Neural FCA
OSDA BIG HW

Ruben Safaryan

HSE

December 13, 2022

Dataset

	Temperature	L	R	A_M	Color	Spectral_Class	Type
0	3068	0.002400	0.1700	16.12	Red	M	0
1	3042	0.000500	0.1542	16.60	Red	M	0
2	2600	0.000300	0.1020	18.70	Red	M	0
3	2800	0.000200	0.1600	16.65	Red	M	0
4	1939	0.000138	0.1030	20.06	Red	M	0

Preprocessing

- OneHotEncoding,
- average properties of different types of stars.

Those averages are governed by the following table:

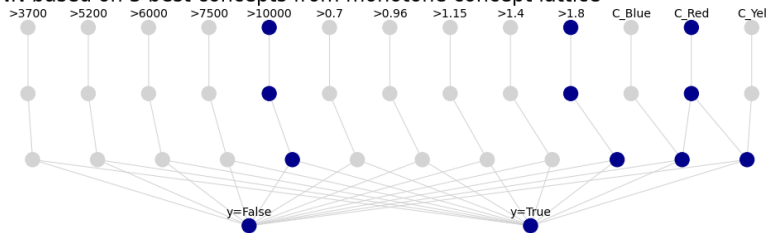| Class	Effective temperature[2][3]	Vega-relative chromaticity[4][5][a]	Chromaticity (D65)[6][7][4][b]	Main-sequence mass[2][8] (solar masses)	Main-sequence radius[2][8] (solar radii)	Main-sequence luminosity[2][8] (bolometric)
O	≥ 30,000 K	blue	blue	≥ 16 $M_\odot$	≥ 6.6 $R_\odot$	≥ 30,000 $L_\odot$
B	10,000–30,000 K	blue white	deep blue white	2.1–16 $M_\odot$	1.8–6.6 $R_\odot$	25–30,000 $L_\odot$
A	7,500–10,000 K	white	blue white	1.4–2.1 $M_\odot$	1.4–1.8 $R_\odot$	5–25 $L_\odot$
F	6,000–7,500 K	yellow white	white	1.04–1.4 $M_\odot$	1.15–1.4 $R_\odot$	1.5–5 $L_\odot$
G	5,200–6,000 K	yellow	yellowish white	0.8–1.04 $M_\odot$	0.96–1.15 $R_\odot$	0.6–1.5 $L_\odot$
K	3,700–5,200 K	light orange	pale yellow orange	0.45–0.8 $M_\odot$	0.7–0.96 $R_\odot$	0.08–0.6 $L_\odot$
M	2,400–3,700 K	orange red	light orange red	0.08–0.45 $M_\odot$	≤ 0.7 $R_\odot$	≤ 0.08 $L_\odot$

Preprocessed Dataset

	>3700	>5200	>6000	>7500	>10000	>0.7	>0.96	>1.15	>1.4	>1.8	C_Blue	C_Red	C_Yel	y
0	False	False	False	False	False	False	False	False	False	False	False	True	False	False
1	False	False	False	False	False	False	False	False	False	False	False	True	False	False
2	False	False	False	False	False	False	False	False	False	False	False	True	False	False
3	False	False	False	False	False	False	False	False	False	False	False	True	False	False
4	False	False	False	False	False	False	False	False	False	False	False	True	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
235	True	True	True	True	True	True	True	True	True	True	True	False	False	True
236	True	True	True	True	True	True	True	True	True	True	True	False	False	True
237	True	True	True	True	False	True	True	True	True	True	False	False	True	True
238	True	True	True	True	False	True	True	True	True	True	False	False	True	True
239	True	True	True	True	True	True	True	True	True	True	True	False	False	True

Next step

•Concept Lattice
•Choosing best concepts
23 concepts with f1_score
28 concepts with accuracy_score
3 concepts with log_loss
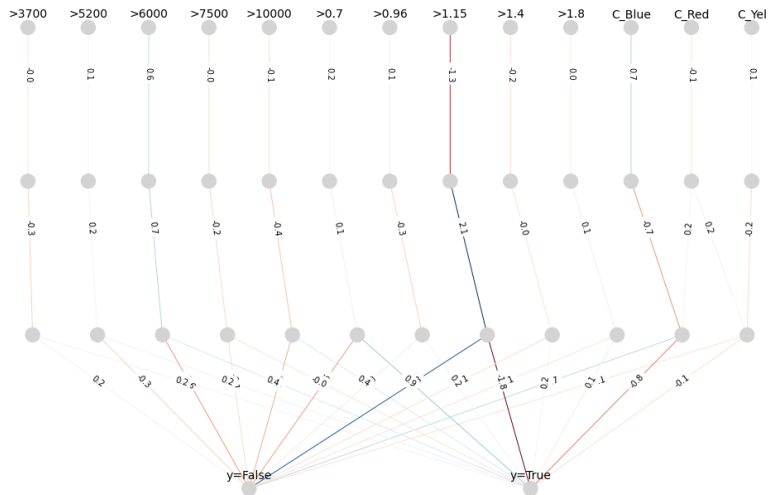3 concepts with roc_auc_score

Visualizations

NN based on 3 best concepts from monotone concept lattice

*Blue neurons are the ones activated by description {'>1.8', 'C_Red', '>10000'}

Visualizations

Neural network with fitted edge weights

Quality Comparison

fca_ac	0.979167
fca_f1	0.977290
lreg_f1	1.000000
lreg_ac	1.000000
scv_f1	0.978336
svc_ac	0.979167
tree_f1	0.995839
tree_ac	0.995833

Overfit?? Very likely

Thank you for your attention!