Mimicry Detection
Automated and Reliable Language Style Matching
Tom Carrick

Mimicry

Mimicry in social interactions

When interacting with others, humans show a tendency to mimic one another's behaviour, including speech, physical actions and written text. Mimicry is also a measure of social cohesion and 'liking'.

Detection of mimicry can be a valuable indicator in conversations between two people as well as group dynamics in collaborative settings.

Potential applications of mimicry detection

Increased levels of mimicry within a pair or group has been shown to indicate:

- Increased rapport and group cohesion
- Likelihood of initiating a romantic relationship
- Happiness and stability of romantic relationships
- Increased engagement and performance in group tasks
- Having a strong emotional involvement in a conversation
- Increased likelihood of confessions in police interviews
- Greater child attachment security in mother/child pairs.

In addition, decreased levels of mimicry compared to the rest of the group has been shown to be an indicator of an insider threat.

Language Style Matching

How LSM works

Language Style Matching (LSM) is a prominent method of classifying linguistic mimicry numerically. First texts or transcripts of verbal communications are aggregated into one file for each member of the group.

Each file is read and the number of words for each of the following classes of function words are counted:

- Adverbs (about, also, maybe, usually)
- Articles (a, an, the)
- Auxiliary verbs (could, do, have, must)
- Conjunctions (also, if, or while)
- Impersonal pronouns (anyone, it, that, what)
- Negations (can't, don't, no, not, without)
- Personal pronouns (he, me, she they, you)
- Prepositions (about, in, of, with)
- Quantifiers (all, less, many, whole).

These counts are normalised for message length and compared with scores from the other person to formulate a final LSM score.

Limitations of LSM

While LSM is a popular and useful measure, it has some significant drawbacks that limit its usefulness. All of these drawbacks relate to LSM's reliance on Linguistics Inquiry and Word Count (LIWC), a linguistics software package.

One major limitation is that LIWC makes LSM difficult to automate and customise for different uses. Another problem is that LIWC uses a static list of function words and classes, which leads to the following limitations:

- As many words have multiple meanings, some words can be counted (or not) incorrectly.
- LSM is intolerant of incorrect spelling, typos and most forms of text-speak, making it unsuitable for analysing online discourse.
- LSM can currently only be used to analyse texts in the English language.

Improving Language Style Matching

The primary expected output of this project is to develop a new variant of Language Style Matching with the following improvements to address the observed limitations.

Automated

Ability to integrate with standard Natural Language Processing tools and libraries means that the entire process can be automated, without first pulling texts into spreadsheets, etc. For example, this could enable LSM to be run as part of a monitoring system.

Context-aware

As LSM matches words against a static word list, it isn't aware of the context the word appears in. There are many words in English that have multiple meanings depending on the context, sometimes as a function word and sometimes not. By moving away from a word list, it is possible to get more accurate results.

More resilient

Related to the use of a static word list, LSM is currently very intolerant to common typos and misspellings — the word list assumes everything is spelled correctly. While some 'text speak' is included in the word list, it's highly inexhaustive.

Moving away from a word list will fix these problems and make the new LSM more resilient and importantly a much more appropriate tool for working with online, typed and poorly transcribed texts.

International

Currently LSM only works with the English language as LIWC's function word lists only exist in English. Currently it is not known if LSM would work in other languages, and creating a function word list for other languages would be a time-consuming process.

However, standard tools already exist that can extract function words with classes in a similar way, which could open up mimicry research with LSM into many different languages.