

Active Learning To Reduce Human Labeling For Automatic Psychological Text Classification

Jahson O'Dwyer Wha Binda (j.t.odwyerwhabinda@student.tudelft.nl)

Supervisors: Willem-Paul Brinkman & Merijn Bruijnes

1. Background

1.1 Schema-Therapy

Psychotherapy for the treatment of personality and psychological disorders. Where **Schema** is known as an unhealthy pattern of thoughts and behaviours.

1.2 Labelling Problem

To train AI models to detect a person's schema training data is necessary, human labelling of user stories can be very expensive.

A possible solution to reducing the quantity of human labeling is through the use of **Active Machine Learning**

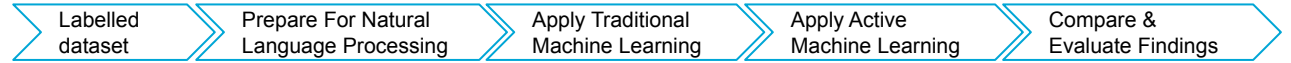
1.3 Active Machine Learning

Special case of machine learning where during training the agent will **query** the user for only the labels of a data point it deems most useful for its classification.

1.4 Research Question

How well can active machine learning be used to support human labelling of a data set?

2. Approach



3. DataSets

1. David Allaart's Schema Data Set [1]
 2. Franziska Burger's Schema Data Set [2]
 3. Sentiment140's Twitter Data Set [3]
- *Schema Data sets pre-processing inspired by Burger's & Allaart's research.

4. Why Twitter Data

Control experiment due to possible limitations on Schema Classification.

Labels Tweets to a **Positive** or **Negative** Sentiment

Tweets: Similar structure to user stories for schema.
Sentiment: Similar to schema.

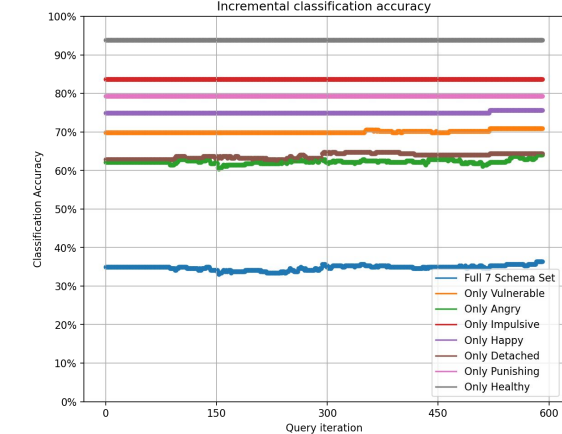
5. Results (Target Accuracy: +/- 3% of Traditional ML)

98%+ Reduction In All Cases

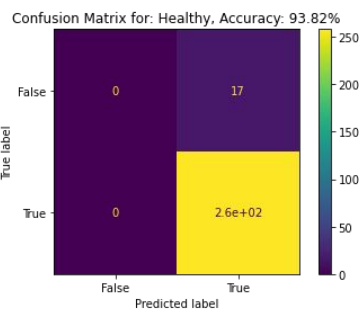
Allart's Data	Accuracy	Labeled Data	Burger's Data	Accuracy	Labeled Data	Twitter Data	Accuracy	Labeled Data
Traditional ML	37%	1100	Traditional ML	21%	4151	Traditional ML	76%	1.28 Million
Active ML	35%	8	Active ML	21%	18	Active ML	73%	20120

6. Findings Allaart's Data

- Target Accuracy Achieved Instantly

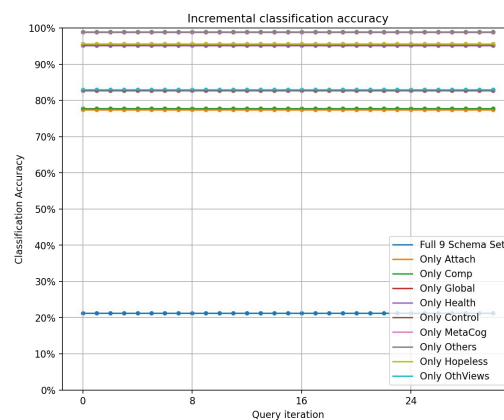


- Target Model is very simple

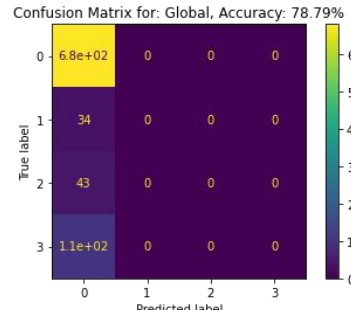


7. Findings Burger's Data

- Target Accuracy Achieved Instantly

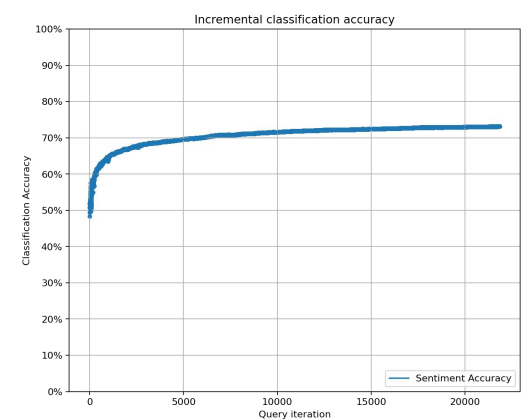


- Target Model is very simple

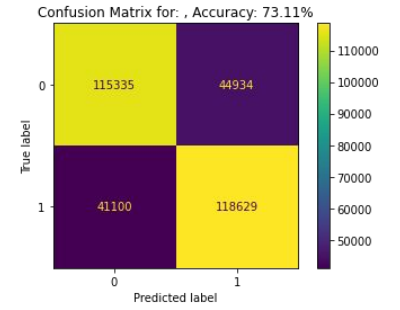


8. Findings Twitter Data

- Target Accuracy Achieved With Logarithmic Trend



- Target Model is not a single label predictor



9. Conclusion

Based on experimental results **Active Learning can greatly reduce the amount of required human labeling for current Automatic Psychological Text Classification.**

Current simplicity of the traditional learning models for schema therapy require very little data to replicate.

Twitter results may indicate the trend of schema label reduction when schema classifiers are trained on a more realistic data set.

10. Limitations of current Schema Classification

Real world accuracy of traditional and active ML classifiers trained on currently available schema data may be lower than experimental accuracy.

Schema Results Show Signs Of:

- Unbalanced data set
- Overfitting

11. Future work

Improvements to Schema Classification

- Research possible flaws in currently available data
- Perform active learning with schema classifiers with more complex models.

Logarithmic Trend Of Accuracy/Data on Natural Language Classification:

- May allow for mathematical estimates of how much labeled data active learning will require to achieve a target accuracy

[1] A. David, "Schema mode assessment through a conversational agent," Master's thesis, Delft University of Technology, 2021

[2] B. Franziska, "Natural language processing for cognitive therapy: extracting schemas from thought records," Master's thesis, Delft University of Technology, 2021.

[3] G. Alec, B. Richa, and H. Lei, Sentiment140 Twitter Data Set, 2009. Available at <http://help.sentiment140.com/>