計測と制御, 42(6):508–511, 200

解説
特集「人工知能の現在と将来」

ウエッブインテジェンスの可能性

ー知識情報基盤としての WWWー

Prospective Directions for Web Intelligence

-- WWW as information/knowledge infrastructure --

武田英明（非会員）

Hideaki Takeda

国立情報学研究所
千代田区一ツ橋 2-1-2
National Institute of Informatics
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430
101-8430 千代田区一ツ橋 2-1-2
電話：03-4212-2543
ファックス：03-3556-1916
メール：takeda@nii.ac.jpT

Keyword: WWW(WWW), エージェント(agent)、セマンティック Web（Semantic Web）

１．はじめに

伝統的に人工知能はコンピュータによる記号処理を中心に発達してきた。その是非については本特集の他の解説に譲るが、記号で表されたものをどう処理するかについての技術（知識処理技術）は蓄積されてきたことは間違いない。ただし、これまでコンピュータによる記号処理に利用できる資源は極めて限られていた。辞書や文書は徐々に電子化されていたが、集める方法はなかった。特定の目的の情報がほしければ知識エンジニアが手作業で入力していくしかなかった。これは大きな障害であったが、またよい言い訳になった。World Wide Web(以下 Web)はこのような状況を一変させてしまった。これまで各所で行われた電子化されて蓄積されたデータは相互につながり、また Web のためにあらたに多様な情報が電子化されて流通するようになった。これは知識処理技術にとって大きなチャンスであると同時に試練でもある。

2. Webインテリジェンスの可能性[1]

Web はもはや我々の生活に欠く事のできない情報インフラストラクチャとなっている．Web はもはや日常世界のありとあらゆる情報を蓄えるようになり，Web は我々の日常世界の投影のようにそれ自身ひとつの世界を構成するようになっている．ただし，Web は我々の世界の投影に留まらない．Web により，限られた人だけであった情報流通活動が多くの人によって可能なものになり，多様なコミュニケーションが可能になった．我々の日常世界も新しいコミュニケーション手段としての Web によって変化を余儀なくされている．

Web にあるコンテンツ（近年は Web を介した活動も含む）は我々の高度な知的活動のスナップショットとであり，このような情報がデジタル情報として手に入るというのは情報処理，ことに知的な記号処理を特徴すする人工知能研究にとってはまたとない資源である．

ただし，Web の特徴は我々の日常世界と表裏一体に発展してきたので，その構造や利用において単に情報システムだけを切り出すと本質的な点を見逃す可能性がある．この点でこれまでの情報システムと大きく異なる．すなわち，Web の問題は Web システムという計算機システムやその上での情報だけを注目するのでは不十分である．日常世界と Web の世界の関係を常に考える必要がある．

以上の関係を模式的にしたものを図１に示す．ここでは 2 重化した 2 層構造として表している．単純化すれば 3 層構造である[2]. 計算機システムとしてのWebとはハードウエアとソフトウエアから構成される計算機ネットワークのことである．Webコンテンツとはその上で蓄積，交換される情報のことである．

まず，計算機システムとしての Web をみた場合，Web コンテンツを要求仕様とするシステムである．すなわち，どのような情報をどのような形で流通されたいのか，といった計算機システムへの要求はコンテンツからくるものであり，その要求に計算機システムは対応する必要がある．もちろん逆もあり，計算機システムにおける新しい方式の開発が Web コンテンツのあり方を決定することもある．

その一方 Web コンテンツは日常世界が要求仕様である必要がある．現在の Web コンテンツは日常世界の情報のうち，いわば“たまたま”Web コンテンツとして顕在化されたものにすぎない．日常世界において情報として顕在化したいといった要求が Web コンテンツに変化させることになる．同じく逆もあり，Web コンテンツによって日常世界が変えられることもありうる．

ではこのようなモデル化の中で人工知能の研究はどこに位置付けられ，どのような貢献が期待できるであろうか．ここでは日常世界での情報活動になぞらえて「通信活動（コミュニケーションする）」,「調査活動（調べる）」,「構造化，組織化活動（整理する）」という３つの活動に分けて考えることにする（表１参照）．

計算機システムとしての Web（I）では基本的にこれまでの人工知能研究の応用であり，分散知識システムをいかに効果的に運用できるかということが中心的課題である．たとえば，ネットワークの自律的構成方法（構造化，組織化活動）といったものや分散知識ベース上のエージェント（通信活動）といったものがある．

Webコンテンツ(II)ではより広範に人工知能の技術が適用されている． Web検索における自然言語処理は人工知能の技術がWebによって大いに適用範囲を広げた典型例であろう．また，構造化，組織化活動では知識ベースシステムの技術がWebコンテンツの構造化で適用されている．Semantic WebにおいてはWebコンテンツはオントロジーによって関連付けられることで人間だけでなく，計算機によって解釈が可能になり，Webの自動化や統合が容易になるという仕組みを提案実装されている[3]．また，またWebマイニングという分野もできている．Webマイニングは Web上のテキストからのマイング（テキスト・マイング），Webのアクセスログの解析（Webログ・マイニング），リンクで構成されたネットワークの解析（Web構造マイニング）に分けることができるが，前者２つではデータマイニングの技術が適用されている．後者ではPageRank3)に代表されるリンク解析という新しい分野が発展しつつある．

---

[1] 本章は 1)での議論を発展，改定したものである．より詳細な議論と参考文献は 1)を参照されたい．

[2] レッシグ 2)はインターネットを物理層，コード層，コンテンツ層の 3 層に分けている．本稿の分類でいれば，計算機システムとしてのWeb層が物理層とコード層に分化しており，逆に日常世界層が追加されている..

[3] Semantic Webに関しては情報処理特集5)および人工知能学会誌特集6)に詳しい．あるいはHhttp://www.w3c.org/2001/sw/H を参照されたい．またSemantic Webの概要については 7)も参照されたい．

日常世界（III）そのものをテーマとするのは Web の研究とは異なってしまうので，本稿のテーマから外れるが，日常世界と Web コンテンツの接点(II-III 層)は重要な研究テーマである．先に述べたように Web の他の情報システムと異なる点は日常世界の問題と深く関係している点にある．例えば，人間の社会がつくる複雑さやダイナミクスが Web にどのように影響しているかということは興味深いテーマであるし，逆にそのような社会での活動を Web から支援できるかといったアプローチもある．
人工知能的視点からみれば，個々の人間の知能を切り離して考えるのではなく，人間の集団，すなわち社会全体としての知能（「社会知能」）を考えることを意味している 4)．これは人工知能にとって新しい挑戦である．

3. 社会と Web の接点を探るケーススタディ

前章では Web と人工知能の接点について３つの層に分けて議論を行った．特に Web コンテンツの層と Web と日常世界の接点の層において人工知能技術の適用と新しい発展のあることを示した．そこで以下ではその研究の例として著者が関係する研究を３つ取り上げ，具体的に研究の可能性をみていくことにする．

3.1 インターネットディレクトリの統合 8)9)

この研究は先の分類でいえば，Web コンテンツ層のおける構造化，組織化活動にあたる．
Webコンテンツの特徴の一つは多数の人が参加して分散的に作られていく点にある．より詳しくいえば，分散的である一方暗黙に協調的に構築されている．Webコンテンツの作成者はお互いに連絡をとりながらつくるわけではないが，他のWebページをリンクとして参照することで，結果的に協調的な振る舞いになっている．当然，Webを知識源とみたてたときも同様に分散(弱)協調的なものである．このような性質をもつ知識をどう使っていくかは知識処理として新しい課題である[4]．この研究では異なる知識間の関係を発見することを目的とする．具体的にはYahoo!に代表されるインターネットディレクトリのディレクトリ構成を知識としてみなして[5]，異なるインターネットディレクトリ間の関係を発見する．

(1) 基本方針と手法

インターネットディレクトリのカテゴリーを概念，カテゴリーに分類されているページ（URL）をインスタンスとみなして，複数の概念体系（インターネットディレクトリ）における概念間の関係を発見する．ここで注目するのはインスタンスの共有である．複数の似て非なる概念体系（例えばYahoo!と Lycos, Open directory）においては共通するインスタンス（ページ）が含まれていることがある．この共通性を利用して概念の関係を求めることにする．このとき階層の上位の概念は下位の概念のもつインスタンスも自身のインスタンスと考えることで間接的に階層性を利用する（図２参照）．
ここではインスタンス共有による類似概念に判定にκ統計量 10)を用いている．また比較する概念組は概念階層の上位からはじめ，類似性が判定された概念の下のみ，さらに比較するという方法をとることで，無駄な探索を抑えている．

(2) 結果

実験例の一つとしてYahoo!とLycosのディレクトリの対応する部分で行った結果を図３に示す．図３上部には利用したデータ数を示している．この実験ではデータを 10 分割し，その９個分を訓練データとして，残りの一つをテストデータとして検証した. その結果が図３下部に示している．だいたい, 70%から 90%の割合で正答になっていることがわかる．この他，Yahoo!と Open directory の比較でも同様な結果が得られた．

(3) 考察

この手法で興味深いのはページの内容を一切使わずに，構造的情報（概念階層とインスタンス）だけで十分な成果が得られている点である．類似研究（例えば 11)）ではページの内容を使った文書解析によって類似関係を判定するのが一般的である．この概念体系も大規模なネットワークであり，この意味でリンク解析による手法と同様のネットワーク構造を利用した手法と位置付けることができる．

3.2 Bookmark の共通性を利用した共通話題ネットワークの構築 12)13)

この研究は社会と Web の接点の層(II-III 層)の通信活動と構造化，組織化活動の両方に位置付けられる研究である。Web 上にある情報だけに注目するのではなく，その背後にある情報を提供する人やコミュニティまで含めて考察することで，Web の持つ複雑さやダイナミクスが明瞭になることが期待される．
3.1 節の研究ではインターネットディレクトリを知識とみなしたが，ここでは個人のブックマークを個人の持つ知識とみなして，複数の個人知識間の関係の発見方法を開発している．
前節の研究と同様，概念(この場合はブックマークのフォルダー)の類似関係を用いる．ただし，こちらでは共通ページの可能性が少ないのでページのテキスト解析を使ってページ間の関連度を定義して，その関連度から概念間の関連度

---

[4] Semantic Webにおけるオントロジーも本質的に分散的構築されるので，オントロジー統合は不可避である．

[5] 各カテゴリを概念とみなせば，インターネットディレクトリはその対象領域のインスタンス（ページ）を階層的に分類するオントロジーであるということができる．インターネットディレクトリのような概念間の階層的な関係があるが，概念や関係の定義がないをlight-weight ontology，そういった定義をもつものをheavy-weight ontologyと分けて言うことがある．

を計算している．複数のブックマークの概念間で類似関係を計算することで複数知識源の概念のネットワークを計算することができる（図４参照）．具体的にはあるページ間類似度のある閾値を越えたら推薦ページとし，フォルダー間で推薦ページがある閾値を越えたら推薦フォルダー，すなわち類似概念として同定する．図４でA,B,Cは個人を示し，A,B,Cに直接つながる要素はそれぞれがもつ概念（フォルダー[6]）を示しており，この概念間を結んでいるリンクが発見された関係である．この例では３人が「研究」関連概念で結ばれていることや，AとCは「Unix」関連の概念で比較的強く結ばれていることがわかる．このようにユーザ間の関係を単に近い，遠いではなく，話題の共通性で結びつけることを可能にした．この概念間関係（フォルダー間関係）はページ毎の関連度よりも受け入れやすいことが被験者の主観評価によって確かめられた．

このような個別の概念間の関係を超える個人間の関係は発見できないであろうか．このためにこの研究ではカテゴライズド近似度という指標を提案している．ユーザａｂ間のカテゴライズド近似度 CB(a,b)は

$CB(a,b)=N_f(a,b)R_f(a,b) / N_p(a,b)$

$N_f(a,b)$：ａｂ間での推薦フォルダー数

$R_f(a,b)$：ａｂ間での推薦フォルダーの平均関連度

$N_p(a,b)$：ａｂ間での推薦ページ数

これは類似しているページ間でその所属フォルダー間が類似しているときに高い値を示す指標である．例えば図５で網のかからない線（フォルダーが類似していないがページが類似している）が少ないほどこの値はよい．

この指標を被験者実験でえら得る数値で検証してみると，被験者が主観評価値とよい相関がえられることがわかった．すなわち，人は自分と同じような情報分類をする人からの情報を信頼するということがわかった．これは情報流通の際の選択基準の一つになりうるだろう．

４．おわりに

本稿では知識情報基盤としてのWWWについて，全体のパースペクティブを示すと共に研究例を紹介した．WWWは計算機が利用可能な知的情報（知識や知的活動）の宝庫であるので，WWWは人工知能のまたとないフィールドである．また人工知能はまたWWWと出会うことにより分散知識ベースやネットワーク構造分析など新しい展開を遂げつつある．今後もWWWの発展に伴い，この分野も発展していくであろう．

1) 武田英明. 知性のネットワークとしてのWWW —Web Intelligence に関する一考察—. *人工知能学会誌*, 17(3):346–351, 2002.

2) Lawrence Lessig, The Future of Ideas: The Fate of the Commons in a Connected World, Random House, 2002.（邦訳：山形 浩生訳，コモンズ，翔泳社，2002）

3) Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, 1998.

4) Toyoaki Nishida. Social intelligence design. In T. Terano, T. Nishida, A. Namatame, S. Tsumoto, Y. Ohsawa, and T. Washio, editors, New Frontiers in Artificial Intelligence, Joint JSAI 2001 Workshop Post-Proceedings, LNAI 2253, pp. 3-10. Springer-Verlag, 2001

5) 特集：セマンティック Web, 情報処理, Vol.43, No.07, 2002

6) 特集：「Semantic Web とその周辺」，人工知能学会誌, Vol.17, No.4,2002.

7) Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001

8) Ryutaro Ichise, Hideaki Takeda, and Shinichi Honiden. Integrating Multiple Internet Directories by Instance-based Learning. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, (IJCAI-03)*, 2003. (to appear)

9) 市瀬龍太郎, 武田英明, 本位田真一. 階層的知識間の調整規則の学習. 人工知能学会論文誌, 17(3):230–238, 2002.

10) J.L. Fleiss, Statistical Methods for Rates and Proportions, Wiley and Sons, NY, 1981 (邦訳：係数データの統計学，東京大学出版会，1975)

11) R. Agrawal and R. Srikant, "On Integrating Catalogs", Proc. of the Tenth Int'l World Wide Web Conference, Hong Kong, May 2001

12) 濱崎雅弘, 武田英明, 松塚健, 谷口雄一郎, 河野恭之, 木戸出正継. Bookmark からの共通話題ネットワークの発見手法の提案とその評価. 人工知能学会論文誌, 17(3):276–284, 2002.

13) Hideaki Takeda, Takeshi Matsuzuka, and Yuichiro Taniguchi. Discovery of shared topics networks among people --- a simple approach to find community knowledge from www bookmarks ---. In Proceedings of the Pacific Rim International Conference of Artificial Intelligence (PRICAI 00), Lecture Notes in Artificial Intelligence, No. 1886, pages 668–678, 2000.

[6] 簡便のため，ここでは概念階層は１階層に縮約している．

表 1　Web インテジェンス研究の分類と研究トピックスの例

活動 / 研究層	通信(コミュニケーションする)	調査（調べる）	組織化，構造化（まとめる）
Web と日常世界の接点(II-III)	情報伝播モデル／解析	Human-Web インタラクション	コミュニティの発見／利用
Web コンテンツ（II）	Web サービス，Semantic Web 言語	Web 検索，Web マイニング	Web 構造解析，オントロジー構築／統合／利用，知識管理
計算機システムとしての Web(I)	エージェント，P2P	P2P クエリ，XML クエリ	

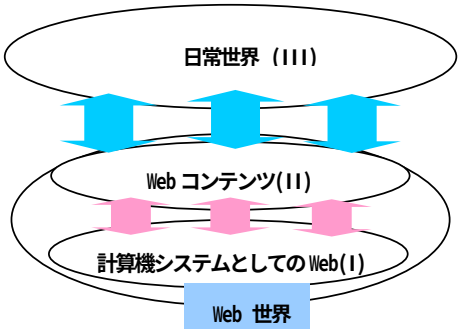

図１　日常世界と Web 世界

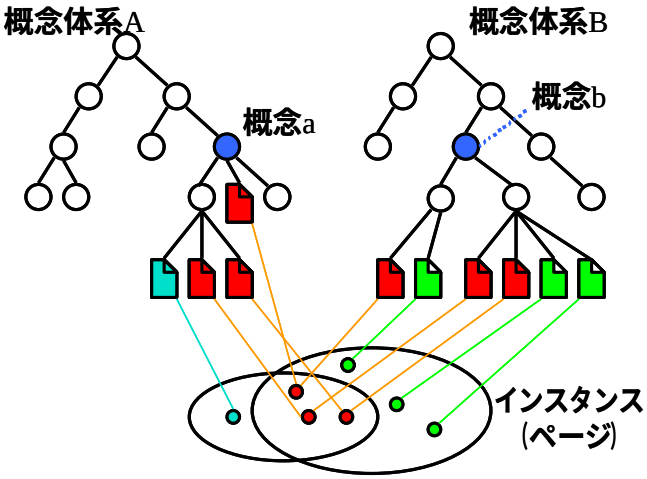

図２　インスタンスに基づく類似概念の同定

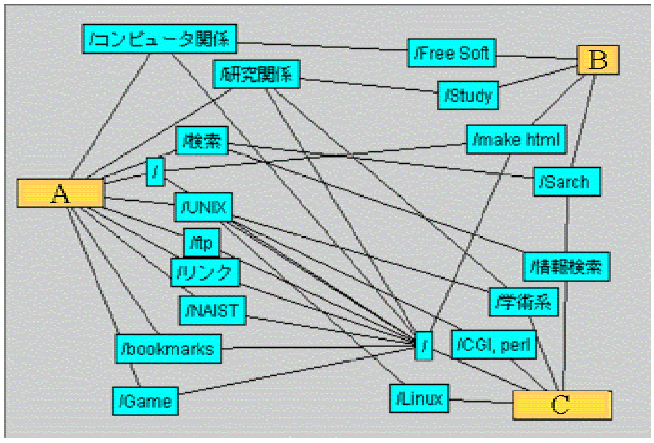

図４：共通話題ネットワークの例

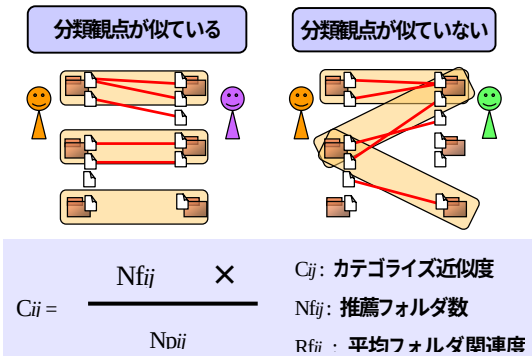

図５：カテゴライズド近似度

	Yahoo!		LYCOS		共通ページ
	カテゴリ	ページ	カテゴリ	ページ	
文学	493	3192	186	1119	468
企業	7554	58609	413	5904	3992
リクリエーション	3164	19609	709	4941	1939

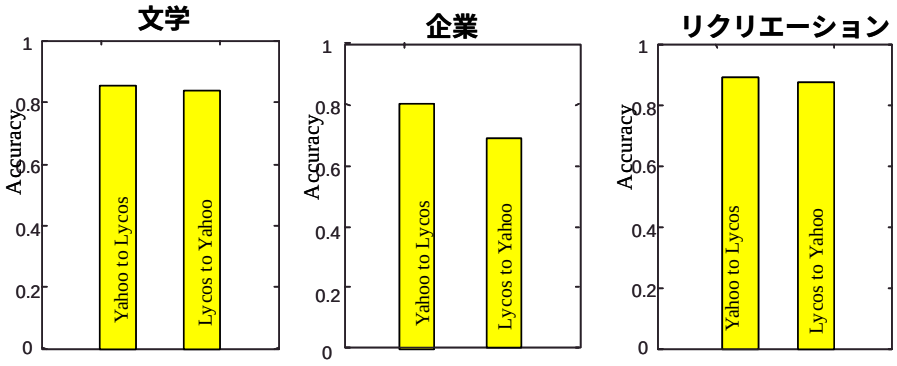

図３　インターネットディレクトリの統合の実験例