Webを利用したファジィデータの取得とその分析・活用

吉川　歩（岡山大学　教育学部）
ayumi@sip.ed.okayama-u.ac.jp, http://sip.ed.okayama-u.ac.jp/

テキストデータを解析応用する上で，程度の表現や概念に見られる境界の不明瞭さや意味の個人差の適切な取り扱いは重要な意味を持つ．ここではファジィ集合を用いてそれらを定量的に扱う方法について概説した．

まず同一概念に対するファジィ集合が全体集合や個人により異なる例を紹介し，全体集合を決定する文脈や個人差の扱いの重要性を示した．特にファジィデータの分析では，データ測定に当たるファジィ集合の同定が重要であることを強調した．

次にファジィ集合の同定法であるメンバーシップ関数同定法とファジィ評定法を取り上げた．まずメンバーシップ関数同定法を測定原理により呈示要素評定法と該当要素応答法に大別し，それらの代表手法である直接推定法とファジィグラフ評定尺度図法を例として，原理，方法および問題点について述べた．次にそれらの問題点を3値の所属の評定と2分木による回答要素の効率的な選択により改善を図った境界漸近推定法の原理と方法を詳述し，JAVA言語で実装されたプログラムを用いたデモを行った．他方，アンケート調査という視点からテキストデータを捉え，ファジィ範疇法とファジィ同時評定法(FCR法)の2種のファジィ評定法を紹介した．ファジィ範疇法はカテゴリー評定尺度図の心理尺度構成をカテゴリーをファジィ集合により表現することで行う手法である．カテゴリーの曖昧さ，個人差の適切な取扱いに加え，尺度構成とデータ測定の分離という特徴がある．またFCR法は，相反する2本の単極グラフ尺度図を用いて評定した結果を合成し，統合値を求める手法である．2本の尺度図による同時
評定で，評定値だけでなく，無関心，矛盾などの付加的な情報も得られるという特徴がある．

最後に，曖昧な概念の間の類似度，一致度をファジィ集合間の演算として実験的に算出した結果を紹介した．それらの主観的な評定を反映する演算は，2つのファジィ集合間の要素軸上の距離に関するものであることを概説した．

テキストデータの特徴

●**定性的**：非数値
　例）昨晩は寒かった⇔昨晩の最低気温は3°C
　⇒数値化が必要な場合が多い
●**冗長**：不要な情報・冗長な情報が含まれる
　例）この計算機は本で見たことがありますが，使ってみるとあまり使いやすくないことが
　　　わかりました．＝「この計算機はあまり使いやすくない」
　⇒不要情報の切り捨て，必要情報の抽出
●**曖昧**：不明確（幅を持つ），多義（一義に決まらない）
　例）程度表現：とても美しくない⇒非常に汚い／やや美しい（曲言法）
　⇒曖昧さの定量化
●**個人依存**：主観により影響
　例）「<u>もうすぐ</u>行きます」：待っている人＝3分，待たせている人＝30分
　⇒主観性の考慮と適切な取り扱い

◎**日常的**：慣れ親しんでいる
◎**ニュアンスの付加**：数値で表せない情報の付加

収集方法の電子化（計算機の援用）

収集方法の電子化＝労力の削減⇒多種・多量のデータを扱うことが可能

・**多量のデータの集積**⇒統計的手法による定量化＝集団に共通する傾向の抽出
◎**多種（異なる個体）のデータの集積**⇒個人差を抽出可能な定量化＋曖昧さの定量化
＝主観性の抽出
⇒本稿で扱う対象．特に**「とても」のような程度修飾語**を対象．

ファジィ集合

概念や程度の曖昧さ（境界の不明瞭さ）を定量的に表現する手法
要素とその要素が対象に属する程度（所属度）の対で表記＝メンバーシップ関数
例）若い⇒年齢を要素として，明確に「若い」に属する境界を定めることができない
⇒日常的な言語や概念は大部分がファジィ集合

ファジィ集合では文脈が重要

文脈＝ある概念が定義されている全体集合＝背景情報
⇒同じ「若い」でも背景が異なれば意味がまったく違う
例）「若い大学生」と「若い政治家」
全体集合：若い大学生＝[18,30]歳，若い政治家＝[30,80]歳
同じ30歳でも若さの度合いは全く異なる：若い大学生　所属度=0，若い政治家　所属度=1

ファジィ集合は個人の主観を反映

同じ「とても」でも，AさんとBさんで意味がまったく同じことはほとんどない
ファジィ集合の形状そのものがいろいろな情報を表している⇒意味の広がり，個人差（主観）
☆適切に形状を定めること（＝メンバーシップ関数の同定）が分析の重要な部分を占める

ファジィ集合の同定法

ファジィ集合の同定法＝メンバーシップ関数同定法とファジィ評定法
同定法の大別⇒要素を与えるか，所属度を与えるか
呈示要素評定法：要素を示してそれに対する所属度を回答させる方法
該当要素応答法：所属度を示してそれに対応する区間を回答あるいは要素を列挙させる方法

直接推定法

提示要素評定法の一種
要素を提示して，その所属度を評定⇒**折線型のメンバーシップ関数**
［問題点］同定精度（形状の表現力）と回答数がトレードオフ
所属度を数値で評定【表現）するのが難しい

ファジィグラフ評定尺度図法

該当要素応答法の一種（ファジィ評定法）
所属度>0と1の範囲を，全体集合を表す数直線上で回答⇒**台形型メンバーシップ関数**
所属度>0：「概念や言葉を表していると見なせる範囲＝全く表していない範囲を除いた範囲」
所属度1：「概念や言葉を最もよく表している範囲」
※所属度>0の範囲＝台集合，所属度1の範囲＝1レベル集合
［問題点］2つの範囲の境界を決めるのが難しい

境界漸近推定法（BASE法）

該当要素応答法と提示要素評定法の折衷法＝ファジィグラフ評定尺度図法の改良版

境界の直接決定を回避⇒間接的に推定：4つの境界が存在する区間を狭めて，位置を推定
所属度の評定＝3値評定：｛属する，属さない，どちらとも言えない｝
計算機による要素のシステマティックな選択（自動）：2分木
⇒両端で評定の異なる区間の中点を評定させる
属する－どちらとも言えない，どちらとも言えない－属さないの区間幅を狭めることで境界を推定
◎境界の直接評定を回避，直接推定の約半分の評定回数で同定可能，回答しやすい3段階評定

ファジィ範疇法＆多重尺度図法

ファジィ評定法
カテゴリー評定尺度図（数値評定尺度図）の評定結果の尺度構成法
※カテゴリー評定尺度図：かなり，非常になどの言語カテゴリーを目盛り
●**系列範疇法**（従来法）：評定結果の分布を正規分布仮定して尺度値を推定
［問題点］個人差が埋没，カテゴリーの程度の広がりが適切に扱えない，など

評定結果とカテゴリーの尺度構成を分離することで回避：**ファジィ範疇法**
⇒(1)尺度図の言語カテゴリーの程度表現語のメンバーシップ関数を同定
(2)カテゴリー評定尺度図で対象を評定
(3)評定結果を程度表現語のメンバーシップ関数として定量化
例）評定結果：かなり高い　→　かなり高いのメンバーシップ関数を尺度値
◎実時間で，心理尺度値が得られる．程度表現語の程度の広がり，個人差が適切に扱える．
検証実験により，距離尺度を満たす．

カテゴリー評定尺度図から得られる結果＝離散値⇒記述力で連続評価に劣る，反面，回答しやすい．
⇒複数のカテゴリー評定尺度図を同時に回答させることで記述力を改善：**多重尺度図法**
2本の異なるカテゴリー評定尺度図から得られた結果（ファジィ集合）を論理和により合成
例）尺度図A：かなり高い，B：けっこう高い　⇒　かなり高い　または　けっこう高い

FCR(Fuzzy Concurent Rating)法

ファジィ評定法
相反する概念を同時に評定⇒多義性を考慮した評定法
個々の概念についてグラフ評定尺度図で1点回答させる
例）ある場合には，良いと判断できるが，見方を変えると悪い部分もある
相反する概念をそれぞれ，0（＝悪い）と1（＝良い）を表すファジィ数（三角型）として表現し，
それぞれの評定を重みとして論理和演算で合成し，重心を求めて評定値を求める

曖昧な概念の間の類似度，一致度

数値データ間の類似度の判定：距離
ファジィ集合間の類似度の判定：指標の候補が多数存在
＝距離，重なり方などの定量化方法が多数考えられる
◎数学的な性質よりも，主観的な類似，一致の判断との整合性が重要
⇒実験による導出および検証が必要

数学的な類似性指標と主観的な類似度判断の比較

導出・検証実験の流れ
(1)かなり高い－非常に高いのようなファジィ概念の間の類似度の評定

(2)各ファジィ概念のメンバーシップ関数の同定
　⇒かなり高いと，非常に高いのメンバーシップ関数に対して，類似性指標の候補の値を算出
(3)それらの値と主観的な類似度の評定結果を比較

実験結果
ファジィ集合間の類似度としてよく使用されてきた重なり方を表現する指標よりも，要素軸方向の距離を表す指標（例えば，ファジィ集合の重心間の距離）が主観的な類似度と整合

参考文献

添付：ファジィ評定とメンバーシップ関数同定法，日本ファジィ学会誌，10，2，pp.184-192，1998
主観を反映したファジィ概念間の類似性指標と一致性指標，
第14回ファジィシステムシンポジウム講演論文集，pp.867-868，1998

その他関連文献はhttp://sip.ed.okayama-u.ac.jp/にPDF形式でアップしてあります.

付録．類似性，一致性指標の候補

所属度関連

所属度最大値＝高さ：$\mathrm{hgt}(A) = \max_{x\ X} \mu_A(x)$

濃度：$\mathrm{power}(A) = \int_{x\ X} \mu_A(x)\,dx$

論理和に対する濃度比：

論理積：$\mu_{A\ B}(x) = \min\{\mu_A(x), \mu_B(x)\}$

代数積：$\mu_{A\cdot B}(x) = \mu_A(x)\cdot\mu_B(x)$

限界積：$\mu_{A\mathrm{o}B}(x) = \max\{0, \mu_A(x) + \mu_B(x) - 1\}$

激烈積：$\mu_{A\,.\,B}(x) = \begin{cases} \mu_A(x)\cdots\mu_B(x) = 1 \\ \mu_B(x)\cdots\mu_A(x) = 1 \\ 0\cdots\mu_A(x), \mu_B(x) < 1 \end{cases}$

対称差：$\mu_{A\Delta B}(x) = \max[\min(\mu_A(x), 1-\mu_B(x)), \min(1-\mu_A(x), \mu_B(x)),]$

絶対差：$\mu_{A\ B}(x) = |\mu_A(x) - \mu_B(x)|$

ユークリッド距離(Minkowskimetric)：$d_2(A,B) = \left(\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)|^2\right)^{1/2}$

距離関連

Housdorff距離：$q(A,B) = \max\{|a_1 - b_1|, |a_2 - b_2|\}$

全レベル集合の総和：$q_1(A,B) = \int_{\alpha=0}^{1} q(A_\alpha, B_\alpha)\,d\alpha$

全レベル集合の最大値：$q_\infty(A,B) = \sup_{\alpha\geq 0} q(A_\alpha, B_\alpha)$

1レベル集合：$q_*(A,B) = q(A_{1.0}, B_{1.0})$

相違指標：$\Delta(A,B) = (|a_1 - b_1| + |a_2 - b_2|)/\ 2(\beta_2 - \beta_1)$

全レベル集合の総和：$\Delta_1(A,B) = \int_{\alpha=0}^{1} \Delta(A_\alpha, B_\alpha)\,d\alpha$

全レベル集合の最大値：$\Delta_\infty(A,B) = \sup_{\alpha\geq 0} \Delta(A_\alpha, B_\alpha)$

1レベル集合：$\Delta_*(A,B) = \Delta(A_{1.0}, B_{1.0})$

重心間距離：$\mathrm{Dist\text{–}gc}(A, B) = \left|\mathrm{FMO}(A) - \mathrm{FMO}(B)\right|$, $\mathrm{FMO}(A) = \left(\int_{x \in X} x\mu_A(x)dx\right) / \mathrm{power}(A)$

ファジィ距離：$\mu_{\tilde{d}(A,B)}(\delta) = \sup_{\delta = |u-v|} \min(\mu_A(u), \mu_B(v))$

1レベル集合の下限値：$\mathrm{GLB\text{–}FDL1}(A,B) = \inf_{a \in A_{1.0}, b \in B_{1.0}} |a-b|$

1レベル集合の上限値：$\mathrm{LUB\text{–}FDL1}(A,B) = \sup_{a \in A_{1.0}, b \in B_{1.0}} |a-b|$

重心：$\mathrm{GC\text{–}FD}(A,B) = \mathrm{FMO}(\tilde{d}(A,B))$

（相対濃度，重心）座標間距離：$\mathrm{V}(A,B) = \sqrt{(\mathrm{power}(A) - \mathrm{power}(B))^2 + (\mathrm{FMO}(A) - \mathrm{FMO}(B))^2}$

［凡例］

$\mu_A(x)$ ：ファジィ集合Aのメンバーシップ関数（所属度）

A_α ：ファジィ集合Aのαレベル集合（所属度α以上の要素の範囲＝通常の集合）

a_1, a_2 ：Aのαレベル集合の下限，上限