わかみず会第424回（2018.4.11）

ダイアグラム(図形言語）について

小島 俊雄

１　ダイアグラム － 事例

天気図

UML

系統図

・・・

２　構成要素と順序関係

文字表現との差異

３　知識の樹の体系 Chain・Tree・Network

４　文献のまとめと今後

系統図の例

棋士系統図

https://www.shogi.or.jp/player/images/diagram.pdf

プログラミング言語

https://www.levenez.com/lang/lang_a4.pdf

コンピュータ

http://ftp.arl.army.mil/ftp/historic-computers/png/comp-tree.png

ニュートンの手書き図

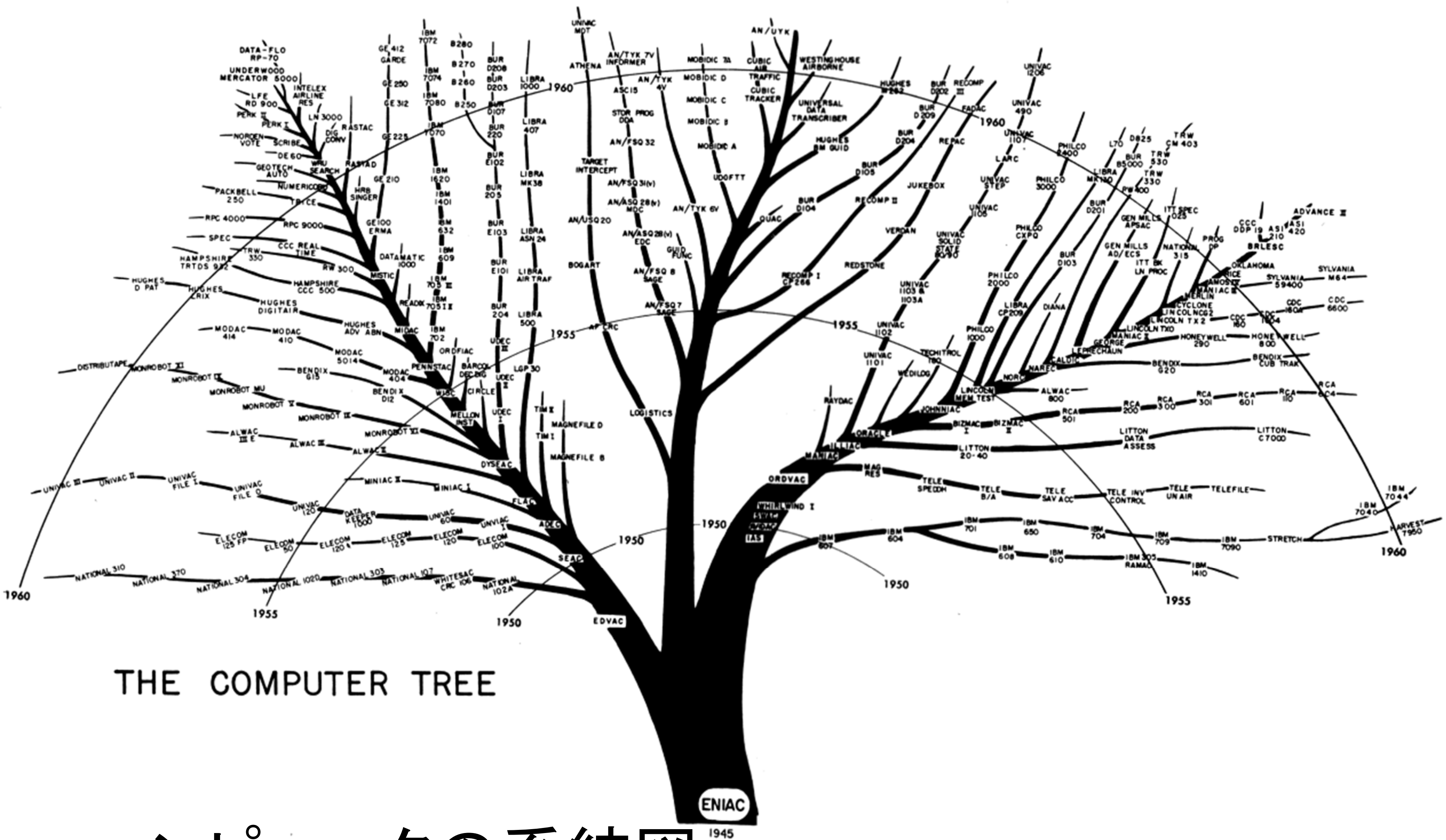
http://evolvingthoughts.net/2011/12/25/happy-newtonmas/

百科全書の知識の樹

https://encyclopedie.uchicago.edu/content/arbre-généalogique

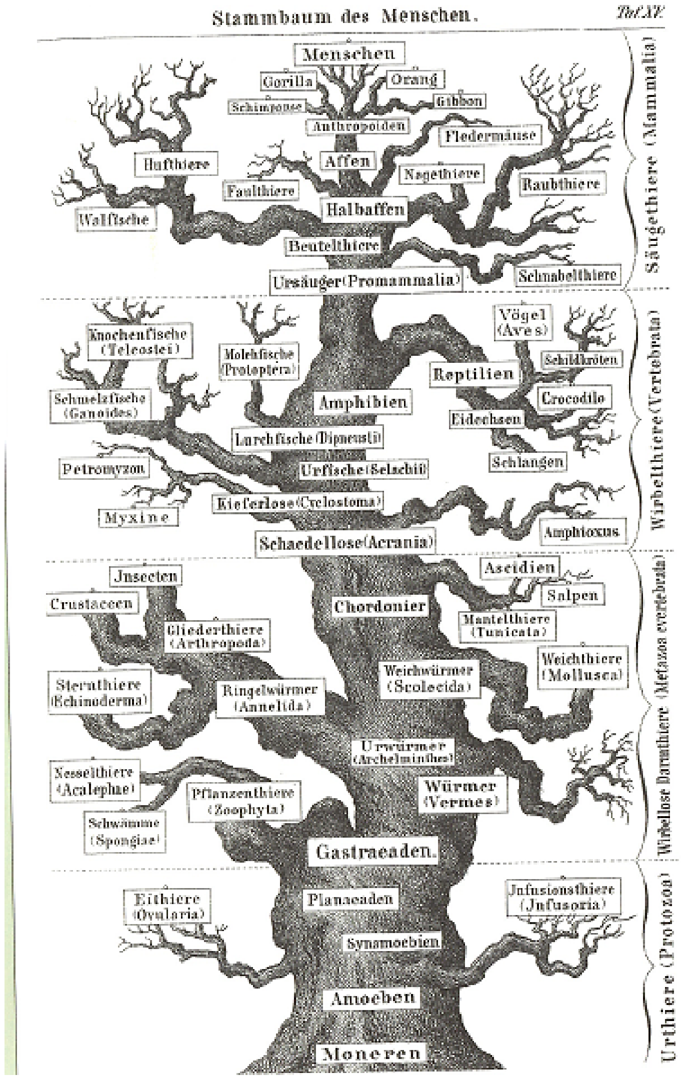
任天堂ソフトのまとめ

http://n-styles.com/main/2006/11/12-223000.php



コンピュータの系統図

http://ftp.arl.army.mil/ftp/historic-computers/drawings/big-comp-tree.gif

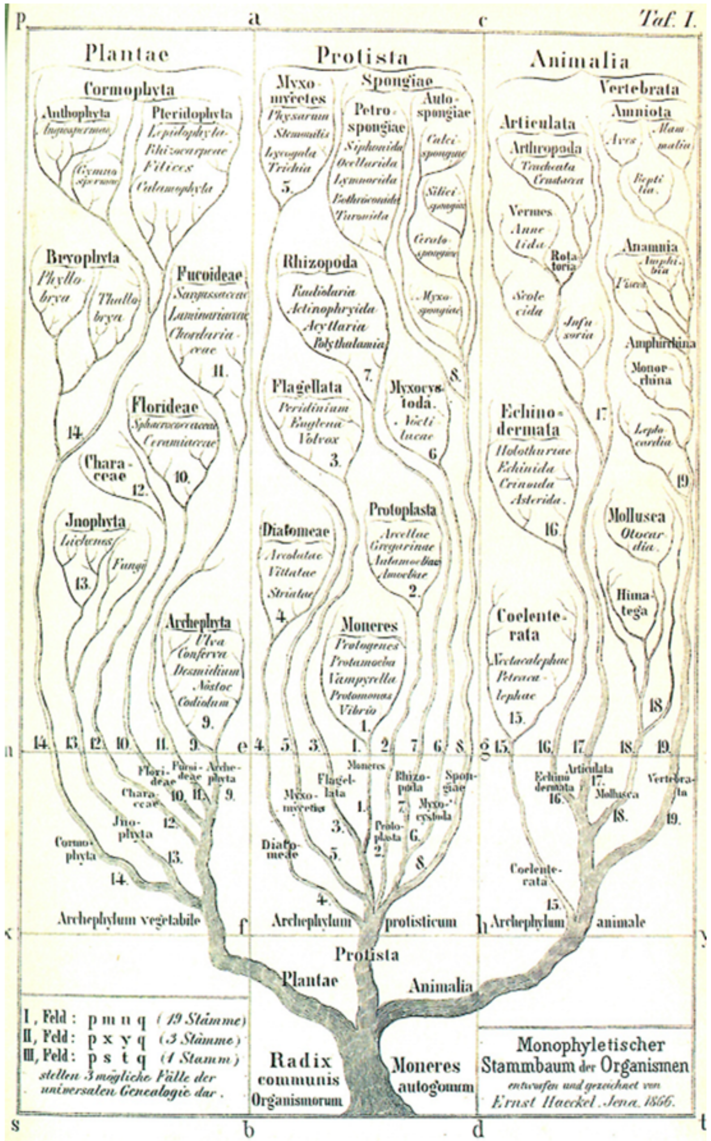


ヘッケルによる
　　人類の系統発生
(1874)

・根は仮想的な
　単細胞生物で頂点にヒト

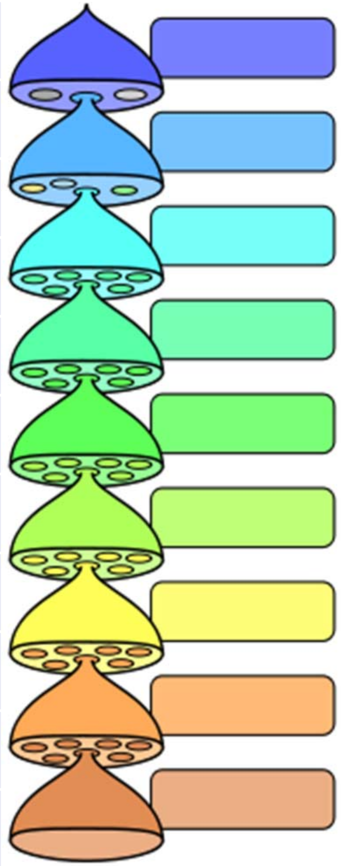
・個体発生は系統発生を繰り返す

ヘッケルの全生物の単系統的系統樹
(1866)



ダーウィンの「種の起源」では1枚の
ダイアグラムのみを含んでいて分岐の
原理を説明する補助的な位置づけ
日記（遺品）として「生命の樹」等が
残されて研究されている。
ダイアグラム自体は説明図的で、ヘッケル
のような芸術的にも評価されるダイアグラ
ムではない。
家系図は、ヨーロッパで幅広く作成され
てきている。最近は、系図作成ソフトも
使われている。数学的な体系化は、19世紀
の進化論に啓発されて広がっていった。
宗教との緊張関係もあった。

和名	英名	例：ヒト
ドメイン:	domain:	真核生物
界:	kingdom:	動物界
門:	phylum	脊索動物門
	／division:	（脊椎動物亜門）
綱:	class:	哺乳綱
目:	order:	サル目
科:	family:	ヒト科
属:	genus:	ヒト属
		Homo
種:	species:	H. sapiens



生物の分類（ウィキペディア）リンネの命名規則

動物界後生動物亜界

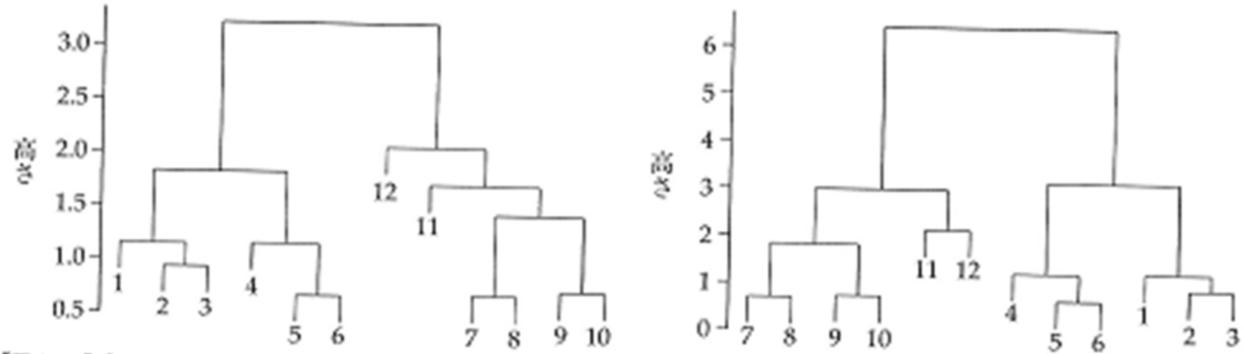
脊索動物門羊膜亜門

哺乳綱真獣亜綱正獣下綱

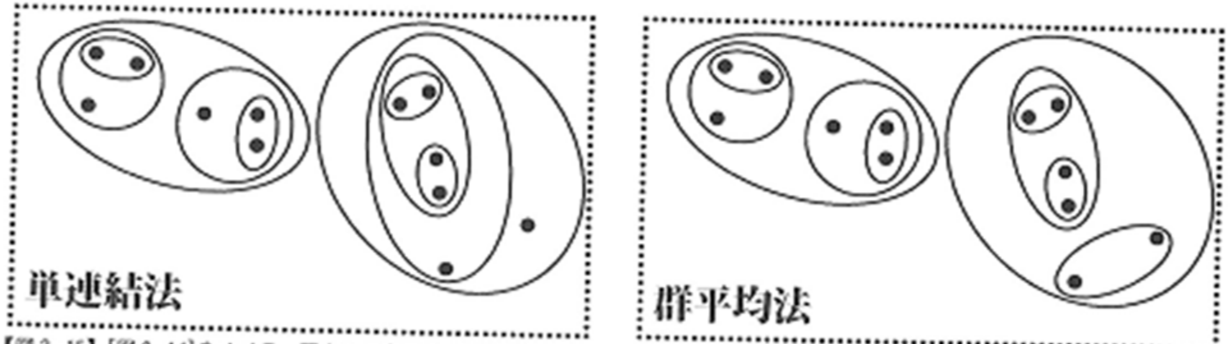
サル目真猿亜目狭鼻猿下目

ヒト上科ヒト科ヒト下科

ホモ属サピエンス種サピエンス亜種

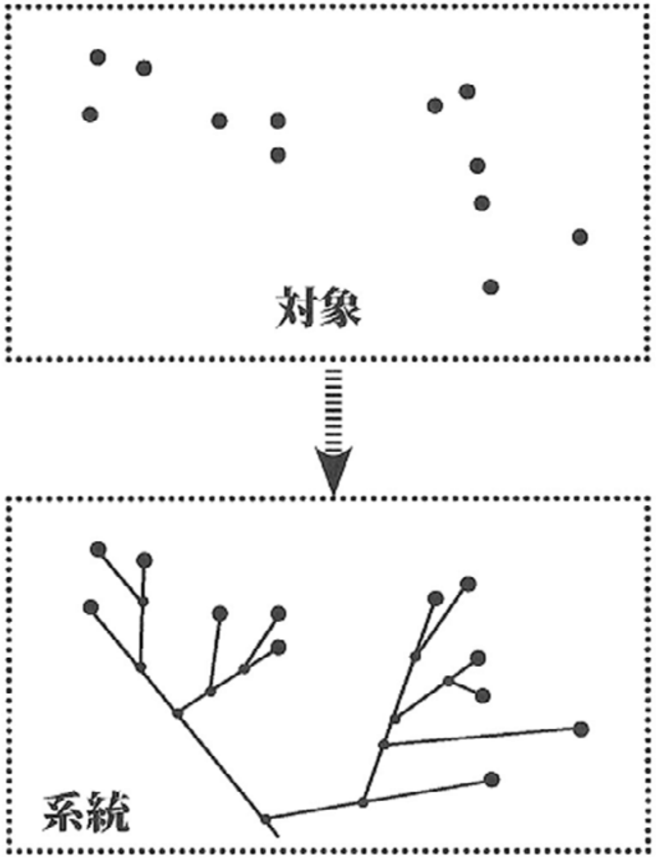


【図 3-14】[図 3-13]のデータを単連結法（左）と群平均法（右）でクラスタリングしたときのデンドログラム。



【図 3-15】[図 3-14]をオイラー図として表示する。単連結法（左）と群平均法（右）。

クラスターリング（単連結法と群平均法）



【図3－3】[図3－1]の点の“配置”に基づいて対象を「つなぐ」操作をした。

対象の背後に仮想的に不可視としてあるものを、対象を可視的に線によって「つないでいる」

分類思考と系統樹思考

分類思考：生得的で、人間の認知性向に合致

←進化研究は分類への十分な知見が必須

系統樹思考：認知心理学的な背景はない場合があり

20世紀後半に新たに出現

（歴史的には19世紀まで遡れる）

←歴史は実践可能な科学

思考法の変化が存在しているのではないか

ダイアグラム（図形言語）

多くのデータを解析し、一般性・規則性・パターンを理解するスキル
直感的な（絵や図より迅速かつ的確）な理解の手段
可視化のツール
オブジェクト（対象物）の複数属性を可視化･･･分野ごとの特色
オブジェクト間の関係性を可視化---グラフ理論的な構造上の特徴
◎ チェイン・ツリー・ネットワーク ･･･離散数学

→ 20世紀後半からコンピュータの発展に支えられ発展

デメリット（限界）
人間が直感的に理解できる次元数は2
応用分野の特色を生かすと個別的にスキルになり適用領域がせまい
メリット
コンピュータ利用によりより大規模複雑な応用分野への適用

今回の話題

分類と系統を主題に応用分野共通の進展をまとめたい！

ダイアグラム（図形言語の基礎）

対象物（オブジェクト）に付随する複数の属性の同時可視化
複数オブジェクト間の関係性をまとめて表示
グラフ理論的な構造的特徴に基づく分類でChan, Tree, Network
に分類して関係を定義する。

[順序関係の定義]

オブジェクト集合Pに属する要素x, yの2項関係Rについて、

反射律：xRxの成立
推移律： xRy かつ yRz ならば xRz、
反対象律 xRy かつ yRx ならば x=y、
比較可能律xRy またはyRx が成立するとして順序関係を定義

擬順序関係　反射律と推移律
半順序関係　反射律、推移律、反対称律は不成立
全順序関係　反射律、推移律、反対称律、比較可能律

[被覆関係の定義]

任意のオブジェクトxとyが①②を満たすとき、被覆関係にあるとし、
x*Ryと表わされる。

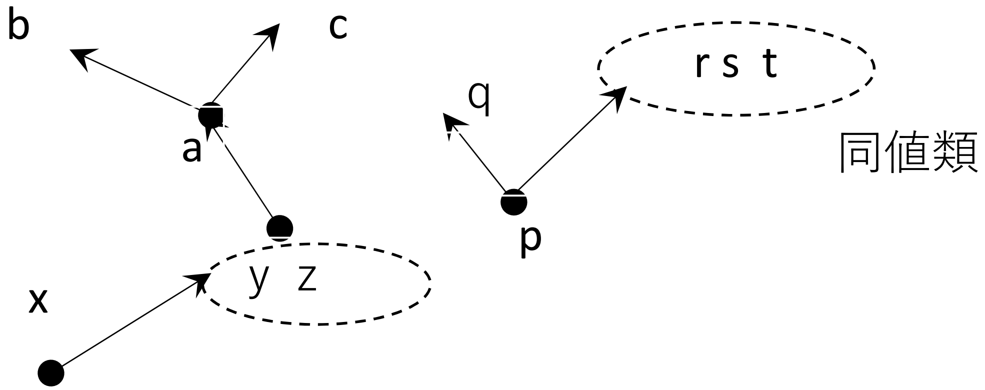
オブジェクトPの要素x,y,zに関して

① xRy かつ x≠y
② xRz かつ zRyは不成立

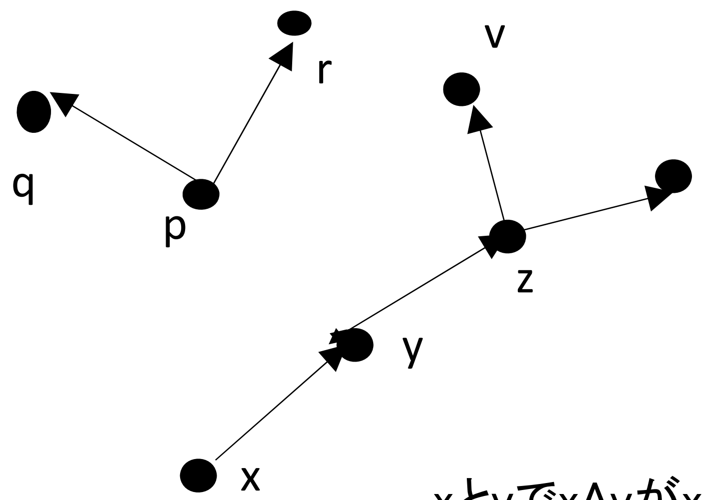
被覆関係の意味：　他のオブジェクトを介在させない直接的な2項関係

例1　オブジェクト{x,y,z}が全順序集合のとき、全順序関係Rは、
例えば、xRy yRz xRz xRx yRy zRz で重複を除くとxRzは推移律から重複となり、被覆関係*R はx*Ry r*Rz のみとなる。このとき、→を用いて、x→yとして図示したグラフをハッセ図という。

例2　{a,b,c,p,q,r,s,t,x,y,z}を擬順序集合とするとき、全ての擬順序をリストアップして、xRy, xRz, yRa, zRa, aRb, aRc, pRq, pRr, pRs, pRt, xRx, yRy, zYz, aRa, bRb, cRc, pRp, qRq, rRr, sRs,tRt, xRa, xRb, xRc, yRb, yRc
となったとする。ここから、反射律aRa, bRb, cRc, pRp, qRq, rRr, sRs,tRt, は被覆関係の条件を満足しない。また、推移律による順序関係
xRa, xRb, xRc, yRb, yRcは被覆関係を満足しない。擬順序から導かれる被覆関係は、xRy, xRz, yRa, zRa, aRb, aRc, pRq, pRr, pRs, pRtとなる。



半順序関係が定義された半順序集合では、**反射律、推移律、反対称律が成立しているが、比較可能律が必ずしも成立していないので、順序関係が定義されていないオブジェクト対があってもよい。**



xとyでxAyがxはyに等しいかまたはYの祖先であるという関係と考えることが可能である。
祖先関係は成立しない場合もある。

つづき

[上界と上限、下界￥と下限]

半順序関係Rをもつある半順序集合Pのオブジェクトxとyに対して、

(1) xとyの上界とは、集合U={z ∈P, xRz かつ yRz}である。

上界Uの任意のzに対してz_0Rzとなるz_0をxとyの上限といいsup{x,y}と書く。

(2) xとyの下界とは、集合L={w ∈P, wRx かつ wRy}である。

下界Lの任意のwに対してwRw_0となるw_0をxとyの下限といいinf{x,y}と書く。

[束と半束]

半順序集合Pに属する任意の要素x, yに関して、

(1) 上限sup{x,y}が存在するとき、Pを上半束と呼ぶ。

(2) 下限inf{x,y}が存在するとき、Pを下半束と呼ぶ。

(3)上限sup{x,y}と下限inf{x,y}がともに存在するとき、Pを束と呼ぶ。

[樹状半順序集合]

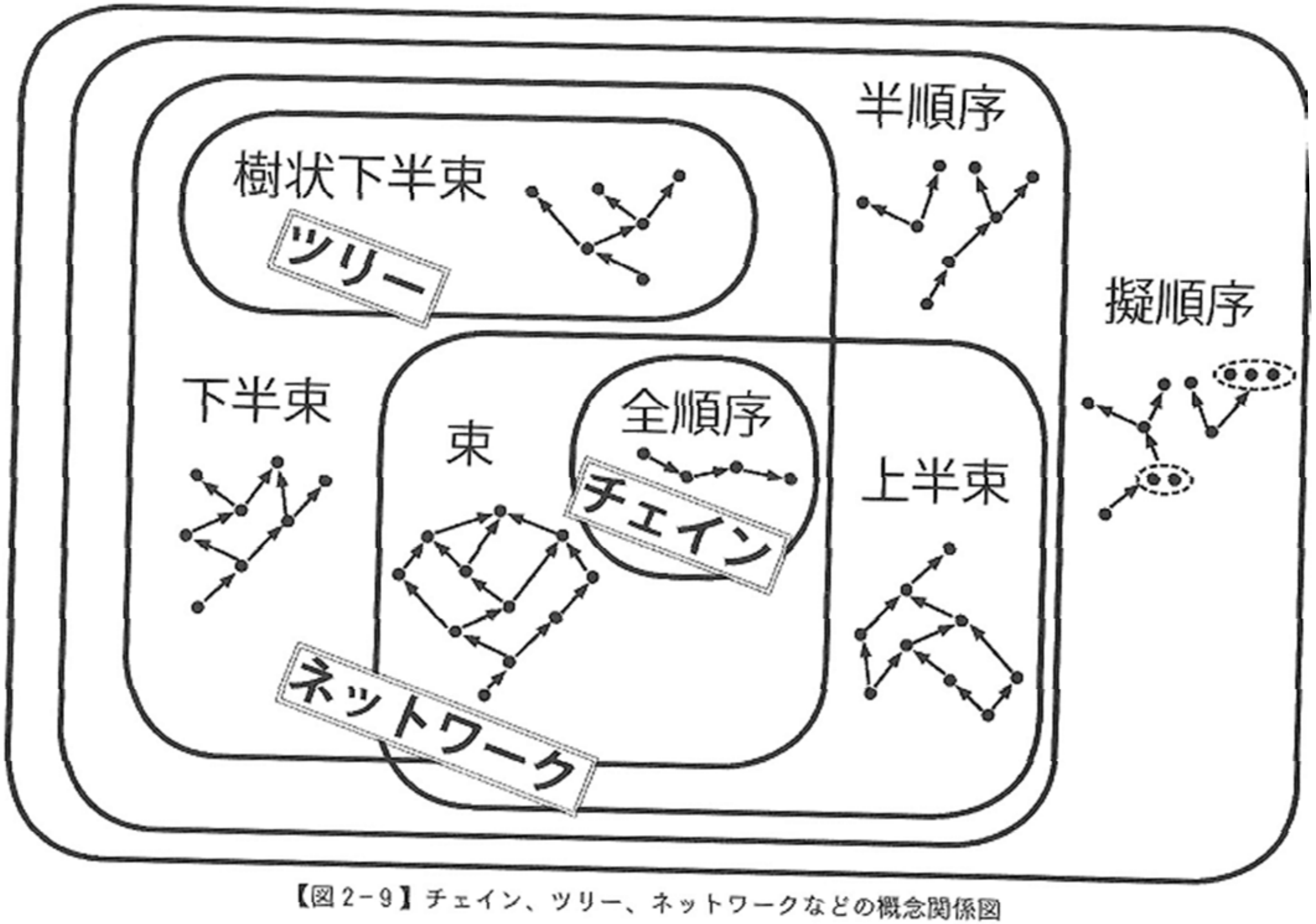
半順序関係Rの要素(例えばa,b,c)をもつオブジェクト集合で

樹状律: aRcかつbRcならばaRbまたはbRa

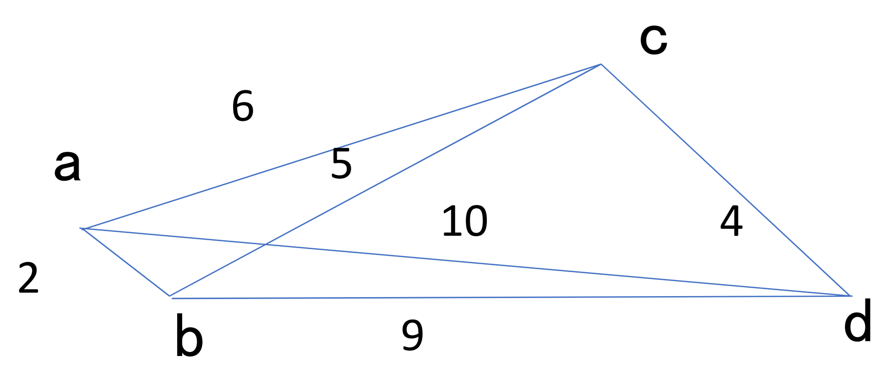
が成立する。

[樹状下半束]

樹状半順序集合Rの任意のオブジェクトx,yが下限inf{x,y}を持つ。



チェイン・ツリー・ネットワークの概念関係図



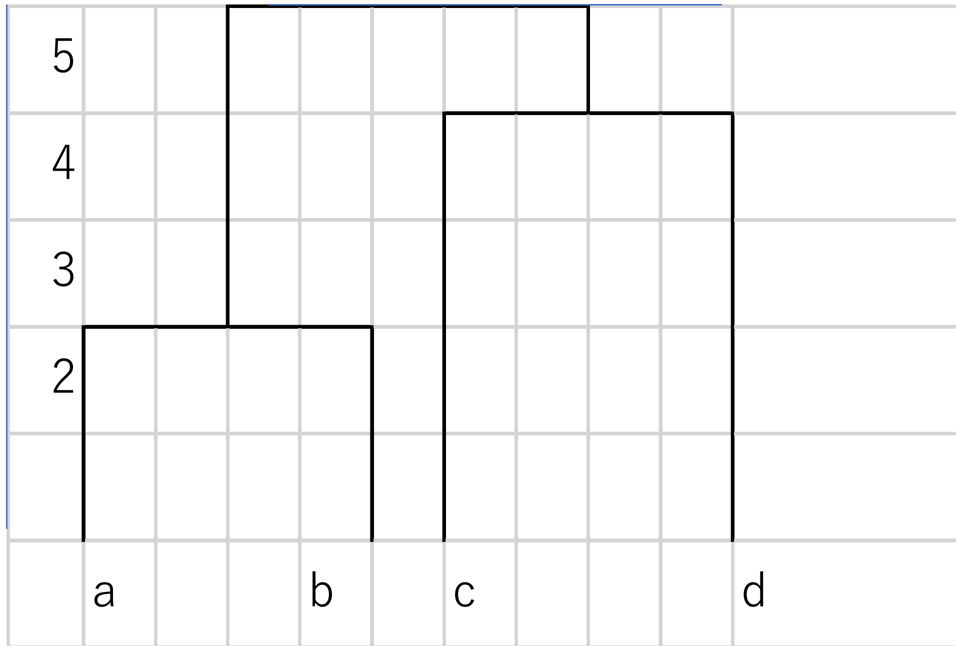
b	2		
c	6	5	
d	10	9	4
	a	b	c

c	5	
d	9	4
	(ab)	c

(cd)	5
	(ab)

クラスター分析:　距離による分類

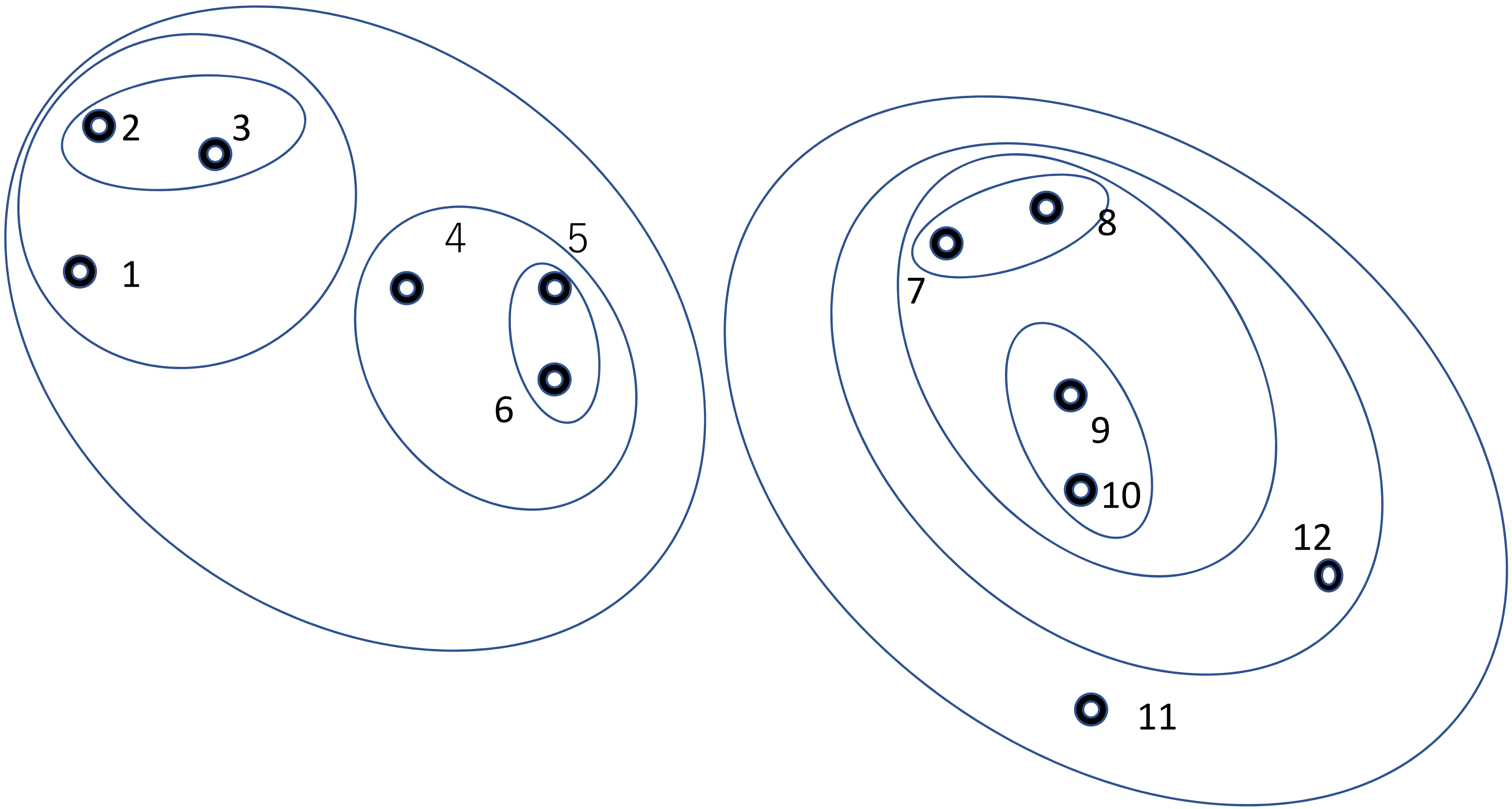
クラスタリングの例題(単連結法)



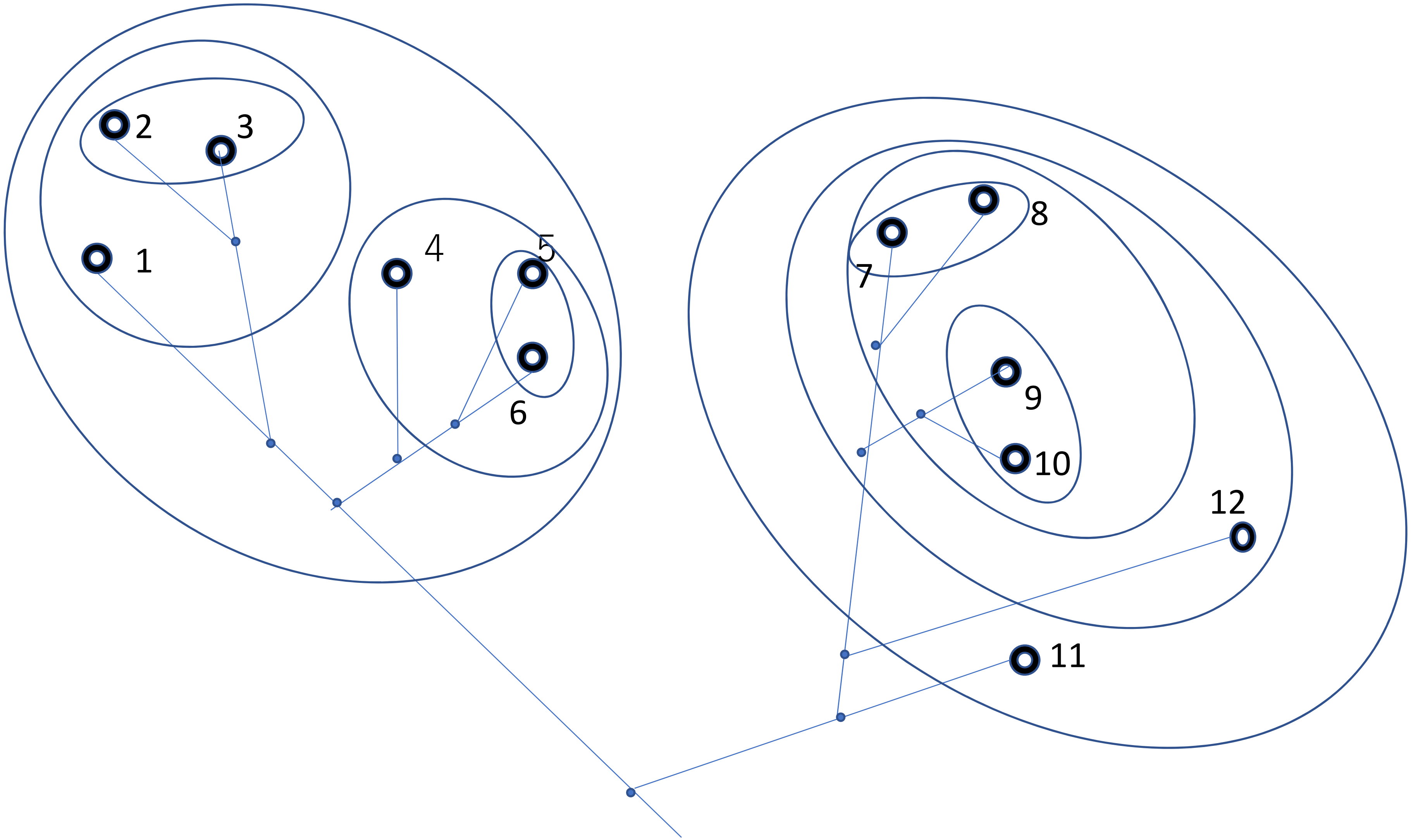
類似性を距離で表現

単連結法によるデンドログラム（樹形図）

注：樹形図：全ノードを結び、ループのないグラフ



オイラー図



分類と系統が一致する状況

集合Xに属するx、yの類似性の条件

距離Φ(x,y)

A1　Φ(x,y)≧0　　非負性

A2　Φ(x,y)= Φ(y,x)　　対称性

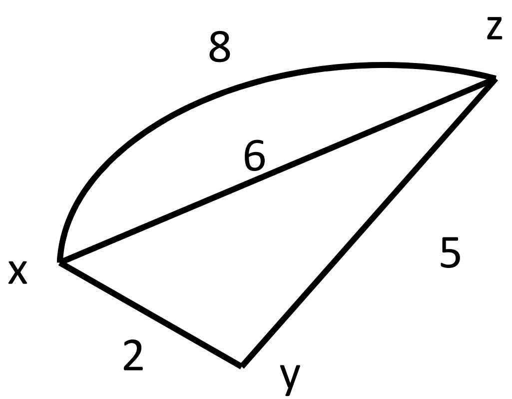
A3　Φ(x,y)=0　となる必要十分条件はx=yである　　確定性

注：　反射性はΦ(x,x)=0で、確定性の方が強い

A4　Φ(x,z)≦ Φ(x,y)+ Φ(y,z)　　三角不等式

A1ないしA4を満足する距離Φ(x,y)を計量という。

注



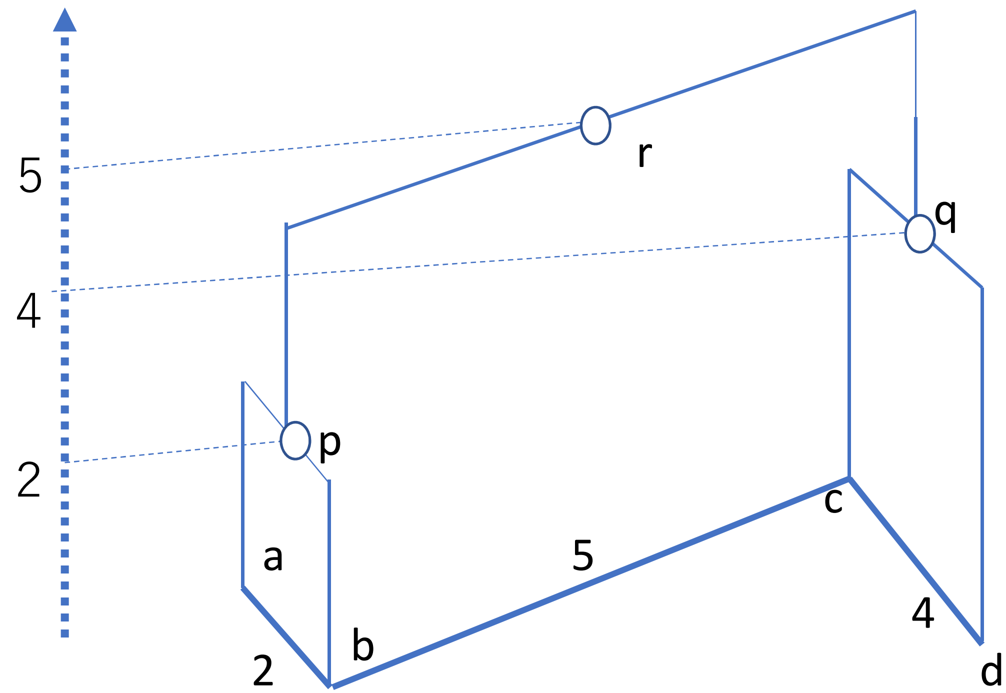
条件A4の成立について

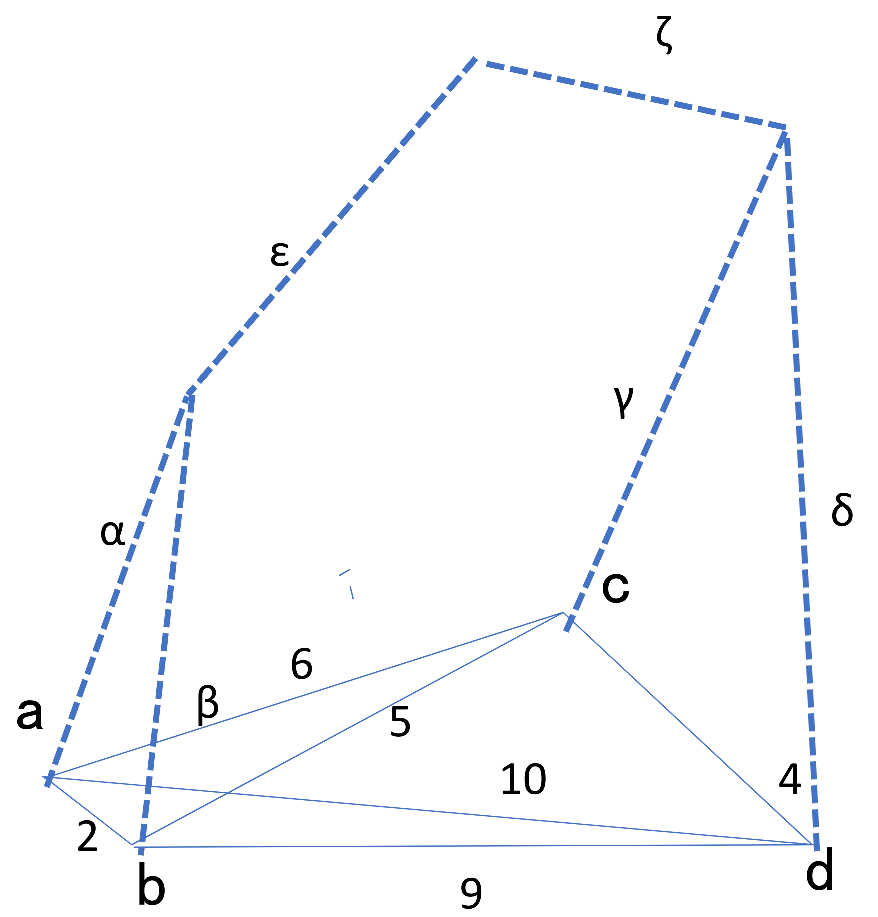
5+2 > 6

円弧で8になると、

8>2+5となり、三角不等式は不成立

樹形図による距離情報の頂点表現と経路表現
距離尺度が樹形ダイアグラムでどのように表示されるか
　白丸は内点（樹形図の分岐点）
　頂点表現で距離は内点に付与される数値：クラスターレベル（高さ）
　経路表現では樹形図の辺の長さ



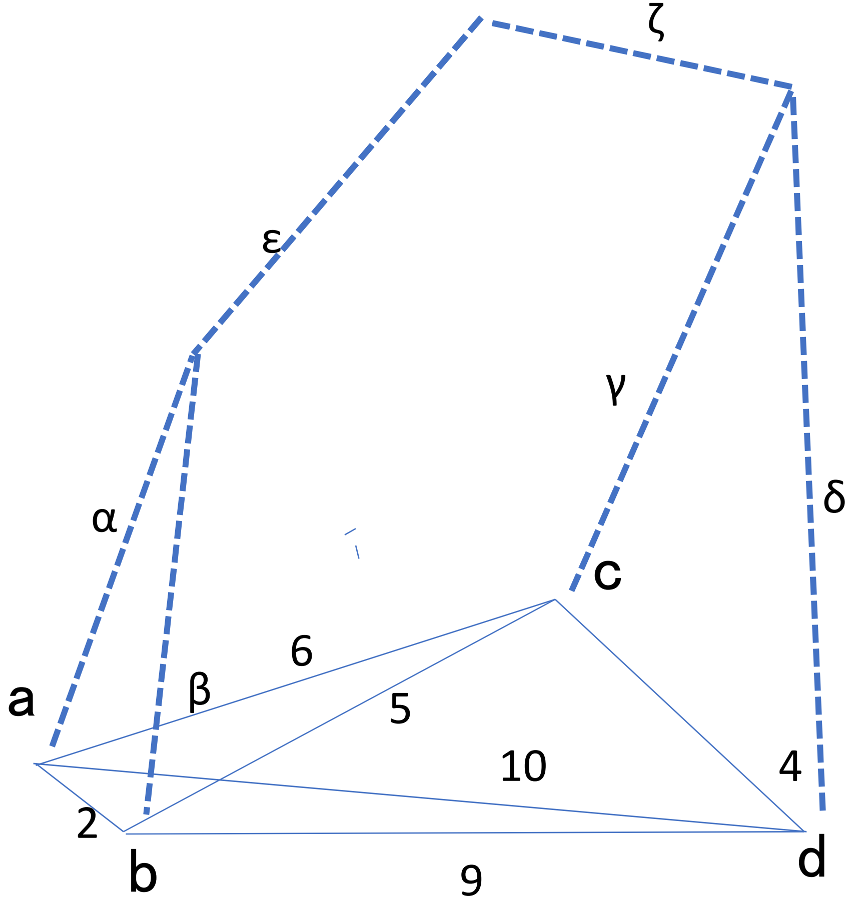


樹形図の頂点(位置)間の距離でなく、

辺(長さ$\alpha\ \beta\ \gamma\ \delta\ \varepsilon\ \zeta$)の経路表現

・・・樹形図の経路長の総和で計算

経路長の総和の式



d(a,b): $2=\alpha+\beta$

d(c,d): $4=\gamma+\delta$

d(a,c): $6=\alpha+\gamma+\varepsilon+\zeta$

d(a,d): $10=\alpha+\delta+\varepsilon+\zeta$

d(b,c): $5=\beta+\gamma+\varepsilon+\zeta$

d(b,d): $9=\beta+\delta+\varepsilon+\zeta$

を計算すると、

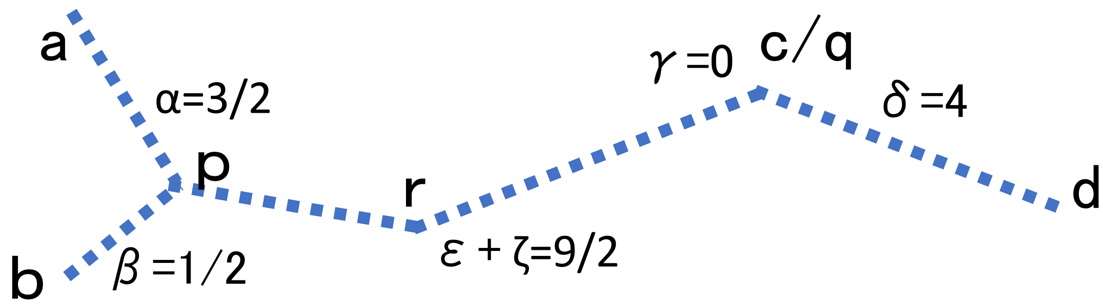
$\alpha=3/2$

$\beta=1/2$

$\gamma=0$

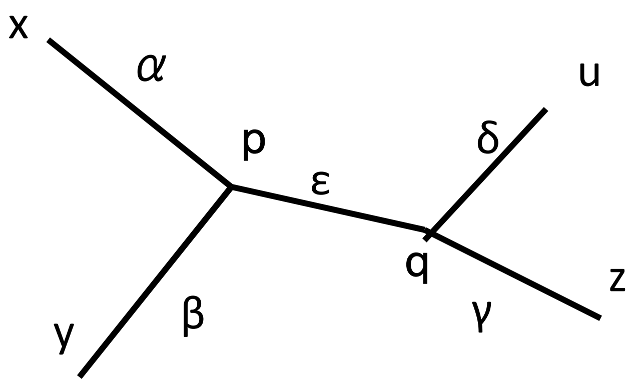
$\delta=4$

$\varepsilon+\zeta=9/2$



樹形図
- デンドログラム（頂点間の距離で辺の情報は実質的な意味を持たない）
- <u>経路表現の樹形図</u>（経路長の和でオブジェクト間の距離を表現する）
  端点（オブジェクトに対応する）間の距離は、<u>仮想的な内点</u>経由の経路長で評価）

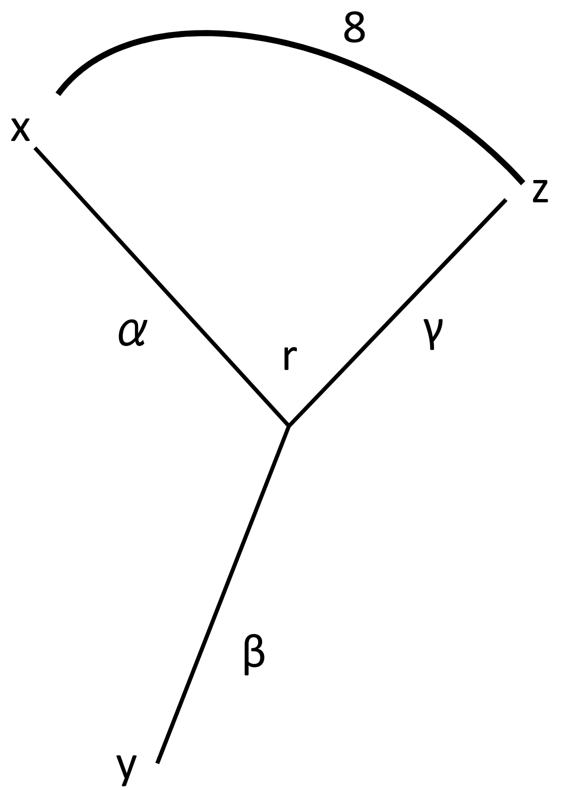
樹形図は分岐的な表現であるが、デンドログラムは分類のためで、経路表現は「つなぐ」ことに関心がある。



	x	y	z	u
x	0	9	7	6
y	9	0	6	5
z	7	6	0	5
u	6	5	5	0

三角不等式は満足するがαからεを求めるとε=-1となり、樹形図では4端点を接続不可である。また、解が存在しない場合がある。

では、$\alpha=5$ $\beta=4$ $\gamma=3$ $\delta=2$ $\varepsilon=-1$ となる。なお、4つのオブジェクトから3つを取り出して構成される組、{x, y, u} 等では三角不等式は成立している。



d(x,y)=2=α+β

d(y,z)=5=β+γ

d(x,z)=6=γ+α

からα,β,γを求めるとα=3/2 β=1 γ=9/2

となり、経路表現が求められる。

ここで、

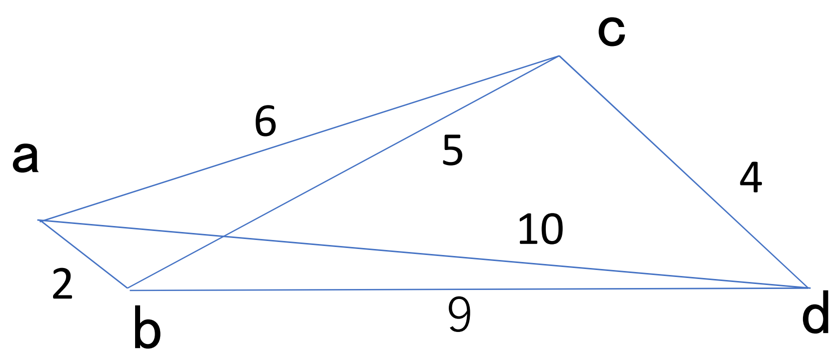
d(x,z)=8=γ+αとすると（円弧で結ぶ）、

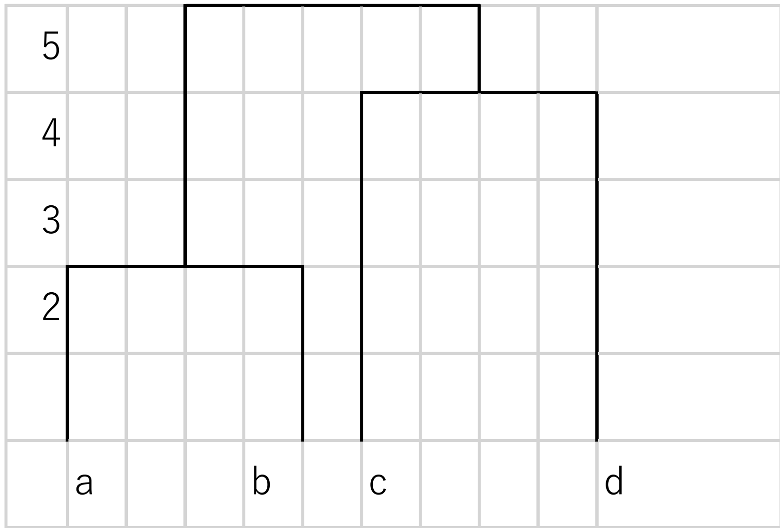
α=5/2、β=-1/2、γ=11/2

で経路表現はできない。

この三角不等式の条件だけでは経路表現できない

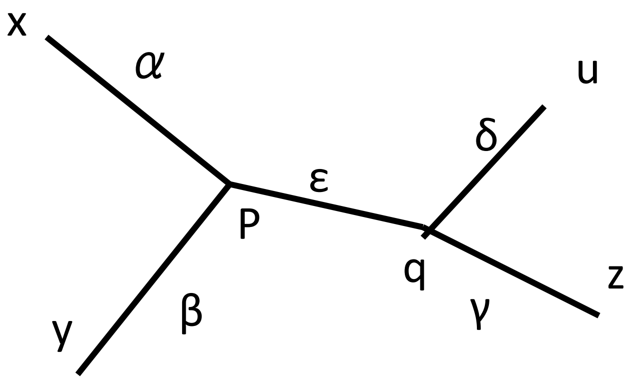
4オブジェクトを例示する。





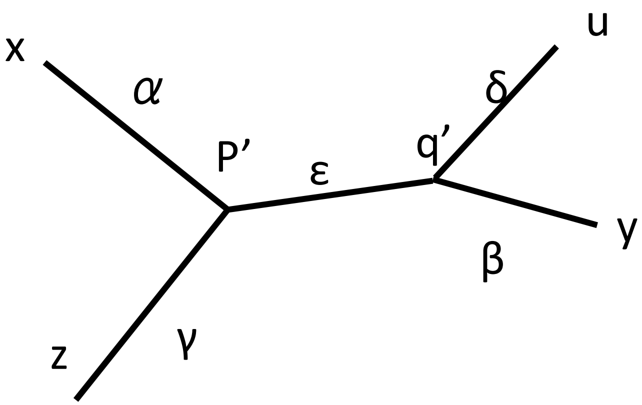
数量分類学：操作的分離単位(OUT)の間の距離を数値化

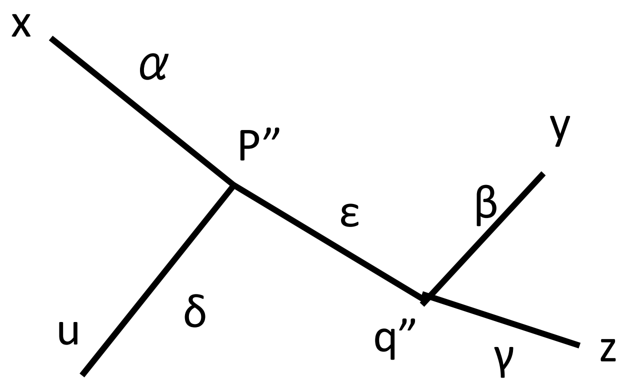
デンドログラム（樹形図のうちの表形図）
単連結法によるクラスタリング



	x	y	z	u
x	0	9	7	6
y	9	0	6	5
z	7	6	0	5
u	6	5	5	0

三角不等式は満足する
がαからεを求めると
ε=-1となり、樹形図では
4端点を接続不可である。
また、解が存在しない
場合がある。

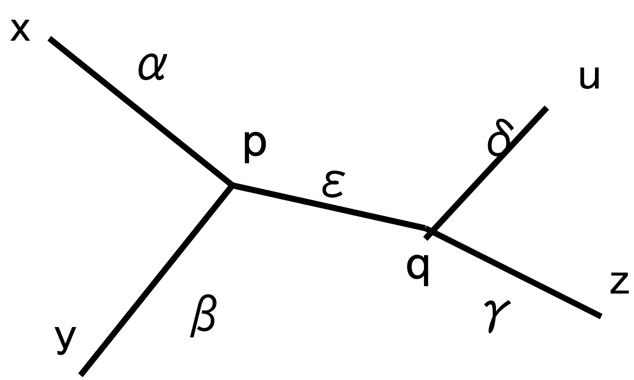




d(z,y): $6=\gamma+\varepsilon+\beta$
d(z,u): $5=\gamma+\varepsilon+\delta$　　$\beta-\delta=1$　　左と同様に解はない
d(x,y): $9=\alpha+\varepsilon+\beta$
d(x,u): $6=\alpha+\varepsilon+\delta$　　$\beta-\delta=3$　解はない

上記のp’q’やp”q”では解は存在しない：<u>εが非負</u>

経路表現が可能になる樹形図を構成するための必要条件



	x	y	z	u
x	0	9	7	6
y	9	0	6	5
z	7	6	0	5
u	6	5	5	0

$d(x,y)=\alpha+\beta$

$d(x,z)=\alpha+\varepsilon+\gamma$

$d(y,z)=\beta+\varepsilon+\gamma$

$d(y,u)=\beta+\varepsilon+\delta$

$d(z,u)=\gamma+\delta$

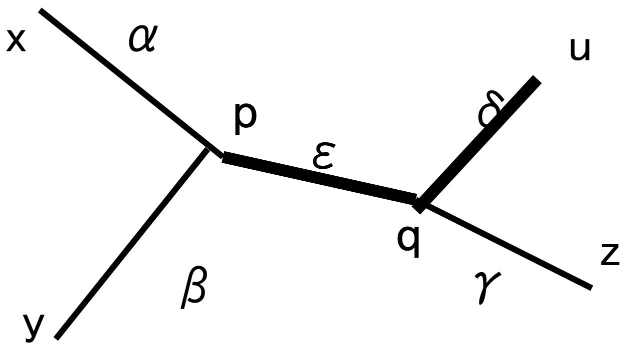
$d(x,u)=\alpha+\varepsilon+\delta$

uを含む3点の距離と経路長を求めるとx,y,uに関して

$d(x,u)+d(y,u)-d(x,y)=2(\varepsilon+\delta)$であるから、

$\varepsilon+\delta=1/2\{d(x,u)+d(y,u)-d(x,y)\}$

となる。



同様にx,z,uの3点に関しては、

$$d(x,u)+d(z,u)-d(x,z)=2\delta$$

となり、

$$\delta=1/2\{d(x,u)+d(z,u)-d(x,z)\}$$

となる。

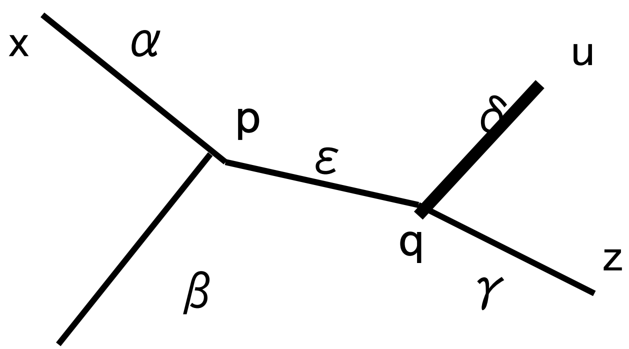
さらに、x,z,uの3点に関しては、

$$d(y,u)+d(z,u)-d(y,z)=2\delta$$

となり、

$$\delta=1/2\{d(y,u)+d(z,u)-d(y,z)\}$$

となる。



さらに、y,z,uの3点に関しては、

$$d(y,u)+d(z,u)-d(y,z)=2\delta$$

となり、

$$\delta=1/2\{d(y,u)+d(z,u)-d(y,z)\}$$

となる。

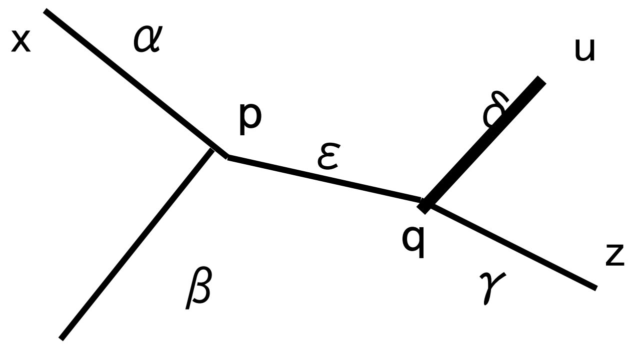
さらに、x,z,uの3点に関しては、

$$d(y,u)+d(z,u)-d(y,z)=2\delta$$

となり、

$$\delta=1/2\{d(y,u)+d(z,u)-d(y,z)\}$$

となる。



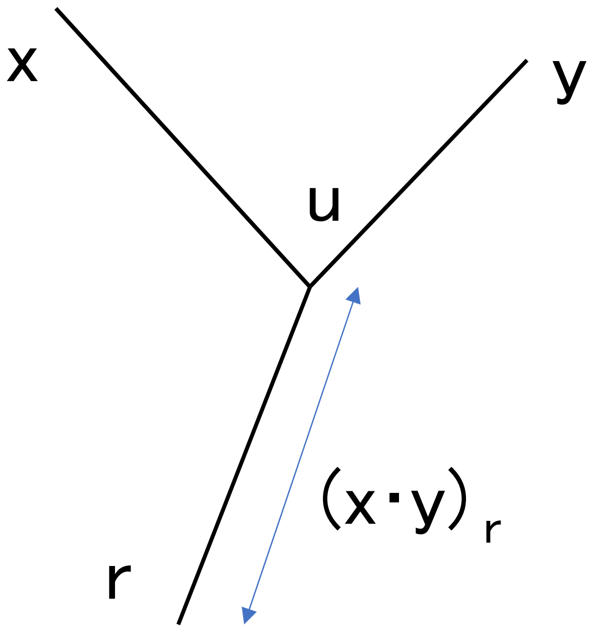
グラモフ積

距離d定義されているとき、

あるrに関するxとyのグラモフ積$(x\cdot y)_r$は

$(x\cdot y)_r = 1/2\{d(x,r)+d(y,r)-d(x,y)\}\ (x \neq y)$

$= 0\ (x=y)$



グラモフ積の説明

端点x,y,rについては距離が与えられているが内点uについての情報はないが、上記のように仮想的な内点uから参照点rへの距離をdの線形結合で計算可能としていって、<u>端点の内点を構築</u>

樹形図が描けるか否かの条件は$\varepsilon \geqq 0$である。

定理: 3つの実変数X,Y,Z（$X \leqq Y \leqq Z$)で

$X \leqq \max(Y,Z)$; $Y \leqq \max(Z,X)$; $Z \leqq \max(X,Y)$

となる必要十分条件は、$X \leqq Y=Z$である。

定義　超計量性：任意の3点X,Y,Zに関する距離dに関して

$d(x,y) \leqq \max\{d(x,z),d(y,z)\}$; $d(y,z) \leqq \max\{d(x,y),d(x,z)\}$;

$d(x,z) \leqq \max\{d(x,y),d(y,z)\}$;

のとき、dを超計量という。

ファリス変換

ある定数Cからグロモフ積を引いた$\{C-(x \cdot y)_u\}$をいう。

$C-(x \cdot y)_u \leqq C-(x \cdot z)_u = C-(y \cdot u)_u$が成立する。

ファリス変換は超計量で、dを用いると

$d(x,y)+d(z,u) \leqq \max\{d(x,z)+d(y,u),d(x,uz)+d(y,z)\}$

そして、x,y,z,uに関する距離dの和

$(x,y)+d(z,u),d(x,z)+d(y,u),d(x,u)+d(y,z)$

が超計量性条件を満足するとき、距離dを相加的計量と呼ぶ。

超計量性が成立すれば相加性は成立し、相加性が成立すれば計量性は成立するが逆は成立しない。

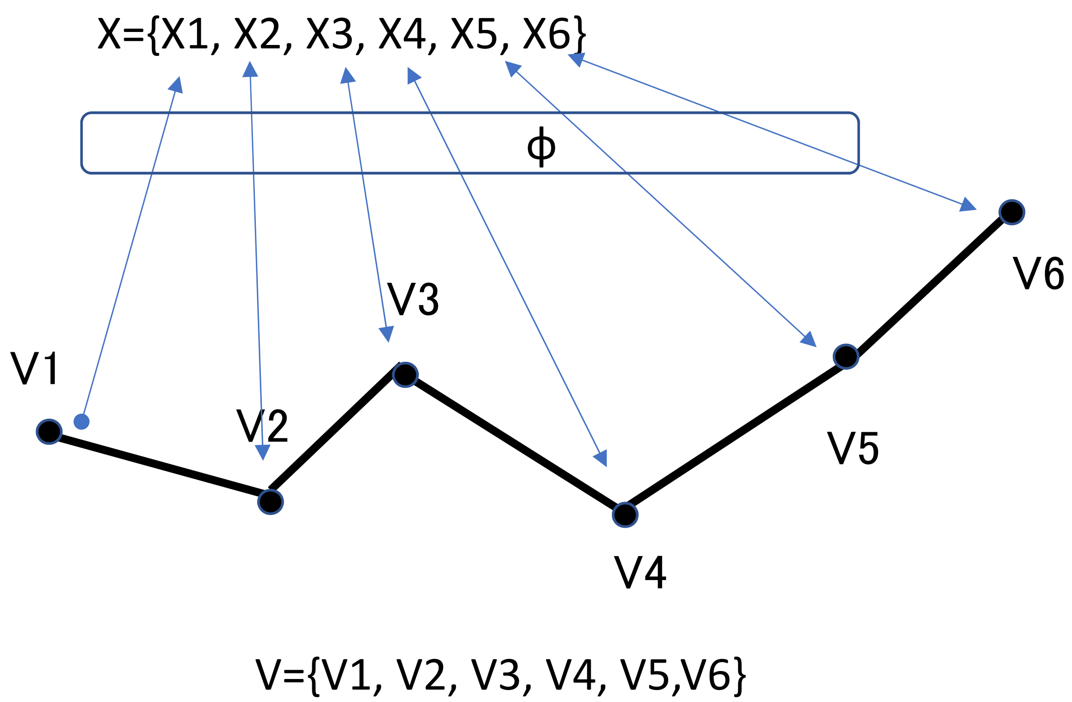
<u>樹形図で接続できる必要十分条件は相加性の成立　（樹形性定理）</u>

X樹
空集合でない有限集合XのX樹
条件1　頂点集合Vと辺集合E
条件2　ラベリング写像φ:X→V度数2以下の全頂点v∈Vは
　　v∈φ(X)である。

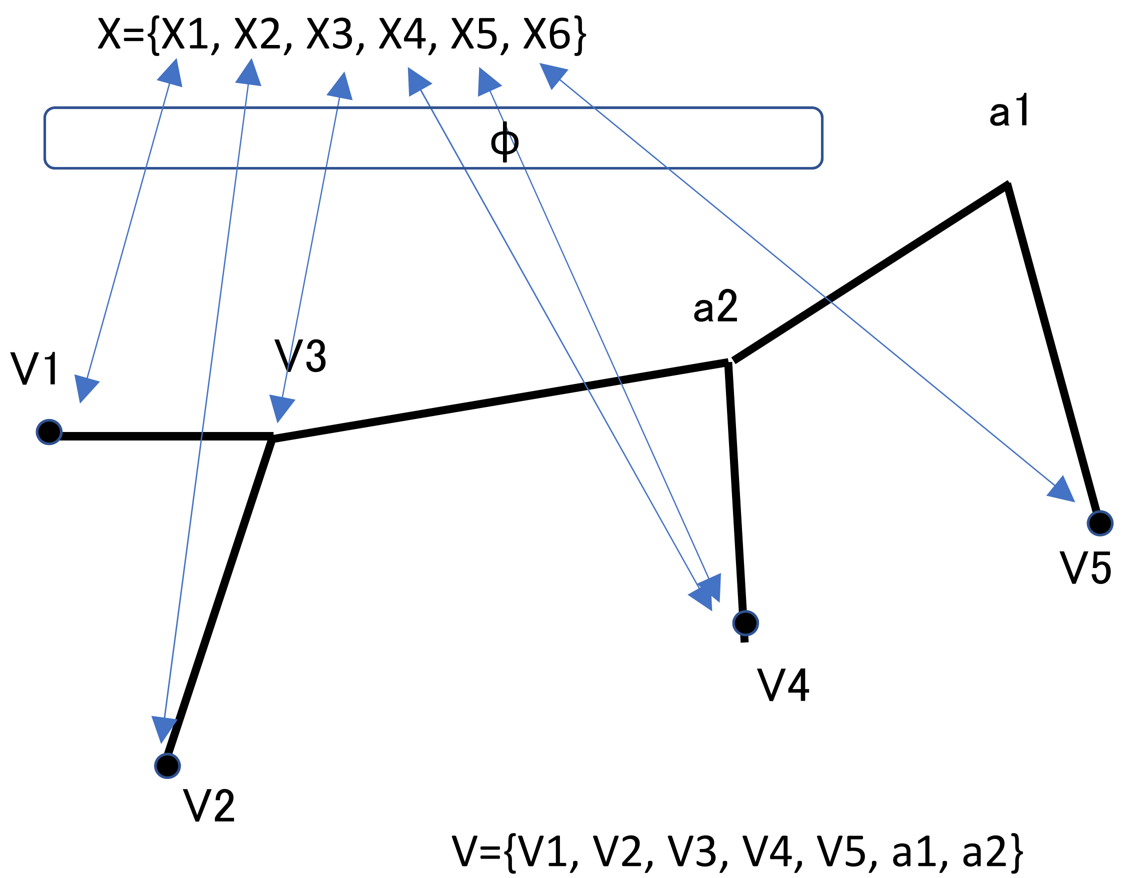
退化したツリー（樹形図）で、両端は度数1でそれ以外は度数2
度数3以上の頂点についての条件はない。

系統X樹
空集合でない有限集合XのX樹
条件1　頂点集合Vと辺集合E
条件2　ラベリング写像φ:X→V度数2以下の全頂点v∈Vは
　　v∈φ(X)である。
条件3　φの値域φ(X)は頂点集合Vに含まれる度数1の頂点集合
条件4　対象物を端点の対応φは全単射

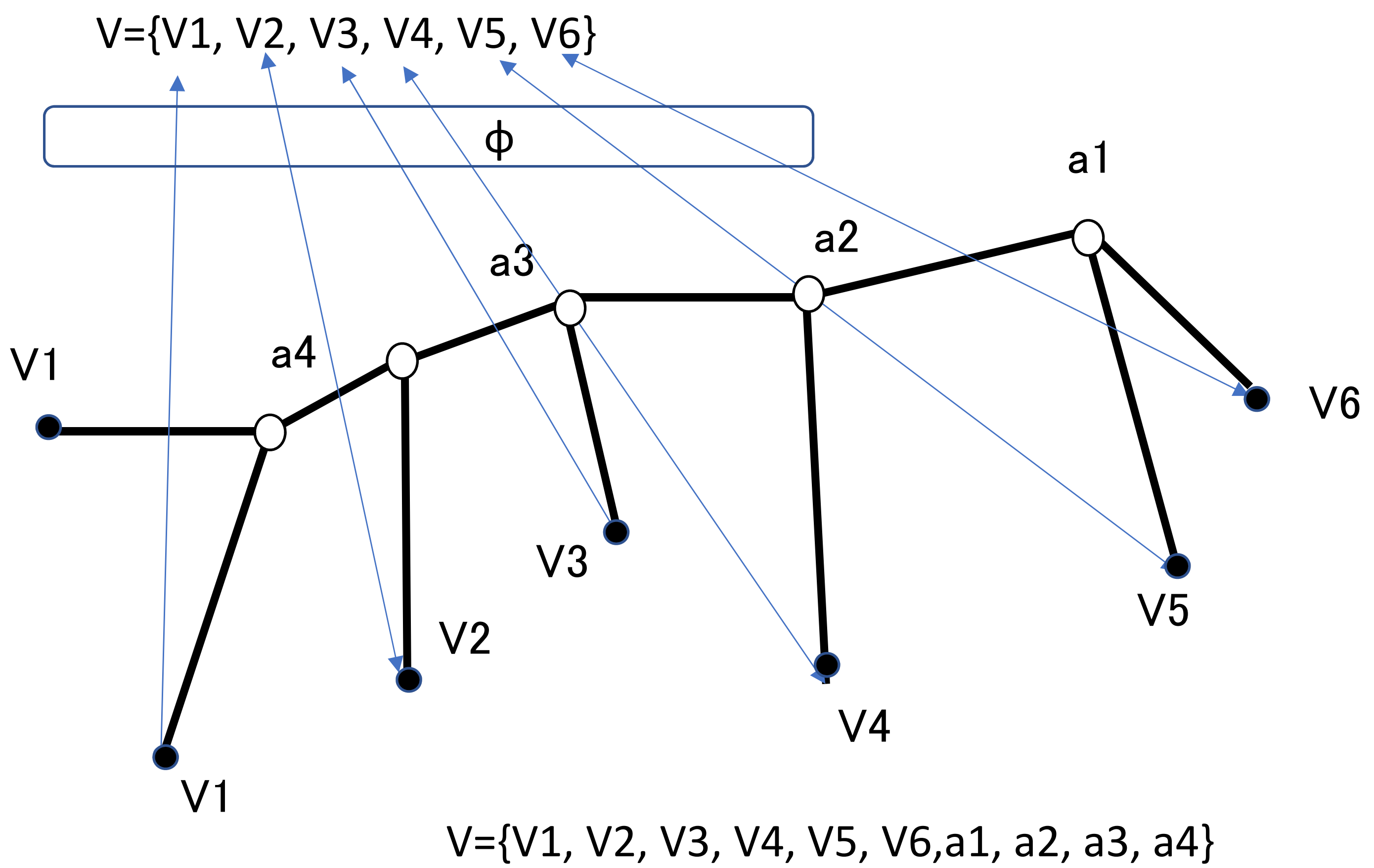
•



最も単純なX樹の例（基底樹は実線で連鎖状のチェインで、全頂点をつなぐ。対象物集合集合Xを定義域とするラベリング写像ϕから頂点集合への対応付けを矢印で示す。



一般の系統X樹の例（基底樹は実線で、分岐的な樹形図でラベリング写像ϕの対象物集合Xと頂点集合の対応付けは1対1でなくてよく、度数3の頂点への対応も可



系統X樹の例（基底樹は実線で、分岐的な樹形図でラベリング
写像ϕの対象物集合と頂点集合の対応付けは↔で示す全単射

系統X樹の利用

実在する可視的な対象物集合
不可視的な内点を含む頂点集合を補って体系化

頂点構造の評価にアブダクションを用いる場合もある。

演繹

演繹は、仮定と規則「aならばbである」から結論bを導く。
妥当な演繹は、仮定が真であれば結論も真であることを保証する。

帰納

帰納は、仮定が結論 b を伴ういくらかの事例を観察した結果として
規則「a ならばb である」を蓋然的に推論する。
帰納は、推論した規則が真であることを保証しない。

アブダクション

アブダクションは、結論 bに規則「a ならば b である」を当てはめて
仮定a を推論する。帰納が仮定と結論から規則を推論するのに対し、
アブダクションは結論と規則から仮定を推論する。
アブダクションは、推論した仮定が真であることを保証しない。

https://ja.wikipedia.org/wiki/アブダクション

タイプとトークン

タイプとトークンの区別（Type-token distinction）は概念そのものとその概念で指示される特定・個別の対象を区別すること。タイプとトークン[1]という対概念はパースによって導入された。
例えばガレージにあるあなたの特定・個別の自転車は「自転車」というタイプの中のトークンと理解される。このトークンとしての自転車は、時空の中の特定の位置を占めているが、タイプとしての自転車はそうではない。「最近自転車の人気が高まっている」と言うとき、この「自転車」はタイプとしての自転車を指して言っているのに対し「あなたのガレージの中の自転車」と言うときの「自転車」はトークンとしての自転車である。
タイプは時空間に位置を持たない抽象的対象と考えられ、これは「特定の対象（particulars）」が時空間に位置を持っているのと対照的である。この「特定」という用語は単純に言えば「具体的で物理的な対象」というような意味である。タイプがトークンをまったく持たない場合も考えられる。例えば「二つの素数の和でないような偶数」というタイプに対するトークンは存在しない。
https://ja.wikipedia.org/wiki/タイプとトークンの区別

三中信宏の著作物系譜
と利用に対するアプローチ

基礎

小野寛晰
情報数学
共立出版
1994..2

生物系統学
東京大学出版会
1997.12
横書き

基礎

D・サダヴァ
アメリカ版大学生物学の教科書
第2巻分子遺伝学
講談社ブルーバックス
2008(2010.5翻訳出版)

系統樹思考の世界
すべてはツリーとともに
講談社現代新書
2006.7
縦書き

分類思考の世界
なぜヒトは万物を「種」に分けるのか
講談社現代新書
2009.9
縦書き

応用

進化思考の世界
ヒトは森羅万象をどう体系化するか
NHK出版2
2010.9
縦書き

系統樹曼荼羅
チェイン・ツリー・ネットワーク
2012.11
NTT出版
横書き

思考の体系学
分類と系統から見たダイアグラム論
春秋社
2017.4
縦書き

系統体系学の世界:生物学の哲
学がたどってきた道
技術評論社
2018.4
横書き(?)

統計思考の世界
勁草書房
2018.4
縦書き(?)

ソフトウェア

系統推定に用いられる形質データの数の巨大化：
形質数だ けでなく、端点（種）の数も増大し、最節約法や最尤法等の目的関数

A.共通祖先の復元は

祖先復元遺伝子や祖先復元タンパク質は、 生物が過去の世界で持っていた機能や活性を 現在の世界では失っている場合がある。 祖先の復元は、失った機能、失った活性を見つける事かできる。 極論すると新しい人工タンパク質の発見を 意味する。

B.近縁関係の重要性は

生物関係者であれば大変に重要な事柄であることは周知の事実である。 「系統樹」の中でも最節約法は近縁関係を正確に出すことで有望である。

C.「系統樹」を使用する意義

大量のデーターから網羅的に調査している時間の評価は、 事の優位性が把握出来なく、時間の浪費にならないかの懸念ではある。 この場合「系統樹」の最節約法に基づく探索を 行えばもっとも効率的に最適な標的を見つけることができる可能性がある。

反省と今後の方向

・精密な理解（数学の勉強））

基礎的な素養の不足

・系統樹関係ソフトウェアの利用可能性

◎出版物の系譜としての理解

・データの収集（加工技能）と体系化

公開情報（ビデオ）の利用が可能

過去の出来事（不可知の現象）にベストな説明を求める

ご清聴とディスカッションに深謝！