may contribute to the success of various tasks like opinion mining and sentiment analysis (Feldman, 2013; Liu, 2012; Yang et al., 2020).

Recommendations for Practitioners Practitioners in mental health and social work should prioritize enriching machine learning models with sentiment lexicons to enhance the accuracy and effectiveness of depression detection in online discourse. By incorporating lexicons capturing emotional nuances, practitioners can improve the sensitivity of their screening processes.

Recommendations for Researchers Future research endeavors should focus on further refining machine learning models by enriching them with sentiment lexicons. Additionally, exploring the integration of sentiment lexicons into deep learning models could provide further insights into the classification of emotional content in textual data.

Impact on Society The findings have significant implications for the development of more accurate and efficient methods for detecting depression in online Hebrew discourse. By leveraging advanced machine learning techniques augmented with psychological feature lexicons, this research contributes to enhancing mental health interventions and promoting well-being in online communities.

Future Research Future research should not only continue exploring the integration of sentiment lexicons into machine learning models but also extend this investigation to deep learning architectures. Investigating the effectiveness of sentiment lexicons in enhancing the performance of deep learning models could advance our understanding of emotional analysis in textual data and improve depression detection algorithms.

Keywords emotional analysis, Hebrew texts, machine learning, psychological feature lexicons, and depression detection

REFERENCES

Aisopos, F., Tzannetos, D., Violos, J., & Varvarigou, T. (2016, March). Using n-gram graphs for sentiment analysis: an extended study on Twitter. *In 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 44-51). IEEE. https://doi.org/10.1109/BigDataService.2016.13

Amram, A., David, A. B., & Tsarfaty, R. (2018, August). Representations and architectures in neural sentiment analysis for morphologically rich languages: A case study from Modern Hebrew. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 2242-2252).

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM, 56*(4), 82-89. https://doi.org/10.1145/2436256.2436274

HaCohen-Kerner, Y., Miller, D., Yigal, Y., & Shayovitz, E. (2018). Cross-domain Authorship Attribution: Author Identification using char sequences, word unigrams, and POS-tags features. *Working Notes of CLEF, 2125(1), 1-25.*

Liu, B. (2012). *Sentiment analysis and opinion mining.* Springer. https://doi.org/10.1007/978-3-031-02145-9

Shapira, N., Atzil-Slonim, D., Juravski, D., Baruch, M., Stolowicz-Melman, D., Paz, A., Alfi-Yogev, T., Azoulay, R., Singer, A., Revivo, M., Dahbash, C., Dayan, L., Naim, N., Gez, L., Yanai, B., Maman, A., Nadaf, A., Sarfati, E., Baloum, A., ... Goldberg, Y. (2021, June). Hebrew psychological lexicons. *In Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access* (pp. 55-69). https://doi.org/10.18653/v1/2021.clpsych-1.6

Tsarfaty, R., Seker, A., Sadde, S., & Klein, S. (2019). What's wrong with Hebrew NLP? And how to make it right. *arXiv preprint arXiv:1908.05453*. https://doi.org/10.18653/v1/D19-3044

Yang, L., Li, Y., Wang, J., & Sherrat, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access, 8*, 23522-23530. https://doi.org/10.1109/ACCESS.2020.2969854

APPENDIX – THE LEXICONS NAMES AND SIZE

Name	Description	Amount
ANGER	Words related to expressing anger or irritation	230
ANXIETY	Words associated with feelings of anxiety and nervousness	241
ASHAMED	Words reflecting a sense of shame or embarrassment	175
CALM	Words indicative of a calm and composed emotional state	160
CONFUSION	Words representing a state of confusion or bewilderment	175
DEPRESSIVE	Words associated with feelings of depression	162
DISGUST	Words expressing a sense of strong dislike or revulsion	191
EMOJI_NEG	Emojis conveying negative emotions	297
EMOJI_NEU	Emojis conveying neutral emotions	226
EMOJI_POS	Emojis conveying positive emotions	511
FATIGUE	Words related to feelings of tiredness or exhaustion	212
GUILTY	Words indicating a sense of guilt or remorse	183
HOSTILE	Words reflecting a hostile or aggressive attitude	179
JOY	Words associated with feelings of joy and happiness	207
NEG	Negative sentiment words	1626
NEG2	Additional negative sentiment words	115
NERVOUS	Words expressing nervousness or apprehension	214
NOTAMUSED	Words conveying a lack of amusement or boredom	111
NOTANTICIPATION	Words indicating a lack of anticipation	106
NOTCALM	Words suggesting a lack of calmness or tranquility	187
NOTCONTENTMENT	Words indicating a lack of contentment	201
NOTINTERESTED	Words conveying a lack of interest or enthusiasm	151
NOTJOY	Words indicating a lack of joy	365
NOTNERVOUS	Words reflecting a lack of nervousness	158
NOTPROUD	Words indicating a lack of pride	117
NOTTRUST	Words suggesting a lack of trust	141
NOTVIGOR	Words indicating a lack of vigor or energy	172
PARALINGUISTIC	Words related to paralinguistic features, such as tone or intonation	150
POS	Positive sentiment words	906
POS2	Additional positive sentiment words	82
PROUD	Words expressing a sense of pride	153
SAD	Words associated with feelings of sadness	203
SURPRISE	Words reflecting a sense of surprise	138
TRUST	Words associated with feelings of trust and confidence	156

AUTHORS

Ron Keinan was born in Ashdod, Israel in 1992. He received a bachelor’s degree in software engineering from the Jerusalem College of Technology (JCT), Jerusalem 2022. Now he is currently working as a firmware developer in Intel corporation and studying for his master's degree in data mining in JCT, Including a thesis research in the area of Big Data Mining and Hebrew NLP, under the supervision of Prof. Dan Bouhnik and Dr. Efraim Margalit.

Efraim Margalit was born in Jerusalem, Israel in 1961. He received a B.A. degree in Natural Sciences and Mathematics from the Open University of Israel, in 2000. He received an MA degree in 2008 and a Ph.D. degree in 2018. In Information Sciences from Bar-Ilan University, Ramat Gan, Israel.

Dr. Margalit has over four decades of experience in development, management, and research in the field of computer science in the private and public sectors and the academy.

He served as a computerized content researcher in a research project conducted at the Faculty of Law at Tel Aviv University and supported by the Israel Science Foundation.

In recent years, Dr. Margalit has served as a lecturer for AI and DM at the Faculty of Engineering at Jerusalem College of Technology.

His research fields are Computational Linguistics and Artificial Intelligence.

Dan Bouhnik received a B.Sc. degree in Computer Science from the Jerusalem College of Technology in 1981 and a M.Sc. Degree in Computer Science from Bar-Ilan University in 1988, and a Ph.D. degree in Computer Science from Pacific Southern University, in 1991.

He is the author of several books used for teaching advanced computer sciences.

He is currently an Associate Professor and serves as the Rector of the Jerusalem College of Technology. His main research interests include Information Science, Knowledge Management. Information Security and Big Data issues and their effect on society.