MoralStrength: Exploiting a Moral Lexicon and Embedding Similarity for Moral Foundations Prediction

Oscar Araque, Lorenzo Gatti, Kyriaki Kalimeri
Intelligent Systems Group, Universidad Politecnica de Madrid, Madrid, Spain
Human Media Interaction Lab, University of Twente, Enschede, The Netherlands
Data Science & Digital Philanthropies Laboratory, ISI Foundation, Turin, Italy

SCAN ME

Link to Preprint!!

Research Question

Can we better predict the moral rhetoric in user-generated text?

Moral values influence the way we rationalize and take a stance upon controversial topics, like a**bortion, homosexuality, climate change,** or even **vaccine hesitancy**.

They are also closely related to our **political views** and the opinion formation mechanisms regarding **immigration, political extremism, and poverty.**

Here we propose the new ***MoralStrength lexicon*** for morality analysis.

Moral Foundations Theory

Care/Harm: virtues of caring and compassion.
Fairness/Cheating: unfair treatment, inequality, notions of justice.
Loyalty/Betrayal: obligations of group membership, loyalty, vigilance against betrayal.
Authority/Subversion: social order, obligations of hierarchical relationships such as obedience, respect
Purity/Degradation: physical and spiritual contagion, virtues of chastity, wholesomeness and control of desires.
Liberty/Oppression: feelings of reactance and resentment people feel toward those who dominate them and restrict their liberty

Moral Foundations Dictionary (MFD)

(i) a limited amount of lemmas and stem of words
(ii) radical lemmas rarely used in everyday language, e.g. homologous, apostasy
(iii) an association with a moral bipolar scale, so-called vice and virtue, but without any indication of **strength**.

MoralStrength Dictionary

(i) contains 5 times more lemmas with respect to the MFD (~1000)
(ii) expansion via WordNet including common use words
(iii) human annotations of "strength" in a Likert-Scale for all lemmas

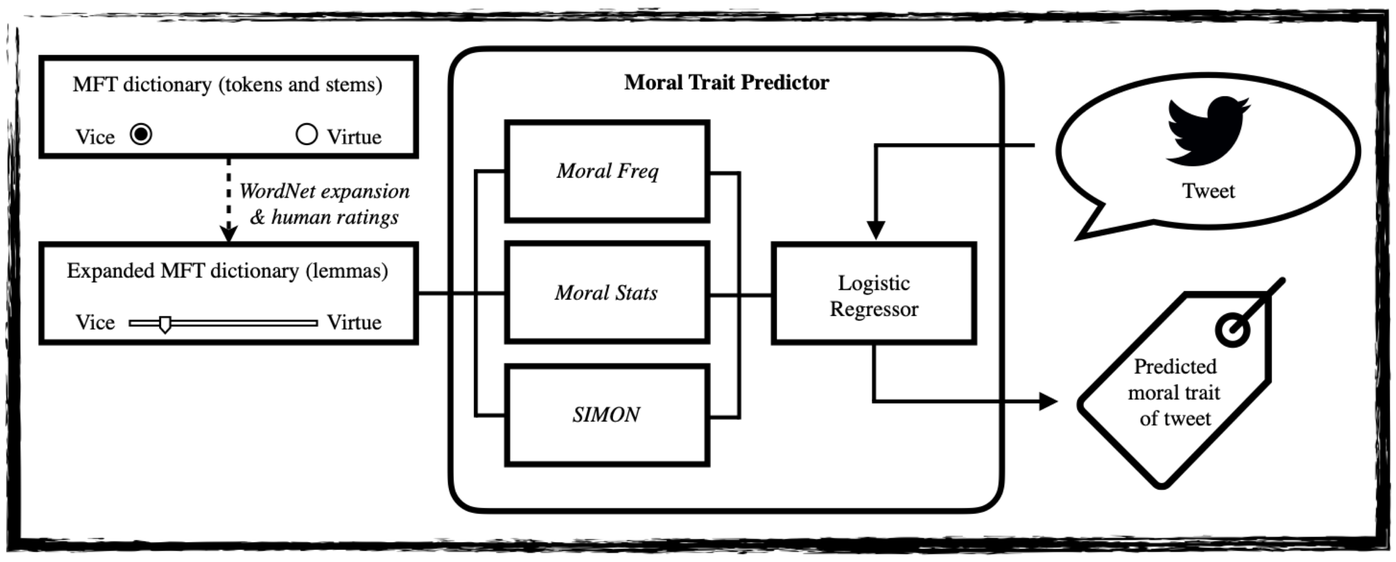
Evaluation

We evaluated our framework on the Moral Foundations Twitter Corpus which conists of 7 datasets of various topics and contains approximately **35,000 annotated tweets.**

We propose three approaches of increasing complexity which employ the MoralStrength lexicon to predict the moral rhetoric:

Moral Freq: frequency counts of the lemmas
Moral Stats: statistical summary of the lemmas
SIMON: word embedding similarity based representations

Our Framework



Baseline Models: Unigrams and frequency counts with MFD
Simple Models: Moral Freq, Moral Stats, SIMON
Combined Models: SIMON + Moral Freq, SIMON + Moral Stats, SIMON + Moral Freq + Moral Stats

Outperforming the current state-of-the-art.
On average F1-score of 86.25% vs 44,30% (p-value< 0.01) over all datasets.

Take Home Message

MoralStrength provides a tool for a more in-depth understanding of the moral narratives.

Still, there are many points for further research since context, culture, and medium may affect the expression of morality.

ISI Foundation
POLITÉCNICA
UNIVERSITY OF TWENTE.

We are at EUROCSS! Come chat with us:
@KyriakiKalimeri & @Friggigatto