Exploring the Capabilities of Geometric Semantic Genetic Programming

In The Context of Parkinson's Disease Diagnosis

Arastun Mammadli

University of Exeter

April 2024

Table of Contents

Background

Technical Background

Design and Development

Results

Conclusions

Demo

Background & Aims

Key Terminology:

- Geometric Semantic Genetic Programming (GSGP)
- Parkinson's Disease (PD) – a prevalent neurodegenerative disorder.
- UPDRS - tool to evaluate the Parkinson's disease

Project Aim:

- To explore the capabilities of novel Geometric Semantic Genetic through comparative analysis.

Technical Aspects

Data Collection & Preprocessing:

- Utilizing the largest Parkinson's dataset (around 6000 recordings from 42 patients).
- Target features: total UPDRS and motor UPDRS

Software and Tools:

- GSGP: using the GSGP-C++ 2.0 framework.
- STGP: implemented form scratch in C++.
- MLBLs: using the C++'s DLIB library.
- Visualization: using Python libraries (Pandas, Matplotlib, Seaborn).

Experimental Design

Design Overview:

- Selected Data Features.
- Performance Metric: MAE.
- Experimental Setup Verification.

Table: Description of some of the data features used in experimentation

Data Feature	Description
Motor-UPDRS	Clinician's motor UPDRS score, linearly interpolated
Total-UPDRS	Clinician's total UPDRS score, linearly interpolated
Age	Age of the participant
Jitter(%)	Percentage of jitter in voice measurements
Shimmer	Shimmer in voice measurements
NHR	Noise-to-harmonic ratio
HNR	Harmonic-to-noise ratio
RPDE	Recurrence period density entropy
DFA	Detrended fluctuation analysis
PPE	Pitch period entropy

Development Approach

Development Approach:

- The Approach to STGP, GSGP and MLBLs development.
- Implementation of k-fold cross-validation.

Comparison Between Models

Table: Average MAE and training time (milliseconds) across all models.

Model	Motor Set 1	Motor Set 2	Total Set 1	Total Set 2
MLR	7.61	7.26	9.40	9.11
SVM	7.31	7.18	9.30	9.36
KRR	7.53	8.14	9.36	10.45
RBF	7.53	8.14	9.36	10.45
MLP	7.31	7.42	10.38	10.39
GSGP	7.75	7.78	11.76	10.53
STGP	10.09	8.22	10.57	15.37

Model	Motor Set 1	Motor Set 2	Total Set 1	Total Set 2
MLR	336.6	544.6	374.2	499
SVM	8459.2	5281.8	13160	10190.6
KRR	1835.8	47143.6	1848.8	47447.8
RBF	1842.8	48083.6	1833.4	48725
MLP	77.2	134	78.8	137.8
GSGP	590.587	718.7248	584.6136	722.9644
STGP	193926.8	160823.4	166833.4	174699.2

An In-Depth Look

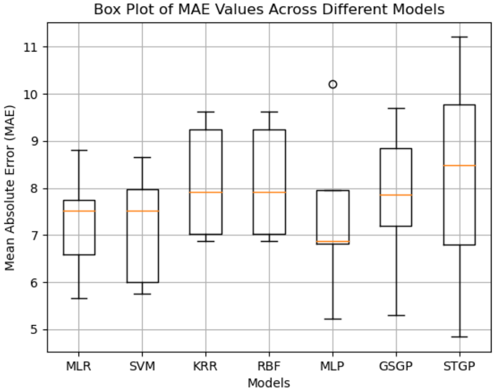

Figure: Averaged MAE results across all models.

Past Works and Originality

- Bakurov's[1] and Castelli's[2] studies find that GSGP often outperforms or matches the best ML methodologies.
- This project employed 2 distinct feature sets for predictions.
- This study also emphasized computational efficiency.

[2] A. Moraglio, K. Krawiec, and C. G. Johnson, "Geometric Semantic Genetic Programming," *PPSN XII*, Springer, 2012, pp. 21–31.

[1] I. Gonçalves, S. Silva, and C. M. Fonseca, "On the Generalization Ability of Geometric Semantic Genetic Programming," *EuroGP 2015*, Springer, 2015, pp. 41–52.

Discussion & Conclusions

Key Findings:

- Only Multi-Layer Perceptron Regression (MLP) fully outperformed GSGP.
- GSGP demonstrates stability and computational efficiency.

Future Directions:

- Exploration of hyperparameter optimization.
- Extending the GSGP technique, and ensemble strategies.

Demo

Link to Presentation: Link

Link to Project Code: Link