Master Thesis - Semester III

Explorative study of explainable artificial intelligence techniques for sentiment analysis applied for English language

by Rohan Kumar Rathore

Advisor: Dr. Anton Kolonin

Scientific workshop "Big Data Analytics"
Novosibirsk State University, Russia

December 8, 2020

Table of Content

Introduction

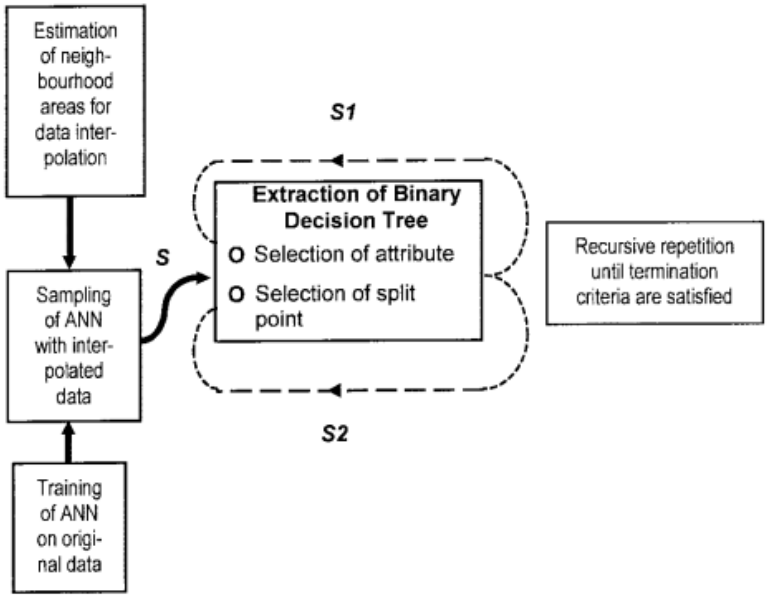
- Artificial intelligence : Artificial agents achieving goals smartly
- Machine learning : Algorithmic models responsible for smartness
- Explainable artificial intelligence : Techniques to explain the models

Recap

Explainable artificial intelligence (XAI) techniques for sentiment analysis

- Sentiment analysis model development
  - Data: IMDB movie reviews
  - Model: multisentiment, context based, neutral-mixed capable, unbiased
- Semester I
  - Local interpretable model-agnostic explanations (LIME): Explaining with surrogate models
- Semester II
  - Layer-wise relevance propagation (LRP): Explaining with propagated weights relevance scores of the network
- Semester III (This work)
  - Artificial neural network decision tree algorithm (Ruleex ANN-DT): Explaining by extraction of decision trees from artificial neural networks

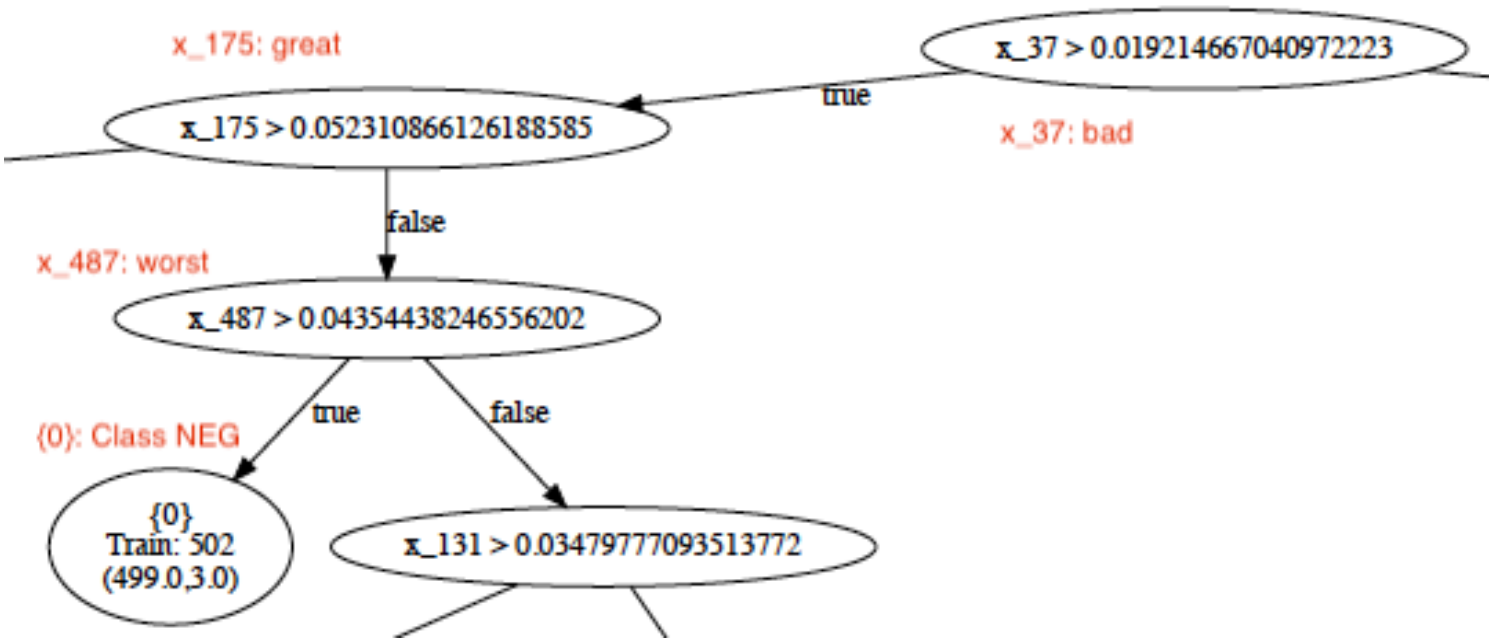
Exploration of Ruleex ANN-DT for sentiment analysis - I



- Selection of Attribute: Similar to CART algorithm of reducing the entropy $$V_w = \sum_{k=1}^{2} \frac{n_k}{n} \mathrm{Var}(O_k)$$
- Stopping criteria: Standard deviation or the variance is zero
- Statistical pruning technique: chi squared

Exploration of Ruleex ANN-DT for sentiment analysis - II

- Decision Tree representation (showing sub-section here) of a simple Sentiment Analysis NN model (features: tf-idf, layers: {500, 100, 30, 2}, training acc: 85.5%, testing acc: 82.8 %)



Next steps

- Improve the model to meet all the functional goal [in-progress]
- Explore bayesian neural networks for sentiment analysis
- Paper readiness

References

- Robert A., Joachim D., Alan B.T., Survey and critique of techniques for extracting rules from trained artificial neural networks, Knowledge-Based Systems, Vol.8 Issue 6, 1995
- Hiroshi T., Extracting Rules from Trained Neural Networks, IEEE Transactions on Neural Networks, Vol.11 No.2, 2000
- LiMin F., Rule Generation from Neural Networks, IEEE Transactions on Systems Man and Cybernetics, Vol.24 No.8, 1994
- Gregor P.J.S., Chris A., Francois S.G., ANN-DT: An Algorithm for Extraction of Decision Trees from Artificial Neural Networks, IEEE Transactions on Neural Networks, Vol.10 No.6, 1999