修論の続き

長瀬　永遠

富山県立大学　情報基盤工学講座

July 28, 2023

今回やりたかったこと

卒論で発表した LiNGAM と DEA を用いた政策における意思決定を支援するシステムについて，卒研時に得られた結果に対してより詳しい検討を行うために GMDH と RF を組み合わせたモデル化の手法を提案し，卒研時のデータに適用する．

進捗報告 1

現在やっていること

GMDH と RF を組み合わせた手法を提案するにあたり，GMDH および RF それぞれ単体で性能評価を行い，それぞれの利点や欠点を理解する必要がある．そのため，それぞれの手法を Python にて実装し，同様のデータに対して適用することでその結果を比較する．

- GMDH（担当：蒲田，中市）
- RF（担当：長瀬）

進捗報告 2

はじめに
進捗報告
おわりに

例題

$$y = 2 - 3x_1^{1.5}x_2^2x_3^{1.5} + x_2^{0.5}x_3^{0.5}x_4^{1.5}x_5^{1.5} \tag{1}$$

不要な説明変数：$x_6, \cdots, x_{15}$,
各説明変数の値：$(0, 1)$ の範囲の一様乱数,
サンプル数：300,
ノイズ：平均 0，標準偏差 0.1, 0.3, 0.5 の正規ノイズ

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_10	x_11	x_12	x_13	x_14	x_15	y
2	0.37	0.5054	0.7106	0.1616	0.9076	0.9798	0.5471	0.7375	0.2835	0.0515	0.7539	0.8549	0.3732	0.0506	0.2447	1.9303
3	0.1129	0.955	0.8035	0.6435	0.4376	0.2078	0.5556	0.8149	0.0278	0.5459	0.6247	0.0166	0.3511	0.0741	0.4028	2.0561
4	0.1136	0.8886	0.2745	0.6553	0.9654	0.615	0.2479	0.4582	0.4874	0.6615	0.7466	0.3737	0.1811	0.0479	0.0612	2.2355
5	0.0329	0.0094	0.2588	0.0345	0.4008	0.7928	0.7368	0.8873	0.5784	0.8212	0.2407	0.0767	0.0051	0.2512	0.6374	2
6	0.3805	0.7699	0.7275	0.1517	0.4636	0.0121	0.0305	0.8303	0.7655	0.0278	0.669	0.6843	0.0609	0.7945	0.4013	1.7548
7	0.7354	0.8672	0.7945	0.9658	0.021	0.3699	0.2567	0.923	0.7832	0.9924	0.7328	0.8333	0.6334	0.4375	0.8107	0.9944
8	0.2968	0.4029	0.5435	0.7498	0.8562	0.5617	0.1903	0.4007	0.0202	0.7563	0.5035	0.8088	0.9299	0.3654	0.5133	2.2091
9	0.5724	0.8763	0.9705	0.6604	0.9533	0.275	0.5914	0.2273	0.9677	0.2411	0.5605	0.393	0.4181	0.3229	0.5266	1.5065
10	0.242	0.9432	0.0029	0.2423	0.8417	0.6887	0.6048	0.9472	0.6514	0.9461	0.2997	0.9198	0.5878	0.7845	0.0906	2.0047
11	0.4567	0.3742	0.7631	0.4521	0.6591	0.382	0.7396	0.9317	0.9952	0.3235	0.7732	0.6383	0.5585	0.6219	0.7368	2.0004
12	0.0225	0.3598	0.9265	0.6693	0.8719	0.045	0.3259	0.2647	0.3065	0.0569	0.4556	0.8095	0.3071	0.979	0.8477	2.2562
13	0.0498	0.8173	0.6264	0.9395	0.138	0.4992	0.1906	0.4107	0.7475	0.346	0.6217	0.0872	0.9325	0.6087	0.6275	2.0223
14	0.4271	0.7218	0.922	0.2995	0.1007	0.6517	0.1601	0.9765	0.5451	0.0207	0.9994	0.2294	0.1211	0.0112	0.2794	1.6178
15	0.6766	0.0811	0.6595	0.5498	0.7094	0.3152	0.9638	0.6999	0.5155	0.6716	0.026	0.8356	0.2123	0.5419	0.8328	2.0504

図 1: サンプルデータ

進捗報告 3

実装

RF に関しては Python のライブラリが存在せず，ネット上にも似たコードはなかったため，一から自分でコーディングしている．とりあえず参考文献をもとに RF5 法を実装している．

BFGS 法 1

同定したい多変量多項式回帰モデル

$$y = f(\mathbf{x};\theta) + \epsilon, f(\mathbf{x};\theta) = w_0 + \sum_{j=1}^{J} w_j \prod_{k=1}^{K} x_k^{w_{jk}} \tag{2}$$

BFGS 法 2

1. θ_1 を初期化し，$\mathbf{H_1} = \mathbf{I}, k = 1$ とする．
2. $\Delta\theta_\mathbf{k} = -\mathbf{H_k}\nabla E(\mathbf{x};\theta_\mathbf{k})$ を計算する．停止条件を、満たせば，反復を終了する．
3. $E(\mathbf{x};\theta_\mathbf{k} + \alpha_k\Delta\theta_\mathbf{k})$ を最小にする α_k を求める．
4. 結合重みを $\theta_\mathbf{k+1} = \theta_\mathbf{k} + \alpha_k\Delta\theta_\mathbf{k}$ に修正する．
5. $k \equiv 0 \mod M$ (M : 全パラメータ数) ならば $\mathbf{H_{k+1}} = \mathbf{I}$ とし，それ以外の場合は，$\mathbf{H_{k+1}}$ を更新する．$k \leftarrow k + 1$ として step2 に戻る．

進捗報告 5

GMDH との比較

正解の数式

$$y = 2 - 3x_1^{1.5}x_2^2x_3^{1.5} + x_2^{0.5}x_3^{0.5}x_4^{1.5}x_5^{1.5} \tag{3}$$

GMDH の方の結果は蒲田の発表参照.
RF5 の結果は以下のようになるはず（参考文献参照）．次回の研究発表までには自分のコードでも同じ結果が出せるようにしておきます.

$$\begin{aligned} y =& 1.9632 - 3.0116x_1^{1.4990}x_2^{1.9951}x_3^{1.4990} \\ &+ 1.0183x_2^{0.4879}x_3^{0.4879}x_4^{1.4990}x_5^{1.4990} \end{aligned} \tag{4}$$

進捗報告 6

提案手法の方針

変数同士の組合せ導出を GMDH で行い，それをもとに各変数の次数を RF5 で求めることで GMDH 特有の層を重ねることで必要な変数の組合せを求める手法を残しつつ計算量軽量化と精度向上を図る（この辺の有効性の出し方は要検討）

おわりに

今後やること

- 自身のデータに適用するための方針を考える
- 作成した RF5 をもとに提案手法を実装
- データの拡充
- 数値実験を行い，卒研時の結果も含めて考察を行う