Using Artificial Intelligence to develop Mapping Algorithms: A Comparison of Symbolic Regression Machine Learning methods for mapping from the EORTC-QLQ-C30 to EQ5D utilities.

MSR223

Ralph Crott[1], Iftekhar Khan[1,2]

[1]Regulatory Scientific and Health Solutions (R-S-S), Shirley, Solihull, United Kingdom, [2]University of Warwick, Coventry, United Kingdom.

INTRODUCTION

Symbolic Regression (SR) is an established Machine Learning technique for identifying optimal mathematical expressions that can describe relationships within a given data structure. However, SR is rarely used in health economic modelling. Several SR algorithms are available from different software.

OBJECTIVES

This research aims at comparing the performance of an artificial intelligence (AI) determined mapping algorithm with traditional regression approaches using data from the EQ-5D-3L and the cancer (non-small cell lung cancer (NSCLC)) specific EORTC-QLQ-C30 (QLQ-C30). This research also compares several different SR approaches with traditional regression estimation.

METHODS

Data from 3 Randomized Controlled Trials (RCTs) in NSCLC reported previously developed mapping models between the QLQ-C30 and the EQ-5D-3L (Jang 2010, Crott 2018, Khan & Morris 2014). The same data were used in SR analyses using four types of software: TuringBot, Mathematica® (through Data Modeler®), GPlearn and PySR (modules in Python). For comparing performance with the standard regression model, the root mean square error (RMSE), 1-R² and mean absolute error (MAE) and the complexity score were used. The model used EQ-5D utilities as the response regressed against 15 QLQ-C30 domain scores.

RESULTS

Table 1: QLQ-C30 Quality of Life Profiles (Mean Scores)

QLQC30 Function	Topical N=2379	Soccar N=994	Jang N=172
Physical (PF)	54.15	76.62	73.37
Relational (RF)	48.49	68.52	67.44
Emotional (EF)	73.63	77.43	75.48
Cognitive (CF)	77.02	82.8	80.03
Social (SF)	65.82	73.75	73.64
Overall QoL (QL)	51.19	65.11	65.89
(FA) fatigue	49.74	34.65	40.82
(NV) nausea/vomiting	10.71	12.09	7.56
(PA) pain	26.50	21.36	25.67
(DY) dyspnoea	51.50	34.44	31.19
(SL) sleep	29.84	25.01	34.87
(AP) appetite	36.89	22.33	22.66
(CO) constipation	20.69	18.64	18.48
(DI) diarrhoea	15.10	6.20	13.36
(FI) financial problems	10.84	20.79	23.05

Table 2: Model Performance across Software and Regression Models

		Topical	Soccar	Jang
Benchmark OLS	R²	0.640	0.630	0.584
	MAE	0.132	0.100	0.141
	RMSE	0.179	0.141	0.199
Symbolic Regression TuringBOT	R²	0.619	0.585	0.642
	MAE	0.137	0.104	0.135
	RMSE	0.184	0.148	0.169
Symbolic Regression GPLearn	R²	0.389	0.312	0.331
	MAE	0.174	0.139	0.178
	RMSE	0.233	0.191	0.232
Symbolic Regression PySR	R²	0.348	0**	0.199
	MAE	0.181	0.167	0.189
	RMSE	0.242	0.231	0.267
Symbolic Regression Mathematica*	R²	**0.656**	**0.670**	**0.735**
	MAE	**0.128**	**0.092**	**0.109**
	RMSE	**0.175**	**0.132**	**0.148**
Published Best values	R²	0.75	0.71	NA
	MAE	0.10	0.13	0.104
	RMSE	0.09	0.11	0.073
	Regression Model	Beta-binomial	Beta-binomial	Piecewise OLS

* Best accuracy results in bold (through DataModeler®).
** regression yielded only a constant mean utility but no equation.
Note: Higher R^2 is considered better, lower RMSE and MAE are considered better.

AI determined methods through SR using the Mathematica® algorithms demonstrated better performance compared to the standard linear regression methods.
SR provided higher R^2 – in some cases (Jang), the R^2 was about 26% higher.

Consequently, mean predicted utilities were closer to the observed mean utilities (as were the standard deviations).

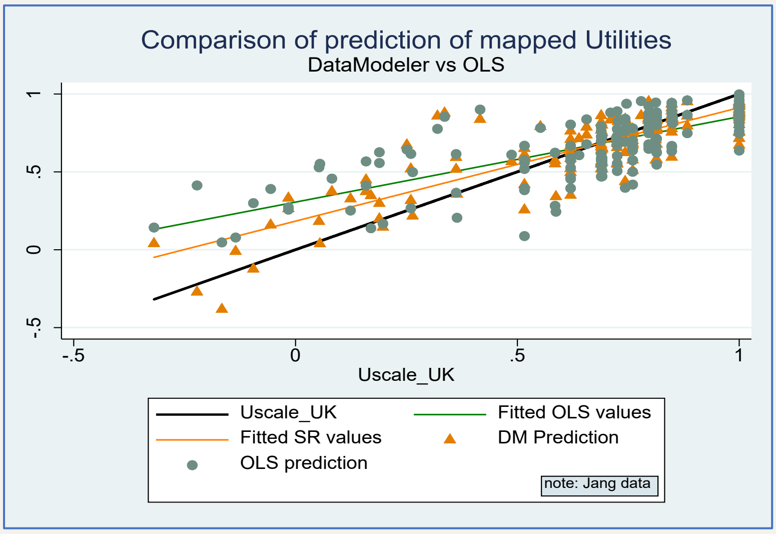
Compared to OLS, Mathematica® was able to improve the accuracy of the prediction while other SR algorithms were less able to reach that goal. It also resulted in the same mean and the closest standard deviation of the estimated utilities compared to the observed utilities.

Table 3: OLS and SR Mean Predicted Utilities

	Topical Mean (SD)	SOCCAR Mean (SD)	Jang Mean (SD)
Observed utilities	0.607 (0.297)	0.750 (0.230)	0.676 (0.284)
Benchmark OLS prediction	0.607 (0.238)	0.750 (0.182)	0.676 (0.210)
SR prediction TuringBot	0.605 (0.248)	0,750 (0.170)	0.671 (0.239)
SR prediction GPlearn	0.618(0.214)	0.741 (0.175)	0.706 (0.153)
SR prediction PySR	0.611(0.148)	0.750 (0) constant	0.670 (0.137)
SR prediction DataModeler	**0.607 (0.241) ***	**0.750 (0.189)***	**0.677 (0.243)***

Note: * closest symbolic regression results to observed mean and standard deviation.

Figure 1: Comparison of Goodness of Fit of Best SR versus OLS in the Jang Data Sample



Best SR Equation (Jang Data)

$$-0.12+\left(2.44\times10^{-3}\right)\text{eFscore}-\left(2.45\times10^{-3}\right)\text{pAscore}+\left(8.57\times10^{-2}\right)\sqrt{\text{pFscore}}-\left(1.41\times10^{-3}\right)\text{pFscore}+\frac{\left(2.85\times10^{-3}\right)\text{pAscore}}{\frac{\text{eFscore}}{6}-\text{fAscore}+\text{nVscore}+\text{pFscore}}+\frac{0.25}{-4-\text{pAscore}+\text{pFscore}}+\frac{\left(3.27\times10^{-3}\right)\left(135.33+e^{3}\sqrt{\text{pFscore}}+3\,\text{pFscore}\right)}{8-\text{cFscore}}+\frac{\left(2.53\times10^{-3}\right)\text{pAscore}}{-6+\text{nVscore}+\text{pFscore}-\text{rFscore}}+\left(1.61\times10^{-3}\right)\text{rFscore}+\frac{\left(2.14\times10^{-3}\right)\left(9.13-\frac{\text{eFscore}}{6}+\frac{\text{eFscore}^{3}}{216}+2\,\text{sFscore}\right)}{6-4\,\text{cFscore}-\text{nVscore}}$$

CONCLUSIONS

Symbolic Regression may outperform standard linear regression. It depends very much on the various SR algorithms. We found that in this case the genetic algorithm in Mathematica provided the best results. Mean predicted utilities were close to observed in most cases. SR could also be helpful in identifying the set of influential explanatory variables. More recent alternative SR methods like Bayesian, neural networks, or transformer-based approaches could improve on the current results. Further research in other disease areas and other generic quality of life measures is warranted as with more advanced regression methods like mixture models or spline regression is warranted.

References

Kronberger, G et al. Symbolic Regression, CRC Press,2024; Dolan, Paul, et al. "The time trade-off method: results from a general population study." Health economics 5.2 (1996): 141-154; Jang, Raymond W., et al. "Derivation of utility values from European Organization for Research and Treatment of Cancer Quality of Life-Core 30 questionnaire values in lung cancer." Journal of Thoracic Oncology 5.12 (2010): 1953-1957; Crott, Ralph, and Andrew Briggs. "Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences." The European journal of health economics 11 (2010): 427-434; Crott, Ralph. "Direct mapping of the QLQ-C30 to EQ-5D preferences: a comparison of regression methods." PharmacoEconomics-Open 2 (2018): 165-177; Khan, Iftekhar, et al. "Comparing the mapping between EQ-5D-5L, EQ-5D-3L and the EORTC-QLQ-C30 in non-small cell lung cancer patients." Health and quality of life outcomes 14 (2016): 1-15. regression methods." PharmacoEconomics-Open 2 (2018): 165-177.