



<https://www.medianeuroscience.org>
<https://github.com/medianeuroscience/emfdscore>

This manual provides an overview of how different moral information metrics are extracted from textual input via eMFDscore and offers guidelines for researchers when interpreting eMFD metrics as returned by the [Global Database of Events, Language and Tone](#) (GDELT; Leetaru & Schrodt, 2013) [Global Content Analysis Measures \(GCAM\)](#) service. We encourage researchers using GDELT GCAM for studying moral content in news articles to consider the guidelines outlined in this manual carefully and further invite suggestions for improvement.

For more detailed background information on the extended Moral Foundations Dictionary (eMFD), please consult the [publication](#).

Foundation Probability Scores Interpretation

The eMFD calculates foundation probability scores at the word level. However, the prevalence of these foundations is reported in GCAM at the document level (e.g. a news article). Thus, codes v.42.2, v.42.3, v.42.4, v.42.5, & v.42.6 return document-level foundation probability scores for each of the corresponding 5 moral foundations (care_p, fairness_p, loyalty_p, authority_p, & sanctity_p). This calculation is formalized as below.

$$d_{p_i r} = \frac{\sum w_{p_i}}{n_m}$$

where d_{p_i} is the document-level foundation probability score for a unique foundation i and unique document r , w_{p_i} refers to the word-level foundation probability score for a unique foundation i , and n_m is the total number of moral words detected across the news article r .

To exemplify this procedure, consider a document that contains 10 moral words. Each word has a uniquely associated foundation probability score across all 5 foundations. If the sum of all care_p scores (at the word level) is equal to 2.4, then this sum divided by the total number of moral words (in this case 10) gives us the document-level foundation probability score which will be equal to 0.24.

This procedure can be confusing for a keen observer, who would intuitively expect the foundation probabilities across all foundations to add up to 1. As the above explanation implies, the sum of probability scores for all foundations could “theoretically” range between 0 - 5. In

the impossible scenario where all detected moral words in a document would have a foundation probability of 0 for each individual foundation, the sum across all foundations would also be 0.

Similarly, in another impossible scenario in which all detected moral words would have a foundation probability of 1 for each individual foundation, the sum across all foundations would be 5. Because all “moral words” in the eMFD possess a foundation probability score of greater than 0 in at least one foundation, and there are maximum foundation probability scores for words in each foundation, the actual possible range for the sum of document level foundation probability scores can be conceptualized as $0 \leq \sum d_{p_i} \leq 5$.

Moral and Nonmoral Words

The greater the size of a document and the more moral words a document contains, the more reliable is the measurement of a document’s moral content. We recommend that researchers monitor the total number of *moral* words detected (n_m), the total number of words in general

(code *wc* in GCAM), and the moral to non-moral word ratio ($\frac{n_m}{(wc - n_m)}$) when interpreting

their results. The word count criteria used to filter for documents with sufficiently reliable measurement of moral content can vary from application to application. In our experience, excluding documents that fall outside the 10th and 90th percentile of the documents’ word count distribution seems advantageous. Of course, this depends on the specific word count distribution in a project and researchers may want to adjust this filter range, especially if the word count distribution is heavily skewed.

Foundation Sentiment Scores Interpretation

Words in virtue categories usually describe morally righteous actions, whereas words in vice categories typically are associated with moral violations. Each moral word in the eMFD is associated with manual annotation highlights (see Hopp et al., 2020) that contain the context within which it appeared. Each word is assigned 5 sentiment scores that denote the average sentiment of the foundation context in which this word appeared. For example, the word "kill" has an average "care_sent" of -0.69, meaning that all "care-harm" highlights in eMFD’s underlying annotation procedure in which "kill" appeared had an average, negative sentiment of -0.69. However, as with foundation probabilities, the prevalence of these foundation sentiments is reported in GCAM at the document level (e.g. a news article). Thus, codes v.42.7, v.42.8, v.42.9,

v.42.10, & v.42.11 return document-level foundation sentiment scores (care_sent, fairness_sent, loyalty_sent, authority_sent, & sanctity_sent). This calculation is formalized as below.

$$d_{s_i r} = \frac{\sum w_{s_i}}{n_m}$$

where d_{s_i} is the document-level foundation sentiment score for a unique foundation i and unique document r , w_{s_i} refers to the word-level sentiment score for a unique foundation i , and n_m is the total number of moral words detected across the news article r .

Based on our experience, foundation sentiment scores are likely to be correlated with foundation probability scores. Ideally, moral sentiment (valence) and moral signal (intensity) should be uncorrelated, but this is an unrealistic assumption given specific text corpora. Therefore, we strongly recommend researchers to avoid the interpretation of moral sentiment independent of moral signal as the analysis of moral sentiment alone provides limited utility and may run the risk of researchers arriving at misleading conclusions.

GDELT and eMFDscore

If you have used the eMFD using the eMFDscore software tool

(<https://github.com/medianeuroscience/emfdscore>), please note the following:

In GDELT, the eMFD is being applied as if you ran the *bow emfd all sentiment* option in eMFDscore. Please check cell 3 & 4 in the [tutorial](#). Note that GDELT returns the scores per DOCUMENT (not word!), and hence you receive the average 5 probabilities and the average 5 sentiments per DOCUMENT. With this data, it is only possible to obtain either VIRTUE or VICE scores per document (because the average sentiment of the document is either positive or negative).

eMFD Historical Backfile

Recently, a 3.5 year eMFD historical backfile was made available that computed eMFD scores for more than 634 million news articles, across 65 languages, that GDELT has monitored since January 1, 2017. More information about this database can be found at <https://blog.gdeltproject.org/a-3-5-year-634-million-article-historical-backfile-for-the-extended-moral-foundation-dictionary/>. Currently, the eMFD historical backfile includes codes v42.2 -

v42.11 which correspond to the eMFD computed foundation and sentiment probability scores . The eMFD historical backfile does not include code c.42.1 (moralwordcount) as it does in the GCAM string returned from the GKG 2.0 API

(<https://blog.gdelproject.org/gdelt-doc-2-0-api-debuts/>). We highlight this observation keeping in mind the influence moral and non-moral word counts might have, as reasoned above, on the reliability of documents' moral content scoring and subsequent interpretations. If researchers want to filter documents based on word counts as explained above, then - at this time - there is no other option than to parse the original GCAM string returned from the GKG 2.0 API.

Translated Articles

We observed that some foreign language news articles being returned by the historical backfile are actually flagged as English language news articles and scored. For example, the language of origin in both of the URLs mentioned below is erroneously identified as “english”.

1. <https://tenasia.hankyung.com/topic/article/2020012129044>
2. <https://www.tagesspiegel.de/berlin/polizei-justiz/tumult-am-unfallort-in-neukoelln-fussgaengerin-von-auto-angefahren-und-schwer-verletzt/25571082.html>)

We have not conducted an exhaustive analysis of this observation, and we are unsure if the scoring by GDEL T uses the “machine-translated in English” versions of the wrongly labeled documents. However, for the sake of ensuring data quality we encourage researchers to remain mindful of this potential limitation.