《短　信》

類似度・共有度・距離

——柴田・熊谷のネットワーク法との関連で——

伊藤　　隆

柴田・熊谷両氏は，「国語学」150集の論文「ネットワーク法における地点間の言語的類似の新しいとらえかたと処理のしかた——言語的特徴による地域分割のためのネットワーク法II——」のなかで，言語的類似の程度を測る尺度として，「共有度」と「距離」のふたつを用いている。小稿では，これらの尺度に含まれる若干の問題点について述べる。

共有度（同じ言語的特徴を共有する程度）と距離（共有度の分布パターン間の類似の程度）は，それぞれ次のように定義されている（135・129ページ）。

任意の2地点 i と j との間の共有度を NC_{ij} とおいて，地点 i が項目 k の特徴を持つならば，$L_{ki}=1$ とし，持たないならば $L_{ki}=0$ とする。項目数を m とすると，共有度は次の式で定義される。

$$NC_{ij}=\sum_{k=1}^{m} L_{ki}\cdot L_{kj} \qquad (1)$$

任意の地点 i と j について，それぞれの地点を中心とした場合の共有度の値の分布パターン相互間の距離 DC_{ij} は次の式を使って定義する。地点数を n とおく。

$$DC_{ij}=\left\{\sum_{k=1}^{n}(NC_{ik}-NC_{jk})^2\right\}^{1/2} \qquad (2)$$

さて，共有度と距離に関するこれらの定義式(1)・(2)は，地点ごとのデータの平均 M_i，分散 S_i^2，標準偏差 S_i，および地点間の相関係数 r_{ij} を用いて，つぎのように書き直すことができる。

$$NC_{ij}=m(M_i\cdot M_j+r_{ij}\cdot S_i\cdot S_j) \qquad (3)$$

$$DC_{ij}=[n\{(M_i-M_j)^2+S_i^2+S_j^2-2r_{ij}\cdot S_i\cdot S_j\}]^{1/2} \qquad (4)$$

式(3)・(4)は，関連性のひとつの尺度である相関係数の値とはかかわりなく，データの平均と分散・標準偏差の大きさによって，共有度と距離の値が変動しうることを示している。

以下に簡単な例をあげる。柴田・熊谷両氏の方法においては，共有度は言語的特徴の有無に関するデータから計算され，距離は，一旦でき上がった共有度のマトリックスから計算されているので，そもそも同じものを測定しているわけではない。が，ここでは，仮りに，言語的特徴の有無に関するデータをもとにして地点間の共有度と距離を求める場合を想定する。したがって，式(2)のなかの NC_{ik} を L_{ki} に，n を m に置きかえる（この置きかえは，むろん，「距離」の数値的性格に関する議論には影響を与えない）。表1に示すのは，a, b,

c, d の4項目の言語的特徴に関し，A, B, C の3つの地点ペアについて計算された共有度，距離および相関係数の値である。表から明らかなように，共有度は地点ペア A と B の違い(項目 d が異なる)を区別しないし，距離は地点ペア B と C の違い(項目 a が異なる)を区別しない。これらは，共有度と距離が，データの分布形を考慮していないこと，言い換えれば，データに含まれる情報を部分的にしか取り上げていないことによっている。ここにあげたのは，ささやかな例にすぎない。が，しかし，データの平均と分散・標準偏差をそろえずにその計算処理を行うことは，結果に予期せざる歪みをもたらさないとも限らない。

		地点のペア					
		A		B		C	
項目	a	1	1	1	1	0	0
	b	1	0	1	0	1	0
	c	0	0	0	0	0	0
	d	0	0	0	1	0	1
共有度		1		1		0	
距　離		1		1.4		1.4	
相　関		0.58		0		−0.33	

表1　地点間の共有度・距離・相関

ここで，もとの定義式(1)・(2)にもどり，地点ごとのデータを平均が0，分散が1（したがって，標準偏差も1)になるように標準化したとするならば，共有度と距離は式(3)・(4)を用いて次のように表すことができる。

$$NC'_{ij}=m\cdot r_{ij} \qquad (5)$$

$$DC'_{ij}=\{2n(1-r_{ij})\}^{1/2} \qquad (6)$$

式(5)・(6)の示すように，このとき，共有度と距離は相関係数の値と項目あるいは地点の数によって決まる値をとる。したがって，データを標準化した場合には，言語的類似の程度を測る尺度として「共有度」もくしは「距離」の代わりに，(標準化していない) もとのデータについて計算された相関係数を用いても実質的には大差ないことになる。

——慶応義塾大学大学院研究生——