Detecting personal attributes through analyzing online forums

Zhilin Wang, Xiaodong Wu, Weizhe Lin and Elena Rastorgueva

Faculty of Education

Introduction

Personal attributes refers to specific ways of thinking/feeling/behavior that individuals have, formed from the interaction between a person's innate tendencies and their need to adapt to their environment. An example of such attributes include various types of problems individuals face in their lives such as particular relationship dynamics as well as reasons for partaking in various social activities.

In the case of understanding problems that individuals face, it might be great value to categorize individual experiences such that individuals can feel that they are not alone in their experiences and can learn how others have attempted to mitigate similar problems (Bandura, 1977).

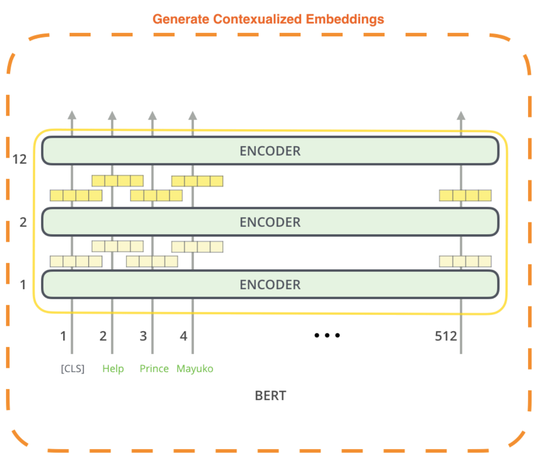
Human judges have long been able to classify many of such attributes (McAdams, Hoffman, Mansfield & Day, 1996; Blagov & Singer, 2004) but automated textual analysis (NLP) techniques have only been successful at a few (Bosson, Swann & Pennebaker, 2000). This is possibly due to the extensive use of the word-count approach (Pennebaker, Boyd, Jordan & Blackburn, 2015), which has largely been replaced by advanced language models in many NLP tasks.

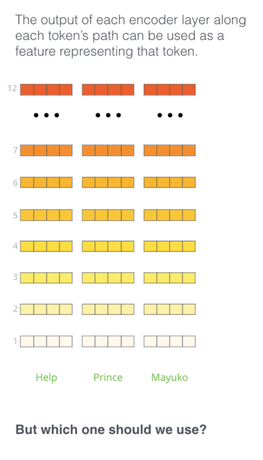
These language models have achieved SoTA performance in many common NLP tasks (Question Answering, Translation etc), demonstrating that they have achieved a great extent of semantic and syntactic 'understanding'. This suggests that these language models might be capable of interpreting and categorizing personal attributes from writing in a similar manner as an experienced counselor/ psychotherapist.

In this research, we will seek to apply recent advances in language models, particularly the BERT model to understand personal attributes of Reddit authors who have sought for advice for their personal situations.

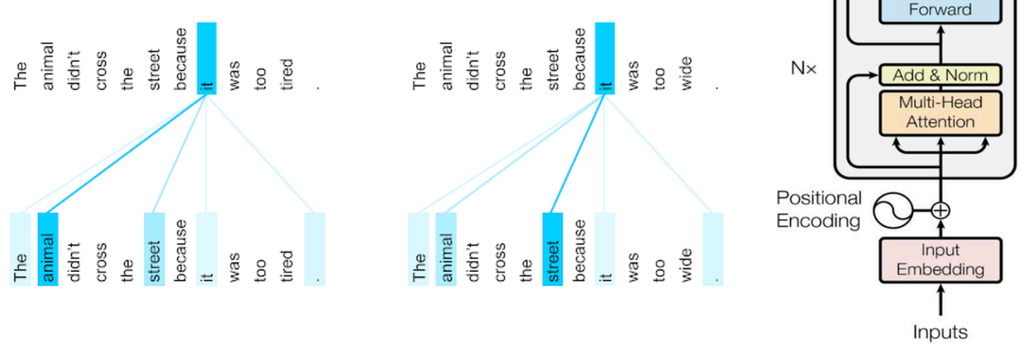
Methods

1. Mining >8000 posts from r/Advice of Reddit
2. Generate 768-length vector embedding of each post using Google BERT



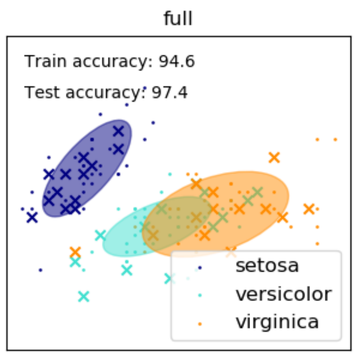


BERT Architecture by Alammar, Jay (2018). The Illustrated Transformer [Blog post]. Retrieved from https://jalammar.github.io/illustrated-transformer/



Detail of encoder: Transformer aka 'Attention' by Google AI Blog

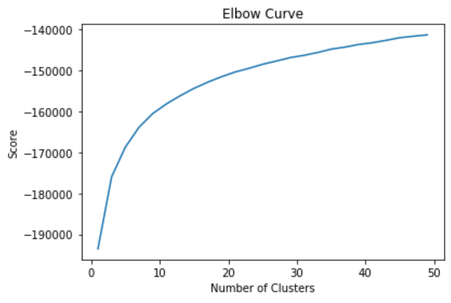
3. Clustering vector embedding using Gaussian Mixture Models



How do Gaussian Mixtures Models work? by Scikit Learn

Relative to KMeans

- Variance in each dimension can be different: 'ellipsis' rather than 'sphere'
- This can detect particular personality attributes that depends on covariance between vector elements
- More even cluster sizes

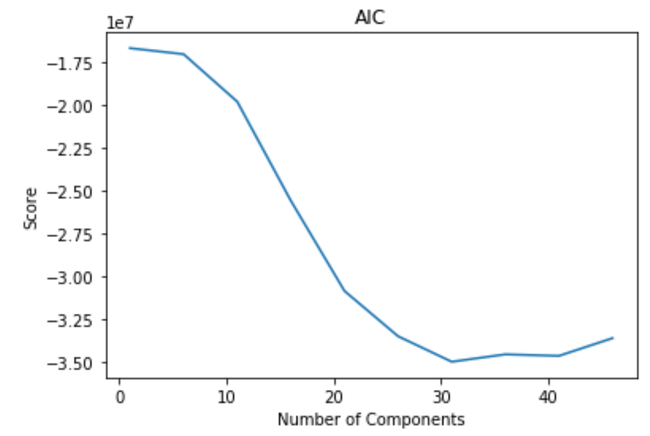


Sum of distance to centroids for KMeans models with varying number of clusters (higher is better). Note lack of distinct elbow-shaped bent

Results

Titles of posts with vectors most similar to "How do I get Vaccinated as a Minor?"

1. Hi, I'm 13, unvaccinated, and damn it seems like I have a lot of stuff to catch up on. What do I do?
2. Lost Health Insurance at 21. Advice?
3. How do I tell my parents I don't believe in god?
4. Should I get vaccinated?
5. My religion is taking away my friends



Akaike Information Criteria for Gaussian Mixture Models with variable number of components (lower is better). Note: plateau at 30 components.

Labels for each component of the Gaussian Mixture Model (Human-labelled)

component 0: toxic friendships
component 1: some form of abuse/harassment by family/friends (author are usually early adolescents)
component 2: best/girl/boy friend issues with some form harassment
component 3: job related but more short term and minor
component 4: girl/boyfriend --> breaking up
component 5: biological related stuff such as illness as well as use of various substances but less severe
component 6: academics/work related but more long term, not knowing how to make a decision
component 7: House/neighbour related
component 8: mental health but more severe
component 9: disappointment with best/girl/boy friend
component 10: supporting best/girl/boy friend
component 11: sexual preferences and behaviors
component 12: dilemma between what the person wants and what significant others including family wants
component 13: issues with trying to get into a relationship
component 14: befriending others
component 15: embarrassment related
component 16: tensions at work/school
component 17: self-improvement
component 18: harassment by people they don't know
component 19: mental health but very severe such as PTSD, depression and suicide
component 20: adolescents unsure about future
component 21: tensions with boy/girlfriend but not yet breaking up
component 22: gaming, videos, internet
component 23: family tensions
component 24: abuse in the family (author are usually adults)
component 25: teenage issues with a focus on relationship
component 26: Transitioning from high school --> university or university to job
component 27: teenage issues with a focus on biological aspects
component 28: Issues relating to money
component 29: motivation (or the lack thereof) for high school/university

Discussion

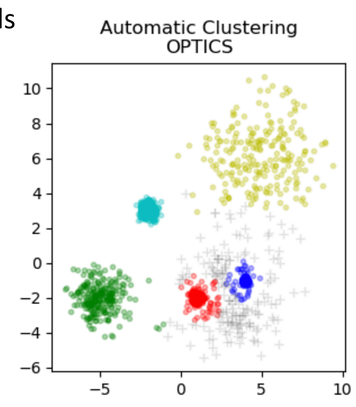
Preliminary experiments suggest that language models can to a large extent categorize personal attributes, at least in terms of personal problems. However, the occurrence of misclassification suggest the need for more investigation to be conducted in optimizing their perform. These are some of the changes we hope to experiment with in the short term.

Alternative embedding methods

We hope to experiment with alternative language models including OpenAI's GPT-2, which is architecturally similar to BERT but pre-trained on a different corpus. In addition, we will also try traditional word count models that emphasize different semantic aspects such as SPLICE (Farnadi et al., 2016), which focuses on author self-evaluation and MRC psycholinguistic database (Coltheart, 1981) which provides semantic, phonological and syntactic information of words

Alternative clustering methods

We hope to experiment with density-based clustering such as DBSCAN/ OPTICS as well as tree-based clustering such as Birch (in place of Gaussian Mixture Models).



Aggregate multiple posts by the same author

This might be able to remove the influence of writing style and other personal attribute-irrelevant 'noise' by establishing a baseline from which the post in question can be removed of.

Use of supervised or semi-supervised learning

Our current approach is wholly unsupervised. This has the advantage of discovering features that are not intuitive to humans. However, a supervised/semisupervised approach might be more accurate because the existence of a clear training objective can lead to more optimized model selection. This can be done by mining for posts from specific subreddits such as r/family or r/ breakups. Alternatively, we can use existing coding schemes for humans as well as their accompanying examples and see if language models can interpret them adequately (Demorest, Crits-Christoph, Hatch & Luborsky, 1999; Hy & Loevinger, 1996; McClelland & Koestner, 1992; McAdams, 1992).