Paper Review on Emotional Intensity Estimation based on Writer's Personality Written by Haruya Suzuki, Sora Tarumoto, Tomoyuki Kajiwara,Takashi Ninomiya, Yuta Nakashima, Hajime Nagahara.

Section 1: Introduction-In the introduction, issues with measuring emotional intensity including the influence of the writer's personality on expressing emotions are considered. The proposed model includes Big Five personality traits outperforming the baselines and achieving state of the art result on the WRIME corpus. This study also creatively integrates personal data, heightening subjective emotional vividness evaluation within SMT.

Section 2: Related Works-The Related Work section succinctly addresses the impact of personal biases on emotional expression, emphasizing factors like language, age, education, gender, race, and personality. The review effectively outlines recent studies, such as Milkowski et al. (2021) and Kajiwara et al. (2021), which personalized emotion analysis by considering annotator labeling variations and the text writer's personality. The paper positions itself as an advancement, proposing a more effective method to model personality information for emotional intensity estimation.

Section 3: Methods-The Methods section introduces a novel model for personality-aware emotional intensity estimation, integrating both text and personality streams. Leveraging a BERT-based text stream for emotional intensity classification, the personality stream employs a multilayer perceptron. Fusion strategies, including difference, product, and concatenation, are explored. The proposed approach achieves emotional intensity estimation by employing a softmax classifier on the fused feature vectors. The detailed and structured methodology reflects a thoughtful design to capture the nuanced interplay between writers' personality and emotional expression.

Section 4: Experiments-In the Experiments section, the paper meticulously details its experimental setup for fair comparison, employing 35,000 SNS posts from 60 writers. The authors split the dataset for training, validation, and evaluation, considering past posts for user representation in PEB. Implementation details cover models such as Bag-of-Words (BoW) and BERT, fine-tuned with writers' personality features. Comparative methods include Pc, Pa, and PEB from previous studies. The chosen evaluation metric is quadratic weighted kappa. These comprehensive details lay the groundwork for a rigorous evaluation, ensuring a thorough assessment of the proposed method against existing approaches.

Section 5: Results-The results reveal consistent improvement in emotional intensity estimation by incorporating the writer's personality. Proposed methods outperform baselines, with the All method achieving notable success. Combining personality with PEB shows synergistic benefits, yielding performance comparable to human annotators. Examples underscore the model's effectiveness in capturing emotional nuances.

Section 6: Conclusion-The conclusion of the paper on emotional intensity estimation based on writers' personality emphasizes the proposed model's success in incorporating writer personality for improved emotional intensity estimation. Future work will explore methods for estimating personality from past posts and integrating them into the current approach.

Paper Review on RedHOT: A Corpus of Annotated Medical Questions, Experiences, and

Claims on social media Written by Somin Wadhwa, Vivek Khetan, Silvio Amir, Byron C. Wallace.

Section 1: Introduction- The Introduction of "RedHOT: A Corpus of Annotated Medical Questions, Experiences, and Claims on Social Media" presents a valuable dataset of 22,000 health-related Reddit posts across diverse medical conditions. Addressing the dual nature of social media—facilitating support and disseminating misinformation—the paper introduces RedHOT-DER, a tool to identify, contextualize, and retrieve scientific evidence for health-related claims. The dataset's sensitivity is acknowledged, emphasizing ethical considerations and IRB consultation. The paper establishes a comprehensive foundation for understanding and mitigating health-related misinformation on social media.

Section 2: Datasets- The dataset section of RedHOT outlines a meticulous two-stage annotation process on health-related Reddit posts. The multi-label scheme captures claims, questions, and personal experiences, with a second stage focusing on Population, Intervention, and Outcome elements within "pure claims." Crowdsourcing via Amazon Mechanical Turk, quality validation measures, and dataset details, including diverse health populations, contribute to a comprehensive and well-structured resource for advancing language technologies in health-related social media analysis.

Section 3: Task and Evaluation- The task and evaluation section of RedHOT delineates a three-step approach, focusing on identifying medical claims on social media and retrieving relevant evidence. Utilizing sequence tagging models and a novel Dense Evidence Retriever (RedHOT-DER), the paper addresses content moderation needs and proposes a method for evidence retrieval, crucial for fact-checking health-related claims. The comprehensive evaluation involves assessing claim identification and PIO extraction, as well as expert-evaluated evidence retrieval. The manual relevance judgments by medical doctors indicate the efficacy of RedHOT-DER, showcasing its potential for accurately retrieving evidence relevant to health-related claims on social media.

Section 4: Related Works- The related works section of RedHOT provides a comprehensive overview of prior research on claim validation, emphasizing the two-step process involving evidence retrieval and subsequent prediction. It highlights scientific claim verification efforts, such as SciFact and COVID-19-related studies, and underscores the reliance on crowdworkers for annotating health-related texts. Additionally, it notes the existence of prior health-related Reddit corpora, showcasing the evolving landscape of research in this domain.

Section 5: Conclusion- In conclusion, RedHOT introduces a valuable dataset for language technology development in health-related social media. While acknowledging dataset limitations, the paper details efforts to ensure ethical data use, including privacy considerations. The authors demonstrate the dataset's utility in training models for post categorization and claim retrieval, emphasizing its potential impact on content moderation and insights into patient experiences. Ethical considerations are addressed, highlighting a conscientious approach to user privacy and consent.