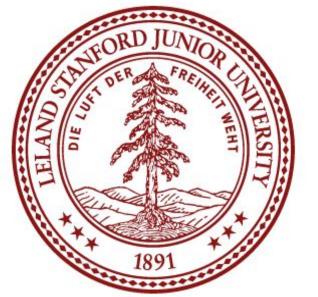
Predicting Myers-Briggs Type Indicator with Text Classification

Team: Ian Knight, Rayne Hernandez

CS 224N: Natural Language Processing with Deep Learning



Problem Description



- The Myers-Briggs Type Indicator (MBTI) is a psychological model of human personality with 16 possible personality
- Personality types are a combination of 4 different binary categorizations (e.g., Introverted vs. Extroverted)

Approach

Preprocessing

Proportionality

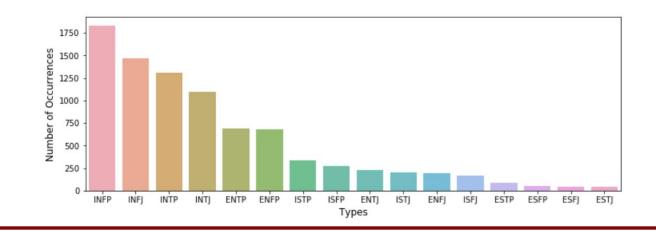
Selective Word Removal



• Using ~430,000 social media posts from the web forum **PersonalityCafe.com**, we trained a text classifier capable of sorting people into their personality type based on their social media posts

Challenges

The data we acquired from PersonalityCafe tended to be skewed towards certain personality types, which forced us to artificially balance our data

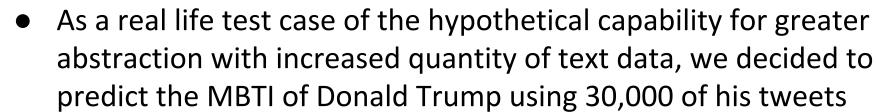


Results

Baseline: Multinomial Naive Bayes

Naive Bayes Accuracy **Recurrent Neural Network** N vs. S I vs. E **RNN Accuracy**

Extrapolation: Donald Trump's Twitter Account

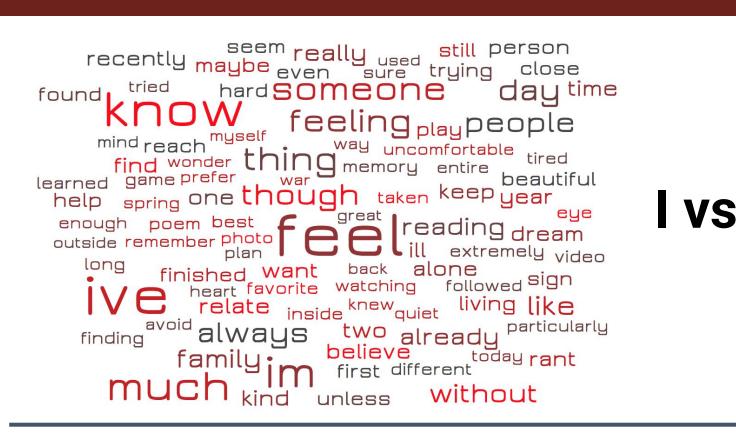


- The result matched what MBTI experts claim to be Trump's type.
- ESTP is the archetype known as "the Entrepreneur"

65 -

0.5533546805381775 0.5134227871894836 0.5419855117797852 0.47002798318862915 Final prediction: ESTP

Extrapolation: Word Clouds



making post thinking S I tsaid welcome reason bility attractive guys Leader extrovert thats read forum

experience life yet social id world idea exactly world around recently world idea exactly term idea exactly term time trying honest seem meaning three people one find believe hard laways understand everything part hope many mind

refuse Want shit dont load sleep keep fit description option. prettycooking option includes math clothing imp home rant google Wellwhats playing car adventure car adventure last hobby illjob functions kid pay WOr Kyeah

practical inferior kill style edit regular

Lose preferlike though

pick bill lose mart set common post

cant effect tried short look new currently sport enjoy thing party fucking high

everything moment together heart

know maybe value life others

think around well right thanks person someone close ever scared anyone much people felt fall definitely beautiful dear especially tend sound sharing wish met loved thats song want sad hear sorry lol husband put year haha lately ivereally methough feeling

good figure nice amazing yes best

depressed bit relate sometimes today boyfriend seem poem hi tired hard going kind lot day different time thing everyone hurt awesome on open open happy help friend world thank always totally always totally

called interest talking ask meaning piss fucking whatever logiceducation

fucking physical logicalengineering actually attention

done power huge vote clearly

said nope issue rule expect money study

worth law one every study

claim here work problem asks

case post

topic fact cant number trait

write certain math rant anything

imagine compare intellectual debate merely imagine compare intellectual debate saying office Source intelligence therefore convince previous pay sense process idea

probably intelligence score make bit hang really fucking smoking example call fun scared affect thats ever fill lack live circumstances pretty smoke playing never like go music fall hit mom rather shit problem stuff sport bible anything still bored lets bought extroversion use list creat weed belief extroversion use list crazy weed bought laugh religion talk whole answering theyre rarely guyill lazy realized normally king actually joke party xd medications spend serious school drive around game neutral watch

going take extraverted christmas eventually made hope definitely position act relationship fully help described experience issue let one vision sorry ive detail long work see nothing

people due public relate glad process right type

world rant yes close order goal living book insight finding spiritual need thank little introverted reading please want cognitive give photo share others first mother recently hear seeing perhaps year keep thanks understand

told truly different

life difficult

Conclusion

- Our work suggests that there is an underlying linguistic component to MBTI that is not fully captured by means of traditional assessment methodology
- Our model offers fast MBTI classification of massive amounts of publicly available text online
- our model represents a trade-off of between two aspects: we achieve lower rates of perfect classification in exchange for higher rates of approximately correct classification (i.e. "good" classification). This latter aspect is a positive feature that multiclass models fail to achieve

Lemmatization **Tokenization Padding** Model OUTPUT **Embedding** Dense **LSTM** Layer Layer

Acknowledgements

We would like to thank Professor Richard Socher for being our advisor.