

# Human Behavior Assessment using Ensemble Models

Abdullah Faiz Ur Rahman Khilji<sup>1</sup>, Rituparna Khaund<sup>2</sup>, Utkarsh Sinha<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering

<sup>2</sup>Department of Electrical Engineering

National Institute of Technology Silchar

Assam, India

{abdullah\_ug@cse, rituparna\_ug@ee, utkarsh\_ug@cse}.nits.ac.in,

## Abstract

Behavioral analysis is a pertinent step in today's automated age. It is important to judge a statement on a variety of parameters before reaching a valid conclusion. In today's world of technology and automation, Natural language processing tools have benefited from growing access to data in order to analyze the context and scenario. A better understanding of human behaviors would empower a range of automated tools to provide users a customized experience. For precise analysis, behavior understanding is important. We have experimented with various machine learning techniques, and have obtained a maximum private score of 0.1033 with a public score of 0.1733. The methods are described as part of the ALTA 2020 shared task. In this work, we have enlisted our results and the challenges faced to solve the problem of the human behavior assessment.

## 1 Introduction

Human behavior assessment is an important computation task that automates the task of detecting human behavior from textual data. The behavior in the text depends on many parameters. Some of these include words of different types including attitude and appraisal (Martin and White, 2003). The use of evaluative language allows for a greater deal of solidarity in the text (Martin and White, 2005). Various rule-based algorithms can be used to evaluate the essence of the sentence. The sentence judgment can be divided into two sections viz. social esteem and social sanction. The former comprises normality, capacity, and tenacity. Whereas the latter includes veracity and propriety. Sentence classes, their meaning, and sample explanation are included in Table 1.

Various approaches include natural language processing tools to extract the sentiment or detect human behavior from the text. The work by (Liu,

2012), describes various aspects of the sentiment analysis and opinion mining problem. Since the above task belongs to the natural processing domain, it brings along various difficulties, including coreference resolution, negation handling among many (Bakshi et al., 2016).

To classify the sentences into the above 5 classes, we have formulated the same into a machine learning multi-classification task. This paper investigates different approaches for the human behavior assessment, as part of the Australasian Language Technology Association (ALTA) 2020 shared task (Mollá, 2020).

The rest of the paper is divided as follows. The related works are enlisted in Section 2. The dataset description is given in Section 3. The experimental setup is given in Section 4. The experimentation details are described in Section 5. Results and analysis are tabulated in Section 6. Finally, we conclude with discussion and conclusion in Section 7.

## 2 Related Works

On experimental investigation of the problem, we have found that the given problem closely resembles the multi-class human sentiment analysis such as the multi-class sentiment analysis using clustering and scoring (Farhadloo and Rolland, 2013). The work by (Farhadloo and Rolland, 2013), uses the semantic analysis and clustering on a bag of nouns to identify the class of the sentiments based on the textual description. Other works show the use of multi-class class SVM<sup>1</sup> (Lavanya and Deisy, 2017) which employs topic adaptive learning method to produce more generic and abstract based systems. There also exists machine learning systems that perform discourse analysis (Oteíza,

<sup>1</sup><https://scikit-learn.org/stable/modules/svm.html>

Table 1: Class Meaning, Category and Examples

S. No.	Category	Class Name	Meaning	Example Sentence
1	Social Esteem	Normality	“How unusual one is.”	“He is unfashionable.”
2		Capacity	“How capable one is.”	“The student is a child prodigy.”
3		Tenacity	“How resolute one is.”	“They are truthful and hard-working.”
4	Social Sanction	Veracity	“How honest/truthful one is.”	“She is hard-working and truthful.”
5		Propriety	“How ethical one is.”	“He is too arrogant to learn from his mistakes.”

2017) on the description to map out the sentiment.

Some works also show that systems are performing better if there is a fusion of more than one architecture like that of the GME-LSTM(A)<sup>2</sup> (Chen et al., 2017) and (Prabowo and Thelwall, 2009), which uses multi-phased architecture and thereby takes the advantage of those methods as well as the concept of word-level and fine level fusion techniques to surpass other state-of-the-art techniques.

As a part of this experimentation, we have used ensemble models to tackle different aspects of the problem. Starting from the XLNet Pretraining as given in Section 4.1 to decision tree classifier is discussed in Section 4.4 and up to XGBoost (in Section 4.5). These were used in different phases of feature generation, multi-class classification, analysis, and validation.

### 3 Dataset

The labeled dataset<sup>3</sup> for the ALTA 2020 shared task was provided by the organizers. The dataset included single, multiple, or no labels for a single sentence as the output label. The train data contains a total of 200 instances of labeled data, whereas the test set contains 100 instances. The dataset provided was based on the Semeval 2018 AIT DISC dataset<sup>4</sup> (Mohammad et al., 2018). For the purpose of experimentation, we have worked with both sets of data, with and without preprocessing. Preprocessing steps include removal of punctuation and stop words.

<sup>2</sup>Gated Multimodal Embedding LSTM with Temporal Attention

<sup>3</sup><https://www.kaggle.com/c/alta-2020-challenge/data>

<sup>4</sup>[https://competitions.codalab.org/competitions/17751#learn\\_the\\_details-datasets](https://competitions.codalab.org/competitions/17751#learn_the_details-datasets)

## 4 Experimental Setup

Since the data provided to us by the organizers is quite small as discussed in Section 3, we employed the use of machine learning techniques instead of data craving deep learning methods. For the word embeddings, we have experimented with the XLNet (Yang et al., 2019) pre-trained embeddings and the freely available spaCy<sup>5</sup> word embeddings.

### 4.1 XLNet Pretraining

XLNet (Yang et al., 2019) is an efficient pretraining method in comparison to the Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018), due to the various improvements in the model. XLNet is a pretraining method based on generalized autoregressors, that learns bidirectional context information. The autoregressive nature overcomes the deficit of the BERT model. We have used the pretrained XLNet model as provided by spaCy and used the generated vectors for the downward classification tasks.

### 4.2 SpaCy Pretraining

Here, we have used the `en_core_web_lg` model as provided by spaCy. The sentence vectors generated by the model is used directly for the multi-classification step.

### 4.3 Polynomial Features

Polynomial features are obtained by raising exponential powers to the existing set of features (James et al., 2013). It can also be termed as a feature engineering task, wherein new inputs are generated based on the current set of inputs. For our experimentation, we have experimented with polynomial features of various degrees.

<sup>5</sup><https://spacy.io/>

Table 2: Sample Predictions from the Model

S. No	Prediction	Text	Actual Behaviour	Predicted Behaviour
1	Correct	Actually be arsed with my sister sometimes, she controls the TV 90% of the time and when I watch one thing she gets in a huff	Normality	Normality
2	Correct	You ever just be really irritated with someone u love it's like god damn ur makin me angry but I love u so I forgive u but I'm angry	Capacity	Capacity
3	Correct	@SaraLuvvXXX : Whaaaaat?!? Oh hell no. I was jealous because you got paid to f**k, but this is a whole new level. #anger #love #conflicted& Propriety	Propriety	Propriety
4	Incorrect	it makes me so f**king irate jesus. nobody is calling ppl who like hajime abusive stop with the strawmen lmao	Propriety	Normality
5	Incorrect	Goddamn headache.	Propriety	Capacity, Tenacity
6	Incorrect	I wanna kill you and destroy you. I want you died and I want Flint back. #emo #scene #f**k #die #hatered	Capacity, Tenacity	Propriety

#### 4.4 Decision Tree Classifier

Decision Tree (Swain and Hauska, 1977) is a machine learning technique based on the supervised approach. This algorithm is commonly used for both classification and regression tasks. It formulates the task as a graphical structure, wherein the features are represented as the internal nodes. The rules are represented by the tree branches. Finally, the outcome of the tree is given by the leaf.

#### 4.5 XGBoost

Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016), is a scalable algorithm frequently obtaining state-of-the-art results in many machine learning tasks with limited dataset size. The given algorithm is a combined model of decision trees, which uses copies of itself to improve the model performance and minimizes error. It is an efficient version of the well known stochastic gradient boosting algorithm.

## 5 Experimentation

As discussed in Section 4, we have used various machine learning techniques for the given multi-classification problem and have used feature vectors generated from different deep learning approaches as discussed in Sections 4.1 and 4.2. The generated sentence vectors of each sentence are fixed to a length of 300. For reasons attributed to computational cost and efficiency, we have used polynomial features of degree 2 in our experiments. The results obtained using different approaches are tabulated in Table 3. Table 3 is sorted based on the private score as provided by the organizers. We have experimented with various approaches, an overview of which is given in Section 4. We have also, experimented with our ensemble model having polynomial features with degree 2 trained on a decision tree classifier. This ensemble model has experimented been on both XLNet and spaCy word embeddings. The model incorporating the use of XGBoost has also been used. Various other approaches are employed and the obtained score is tabulated in Table 3.

Table 3: Techniques Employed with corresponding Public and Private Mean F-Score

S. No.	Approach	Private Score	Public Score
1	XGBoost with spaCy pretrained embeddings	0.1033	0.1733
2	Polynomial features with degree 2 together with decision tree classifier, using pre-trained XLNet embeddings	0.1000	0.1600
3	Using polynomial features and decision tree regressors, with spaCy pretrained embeddings.	0.0593	0.1866
4	Decision tree with spacy embeddings	0.0533	0.2066
5	Polynomial features with degree 2 together with decision tree classifier, using pre-trained spaCy embeddings	0.0533	0.2200
6	Decision tree classifier along with polynomial features of degree 2 , incorporating removal of stopwords	0.0533	0.2033

## 6 Results and Analysis

The result from the experimentation, as discussed in Section 5 are tabulated in Section 3. As we can see from Table 3, the highest score of 0.1033 on the private dataset is using the XGBoost approach with pretrained spaCy embeddings. The highest score of 0.2200 on the public leaderboard is using a decision tree classifier with polynomial features of degree 2.

## 7 Discussion and Conclusion

In our work, we have worked with various deep learning algorithms and fusion techniques to study and investigate human behavior. We have also set up the analogy between the human sentiment analysis and behavior in Section 2. We have also trained our system based on various architectures and the best results can be referred to in Section 2. As the dataset size was not so significant, the system is not trained on complex deep learning-based architectures. From Table 2 we can see that the first three predictions go with the original analysis and the last three contradicts the original interpretation, we can also see that the actual output contains more than one class (as shown in Table 2), our analysis engine can replicate the same, as can be seen from Table 2, but since the textual description was so short, the system was not able to properly analyze and map it with the output.

Thus, from the above observations, we can infer that a less complex framework can sometimes perform better than complex architecture, moreover, if the dataset size would be significantly more, then a more complex architecture could have been devised and incorporated. The semantic analysis could have been carried out using those datasets.

Future works can involve a rule-based approach for the same problem statement. Such an approach would be able to provide much better results even on a smaller dataset. Various techniques could be used to improve on the dataset size, and a deep learning architecture can be developed to cater to the same.

## Acknowledgment

The authors are thankful to the organizers of the ALTA 2020 shared task for organizing the event and Kaggle InClass for hosting the competition. Special thanks to Dr. Diego Mollá-Aliod for his support.

## References

- Rushlene Kaur Bakshi, Navneet Kaur, Ravneet Kaur, and Gurpreet Kaur. 2016. Opinion mining and sentiment analysis. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 452–455. IEEE.
- Minghai Chen, Sen Wang, Paul Pu Liang, Tadas Baltrušaitis, Amir Zadeh, and Louis-Philippe Morency. 2017. Multimodal sentiment analysis with word-level fusion and reinforcement learning. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 163–171.
- Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Mohsen Farhadloo and Erik Rolland. 2013. Multi-class sentiment analysis with clustering and score

- representation. In *2013 IEEE 13th international conference on data mining workshops*, pages 904–912. IEEE.
- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An introduction to statistical learning*, volume 112. Springer.
- K Lavanya and C Deisy. 2017. Twitter sentiment analysis using multi-class svm. In *2017 International Conference on Intelligent Computing and Control (I2C2)*, pages 1–6. IEEE.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.
- James R Martin and Peter R White. 2003. *The language of evaluation*, volume 2. Springer.
- JR Martin and PR White. 2005. R 2005: The language of evaluation: Appraisal in english. *Hampshire: Palgrave Macmillan*.
- Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.
- Diego Mollá. 2020. Overview of the 2020 alta shared task: Assess human behaviour. In *Proceedings of the 18th Annual Workshop of the Australasian Language Technology Association*.
- Teresa Oteíza. 2017. *The appraisal framework and discourse analysis*, pages 457–472. Routledge Handbooks.
- Rudy Prabowo and Mike Thelwall. 2009. Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2):143–157.
- Philip H Swain and Hans Hauska. 1977. The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics*, 15(3):142–147.
- Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems*, pages 5753–5763.