

TM-0860

グループ知識獲得支援システム
GRAPEにおける初期知識ベース獲得機能

国藤 進, 上田晴康, 須永知之,
井深克憲, 岩内雅直(富士通)

February, 1990

©1990, ICOT

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

グループ知識獲得支援システム G R A P E
における初期知識ベース獲得機能
An Initial Knowledge Base Acquisition Mechanism
in Group Knowledge Acquisition Support System GRAPE

國藤 進 上田晴康
Susumu Kunifuji Haruyasu Ueda

富士通国際情報社会科学研究所
International Institute for Advanced Study of Social Information Science
FUJITSU LIMITED

須永知之 井深克憲 岩内雅直
Tomoyuki Sunaga Katsunori Ibuka Masanao Iwauchi

富士通ソーシャルサイエンスラボラトリ
FUJITSU SOCIAL SCIENCE LABORATORY Ltd

Abstract: We are researching and developing a new group knowledge acquisition support system GRAPE (Group knowledge Acquiring and Processing Environment), to overcome knowledge acquisition bottleneck in building knowledge-based systems. GRAPE has two functions, i.e. an initial knowledge-base acquisition function and a planning problem support function to solve a planning problem. This paper presents a design philosophy of the initial knowledge-base acquisition function of GRAPE. The function supports the problem formalization steps and the knowledge representation step of the knowledge-based system methodology from the participants of given macro-decision problem.

1. まえがき

知識システム構築最大のボトルネックである知識獲得ボトルネック [Lenat 86] を解消するために、著者らはグループ知識獲得支援システム GRAPE (Group knowledge Acquiring & Processing Environment, 「葡萄」) を研究開発中 [Kunifuji 89] である。本論文では、GRAPE の初期知識ベース獲得機能について概説する。GRAPE を設計するに当たって、知識システム構築方法論と知識獲得支援システム/ツールのサーベイ [Kunifuji 88] を行った。その結果、知識システム構築の上流工程支援およびグループ知識獲得支援に、研究開発の立ち遅れがあることが分かり、従来の知識獲得支援ツールとの接続を考慮した上で、新たにグループ知識獲得支援システム GRAPE を提案することになった。

知識システム構築方法論は、一般に次のような問題解決過程 [Kunifuji 88] からなる。すなわち、①問題の設定、②既存技術の評価、③知識源の同定、④専門家モデルの同定、⑤ユーザモデルの同定、⑥知識表現の選択、⑦知識の抽出、⑧知識の変換、⑨知識ベースの管理、⑩性能の評価である。すなわち、①～⑤が問題の定式化段階、⑥が知識表現の選択段階、⑦～⑩が知識の獲得段階に相当する。それぞれに対応し、問題定式化ボトルネック、知識表現ボトルネック、知識獲得ボトルネックという三種のボトルネックが指摘さ

れている。

著者らのサーベイによると、⑦～⑩の後半の工程を支援するツールは多々存在するが、①～⑥の前半の工程を支援するツールはほとんど存在しない。また一人の専門家からの知識獲得支援ツールは存在するが、多くの人々（グループ）からの知識獲得支援ツールは存在しない。そこでGRAPE では、与えられた問題に関連するグループ・メンバの知見を集集し、システム分析の各種手法を駆使して、②～⑥の過程を支援する。特に、④と⑤のモデル同定を通じて、⑥の知識表現の決定を支援する。結果的に、GRAPE は与えられた問題の初期知識ベース獲得機能と問題解決支援機能とからなる。後者としては、特定の問題領域を設定しないとツール化しにくいので、計画問題支援機能を付与することにした。これは、計画問題（プランニング問題、スケジューリング問題等）への適用といった実際の問題解決に、GRAPE が役に立つことを実証するためである。

2. GRAPE 構想

本研究は知識獲得支援ツール／システム面の研究開発から協調問題解決向きグループウェアに問題接近する第一歩となることを目標としている。最初に、「葡萄」GRAPE の研究開発環境を規定する。「葡萄の一粒」に相当する個々のワークステーションが、ローカルネットワークを経由し、互いに通信しあい、共通の問題解決のための合意を形成しあうという環境を想定する。「葡萄の一房」は各地に散在し、サイトを形成すると考える。各サイト間は、「葡萄の棚」に相当するグローバルネットワーク経由で通信される。著者らが研究開発中のシステムは第5世代コンピュータプロジェクトの一環で行われているので、「葡萄の一粒」はオブジェクト指向論理プログラミング言語ESP マシンであるPSI-IIを用いる。複数台のPSI-IIマシンの複合体が各地に散在し、サイトを形成し、各サイト間は、イーサネット経由で通信されることになる。サイトのあるもの、あるいは「葡萄の一粒」のあるものは、SUN あるいはMacintosh 上でインプリメントした方が効率的であろう。実際、グラフの表示のための計算やグリッド [Fransela 77] の固有値計算等は、SUN あるいはMacintosh 上でインプリメントした方が得策である。

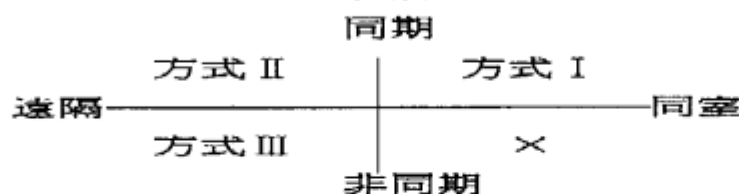


図1 グループウェアの分類

問題解決へのグループ参画を考える場合、グループデザインに参加する参画者 (participant) が複数のワークステーションで同期／非同期で参加するか、同室／遠隔の形態で参加するかによって、システムの構築法 [Ishii 89] が異なってくる。実際、図1に見られるように、三種のシステムが存在しうる。第IV象限に属するシステムは、概念的に存在しない。GRAPE も、図1のどの象限に属するかによって、システム構築法が異なってくる。ここでは便宜的に、方式Iのグループ知識獲得支援システムをGRAPE、方式II、IIIのグループ知識獲得支援システムをGRAPEsと呼ぶことにする。著者らは簡単のため、電子黒板システムColab [Stefik 87] と同様に、現段階では方式I (同期・同室) のインプリ

メンテーションを考える。順次、方式Ⅱや方式Ⅲのインプリメンテーションへと拡張していく。

次に、GRAPE の初期知識ベース獲得機能の基本構想を述べる。GRAPE 研究開発の目的は、知識システム開発の上流工程（問題定式化、知識表現選択）支援とグループ知識獲得支援の本質を明らかにすることである。前者については、知識ベースのラビッドプロトタイプ支援機構の構築を通じて、主観的な知識ベース構造を決定することに特徴がある。後者については、グループウェアに対する先行研究であり、KJ法〔Kawakita 87〕を中心とする協調問題解決支援環境の構築に特徴がある。

GRAPE の精神はKJ法で、その技術としては「Colab+ AQUINAS+ α 」〔Stefik 87, Boose 87〕を活用することをモットーする。 α としては、諸々のシステム分析技法を活用する。GRAPE は初期知識ベース獲得機能と計画問題支援機能からなる。初期知識ベース獲得機能は、与えられた問題のマクロ決定を支援し、与問題に対する主観的評価のガイドライン（枠組み）を決定する。これに対して、計画問題支援機能は、与えられた問題のミクロ決定を支援し、既存知識を再利用しつつ、前記ガイドラインを満足するプランを生成していく。本論文では、GRAPE の初期知識ベース獲得機能をグループウェア的側面から明らかにしていく。GRAPE の初期知識ベース獲得機能は、図2に示されているように、仮説抽出、仮説構造化、属性抽出、属性構造化、グリッド分析・統合を経て、主観的概念木の生成を支援する。それぞれの機能の詳細は、次章以降で述べる。

3. 初期知識ベース獲得機能のシステムフロー

GRAPE 初期知識ベース獲得機能は、いわゆる前処理、基本処理、後処理に相当する問題決定部、基本（グループ）決定部、決定調整（コンサルテーション）部の三つの処理部から成る。標準的には、基本決定部で、その問題に関与する全てのグループメンバが参加し、決定調整部でグループメンバと専門家とのコンサルテーション・プロセスを通じて、グループ決定の結果を調整していく。問題の種類によっては、基本決定部で専門家の知見を整理し、決定調整部でグループメンバとのコンサルテーションを行う。

(1) 問題決定部

GRAPE ユーザは最初、解くべき問題を事前に決定しておく。解くべき問題の決定そのものを支援するグループウェア構築も、極めて魅力的な研究開発課題であるが、KJ法そのものを支援するツールを作ることと同様、現状では極めて困難なのでGRAPE 向きの問題が所与であると仮定する。我々の目標は「コンピュータの得意なことはコンピュータに、人間の得意なことは人間に！」という設計方針のもと、コンピュータと人間との協調型アーキテクチャを明らかにすることにある。また随所で、ユーザのコンピュータ入力のみられるが、誰でも気楽に参加できるように、マウスのクリックによるメニュー選択とキーワード入力を基本とする。

(2) 基本（グループ）決定部

GRAPE の基本決定部は、以下に述べるプロセスから成る。

①最初、その問題のグループデシジョンに参加するグループメンバを決定する。ここでは各メンバの役割は対等であり、その問題の解決のための計画に積極的に参加していることを強調し、彼らのことを参画者と呼ぶ。参画者の人数は、2～10人位を想定しているが、

4～6人位が理想的であろう。参画者の人数がやたら大きいと、会議で合意をとるのが困難と同じ現象を呈する。このプロセスは電子会議を知的かつ生産的に進めるにはどうするか的设计であり、基本的にはグループ問題解決の要領で決定していけばよい。

②上記参画者の中から、コーディネーターとなる司会役を決定する。当面、司会役の役割は通信の同期をとることであり、ネゴシエーションをガイドすることではない。従って、司会役にコーディネーション特有の能力を要求するものではないが、将来的には生産的なデバートのガイド能力を要求することになるであろう。

③司会役からみて参画者が通信し易いマシン接続を決定し、それに伴いマシン間がトークできるように立ち上げる。ここに、マシンの接続方式には、階層結合とネットワーク結合が考えられるが、司会役が通信の同期をとる関係上、ここでは階層結合をとる。PSIのプロセス立上げが遅いため、ネットワーク結合は実用的でない。参考までに、ネットワーク結合の場合、プロセッサの個数の二乗のオーダーに比例したネットワーク設定時間が、階層結合の場合、プロセッサの個数のオーダーに比例したネットワーク設定時間がかかる。ただし、方式Ⅲの場合はネットワーク結合の方が現実的である。

④問題の解の候補集合を仮説（あるいは対象）と呼ぶ。全ての参画者が考えられる仮説を入力していく。入力された情報はWYSIWIS(What You See Is What I See)原則〔Stefik 87〕で、全ての参画者がウィンドーで見れるようにする。またサブグループ間コミュニケーションの重要性に鑑み、サブウィンドーを共用するコミュニケーションも許容することにする。理想をいえば、Colabのように、大型の共通スクリーンBoardnoterの欲しい所である。

⑤仮説の数がある値(7 ± 2)以上の場合、仮説間のクラスター分析を行い、仮説間構造を抽出していく。クラスター分析としては、(KJ法を含む)マニュアル法、ファジィクラスターリング、林の数量化理論等、種々のものが考えられるが、GRAPEではデフォルトとして、ファジィクラスターリング〔Mizumoto 88〕を標準装備している。仮説の総数としては、一人当たり10個以内、最大100個程度の仮説が、快適な対話環境を保証する上で、妥当なリミットと考える。

ただし、ファジィクラスターリングの場合、グループ全体の値は算術平均を採用する。ファジィ行列の整合性をとるためには、ある種の制約条件を満足しなければならない。そこで、グループ入力の整合性を取るには、付録で述べる幾つかの工夫を要する。

⑥上記の構造化された仮説のトップレベルから順次、それらを差別化する対となる属性を全ての参画者が入力していく。入力形式は((属性, 対属性), 統合属性)というタプルで与える。対属性入力の理由は、後述の対比較を円滑に進めるためである。入力された情報はWYSIWIS原則で、ウィンドー表示されるが、階層構造の各レベルを区分する属性が抽出されていく。

⑦属性の数がある値(7 ± 2)以上の場合、属性間のクラスター分析を行い、属性間構造を抽出していく。クラスター分析としては、(KJ法を含む)マニュアル法、拡張ISM(Interpretive Structural Modelling)〔Kunifuji 79〕、属性クラスターリング等、種々のものが考えられる。GRAPEではデフォルトとして、拡張ISMを標準装備している。属性の総数としては、階層の各レベル当たり、最大20～30程度の属性が、付与されると考える。

ここでの処理の本質は属性間の従属性・独立性の分析にあるが、拡張ISMをデフォール

トとする理由は、それが基本的に属性間の従属性分析の一手法を提供するからである。お互いに併存到達可能な属性集合同志を等価な属性集合とみなし、縮約した骨格構造を抽出するプロセスを経て、属性間構造が抽出されたと考える。

⑧AHP [Saaty 80, Tone 86, Sawai 89] 利用を想定し、仮説—属性間の必要な部分構造のグリッド値 [Fransella 77, Hart 86] を入力していく。仮説—属性構造間の一対比較に関しては、グループ全体の値は幾何平均を採用する。グリッド値の入力も、全体の整合性を保持するためにある種の制約を満足しなければいけない。グループ全体の整合性を保つために、不完全一対比較行列 [Harker 87, Takeda 89] を用いた不完全情報の補完法を利用する。

(3) 決定調整 (コンサルテーション) 部

①グループ決定の結果を修正するコンサルタント役の専門家に参加してもらう。その問題の専門家であるコンサルタントの持っている情報は悔りがたいものがある。専門家の意見に謙虚に耳を傾け、上記グループ決定の過程では考慮し損ねた視点や指摘された現実的制約を用いて、実行不可能な候補を消去するのに利用できる。

②専門家と上述の参画者 (場合によっては、司会役のみ) による基本処理の繰り返しを行い、主観的概念木の修正を行う。それにより、属性の見直し、仮説の追加、評価値の修正が起こり、専門家の意見を尊重しつつ、参画者全員の意見を反映した非単調な変化が起こる。どういうメンバーが参加するかについては、プラグマティズムの精神で決定していく。例えば、参画者が少数の場合、専門家と参画者全員が参加するのが望ましいが、参画者が多数の場合、専門家と司会役のみが参加するのが現実的であろう。

上記フローに関して配慮された幾つかの検討事項を指摘しておく。属性の尺度が、名義尺度・順序尺度・距離尺度・比尺度のどれであるかによって、グリッド統合の方法が異なってくる。仮説間の親近度は距離尺度、属性間の一対比較値は比尺度と見なすのが自然なので、前者は算術平均、後者は幾何平均を採用した。これら尺度間の変換を行い、尺度統一をするのも一つのアプローチである。実際、順序尺度を距離尺度に変換する展開法が知られているが、一般にこの種の変換は極めて難しい。万一、尺度統一できたとしても、既存の多目的決定理論との違いが無くなる危険性があるので、当面、尺度統一は考えない。属性によっては、主観値を与えるのが適切なものと、客観値を与えるのが適切なものと、2種類ある。これらの属性の区分をし、初期知識ベース獲得機能では前者を、計画問題支援機能では後者を利用するのが素直である。分類や決定に利用できる客観的データベースが存在する場合、たとえばQuinlanのID3アルゴリズムを利用するのが自然である。主観的評価と客観的評価との統合は極めて大切な課題であるが、当面は別個に出来た概念木同志の整合性をチェックするのに用いる。

4. 構造化のための諸手法

本節では、GRAPEでグループの初期知識を獲得しながら、それを知識ベース向きに構造化する過程で使用される諸々のシステム・モデリング技法のうち、特にグループ知識獲得特有の全体の整合性を保持するための技法を提案する。

(1) 仮説構造化とファジィクラスタリング

仮説間の構造解析をする際、類似度という距離尺度を手掛かりに、ファジィクラスタリ

ングを行い、仮説空間の構造を抽出していく。ファジィクラスタリングの入力データは類似度行列で与えられるが、ファジィ行列は付録Aに見られるように、その整合性を維持するために推移性と対称性を満足しなければならない。そこで、GRAPE ユーザはファジィ行列の値を $n(n-1)/2$ 個付与する必要はなく、高々 n 個の値のみを付与すれば良い。このことは、ユーザの知識獲得負担を軽減する。

付録の定理2 をベースに、個々の参画者から与えられた部分ファジィ行列をマージするアルゴリズムを構築できる。たとえば、二つの類似度行列をマージするアルゴリズムを付録Bに与えてある。このアルゴリズムでは、グループの類似度を付与された入力データの算術平均とし、全体の整合性を保ちつつ類似度の補正を行っていく。

(2) 属性構造化と拡張ISM

属性間の従属性分析を行うには、著者らの開発した拡張ISM を用いる。拡張ISM は射影則と擬似推移則を満足する2項関係を分析する技法であるが、その数学的構造はブール方程式で表現できることが証明 [Kunifuji 79] されている。このブール方程式はホーン論理で表現できるので、直ちにPrologに翻訳できる。

まず、属性間従属性のグループ間マージについては、ここでは簡単のため、それぞれのメンバで与えられた2項関係を、単純に追加することにする。その理由は、従属性のみを与えた2項関係の全体に矛盾が生じないからである。もし、従属性のみならず非従属性を与えることを許容すると、このような単純なやり方では整合性が保てない。我々の関心は、拡張ISM の擬似推移的縮約アルゴリズムを適用することである。属性間構造を抽出する際、一番問題になるのは互いに併存到達可能な属性同士を濃縮するプロセスである。我々は実用的計算量のアルゴリズムを作ることに関心があるので、互いに併存到達可能な単純属性間の濃縮は行うが、複合属性間の濃縮は行わない。このアルゴリズムを経て、射影的・擬似推移的閉包の非冗長最小被覆に相当する骨格構造を得る。この骨格構造をもとに、ある仮説クラスを特徴付ける属性の詳細構造を展開することができる。

(3) グリッド統合とAHP

複数のグリッドのマージを取る方法として、相乗平均を取る方法が提案されている。仮に個々人のグリッド行列の整合性が保たれているとしても、相乗平均を取った際、上述のファジィ行列と同様に、いかにしてグリッド行列全体の整合性を取るかといった問題が存在する。そこで、この問題を解決する技法として、不完全一対比較行列のAHP ウェイト計算法を利用する。この方法はHarker [Harker 87] によって与えられ、分かっているウェイト部分の整合性があるときに、分かっていないウェイト部分を整合性があるように埋めて行き、全体として整合性のある完全な一対比較行列を構成する技法である。従って、理想的な推移関係を想定し、欠落情報を補完しようというものであり、全体として整合性は改良されるが、生成された完全一対比較行列の信頼性は落ちている。

5. おわりに

本論文では、GRAPE の初期知識ベース獲得機能の概略について報告した。特に、グループの意見をマージしていくプロセスで採用するシステム分析技法について、その概要を述べた。GRAPE はその応用領域として計画問題（スケジューリング問題、プランニング問題等）にターゲットを絞った計画問題支援機能を付与する予定である。これについては、別

の機会に報告する。またGRAPE 完成時には、主観的評価と客観的評価の入り乱れる問題の解決に寄与する。具体例としては、旅行プランニング、新婚旅行の相談、スケジューリング用PERT作成、ソフトウェアの選定、就職先の決定、結婚相手の選定、家庭のメニュー選定、労使争議の解決、研究開発予算の分配、等に利用できる。

本研究の一部は、第5世代コンピュータ・プロジェクトの一環として行われた。本研究について貴重なヒントをいただいた当研究所戸田光彦部長、新谷虎松研究員、およびファジィ理論について一緒に学習した大森晃、渡部勇研究員に感謝する。

[参考文献]

- [Boose 87] Boose, J.H. and Bradshaw, M.: Expertise Transfer and Complex Problems: using AQUINAS as a knowledge-acquisition workbench for knowledge-based systems, Int. J. of Man-Machine Studies, Vol. 26, 1987, pp. 3-28.
- [Fransella 77] Fransella, F. and Bannister, D.: A Manual for Repertory Grid Technique, Academic Press, 1977.
- [Harker 87] Harker, P.T. and Vargas, L.V.: The Theory of Ratio Scale Estimation: Saaty's AHP, Management Science, Vol. 33, pp. 1383-1403, 1987.
- [Hart 86] Hart, A.: Knowledge Acquisition for Expert Systems, McGraw-Hill, 1986.
- [Ishii 89b] Ishii, H.: Cooperative Work Models and Groupware, FRIEND21 Int. Symp on Next Gen. Human Interface Technologies, Sept. 1989.
- [Kawakita 87] 川喜田二郎: KJ法, 中央公論社, 1987.
- [Kunifuji 79] 國藤 進、竹島 卓: 複合頂点間の到達可能性解析、電子通信学会、信学技報CAS79-110, 1979.
- [Kunifuji 88] 國藤 進: 知識獲得と学習研究の新しい流れ、人工知能学会誌, Vol. 3, No. 6, pp. 741-747, Nov. 1988.
- [Kunifuji 89] 國藤 進、上田晴康、須永知之、井深克憲、岩内雅直: グループ知識獲得支援システムGRAPE 構想、計測自動制御学会、第10回知識工学シンポジウム講演論文集、北海道大学, pp. 47-52, Oct. 19, 1989.
- [Lenat 86] Lenat, D., Prakash, M. and Shepherd, M.: CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks, AI Magazine, Vol. 6, No. 4, pp. 65-85, 1986.
- [Mizumoto 88] 水本雅晴: ファジィ理論とその応用、サイエンス社、1988.
- [Saaty 80] Saaty, T.L.: The Analytic Hierarchy Process, McGraw-Hill, 1980.
- [Sawai 89] 沢井 進、橋本恵司、泉理恵子、宮野敏浩、寺野隆雄: 知識獲得ツールYUAI、情報処理学会第38回(昭和64年前期)全国大会2C-9, pp. 127-128, 1989.
- [Stefik 87] Stefik, M., Foster, G., Bobrow, D.G., Kahn, K., Lanning, S., and Suchman, L.: Beyond the Chalkboard: Computer Support for Collaboration and Problem Solving in Meetings, CACM, Vol. 30 No. 1, Jan. 1987, pp. 32-47.
- [Takeda 89] 竹田英二: 不完全一対比較行列におけるAHP ウェイトの計算法、オペレーションズ・リサーチ、pp. 169-172, 1989年4月号.
- [Tone 86] 刀根 歳: ゲーム感覚意志決定法、日科技連, 1986.

[付録 A] 類似度行列に関する性質

ここでは推移性及び対称性を満たす類似度行列 $A = \{a_{ij}\}$ についての定理を述べる。

定義 1 (推移性) ある行列 $\{a_{ij}\}$ が推移性を満たすとは次式が成り立つ場合である。

$$\forall ij \ (a_{ij} \geq \max_k (\min(a_{ik}, a_{kj}))) \quad (1)$$

推移性は次のように言い換えることができる。

$$\forall ij \ (\forall k (a_{ij} \geq \min(a_{ik}, a_{kj}))) \quad (2)$$

定義 2 (対称性) ある行列 $\{a_{ij}\}$ が対称性を満たすとは次式が成り立つ場合である。

$$\forall ij \ (a_{ij} = a_{ji}) \quad (3)$$

定理 1 類似度行列は、その要素の値として高々 n 種類の値しか採らない。ただし、 n は行列の大きさである。

証明 類似度行列は、一意に分割木にすることができる。こうして作られた分割木は、深さが n よりも深くなることはない。すなわち、分割点 α_i の添え字 i は、 n より大きくなれない。この分割木から元の類似度行列を構成することを考えると、 α -レベル行列は分割木の深さである高々 n 個できる。この高々 n 個の α -レベル行列の和は、行列の要素として各 α_i のいずれかの値しか取らない。

よって、元の類似度行列は高々 n 種類の値しか取らない。

さらに強い定理として次のものがある。

定理 2 類似度行列 $\{a_{ij}\}$ では、次式が成り立つ。

$$\forall ijk \ ((a_{ij} = a_{jk} = a_{ki}) \vee (a_{ij} = a_{jk} < a_{ki}) \vee (a_{jk} = a_{ki} < a_{ij}) \vee (a_{ki} = a_{ij} < a_{jk})) \quad (4)$$

すなわち a_{ij}, a_{jk}, a_{ki} のうち少なくとも二つは値が等しく、他の一つはこれらと等しいか大きい。

証明 適当な添え字 ijk を選んだとき、 $\{a_{ij}\}$ が対称性を満たしているため、一般性を失うことなく式 (5) のように適当に並べ変えることができる。

$$a_1 \leq a_2 \leq a_3 \quad \text{但し } a_1, a_2, a_3 \text{ は } a_{ij}, a_{jk}, a_{ki} \text{ を並べ変えたもの} \quad (5)$$

ここで、 $a_1 \geq a_2$ を示せば良い。式 (2) より、

$$a_1 \geq \min(a_2, a_3) \quad (6)$$

$$a_2 \geq \min(a_1, a_3) \quad (7)$$

$$a_3 \geq \min(a_1, a_2) \quad (8)$$

が言える。

式(5)のように並べ直えたため、式(8)は常に成り立っている。式(7)についても、 $a_1 \leq a_3$ なので、 $a_2 \geq \min(a_1, a_3) = a_1$ となり、既に成り立っている。式(6)については、 $a_1 \geq a_2$ なので、

$$a_1 \geq \min(a_2, a_3) = a_2 \quad (9)$$

式(9)及び式(5)より、式(10)が成り立つ。

$$a_1 = a_2 \leq a_3 \quad (10)$$

[付録 B] 二つの類似度行列をマージするアルゴリズム

二つの類似度行列をマージして新しい類似度行列を作るときに、対称性は簡単に満たすことができるが、推移性を満たす方法は自明ではない。

以下のアルゴリズムを用いることにより、推移性を満たした類似度行列を作ることができる。

- Step 1 二つの類似度行列を $A = \{a_{ij}\}, B = \{b_{ij}\}$ とする。
ただし、 a_{ij}, b_{ij} の値が定まっていなかったことがあれば、既定値として 0 を用いる。
- Step 2 新しい行列 C を作る。 C の要素 c_{ij} は、 $c_{ij} = \frac{a_{ij} + b_{ij}}{2}$ によって定める。
- Step 3 C の全要素を調べて大きい順に並べてリストを作る。
- Step 4 新しい行列 D を作る。 D の各要素を定めるために、Step 3 で作られたリストの大きい方から順に取り出して x とし、Step 4.1 ~ Step 4.2 を繰り返す。
- Step 4.1 x と同じ値を持つ要素を行列 C から全て探しだし、 D の同じ添え字の要素に x を代入する。
- Step 4.2 Step 4.1 で D に代入された全要素に関して、推移性をチェックして値の定まる部分については値を代入する。 $(d_{ij}$ に値をいれたときには、全ての k に対して d_{ik} を調べ、 d_{ik} に値が定まっていたら d_{jk} の値を定める。)
- 値の定め方は次の通り。
- d_{jk} の値も定まっている場合:
推移性を満足している。
- $d_{ij} = d_{ik}$ の場合:
行列 D は、行列 C の値の大きい要素から順に転写して定めるので、 d_{ik} の値が定まっていれば d_{jk} の値が未定のときは、 $c_{jk} < x = d_{ik}$ 。
一方定理 2 より、推移性を満足するには $d_{jk} \geq d_{ik} = d_{ij}$ でなくてはならない。このため c_{jk} を d_{jk} の値とすることはできない。
そこで d_{jk} をその採りうる最小の値である $d_{ij} = d_{ik}$ にする。
- $d_{ij} \neq d_{ik}$ の場合:
定理 2 より $d_{jk} = \min(d_{ij}, d_{ik})$
- Step 5 終了