

弱い情報構造に関する評価実験

Evaluating Weak Information Structures

前田 晴美, 西田 豊明

Harumi Maeda and Toyoaki Nishida

奈良先端科学技術大学院大学 情報科学研究科

Graduate School of Information Science, Nara Institute of Science and Technology

Abstract: We evaluate the usefulness of *Associative Structures* for *human information sharing* which needs to *integrate heterogeneous information*. The following results are obtained in several experiments: (1) associative structures are easy to generate from raw data, (2) human tacit knowledge helps humans to understand the semantics of the associative structures, (3) associative structures are useful for exchanging ideas and thoughts between groups, (4) distance 1+2 neighbor search is appropriate to find information from the information base based on the associative structures, (5) associative structures are useful to extract and reorganize diverse information obtained from multiple information sources and (6) keyword search and the use of associative structures are useful for information retrieval in community.

1 はじめに

これまで我々は「大量の不均質情報を扱う人間の情報共有のためには弱い情報構造を用いたゆるやかな情報の関連付けが有効である」という作業仮説をたて、自然言語テキストやハイパーテキストやイメージなどの多様な情報メディアを意味を定義することなく関連づける連想構造(図1)とよぶ弱い情報構造を提案してきた([1],[2]). 連想構造では、「何が与えられると何が想起されるか」というように情報を関連付け、情報の構成要素であるユニット間の関係を点と矢印を用いて表す。与えられるユニットをキー、想起されるユニットをバリューと呼ぶ。ユニットには(a)概念ユニットと、(b)外部のテキスト・イメージファイルを表す外部参照データユニットがある(詳細は[2]参照).

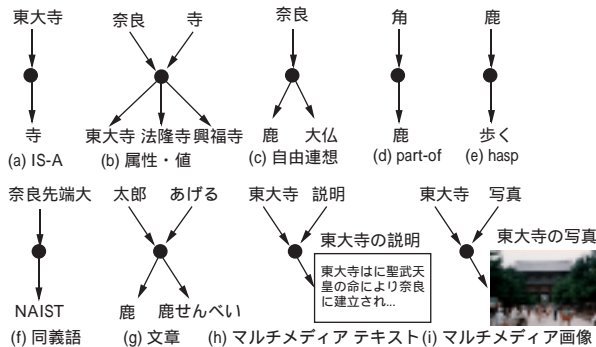


図 1: 連想構造の例

連想構造は意味ネットワークや述語論理のような従来の知識表現よりも、人間やコンピュータにとって生成が容易であるため、既存の不均質な情報源からの情報の獲得を促進できると考える。情報の解釈を受け手に委ね

表 1: 本研究の対象範囲

| 課題 人間集団 | 生成 | 理解 | 検索 | 抽出・整理 |
|------------|-----------------------------------------------------------------|---------------------------------------------|----------------------------------|--------------------------------------------|
| 個人 | 実験1 (自由連想の記述) 実験2 (自由連想, 新聞記事, 個人のメモの記述) 実験3 (ガイドブックから記述) | 実験1 (自由連想の理解) 実験2 (自由連想, 新聞記事, 個人のメモの理解) | 実験3 (ガイドブックから記述した情報の検索) | 実験4 (ガイドブック, 新聞記事, WWW上のHTML文書からの情報の抽出と整理) |
| グループ | | 実験1 (自由連想の理解) 実験2 (自由連想, 新聞記事, 個人のメモの理解) | 実験3 (ガイドブックから記述した情報の検索) | |
| コミュニティ | | | 実験5 (個人情報, 会議情報, 地域情報, 口コミ情報の検索) | |

るため情報の伝達が不完全な場合もあるが、人間が暗黙的に持つ背景知識を利用することにより、かなりの程度理解できると考える。

本研究では、(a) 人間の情報共有のために、(b) 大量の不均質情報を扱うために連想構造がどの程度有用か調べる。基本的な課題として、(1) 人間にとって記述、理解が容易か、(2) 大量の不均質な連想構造の中から必要な情報を見つけ出せるか、(3) 既存の雑多で不均質な情報源の情報をうまく扱えるか、(4) どの程度の間人集団(個人、グループ、コミュニティ)であれば有効か、などの疑問に答えることを試みる。本研究の対象範囲を表1に示す。

以下は日常記憶の共有を支援するシステム CoMeMo¹[1] において行なった評価実験について述べる。

¹前身は CM-2



図 2: 実験 1 で生成された CoMeMo 画面例

2 実験 1: 連想構造の生成実験

2.1 概要

目的 人間が自由連想においてどのように連想構造を生成するか調べる。

被験者 研究室の博士前期課程 2 年 (M2), 博士後期課程 (D) の学生 5 名。

方法 (1) 連想構造の定義と記述方法を説明する。

(2) 以下の指示を行なう。「CoMeMo ワークベンチを用いて“エージェント”というキーワードから連想されることを好きなように記述して下さい。ただし、記述する言葉は最低 10 個以上、20 個程度を目標とします。30 分を経過すると打ち切ります。」

(3) 他人が自由連想した画面を被験者に見せ、以下の質問を行なう。「何が書いてあるか理解できますか」「誰が作成したのかわかりますか」「わかるとすればそれはなぜですか」

(4) 研究室の教官に上記の画面を見せて感想を求める。

2.2 結果

生成された画面例を図 2 に示す。

連想構造を用いて自由連想の記述が容易に行なえる被験者全員が制限時間内に CoMeMo ワークベンチを使って自由連想を記述できた。作成したユニットの平均個数は 27 個、連想構造の平均個数は 16 個である。

相互知識のあるグループ間では連想構造を用いて思考内容を伝達できる。被験者全員が、他人の作成した画面の内容を正しく理解できた。また、ある被験者が作成した画面に関しては全員が作者をあてることができた。1 対 1 対応のリンクを使う方法と比較すると、連想構造では多対 1、多対多対応を用いてキーとなるユニットに連想の意味を付加することにより、より詳細に自由連想の意味を伝達できる。

表 2: 実験 2 テスト結果 (復元率)

| 被験者 | テスト 1 | テスト 2 | テスト 3 | テスト 4 | テスト 5 | 平均 |
|-----|-------|-------|-------|-------|-------|-----|
| D | 100% | 66% | 100% | 100% | 97% | 93% |
| M2 | 100% | 56% | 80% | 100% | 56% | 78% |
| M1 | 60% | 70% | 40% | 93% | 44% | 61% |
| 平均 | 87% | 64% | 73% | 98% | 66% | 78% |

復元率: 連想構造毎に情報の意味を正しく復元できた割合

グループ間で思考内容を伝達することはコミュニケーションに役立つ。他人の作成した画面を見て被験者全員から、笑いがあったり感心の声があがった。80%(4/5 人) が実験そのものが面白いと答えた。教官からは「研究テーマに関する知識量がわかる」、「文献調査を頼むならこの人」、「この人は面白そうなので少し話してみたい」などの感想があった。グループ内で思考内容を伝達することは、人間を知ることにつながり、コミュニケーションに役立つと考える。

3 実験 2: 連想構造の理解実験

3.1 概要

目的 自由連想、個人のメモ、新聞記事などから、連想構造を記述できるかどうか調べる。また、連想構造の定義を説明しなくても、連想構造を理解できるかどうか調べる。

被験者 研究室の (1) 博士後期課程学生 (D)(在籍年数 2-4 年) 3 名, (2) 博士前期課程 2 年 (M2) 学生 (在籍年数 1 年) 5 名, (3) 博士前期課程 1 年の学生 (M1)(配属直後の新入生) 4 名。

方法 (1) CoMeMo ワークベンチを用いて連想構造に基づく 5 つのテスト画面を作成する。

(2) 以下のテスト画面を 1, 2, 3, 4, 5 の順番に被験者に 1 つずつ見せ、以下の質問を行なう。「全体に何が書いてあるかわかりますか」「1 つずつの意味を言っして下さい」。

テスト (1) テスト 1: 大学近辺の店情報, (2) テスト 2: ペルー事件の新聞記事, (3) テスト 3: 研究室の CFP 案内, (4) テスト 4: エージェントの研究メモ, (5) テスト 5: 知識メディアの研究メモ (英語)

3.2 結果

テスト結果を表 2 に示す。

連想構造は直観的に理解できる。連想構造の定義を説明しなくても、被験者全員が、最初に見せた画面 (テスト 1) について、全体として何が書いてあるか答えることができた。連想構造の復元率が、平均で 78% であることから、連想構造は直観的に理解できると評価する。

背景知識が多いほど連想構造を理解できる。D(93%), M2(78%), M1(61%) の順番で復元率が良い。また、テスト 1(大学近辺の店情報) やテスト 4(エージェントの研究メモ) については、大学に 1 年以上在籍する被験

者全員の復元率が100%となっている。新入生のM1は、上級生と比べてテスト1(大学近辺の店情報)やテスト3(研究室のCFP案内)が低いが、これはまだ背景知識が十分でないためと考える。テスト5(知識メディアの研究メモ)についてはD以外の成績が悪いが、これは表示されたユニットが一部の学生にしかなじみのない研究関連用語(英単語)のためである。

文脈情報が連想構造を理解する手がかりとなる。被験者はユニットの単語の意味がわからない場合でも、近くにある理解可能な連想構造の意味を参考にしながら復元を行なった。またテスト3(研究室のCFP案内)では比較的均質な連想構造(国際会議名称とその属性と値)が多かったため、よどみなく復元ができた。

述部に修飾がある文章は記述、復元しにくい。テスト2(新聞記事)において、「セルパ容疑者は一日、英テレビWTNとの無線インタビューで、橋本首相とフジモリ大統領の会談について論評した」という文章に対してキー「セルパ容疑者、2月1日、英テレビWTN、論評、無線インタビュー」、パリュウ「会談、橋本首相、フジモリ大統領」という連想構造を記述したところ、完全に正確に復元した被験者は1名のみであった。

4 実験3: 連想構造の検索実験

4.1 概要

目的 不均質な連想構造の中から必要な情報を検索できるかどうか調べる。

方法 (1) 奈良の観光ガイドブックを参考に、手作業で奈良観光情報ベース(ユニット+連想構造2000個規模)を構築する。

(2) 情報ベース内に解が含まれるような日本語の質問文を50問作成する。

(3) CoMeMo ワークベンチにおいて、指定された複数のユニット間の連想構造の連鎖をたどることにより情報を検索するパス探索[3]の実験を行なう。

事前の予備実験(10問)から、(a) 最短距離で正答(抽出されるべき解がすべて得られ、かつ得られた解に誤りを含まない解答)が得られた場合、探索距離を増加しても正解がふえない、(b) 距離3近傍まで探索すると不要解が多量にまざる、の2点が判明したので、距離1近傍探索と距離1+2近傍探索についてだけ評価を行うことにする。ここで、距離1+2近傍探索とは、距離1近傍の探索で出力が得られない場合のみ距離2近傍の探索を行うという2段階探索を指す。

4.2 結果

解が情報ベースに含まれる場合、距離1+2近傍探索において必要な情報を見つけることができる。距離1+2近傍探索では、距離1近傍探索で見つけられなかったユニットをかなり高い精度で多く見つけることができた。そのため、再現率ばかりでなく適合率の向上もみられた(表3)。

5 実験4: 既存の情報源からの情報抽出・整理実験

実験の詳細は[2]を参照。

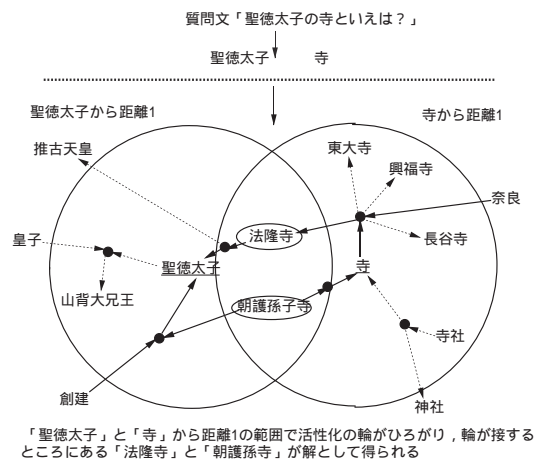


図3: パス探索

表3: 実験3テスト結果

| 情報ベース | 距離 | 適合率 | 再現率 |
|-------------------------------------------------------------------------------------|-------|-----|-----|
| 奈良観光情報ベース | 距離1 | 64% | 64% |
| | 距離1+2 | 76% | 92% |
| 適合率: $\frac{\text{正答(正解がえられ、誤りを含まない解答)数}}{\text{正解が情報ベースに含まれる質問数}} \times 100 (\%)$ | | | |
| 再現率: $\frac{\text{正解がえられた解答数}}{\text{正解が情報ベースに含まれる質問数}} \times 100 (\%)$ | | | |

5.1 概要

目的 既存の雑多で不均質な情報源から連想構造を用いて必要な情報を抽出・整理できるかどうか調べる。

方法 CoMeMo インコーポレータを用いて以下の情報源に対して実験を行なう。基本的なアルゴリズムは、(1) 形態素解析を用いた名詞句の抽出、(2) テキスト構造解析に基づく連想構造の生成、(3) ヒューリスティックを用いたIS-A情報の復元、(4) ヒューリスティックとパス探索を用いた情報の整理である。

情報源 (1) 実験3で手作業で作成した奈良観光情報ベースに対して、「寺」毎(テスト1)と「花」毎(テスト2)に情報を抽出・整理する。

(2) 日本経済新聞全文記事データベースから「インターネット」という単語を含む285件の記事を対象として、記事見出しから名詞を抽出して概念ユニット(キー)として、記事本文を参照する外部参照データユニットをパリュウとする連想構造を生成してから、「大学」毎(テスト1)と「官公庁」毎(テスト2)に情報を抽出・整理する。

(3) WWW上の人工知能研究者のホームページ100個のHTML文書(英語)から情報を抽出し、「people」「project」など7つの整理項目に関して整理を行なう。

5.2 結果

結果を表4に示す。

生データから連想構造の生成が容易である。ガイドブックのような書籍からの連想構造の記述が容易であることと、HTML文書のようなタグ付きテキストや新

表 4: 実験 3 テスト結果

| 情報源 | テスト | 適合率 | 再現率 |
|---------------------------------------------------------------------------------|------|-----|-----|
| (1) 奈良観光ガイドブック | テスト1 | 92% | — |
| | テスト2 | 86% | — |
| (2) 日本経済新聞全文記事データベース | テスト1 | 63% | 91% |
| | テスト2 | 83% | 71% |
| (3) WWW 上の HTML 文書 | テスト1 | 90% | 83% |
| | テスト2 | 68% | 73% |
| 適合率: $\frac{\text{正しく生成された連想構造のユニット数}}{\text{生成された連想構造のユニット数}} \times 100 (\%)$ | | | |
| 再現率: $\frac{\text{正しく生成された連想構造のユニット数}}{\text{生成されるべきユニット数}} \times 100 (\%)$ | | | |

聞記事のような見出し付きテキストから連想構造を機械的に生成することが容易であることがわかった。

連想構造を用いることにより、タグ付きテキストからの情報の抽出と整理を低コストで行なえる。連想構造を用いることにより、自然言語処理を使用しない比較的簡単なアルゴリズムでもある程度の確度で情報の抽出と整理が行なえることがわかった。

6 実験 5: コミュニティの情報共有実験

実験の詳細は [5] を参照。

6.1 概要

目的 連想構造がコミュニティの情報共有に有用かどうか調べる。

方法 (1) CoMeMo モバイル端末のプロトタイプであるコミュニティの情報共有を支援するシステム InfoCommon を試作し、ICMAS'96 モバイルアシスタントプロジェクト [4] で国際会議支援に適用し、会議参加者 100 名に自由に使用してもらう。

(2) 事後アンケート及びログの解析を行なう。

システム概要 InfoCommon は、会議の参加者が発信するメッセージなどの動的な情報や会議の案内などの静的な情報を関連付けて可視的情報として会議参加者にフィードバックする。図 4 に画面例を示す。InfoCommon 情報ベースは、キーワードや情報カードを緩やかに関連付ける連想構造の知識ベースと、コンテンツを蓄えた情報カードベースから構成される (図 5)。システムの主要な機能は (1) 情報検索: キーワードに基づく関連情報の可視化表示, (2) 情報発信: 既存の情報と関連付けたメッセージの発信である。

6.2 結果

連想構造に基づくキーワード検索がコミュニティの情報検索に有用である。アンケート調査の結果、(a)81% が検索結果をよいく感じ、(b)51% がシステムに満足し、(c)55% がシステムが必要な情報を得るために有用であったと回答したことより、InfoCommon がコミュニティの情報検索ツールとして有用であると評価する。

コミュニティではメンバーの持つ背景知識の利用が期待できる。上記の結果は、コミュニティにおいてメンバーが暗黙的に持つ背景知識の利用が期待できることを示していると考えられる。

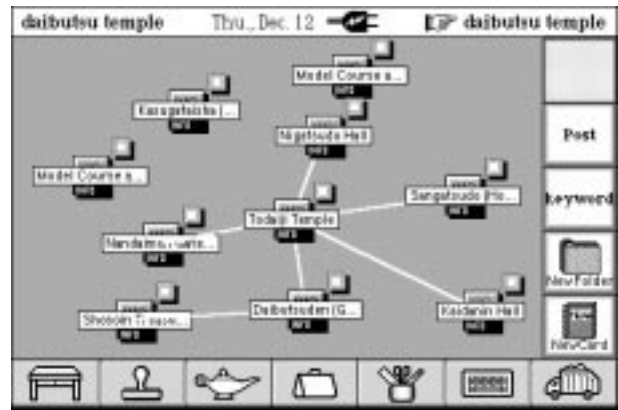


図 4: InfoCommon 画面例

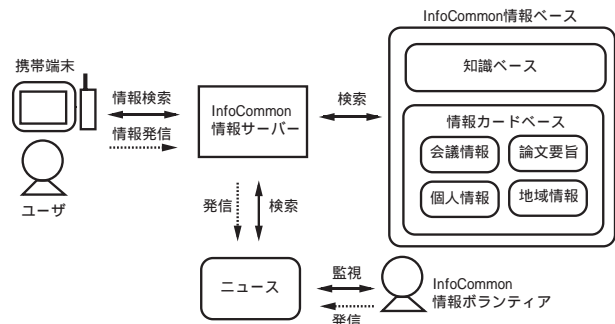


図 5: InfoCommon 概要

7 おわりに

大量の不均質情報を扱う人間の情報共有のために弱い情報構造が有用かどうか調べた。設定した研究対象範囲において以下の結果を得た。(1) 連想構造は生データから生成が容易である。(2) 背景知識が多いほど連想構造がよく理解できる。(3) 連想構造はグループの思考伝達に役にたつ。(4) 不均質な連想構造の中で、距離 1 + 2 近傍探索において必要な情報が検索できる。(5) 連想構造を用いることにより、既存の情報源からの情報の抽出と整理が容易に行なえる。(6) 連想構造に基づくキーワード検索はコミュニティの情報検索において有用である。

参考文献

- [1] 平田高志, 前田晴美, 西田豊明: 弱い情報構造を用いた情報整理支援, 1997 年度人工知能学会全国大会 (第 11 回) 論文集, 16-05, 1997.
- [2] 前田晴美, 梶谷和人, 西田豊明: 連想構造を用いた情報整理システム, 情報処理学会論文誌, vol.38, no.3, pp.616-625, 1997.
- [3] 前田晴美, 梶谷和人, 西田豊明: 情報ベースのユーザフレンドリなインタフェースのための連想構造の提案, Progress in Human Interface, vol.5, no.1, pp.49-56, 1996.
- [4] 西部喜康, 武石英二, 森原一郎, 服部文夫, 石田亨, 西田豊明, 携帯端末による国際会議支援 - ICMAS96 Mobile Assistant Project -, 情報処理学会第 54 回全国大会講演論文集 (4), pp.385-386, 1997.
- [5] 前田晴美, 梶原史雄, 足立秀和, 沢田篤史, 武田英明, 西田豊明: 弱い情報構造を用いたコミュニティの情報共有支援, 情報処理学会第 1 回モバイルコンピューティング研究会, 1997.