

— 第6回：インターネットと知識推論 —

0. はじめに

今回は、インターネット上の知識を使った推論システムという視点から、知識推論にまつわるサイトを紹介します。1章では、ネット上の知識をいかにして利用するかについて、2章では、推論にあたって常識を利用するという試みについて、3章では、ネット上の不確実な知識を扱うための推論法について、それぞれサイトを紹介します。

1. インターネット上の知識の利用

インターネットから事実知識（ファクト）を獲得し、インターネット全体を一つの巨大な知識ベースとして扱おうという試みがあります。WebBroker [1.1], WebKB [1.2], IBROW3 [1.3] 等はその代表的な例です。一方、一つの知識ベースにまとめるのではなく、ユーザの要求に応じてネット上の適切な資源を選択し、その中から情報を抽出するシステム（Infomaster [1.4] 等）もあります。このようなシステムは Facilitator と呼ばれます。これらのシステムを構築する際、ネット上のテキストに含まれる情報を如何にして取り出すか、そして如何に統合するかが重要となります。元々半構造化されている情報に対し、テキスト毎に異なる構造の違いを吸収して統合するシステムは Mediator と呼ばれ、Ariadne [1.5], HERMES [1.6] 等があります。一方、構造化されていない情報に対しては、HTML 構造あるいは文法的な情報を基にした抽出ルールを作成して獲得する事が出来ます。このようなツールは一般に Wrapper と呼ばれ、その開発では人手で事実知識を示すタグを付けた正解例を準備しておいて学習するものが主流です（WebKB [1.2], OntoBroker [1.7]）。また、GUI ツールを提供する事で多様なジャンルに対応した Wrapper を共同開発しようとする動きもあります（W4F [1.8]）。

- [1.1] WebBroker : <http://www.w3.org/TR/1998/NOTE-webbroker/>
- [1.2] WebKB : <http://www.cs.cmu.edu/~WebKB/>
- [1.3] IBROW3 : <http://www.swi.psy.uva.nl/projects/IBROW3/home-ibrow.html>
- [1.4] Infomaster : <http://infomaster.stanford.edu/infomaster-info.html>
- [1.5] Ariadne : <http://www.isi.edu/ariadne/>
- [1.6] HERMES : <http://www.cs.umd.edu/projects/hermes/>
- [1.7] On2broker : <http://www.aifb.uni-karlsruhe.de/WBS/broker/project.html>
- [1.8] W4F(World Wide Web Wrapper Factory) : <http://db.cis.upenn.edu/W4F/>

ネット上の多様なテキストからの推論知識（ルール）獲得の手法として、テキストマイニングが注目されます。テキストマイニングは比較的新しい研究分野であり、この分野に特化した国際会議は筆者の知る限り存在しません。しかし、知識発見全般に関する研究発表の場 Int'l Conf. on Knowledge Discovery and Data Mining (KDD) で、テキストマイニングに関する発表が数多くなされています。今年 (KDD99) はテキストマイニング研究で著名な Ronen Feldman [1.9] によるチュートリアルも行われました。KDD の主催団体である ACM SIGKDD のサイト [1.10] には、KDD 関連の会議の情報も掲載されています。

研究の全体像をつかむには、知識発見およびデータマイニングのポータルサイト KDNuggets [1.11] が、様々なリソースへリンクが張られていて便利です。また、テキストマイニングに特化したリンク集としては Links zum Vortag Text-Mining [1.12] があります。

- [1.9] Ronen Feldman : <http://www.cs.biu.ac.il:8080/~feldman/>
- [1.10] ACM SIGKDD : <http://www.acm.org/sigkdd/>
- [1.11] KDNuggets : <http://www.kdnuggets.com/>
- [1.12] Links zum Vortag Text-Mining : <http://www.cs.uni-bonn.de/III/lehre/seminare/DataMining/SS98/bookmark-sammlungen/Martini.html>

* HTML 版は、<http://wwwsoc.nacsis.ac.jp/jsai/journal/mybookmark/>で公開されています。

2. 常識の利用

現在構築されている常識ベースは、語彙知識を中心としたものが主です。例えば、プリンストン大の WordNet [2.1] や、日本電子化辞書研究所 (EDR) の電子化辞書 [2.2] などがよく知られています。語彙知識を用いて単語や文書の類似性や関連性を判断するシステムとしては、カーネギーメロン大の Lexical FreeNet [2.3]、ジャストシステムの ConceptBase [2.4] 等があります。また、NTT の概念ベースによることば遊びの部屋 [2.5] なども面白い試みです。因果知識等も含めた大規模なものとしては、Cycorp [2.6] による Cyc が有名ですが、残念ながら一部しか公開されていません。

常識ベースとは若干違いますが、Internet Dictionary Project [2.7] のように世界のインターネットユーザで協力して翻訳辞書をつくる、といったものも見受けられます。今後、語彙知識だけにとどまらない常識ベースを共同でつくりたいという動きも出てくれば面白いと思います。

- [2.1] WordNet : <http://www.cogsci.princeton.edu/~wn/>
- [2.2] EDR 電子化辞書 : <http://www.iijnet.or.jp/edr/>
- [2.3] Lexical FreeNet : <http://www.link.cs.cmu.edu/lexfn/>
- [2.4] ConceptBase : <http://www.justsystem.co.jp/cb/>
- [2.5] ことば遊びの部屋 : <http://www.kecl.ntt.co.jp/banana/>
- [2.6] Cycorp : <http://www.cyc.com/>
- [2.7] Internet Dictionary Project : <http://www.june29.com/IDP/>

3. 不確実な知識を扱う推論法

テキストから自動獲得された推論知識は確実なものとはいえず、多くの場合不足や矛盾があります。そこで、不確実で不完全な推論知識から得られた結論について、その妥当性を確率的に評価できる確率推論の枠組が有力となると考えられます。

確率推論についての Society としては、Assoc for Uncertainty in AI [3.1] が有名です。ここでは、年一回の国際会議がもたれています。また、メーリングリストもあり、そのアーカイヴが UAI maillist archive [3.2] で得られます。確率推論についての解説は、AFIT Artificial Intelligence Laboratory [3.3] のページが詳しいです。さらに、様々なフリーウェアも提供されるようになっており、Software for Belief Networks [3.4] などにまとめられています。Decision-theoretic なシステムの開発環境としては、GeNIe [3.5]、analytica [3.6] などがあります。また、ネット上の資源を使った推論という意味では、資源が限られたもとの推論 Resource-Bounded Reasoning [3.7] の考えが有力でしょう。

- [3.1] Assoc for Uncertainty in AI : <http://www.auai.org/>
- [3.2] UAI maillist archive : <http://www.cs.orst.edu/dambrosi/uai-archive/>
- [3.3] AFIT Artificial Intelligence Laboratory : <http://www.afit.af.mil/Schools/EN/ENG/LABS/AI/>
- [3.4] Software for Belief Networks : <http://bayes.stat.washington.edu/almond/belief.html>
- [3.5] GeNIe Homepage : <http://www2.sis.pitt.edu/~genie/>
- [3.6] Lumina Decision Systems : <http://www.lumina.com/>
- [3.7] Resource-Bounded Reasoning : <http://anytime.cs.umass.edu/>

4. おわりに代えて

推論システムに求められる機能のレベルは対象とするドメインによってまちまちですが、ごく単純な機能を装備したお遊び的なものについては、インターネット上でも多くみられます。こういったニーズの向こう側に、一つの新しい人工知能の姿が見えてくるかもしれません。

- [4.1] 猫になれ : <http://www2d.biglobe.ne.jp/~inva/neko/>
- [4.2] 伝説 : <http://cgi.tky.3web.ne.jp/~hiya/legend/index.shtml>
- [4.3] 今日のあなたのキャッチフレーズ : <http://www.rikkyo.com/spc/catch/index.html>
- [4.4] 人工無能人生相談所 : <http://www1.mirai.ne.jp/~panther/muno/chinpun.html>

〔佐藤浩史・廣田啓一・賀沢秀人・藤本和則 (NTT コミュニケーション科学基礎研究所)
<http://www.kecl.ntt.co.jp/DSIU/>〕