Predicting Anxiety Severity from Social Media Use: Differential Evolution as an Optimizer

Micah Tracy, Dr. Ignacio Segovia-Dominguez

Introduction

- Anxiety disorders affect an estimated 4% of the global population: the most common mental health disorder.
- Social media use may contribute to heightened anxiety.
- Generalized Anxiety Disorder (GAD) can be estimated through a 7-item anxiety scale (GAD-7). Moderate to severe anxiety is associated with a score of 10 or above.

- Differential evolution: a population-based evolutionary algorithm that works in two phases.
  - Initialization phase: randomly generated a uniformly distributed population.
  - Evolution: the population undergoes mutation, crossover, and selection until a stopping criterion is met.

Problem Statement

Can we predict if an individual has moderate to severe anxiety based on data regarding their social media use?

We use different optimizers to estimate the parameters for logistic regression. Is differential evolution's performance comparable to other optimizers?

Dataset

- Survey responses from 15-40-year-old individuals in Bangladesh.
- Narrowed to questions regarding social media use and anxiety.
- Calculate a total score for GAD-7: if an individual scores a 10 or above, we classify them as having moderate to severe anxiety.
- We end up with 37 features related to social media use (input), and one label indicating if the individual has moderate to severe anxiety (output).
- 80% training set with 632 observations, and 20% testing set with 159 observations.

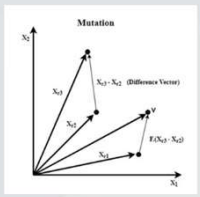
Computational and Mathematical Methods

Logistic Regression:

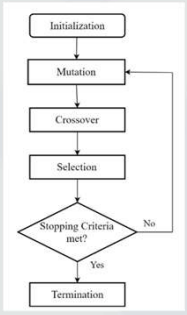
$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 + \ldots + \beta_{36})}}$$

$$\text{Log-Loss} = \sum_{i=0}^{n} -(y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i))$$

Differential Evolution Mutation and Crossover Schemes



Differential Evolution Flowchart



Scipy Optimizers

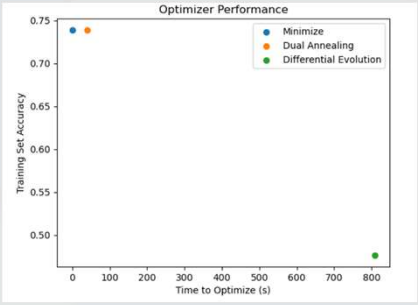
- Minimize: provide the gradient of the function, BFGS method only uses the first derivative (which is provided).
- Dual Annealing: stochastic approach combining Classical Simulated Annealing and Fast Simulated Annealing.
- Differential Evolution: used population size of 15 and a recombination value (CR/crossover probability) of 0.5.

Optimizer Metrics

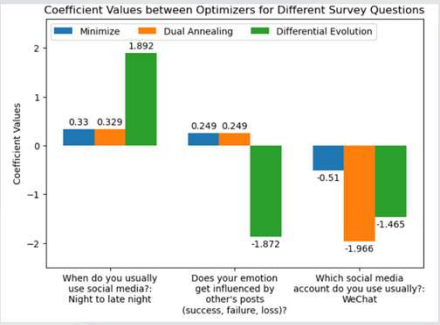
Optimizer	Time	TrainAccuracy	TrainPrecision	TrainRecall
Minimize	0.207152	0.738924	0.674157	0.306122
Dual Annealing	39.716378	0.738924	0.674157	0.306122
Differential Evolution	808.045457	0.476266	0.297297	0.505102

Optimizer	TestAccuracy	TestPrecision	TestRecall
Minimize	0.691824	0.583333	0.264151
Dual Annealing	0.691824	0.583333	0.264151
Differential Evolution	0.440252	0.285714	0.452830

Results



- Differential Evolution was significantly slower than the other optimizers, with lower accuracy and precision.
- However, differential evolution had higher recall than the other two optimizers.



- These survey questions were all in the top 15 coefficients (highest absolute value) for all three of our optimizers tested.
- For the first and last questions shown, the coefficients had the same sign between all three optimizers tested.
- Interestingly, differential evolution had a negative coefficient value for the middle question, whereas the other two optimizers had positive values.

- While our logistic regression models could not perfectly predict if an individual had moderate to severe anxiety, they performed better than random guessing.

References

Bilal, Pant, M., Zaheer, H., Laura Garcia-Hernandez, L., & Abraham, A. (2020). Differential Evolution: A review of more than two decades of research. *Engineering Applications of Artificial Intelligence*, *90*.

Islam, M., Tushar, M., Jannath, S., Moona, A., Akter, S., & Islam, S. (2021). Data set concerning the use of social networking sites and mental health problems among the young generation in Bangladesh. *Data in brief*, *39*.

Hansen, C. (2022). Implementing logistic regression from scratch in Python. *IBM Developer*.

GitHub

Conclusions

- While not completely able to predict anxiety, social media use does seem to impact anxiety.
- Social media users should be mindful of how their mental health may be affected, specifically if they use social media late at night or their emotions are easily influenced by others' posts.
- It may be necessary to reevaluate social media use or see a medical professional.
- In this specific problem, differential evolution did not work well: low accuracy and precision, long runtime. However, the higher recall may be beneficial in a healthcare context.

Future Work/Improvements

- Implementation of differential evolution as an optimizer in PyTorch for use in neural networks.
- Test differential evolution on a dataset with fewer features: since our data had so many features, the algorithm took too long to test different hyperparameters.