言語則を橋梁とした知識融合最適化によるデータ解析手法の開発と高度化

代表研究者　　遠　藤　靖　典　　筑波大学・システム情報系・教授

1 研究調査の要旨

本研究調査では、人間の知識や推論機構を言語則で表現できるファジィ推論をクラスタリングアルゴリズムに組み込むことによってデータ解析を実現する手法を開発し、その理論的性質の解明および数値例を通した有効性の検証を行う.

2 目的と意義

近年，情報通信技術等の発展に伴いデータが高度化・膨大化し，クラスタリングを代表とするデータ解析技術の必要性が高まってきている．ハード c-平均法や階層的クラスタリングといった既存のクラスタリング手法の多くは，目的関数や更新式といった数理モデルの仮定に基づくモデルベースクラスタリングと呼ぶことができる.

これらモデルベースクラスタリングでは，そのアルゴリズムに仮定された数理モデルにそぐわないデータに対してクラスタリングを行った場合，適切な分類結果を得られない．このためクラスタリングを行う際には，データに対する手法の選定が非常に重要になる．この手法の選定には，事前にデータ構造を知ることが必要となるが，近年データの高度化・膨大化が進み，これは容易に行えることではない.

このクラスタリングと同様な状況が制御の分野でも起きていた．制御分野では，制御対象であるプラントなどが高度化したため，数理モデルで記述できずモデルベースな自動制御は行えなかった．そのため，人間が手動で制御を行っていた．この自然言語によって記述される人間の手動制御則を自動処理するために用いられたのがファジィ推論である．このファジィ推論により，数理モデルを仮定せずに人間の手動制御則をコンピュータ上で処理可能となり，高度化されたプラント等に対しても自動制御が可能となった.

制御分野での流れを汲んだ上で，ファジィ推論を用いて数理モデルを仮定せずに人間の分類方法を言語的に記述することにより,データ構造を知らずとも適切なクラスタリング結果が得られるのではないだろうか.現状，ファジィ推論を用いたクラスタリング手法は存在せず，このような言語を用いたアプローチ方法は用いられていない．そのような状況下で，このような言語ベースクラスタリング手法を提案することは非常に難しいだろう．そこで本研究では，言語ベースクラスタリング手法の礎としてファジィ推論を用いた階層型言語ベースクラスタリング手法を提案する．また本手法の理論的性質の解明や数値例を通した有効性の検証を行う.

2 階層型言語ベースクラスタリングの開発

本章では,提案手法である階層型言語ベースクラスタリング（Hierarchical Linguistic-based Clustering, HLC）とそれに用いる推論規則，メンバーシップ関数について述べる．また，提案手法の特性として，クラスタ代表点のすれ違い，また AHC 重心法との等価性を理論的に明らかにする.

HLC では，AHC と同様に各個体 $x_i \ \ (i=1,...,n)$ をクラスタとみなし，クラスタ同士を結合する．その際，時刻 t におけるクラスタ G_i の代表点を $M^{(t)}(G_i)$ とする．このクラスタ代表点を制御対象とみなし，ある速度 v_i と質量 m_i を計算する．特に v_i はファジィ推論によって求められる．この速度 v_i を利用して自身に最も近いクラスタ代表点に近づいていき，クラスタ間非類似度 $d(M^{(t)}(G_i), M^{(t)}(G_j)) = \mid M^{(t)}(G_i) - M^{(t)}(G_j) \mid \ \ (j=1,...,n)$ が定数 ε 以内ならば，その 2 つのクラスタ

を結合し，新たなクラスタを生成する．このクラスタリングプロセスは，AHC と同様に樹形図として表現することも可能である．また本アルゴリズム，規則ではデータ集合 X ，質量 m_i は正規化を仮定している．

2-1 アルゴリズム

HLC のアルゴリズムは，時刻を t，離散計算幅を Δt，個体結合範囲を ε，クラスタ集合を G とすると以下のように表現される．

アルゴリズム

Step 1：以下のように初期状態を設定する．

$$t = 0$$
$$G = \{G_i \mid i = 1,...n\}$$
$$G_i = \{x_i\}$$
$$M^{(0)}(G_i) = x_i$$
$$m_i = 1$$

Step 2：ファジィ推論規則に従い v_i を計算する．

Step 3：$M^{(t+\Delta t)}(G_i) := M^{(t)}(G_i) + v_i \Delta t$ とする．

Step 4：$\min_{1 \le i, j \le n, i \ne j} d(M^{(t+\Delta t)}(G_i), M^{(t+\Delta t)}(G_j))$ を満たすクラスタ G_i と G_j を G から取り除き，新たなクラスタを

$$G' = G_i \cup G_j$$
$$M^{(t+\Delta t)}(G') = \frac{m_i M^{(t+\Delta t)}(G_i) + m_j M^{(t+\Delta t)}(G_j)}{m_i + m_j}$$
$$m' = m_i + m_j$$

とし G に加える。 $n := n - 1$ とする。

Step 5： もし $n = 1$ なら終了．違うのなら $t := t + \Delta t$ とし、Step 2 へ．

このアルゴリズムは比較的少ないステップ数でアルゴリズムが終了するため，推論規則等の差異によるクラスタリング結果の差は発生しにくいと考えられる．

2-2 推論規則

本研究では，数値例による比較のため 5 つの推論規則（Rule Group，RG）を提案する．それぞれの推論規則では，クラスタ代表点 $M^{(t)}(G_i)$ の次元毎の速度 $v_i^l \quad (l = 1,..., p)$ を非類似度 $d_i^l = M^{(t)}(G_j)^l - M^{(t)}(G_i)^l$ $(j = \arg\min_k \mid M^{(t)}(G_i) - M^{(t)}(G_k) \mid)$ を用いてファジィ推論により求める．また，推論規則内では，下表の略称を利用する．またデータ集合 X，質量 m_i は正規化を仮定しているため，$-1 \le d_i^l, v_i^l \le 1$，$0 \le m_i \le 1$ となる．

表 1：言語変数の省略表

vPB	Very Positive Big
PB	Positive Big
PM	Positive Medium
PS	Positive Small
vNB	Very Negative big
NB	Negative Big
NM	Negative Medium
NS	Negative Small

（1）推論規則 1

推論規則 1(RG1)は以下のようになる.

Rule1: If d_i^l is PB, then v_i^l is PB.

Rule2: If d_i^l is NB, then v_i^l is NB.

言語変数に対応するメンバーシップ関数は以下の通りである。

PB: $u(x) = \max(x,0)$

NB: $u(x) = \max(-x,0)$

これは最も単純な推論規則であるため，非ファジィ化手法との相性が重要になる．例えば，非ファジィ化された時の出力値の範囲を考慮すると，Min-Max 法で最小絶対法とは相性が良いと考えられるが，重心法との相性はあまり良くないと考えられる．また，Product-Sum 法と最小絶対法の組み合わせでは，常に$v_i^l = 1$が採用されてしまう.

（2）推論規則 2

推論規則 2(RG2)は以下のようになる.

Rule1: If d_i^l is PB, then v_i^l is PB.

Rule2: If d_i^l is PM, then v_i^l is PM.

Rule3: If d_i^l is PS, then v_i^l is PS.

Rule4: If d_i^l is NB, then v_i^l is NB.

Rule5: If d_i^l is NM, then v_i^l is NM.

Rule6: If d_i^l is NS, then v_i^l is NS.

言語変数に対応するメンバーシップ関数は以下の通りである。

PB: $u(x) = \max(0, \min(4x - 2,1))$

PM: $u(x) = \max(-|4x - 2| + 1,0)$

PS: $u(x) = \max(-|4x - 1| + 1,0)$

NB: $u(x) = \max(0, \min(-4x - 2,1))$

NM: $u(x) = \max(-|4x+2|+1,0)$

NS: $u(x) = \max(-|4x+1|+1,0)$

前件部 $d_i^l = 0$ 付近の入力に対するメンバーシップ関数への帰属度の総和は 1 としていない．これは非ファジィ化手法との相性を考慮した上での設計であり，$d_i^l = 0$ 付近の入力に対するメンバーシップ関数への帰属度の総和を 1 としても，適切な分類結果を得られるわけではない．

（3）推論規則 3

推論規則 3(RG3)は以下のようになる．

Rule1: If d_i^l is PB and m_i is Heavy, then v_i^l is PS.

Rule2: If d_i^l is PB and m_i is Light, then v_i^l is PB.

Rule3: If d_i^l is PS and m_i is Heavy, then v_i^l is PS.

Rule4: If d_i^l is PS and m_i is Light, then v_i^l is PB.

Rule1: If d_i^l is NB and m_i is Heavy, then v_i^l is NS.

Rule2: If d_i^l is NB and m_i is Light, then v_i^l is NB.

Rule3: If d_i^l is NS and m_i is Heavy, then v_i^l is NS.

Rule4: If d_i^l is NS and m_i is Light, then v_i^l is NB.

言語変数に対応するメンバーシップ関数は以下の通りである。

PB: $u(x) = \max(x,0)$

PS: $u(x) = \begin{cases} \max(-x+1,0)(x \geq 0) \\ 0(x < 0) \end{cases}$

NB: $u(x) = \max(-x,0)$

NS: $u(x) = \begin{cases} 0(x > 0) \\ \max(x+1,0)(x \leq 0) \end{cases}$

Heavy: $u(x) = \max(x,0)$

Light: $u(x) = \max(-x+1,0)$

（4）推論規則 4

推論規則 4(RG4)の推論規則は RG3 と同じである。言語変数に対応するメンバーシップ関数は以下のようになる．

PB: $u(x) = \max(x,0)$

PS: $u(x) = \begin{cases} \max(-x+1,0)(x \geq 0) \\ 0(x < 0) \end{cases}$

NB: $u(x) = \max(-x,0)$

NS: $u(x) = \begin{cases} 0(x > 0) \\ \max(x+1,0)(x \leq 0) \end{cases}$

Heavy: $u(x) = \max(0, \min(2.5x - 0.25,1))$

Light: $u(x) = \max(-2x+1,0)$

（5）推論規則 5

推論規則 5(RG5)のみ規則数が多いため，規則表として下表のように記述する．この際，規則前件部の d_i^l と m_i は RG3 と同様に and で繋がれる．

表 2：RG5 規則表

		m_i	
		Light	Heavy
d_i^l	PB	vPB	PM
	PM	PB	PS
	PS	PM	vPS
	NB	vNB	NM
	NM	NB	NS
	NS	NM	vNS

また対応するメンバーシップ関数は以下のようになる．RG5 は，前件部 d_i^l には RG2 と同じメンバーシップ関数を，前件部 m_i には RG3 と同じメンバーシップ関数を採用している．

d_i^l に関するメンバーシップ関数：

PB: $u(x) = \max(0, \min(4x - 2,1))$

PM: $u(x) = \max(-|4x - 2| + 1,0)$

PS: $u(x) = \max(-|4x - 1| + 1,0)$

NB: $u(x) = \max(0, \min(-4x - 2,1))$

NM: $u(x) = \max(-|4x + 2| + 1,0)$

NS: $u(x) = \max(-|4x + 1| + 1,0)$

m_i に関するメンバーシップ関数：

Heavy: $u(x) = \max(x, 0)$

Light: $u(x) = \max(-x + 1, 0)$

v_i^lに関するメンバーシップ関数：

vPB: $u(x) = \max(4x - 3, 0)$

PB: $u(x) = \max(-|4x - 3| + 1, 0)$

PM: $u(x) = \max(-|4x - 2| + 1, 0)$

PS: $u(x) = \max(-|4x - 1| + 1, 0)$

vPS: $u(x) = \begin{cases} \max(-4x + 1, 0)(x \geq 0) \\ 0(x < 0) \end{cases}$

vNB: $u(x) = \max(-4x - 3, 0)$

NB: $u(x) = \max(-|4x + 3| + 1, 0)$

NM: $u(x) = \max(-|4x + 2| + 1, 0)$

NS: $u(x) = \max(-|4x + 1| + 1, 0)$

vNS: $u(x) = \begin{cases} 0(x > 0) \\ \max(4x + 1, 0)(x \leq 0) \end{cases}$

3 階層的クラスタリング (AHC)重心法との等価性

本節では，提案手法 HLC 2 と AHC 重心法との理論的等価性を示す．

HLC2 におけるクラスタ間非類似度を$d(M^{(t)}(G_i), M^{(t)}(G_j)) = \| M^{(t)}(G_i) - M^{(t)}(G_j) \|^2$と変更する．また，上節と同様な推論規則 RG7 とメンバーシップ関数を考える．また合成規則を Product-Sum 法，非ファジィ化手法を最小絶対法とする．

推論規則 7 (RG7)は以下のようになる．

Rule1: If d_i^l is PB, then v_i^l is PS.

Rule2: If d_i^l is NS, then v_i^l is NS.

言語変数に対応するメンバーシップ関数は以下の通りである。

PS: $u(x) = \begin{cases} \max(-x + 1, 0)(x \geq 0) \\ 0(x < 0) \end{cases}$

NS: $u(x) = \begin{cases} 0(x > 0) \\ \max(x + 1, 0)(x \leq 0) \end{cases}$

入力であるクラスタ代表点間距離を用いて前件部から適合度を算出し，合成規則 Product-sum 法により計算すると，後件部のメンバーシップ関数は頂点を $v_i^l = 0$ に持つ三角形型メンバーシップ関数が得られる．これより非ファジィ化の最小絶対法を用いると，必ずクラスタ代表点の速度が $v_i^l = 0$ となるためクラスタ代表点が移動しない．

上で述べた議論と提案アルゴリズム HLC より，必ず非類似度が最小となるクラスタ対のみが順次結合していき，個体は全て移動しない．また新たに加えられるクラスタ G' は，G' を生成したクラスタ代表点対の重心となる．

これよりクラスタの重心を $M(G_i)$ とし，生成されたクラスタ G' とその他のクラスタ G_i との非類似度は

$$d(M^{(t)}(G_i), M^{(t)}(G_j)) = \| M^{(t)}(G_i) - M^{(t)}(G_j) \|^2 = \| M^{(t)}(G') - M^{(t)}(G_i) \|^2$$

となる．従って AHC の重心法の再計算法と HLC の非類似度の計算方法が一致する．
このため HLC で RG7 および上述のメンバーシップ関数，合成規則として Product-sum 法，非ファジィ化手法として最小絶対法を用いれば，AHC の重心法と等価となる．

4 数値例

4-1 準備

（1）使用データセット

本研究で実験対象とするデータは，全て教師付きデータで人工データ 2 つと実データ 1 つである．人工データのみ 2 次元である．人工データ 1 は 200 個の正円 3 つからなる。人工データ 2 はクラスタサイズを大きく変化させており，それぞれのクラスタは 300 個の正円，200 個の楕円，100 個の正円となる．実データは，UCI Machine Learning Repository の Breast-Cancer dataset で乳がん患者を陽性（個体数 444)と陰性（個体数 239)の 2 クラスタに分類してある．本来の実データは個体数 699 だが，欠損データを取り除いたため個体数は 683 となる．

表 3：実験対象となるデータセットの特徴

	個体数	次元数	クラスタ数	クラスタサイズ
人工データ 1	600	2	3	均等
人工データ 2	600	2	3	不均等
Breast-Cancer dataset	683	9	2	不均等

（2）実験パラメータ・比較手法

本研究では，パラメータを $\Delta t = 0.0001$，$\varepsilon = 0.001$ と設定した．また提案手法において，非ファジィ化処理の際に離散処理を行うためのサンプリングが必要となる.そのサンプリング間隔は 0.01 とする.加えて，簡略化推論法を用いる場合は，シングルトンの値を決定しなければならない．本研究では，下表のように設定した.

また本研究で提案手法と比較する手法は，ハード c-平均法(HCM)，AHC(最短距離法)，AHC(重心法)とする．HCM は初期クラスタ中心をランダムに選択するため，初期値依存性を有している．そのため 1000 回試行を行い，目的関数が最小となる結果を採用している．

表 4：簡略化推論法に関するパラメータ設定

ファジィ集合	vPB	PB	PM	PS	vPS	vNS	NS	NM	NB	vNB
v_i^l	1.0	0.8	0.5	0.2	0.01	-0.01	-0.2	-0.5	-0.8	-1.0

（3）評価指標

本研究では，クラスタリング結果に対する定量的評価の指標として Rand Index とプログラムの実行時間を用いる．Rand Index は正解の分割とクラスタリング結果の類似性を評価する指標であり，正解分割とクラスタリング結果が完全に一致する場合，1 となる．正解の分割を L，クラスタリング結果を C とすると，それぞれの個体は 4 種類に分類される．L と C の両方において同じクラスに属する個体対数を A，L において同じクラスだが C において違うクラスに属する個体対数を b,C において同じクラスだが L において違うクラスに属する個体対数を c，L と C の両方において違うクラスに属する個体対数を d とすると，Rand Index は以下のように計算される．

$$RandIndex = \frac{a+d}{a+b+c+d}$$

教師付きデータに対する評価指標は Rand Index 以外にも様々なものがあるが，本研究では Rand Index が最も一般的な指標であること，多くの指標を用いると結果に対する考察の焦点がぼやけること，この２点の理由から Rand Index のみを採用した.Rand Index 以外にも Entropy という指標で評価を行ったが,Rand Index と比べ，かなり低い値が得られた.

4-2 実験結果と考察

（1）人工データ 1 に対する実験結果・考察

実験結果の RandIndex の値を表 5 に示す．表中の NULL は，プログラムが収束せず結果が得られなかったことを示す．これは RG5 で合成規則に関係なく，非ファジィ化手法として FOM を用いた際に発生する．プログラムが収束しない原因は， RG5 の推論結果のメンバーシップ関数が $v_i^l = 0$ で頂点を持つことと考えられる．$v_i^l = 0$ で頂点を持つメンバーシップ関数に対して FOM を用いた場合，$v_i^l = 0$ を出力として採用してしまうため，クラスタ代表点が移動せず結合が進まない．このためプログラムが収束せず結果が得られなかったと考えられる．これは人工データ 2 と実データについても同様なことがいえる．

表 5 より，Rand Index の値が推論規則や合成規則，非ファジィ化手法を変化させたとしてもあまり変動しない．これより，提案手法においてはクラスタサイズが等しく，線形なクラスタ境界を持つクラスタが明確に分離可能なデータセットに対しては，推論規則等が劇的に変化しない限りは結果の変動が少ないと考えられる．

また、RG1 および RG2 で Min-Max と FOM を用いた際に実行時間が他と比較してかなり大きくなっている．これは，後件部メンバーシップ関数の形が原因である．入力に対してメンバーシップ関数の合成を行った際に，$v_i^l = 0$ 付近に最大の帰属度を持つメンバーシップ関数が生成されやすい．そのため，限りなく小さい $v_i^l = 0$ が採用されてしまい，クラスタ代表点の位置の変化量が微小になってしまったため，実行時間が長くなってしまったと考えられる．一方で，RG3 および RG4 で Min-Max と FOM を用いた際は，推論規則を考慮すると Rule4, Rule8 の影響を最も受けやすくなる.そのため,推論結果として得られるメンバーシップ関数は，比較的大きい v_i^l を頂点に持つ形となり，このため FOM で比較的大きい v_i^l が採用され，実行時間が短くなっていると考えられる．

加えて,簡略化推論法を用いた際には,実行時間がモデルベースクラスタリングとほぼ等しくなっている．これは，ファジィ推論において合成ステップを省略しているからである．簡易な推論法である半面，シング

ルトンの値がクラスタリング結果に大きく影響してくるため，パラメータ調整を十分に考慮して行う必要がある．

HLC2 では HLC1 のような実行時間の差は見られない．これは，毎回の結合ステップにおいて AHC と同様にクラスタ対を必ず結合するためである．このため，HLC1 で見られたクラスタ代表点が移動せず，アルゴリズムが収束しない，といったことはなくなる一方で，推論規則等の差異を結果に反映しにくくなっている．また，推論規則別の実行時間の違いは，推論規則の規則数の増加によるものである．これと同様な傾向は人工データ 2 と実データにも言うことができる．

表 5：HLC2 の人工データ 1 に対する RandIndex 値

合成規則	非ファジィ化	RG1	RG2	RG3	RG4	RG5
Min-Max	FOM	0.957719	0.957719	0.951105	0.895785	0.956054
	COG	0.946225	0.985412	0.935110	0.935110	0.957719
	MOM	0.954397	0.985412	0.919889	0.919889	0.957719
Product-Sum	FOM	0.941416	0.985412	0.905440	0.905440	0.922871
	COG	0.947844	0.985412	0.946225	0.935110	0.985412
	MOM	0.941416	0.985412	0.905440	0.905440	0.893096
Simplified	COG	0.941416	0.964457	0.935110	0.911127	0.957719

（2）人工データ 2 に対する実験結果・考察

実験結果の RandIndex の値を表 6 に示す．表 5 より推論規則，合成規則と非ファジィ化手法の組み合わせが重要になることがわかる．この相性の良さは，一般に制御の分野でも言われており，クラスタリングでも同様の傾向が見られるのではないかと考えられる．また，推論規則等の組み合わせの選定さえ間違えなければ，人工データ 2 のようなクラスタ毎に大きくクラスタサイズが変化するデータに対しても有効なクラスタリング手法となる可能性を秘めている．

表 6：HLC2 の人工データ 2 に対する RandIndex 値

合成規則	非ファジィ化	RG1	RG2	RG3	RG4	RG5
Min-Max	FOM	0.995570	0.995570	1	1	0.993372
	COG	1	0.993372	1	0.896027	1
	MOM	1	0.993372	0.953723	0.984691	0.993372
Product-Sum	FOM	0.810150	0.993372	1	1	0.921653
	COG	0.896027	0.993372	0.897657	0.875843	1
	MOM	0.810150	0.993372	1	1	0.883383
Simplified	COG	1	0.993372	1	0.886477	1

（3）実データに対する実験結果・考察

実験結果の RandIndex の値を表 7 に示す．RandIndex の最大値は 0.948605，最小値は 0.932092 となる．前件部変数として速度のみを用いた RG1，2 に関して，HRand Index 値が高くなっている．これは，提案アルゴリズムの推論規則等への依存が弱いからである．HLC は AHC(重心法)と考えが非常に近いため，AHC(重心法)の結果と類似した結果が出やすくなる．これにより，HLC では RG1，2 でも高い Rand Index の値を出すことができたと考えられる．

また、実行時間は人工データ 1 と同様な傾向が得られている．

これまでの実験結果から，必ずしもルール数の増加がクラスタリング精度を向上させるわけではないことがわかる．むしろ RG5 と COG の組み合わせのように悪化する傾向になる場合さえあり得る．これは，クラスタリングに対する制御対象がデータセットであるからである．一般にファジィ制御においては，プラントなどの制御対象は変化せず，その制御対象に合わせてメンバーシップ関数のチューンアップや合成規則，非ファジィ化手法の選定を行う．一方でクラスタリングでは制御対象がデータセットであるため，クラスタリン

グをするデータセットを変えるごとにメンバーシップ関数のチューンアップ等を行うのが望ましい.しかし，これはデータ解析の場で用いられるような正解が無いデータに対しては行えないため，現実的でない．そのため，推論規則等の組み合わせは特異な形式のものでなく，ある程度のルール数，シンプルなメンバーシップ関数を持つ推論規則が汎用的であると考えられる．実際に数値例からは，差は小さいものの，RG5, Product-Sum 法，COG の組み合わせが最も良いということができる.

表 7：HLC2 の実データに対する RandIndex 値

合成規則	非ファジィ化	RG1	RG2	RG3	RG4	RG5
Min-Max	FOM	0.945832	0.943066	0.945832	0.951387	0.948605
	COG	0.948605	0.943066	0.940310	0.940310	0.945832
	MOM	0.948605	0.943066	0.945832	0.948605	0.940310
Product-Sum	FOM	0.943066	0.943066	0.948605	0.948605	0.932092
	COG	0.943066	0.943066	0.948605	0.943066	0.948605
	MOM	0.943066	0.943066	0.948605	0.948605	0.948605
Simplified	COG	0.932092	0.943066	0.945832	0.948605	0.948605

5 おわりに

本研究ではファジィ推論を用いた階層型言語ベースクラスタリング手法を提案し，理論的な考察，および数値例を通した有効性の検証を行った．具体的には，階層型言語ベースクラスタリングアルゴリズムを提案し，それに関する推論規則を 5 つ提案した．また，クラスタ代表点のすれ違いや AHC 重心法との理論的等価性を示した.

数値例より主に以下の 2 点，1)非類似度だけでなく質量の概念を導入した推論規則の方がより多くのデータセットに対しても有効性を示すこと． 2) 推論規則や合成規則，非ファジィ化手法の選定さえ間違えなければ，従来のモデルベースクラスタリングよりも良い結果が得られることを明らかにした．また，本研究では推論規則等の組み合わせとして，RG5，Product-Sum 法，COG が最善であると考えられる.

第 5 章でも述べたが，提案手法のアルゴリズムでは制御対象がクラスタ代表点となるため，理想的にはデータセット毎にメンバーシップ関数等をチューンアップすることである．しかし，実際に存在する教師なしのデータセットに対して，それは困難である．今後はより多くのデータセットに有効な汎用的な推論規則等の組み合わせを提案する必要がある．また，モデルベースクラスタリングと比べて解析が非常に行いにくいため，従来より多くの数値例を通してその有効性を示す必要があるだろう.

本研究はファジィ推論を用いた言語ベースクラスタリング手法の礎と位置づけることができる.そのため，なるべく簡易な推論規則・アルゴリズムを用いて，その有効性を検証した．また HLC はアルゴリズム内部にモデルベースと捉えることも可能な更新式が存在している．今後は，ファジィ推論に立脚したより言語ベースなアルゴリズムを構築すべきであろう．その際に，クラスタの概形を推論するなど推論の対象は少なくとも HLC 1 のようなものだけではないと考えている．そのため，多様な角度から言語ベースクラスタリング手法について，なにを推論すべきか考察を行う必要がある.

【参考文献】

[1] 赤穂昭太郎，カーネル多変量解析 非線形データ解析の新しい展開，岩波書店 (2008).

[2] 石岡恒憲, x-means 法改良の一提案 －k-means 法の逐次繰り返しとクラスターの再併合, 計算機統計学, 第 18 巻，第 1 号，pp.3-13 (2006).

[3] 伊藤貴之，山口裕美，小山田耕二，長方形の入れ子構造による階層型データ視覚化手法の計算時間および画面専有面積の改善，可視化情報学会論文集，Vol.26，No.6 (2006).

[4] 大沢哲，久永隆治，井上泰助，星野貴，志村一男，画像診断を支援する類似症例検索システム「SYNAPSE Case Match」の開発，2013 年富士フイルム株式会社研究報告書 (2013).

[5] 川田量久，石本一生，植田和憲，P2P ネットワークにおけるクラスタリング手法の提案，電子情報通信学会技術研究報告，vol.107，No.30，pp.49-54 (2007).
[6] 日本学会編，ファジィ集合，講座ファジィ第２巻，日刊工業新聞社 (1992).
[7] 本田中二，大里有生，ファジィ工学入門，海文堂 (1995).
[8] 宮本定明，クラスター分析入門 ファジィクラスタリングの理論と応用，森北出版 (1999).
[9] 安信誠二，ファジィ理論の実システムへの応用-仙台市地下鉄列車自動運転-，日本機械学会誌，Vol.96，No. 836，pp. 639-644 (1988).
[10] Costas P. Pappis, Constantinos I. Siettos, Fuzzy Reasoning, Springer, Search Methodologies chapter 15 (2005).
[11] D.Arthur,S.Vassilvitskii, k-means++: The Advantages of Careful Seeding, ACM-SIAM Symposium on Discrete algorithms, pp.1027-1035 (2007).
[12] E. H. Mamdani, Application of fuzzy algorithms for control of simple synamic plant, Proceedings of Institution of Electrical Engineers, Vol. 121, No. 12, pp.1585-1588 (1974).
[13] E. Rendon, I. Abundez, A. Arizmendi, E. M. Quiroz, Internal versus External cluster validation indexes, International journal of computers and communications, vol. 5, Issue 1 (2011)
[14] L. Hubertm, P. Arabie, Comparing Partitions, Journal of classification, vol.2, pp.193-218 (1985).
[15] Lotfi A. Zadeh, Fuzzy sets, Information and Control, Vol.8, Issue 3, pp.338-353 (1965).
[16] M. Mizumoto, Fuzzy controls by product-sum-gravity method, in: Liu and Mizumoto, Eds., Advancement of Fuzzy Theory and Systems in China and Japan, pp.1-4 (1990).
[17] S. R. Kannan, S. Ramathilagam, P. Devi, A. Sathya, Improved Fuzzy Clustering Algorithms in Segmentation of DC-enhanced breast MRI, Journal of Medical Systems, vol. 36, No.1 pp.321-333 (2012).
[18] S. Miyamoto, H. Ichihashi, K.Honda, Algorithms for Fuzzy Clustering, Springer (2008).
[19] UCI Machine Learning Repository, ``http://archive.ics.uci.edu/ml/''.
[20] Werner Van Leekwijck, Etienne E. Kerre, Defuzzification: criteria and classification, Fuzzy Sets and Systems 108, pp.159-178 (1999).

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
On Linguistic-based Clustering	Proc. of The 2014 IEEE International Conference on Granular Computing (GrC 2014)	2014
力学モデルに基づく階層型言語ベースクラスタリング	第 41 回ファジィワークショップ講演論文集	2015
On Hierarchical Linguistic-based Clustering	Journal of Advanced Computational Intelligence and Intelligent Informatics	To be published
On Cluster Extraction from Relational Data Using L1-Regularized Possibilistic Assignment Prototype Algorithm	Journal of Advanced Computational Intelligence and Intelligent Informatics	2015
On Objective-based Rough Hard and Fuzzy c-Means Clustering	Journal of Advanced Computational Intelligence and Intelligent Informatics	2015
On Objective-based Rough c-Regression	Journal of Advanced Computational Intelligence and Intelligent Informatics	2015
力学モデルに基づく階層型言語ベースクラスタリングについて	第 31 回ファジィシステムシンポジウム講演論文集	To be published
A Note on Non-Hierarchical Linguistic-based Clustering	Proc. of The 12th International Conference on Modeling Decisions for Artificial Intelligence (MDAI 2015)	To be published