

国際会議ネットワークを活用した研究動向分析の試み

An Analysis of Global Research Activity based on Conference Map

篠田 孝祐*1

Kosuke Shinoda

*1 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology

国際会議ネットワークとは、公開されている大規模な論文データベースから、著者と国際会議との関係に着目した、国際会議間の関係性を示したネットワークである。本稿では、抽出したネットワークをもとに、計算機科学の研究における、分野の変遷や、各分野において核となる会議などを概観し、そして、抽出したネットワークの実社会への応用を検討する。

1. はじめに

研究者にとって、研究動向を把握することは自身の研究の方向性を決める上で重要な活動である。人工知能研究、ひいては情報科学の分野では、それは、国際会議を知り、どこに投稿するかを決めることと言い換えることもいえるのではないだろうか。

では、我々は、それら国際会議をどのような経緯で知るのであるのか。ここでいう”知る”とは、名前だけでなく、質や評価も含める。まず、考えられるは、自身の経験ではないだろうか。日々の活動のなかで、関連する国際会議の存在を知り、その会議のメイントピックや会議の重要性を認識する。その認識は、自身が経験を重ねることや、自身の知人(社会ネットワーク)を介することで、おそらく確かなものとなる。

では逆に、経験や社会ネットワークに乏しい学生や若手の研究者は、自分に関連する会議の存在を知るための、国際会議に関する良質な情報を得るにはどうしたらよいのだろうか。また、経験が豊富な研究者であっても、既存の”社会ネットワーク”があるだけに、得られにくい自身のネットワークではカバーできない国際会議の情報を得るにはどうすべきだろうか。

我々は、研究動向を知るために、ひいては戦略的に国際会議に投稿していくためには、それぞれの国際会議の評価状況を把握し、そして、関連する国際会議を俯瞰できることが重要と考える。

本論文では、国際会議の関係性を総体的に把握するツールとして国際会議マップの生成を試みる。

2. 国際会議マップ

国際会議マップは、[篠田 08] にて掲載しているマップを利用する(1, 2 参照)。本マップは、DBLP *1で公開されている論文メタデータ(2008年3月時点)を用いており、2²⁰のレコード(2009年4月現在)が登録されている。マップの作成は、国際会議名と研究者の氏名、2種類をノードとした2部グラフを作成し、同一の国際会議で発表した研究者の共起関係をもとに、国際会議のネットワークを生成したものである。つまり、同じ発表者が多い会議は関係の強い(ベクトルコサイン類似度の高い)国際会議としている。なお、本稿に記載しているネット

ワーク図は、graphviz*2を用いて描画したものであり、各ノードの絶対的な座標には得に意味はなく、エッジの有無と、ノード間の相対的な距離が、国際会議間の類似度の強さを意味している。これらの関連性はあくまでも、論文著者の共通性によるものであって、必ずしもトピックが近いというわけではない。しかし、同じ著者が2つの国際会議で発表することが多いという関連性は、文献調査において調査対象とする国際会議の選択の際には、有用な判断材料になると考える。

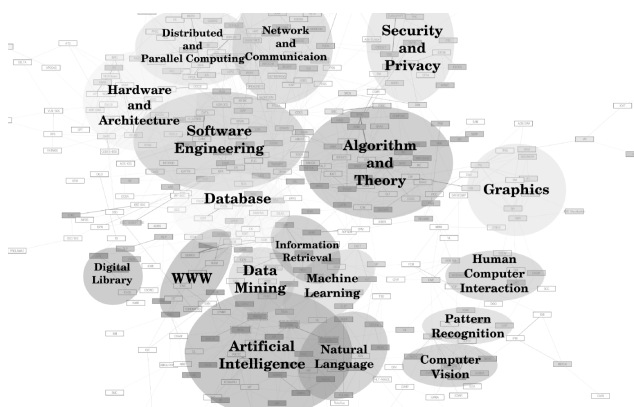


図 1: DBLP からの国際会議マップ (全体図)

参考までに 1 に、2 に含まれる各分野での主要な国際会議を示す。なお、どの国際会議が「主要な」会議かはいろいろな意見があるだろうし、また主要な国際会議に投稿することが良いことであるというわけではない。ここで挙げた会議は、主に Libra AcademicSearch によって Computer Science Directory として分類された分野に登録されている会議のなかで 100 以上の citation がある会議を列挙したもので、著名な国際会議が網羅されていることを必ずしも保障するものではない。

3. 国際会議マップの詳細

1 は、DBLP から抽出した国際会議をネットワークとして表現し、研究トピックのクラスターを重ねて描いた図である。このネットワーク図を俯瞰してみると、人工知能 (Artificial Intelligence) の近隣には、グラフィックス (Graphics)、パターン認識やコンピュータビジョン (Pattern Recognition, Computer

連絡先: 篠田 孝祐, 産業技術総合研究所, 茨城県つくば市梅園 1-1, Kosuke.Shinoda@aist.go.jp

*1 <http://www.informatik.uni-trier.de/~ley/db/>

*2 <http://www.graphviz.org/>

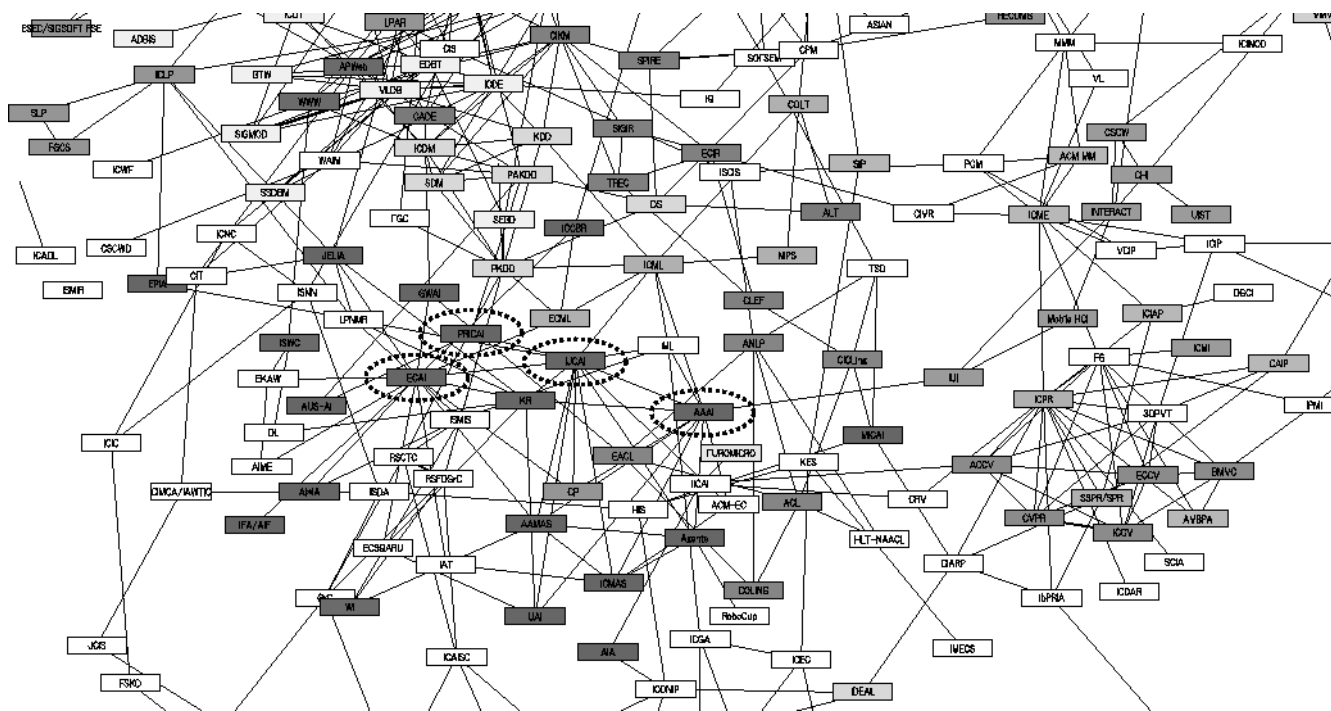


図 2: 国際会議マップ (人工知能に関連する国際会議近辺の拡大図)

表 1: 2 に含まれる分野とその分野の主要な国際会議

分野	主要国際会議
Artificial Intelligence	AAAI, ICGA, IJCAI, KR, UAI, ICMAS, AAMAS, ICAPS, ICRA, ...
Data Mining	KDD, DMKD, PKDD, ICDM, SDM, ...
Machine Learning, Pattern Recognition	ICML, UAI, COLT, NIPS, ICPR, ICDAR...
World Wide Web	WWW, WebDB, ISWC, WISE, APWeb...
Database Systems	SIGMOD, VLDB, ER, EDBT, PODS, ICDE, ICDT, DEXA, DS, ICDAS, ...
Computer Vision	ICCV, ECCV, CVPR, ACCV, ...
Human-Computer Interaction	CSCW, CHI, IUI, INTERACT, Mobile HCI, ...
Information Retrieval	SIGIR, CIKM, DL, TREC, ...
Natural Language	ACL, ANLP, HLT-NAACL, COLING, EACL, CICLing, ...

Vision), データマイニング (Data Mining), データベース (Database), ウェブ (WWW), 情報検索 (Information Retrieval), インタフェース (Human Computer Interaction) などの分野がある。そして、機械学習 (Machine Learning) はほぼ人工知能と隣接している。一方、ハードウェア (Hardware and Architecture) や通信ネットワーク (Network and Communication) などは、人工知能からはやや離れた位置にある。なお、この分野わけは、DBLP にて分類している情報を参考にし、会議ごとに付与している。

さらに、2 で、人工知能関係の国際会議周辺を拡大してみた。人工知能 (AI) の主要国際会議である IJCAI^{*3} があり、それらを囲むようにアジアでの人工知能会議 PRICAI, ヨーロッパでの人工知能会議 ECAI, 米国の人工知能会議 AAAI がある。少し目を上に移すと、ICML や ECML 等の機械学習の会議があり、さらに左上にはデータマイニングの国際会議である KDD や PKDD, PAKDD がある。その先に目を移すと、

データベースの国際会議である VLDB や SIGMOD, ウェブの国際会議 WWW 等が密に結び付いている。WWW と人工知能をつなぐパスの一つにセマンティックウェブの会議 ISWC がある。また、一度人工知能の周辺に視線を戻した右側には、自然言語処理の国際会議 ACL や COLING, また、右上にはインタラクション系の国際会議 CHI や CSCW などもある。

続いて、研究分野の関係の遷移をみるために、5年区切りでネットワークを作成した (図3参照)。これらの図から、人工知能関係の周辺には、自然言語、データベース関係の国際会議が近隣にあるが、一方で、アルゴリズムと理論に関する国際会議は、年を追うにつれ離れていきセキュリティやハードウェア関係の国際会議と近い関係となっていることがわかる。また、人工知能関係を中心として、ウェブや HCI など情報発見など分野が細分化している様子もみてとれる。

4. 分野と重要会議の抽出

図1のネットワークデータを用いて、研究分野の分類をクラスタリングなどで試みたが、図1で示してあるほど綺麗なク

*3 正式名称は、International Joint Conference on Artificial Intelligence. 本稿では、会議の略称が多いため正式名称の記述は一部に留める。

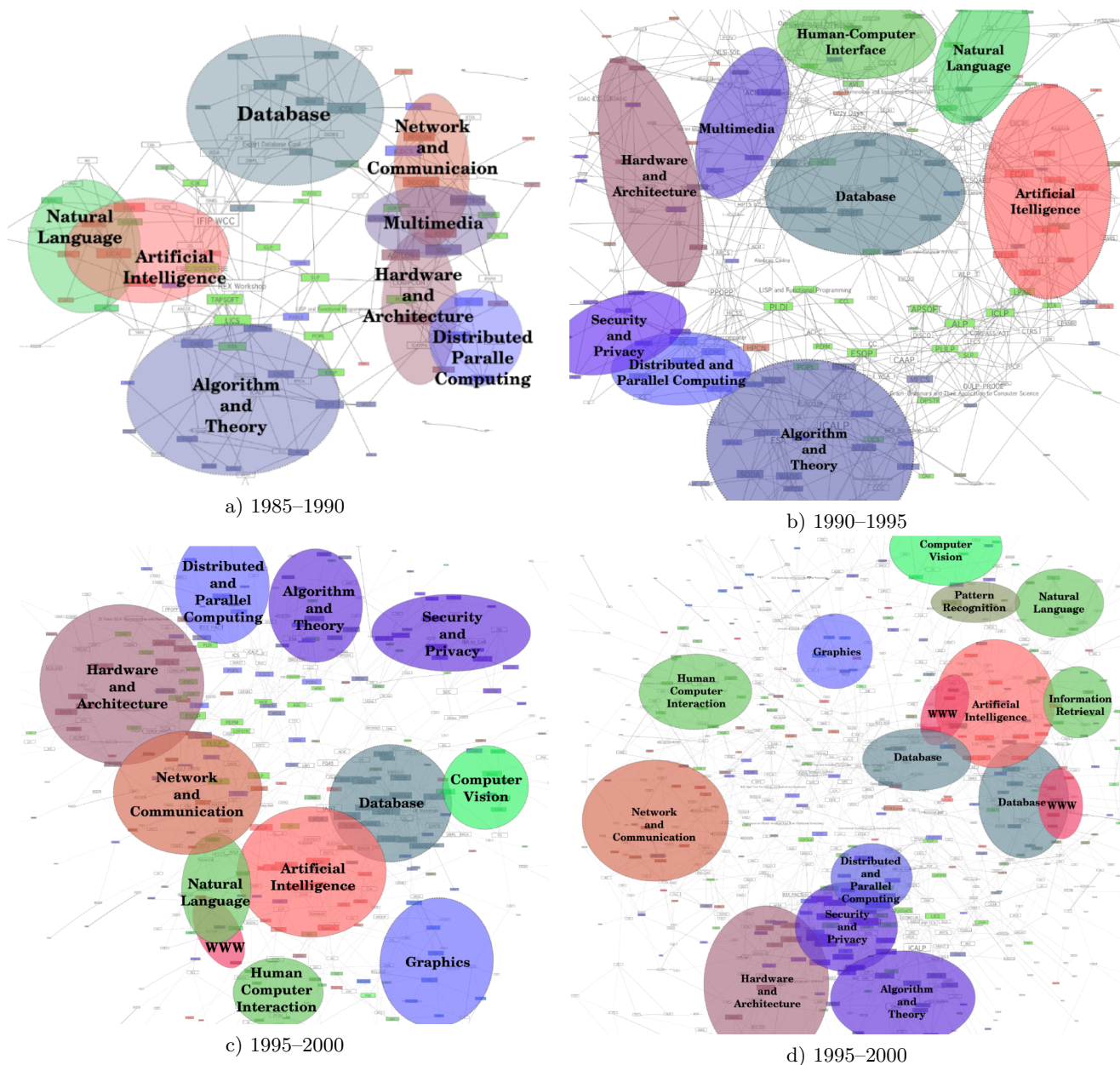
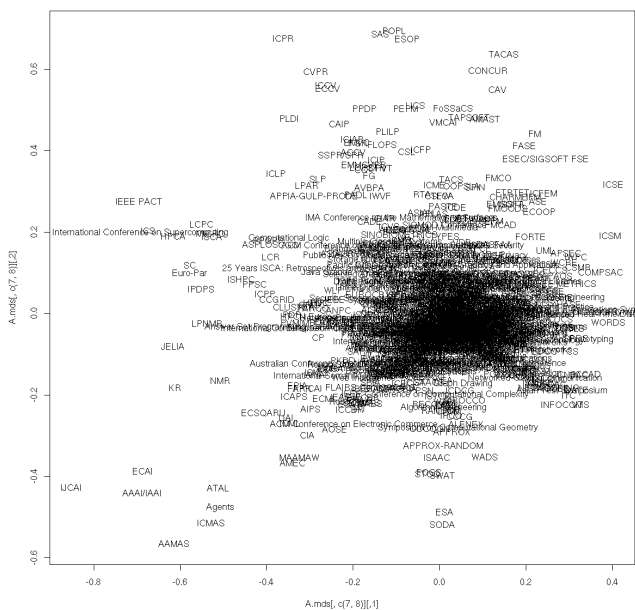


図 3: 5 年間隔の国際会議ネットワークの状態

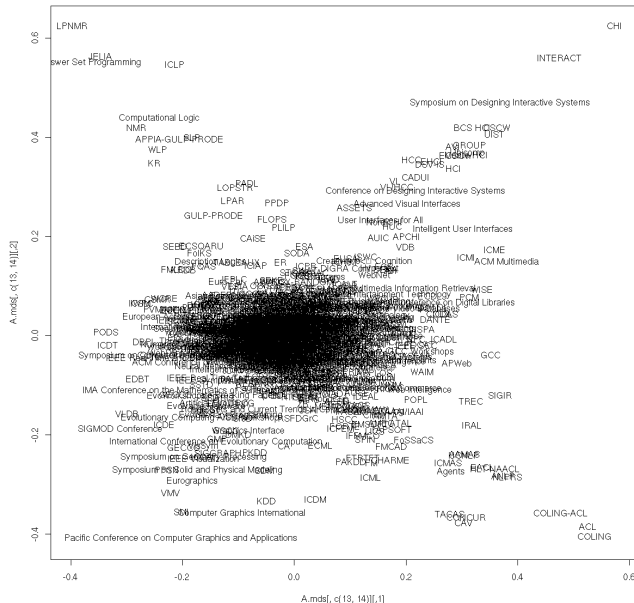
ラスタに分けることができなかつた。その理由として考えられるのは、我々が考えているほど、国際会議同士の関係に分野ごとの区切りが明確ではない可能性があるのではないだろうか。また、全体で 30 年近い期間のデータを扱っているため、経験の長い研究者は複数の分野に携わっている可能性があり、ネットワーク生成規則の関係上経験の長い研究者が強く影響していることも考えられる。また、重要会議に関しても、PageRank などの中心性を評価基準として、国際会議のランキングを試みたが、こちらも多数の投稿を受け付けている国際会議ほど複数の国際会議とのリンクを持ちやすく、必ずしも重要とされている会議がランキングのトップになるとは限らなかつた。これは、おそらく、重要な会議ほど採択率が低く、その重要な会議に採択される研究者は、幅広く採択している会議には投稿していない可能性などが考えられる。本稿におけるネットワーク作成は、採択された論文データをもとに作成したが、より適切な評価をするには、データには決して現れない投稿したが採

択されなかつた論文データを推定できることが、重要会議の抽出には重要ではないかと考える。

本章では、別の試みとして、多次元尺度構成法 8(MDS)[斎藤 80] を用いた分野と重要会議の抽出を試みた。図 4 は、統計解析ソフト R を用いて軸を 20 としたときのプロット図のうちの 2 つである。MDS を用いて図示することで、それぞれ関連ある会議、つまり分野の軸がある程度抽出できていることがわかる。例えば、図 4 a) では、左下部分に IJCAI, ECAI, AAIL, AAMAS など主要な国際会議が突出した位置にある。そして、図 4 b) では、右上方向に CHI, INTERACT などのインターフェース関係の国際会議が、そして、右下方向には、ACL, COLING などの自然言語系の国際会議の軸が確認でき、自然言語と関連度が高い人工知能関係の国際会議も同じ軸にあるが、自然言語という軸の中では低い(中心に近い)位置にあることがわかる。これらから、現在は、軸の数を固定して行っているが、何かしらの手法で軸



a) 人工知能関係の国際会議と重要会議の抽出 (左下)



b) アルゴリズム (左上) ユインタラクション (右上)、自然言語と人工知能 (右下)

図 4: 多次元尺度構成法による分野と重要会議の抽出

の数を決定できるのであれば、分析手段として利用できる。

5. まとめ

近年、情報取得のコストの低下から、大規模データからトピックの動向を抽出する試みは多数行われている。本論文では、そのなかでも国際会議に焦点をあて試みた。論文誌に関しては、非常に有名な指標として Impact Factor が存在するが、論文誌と同様に国際会議についても活動が活発なコンピュータサイエンスの分野では、国際会議の定量的な評価や解析が試みられている [Kuhn 08, Sidiropoulos 05, Zhuang 07, 野呂 05]。本研究も、実用的な形で提供できるように研究をすすめることを考えている。

参考文献

[Kuhn 08] Kuhn, M. and Wattenhofer, R.: he Layered World of Scientific Conferences, in *10th Asia Pacific Web Conference (APWeb)*, pp. 81–92 (2008)

[Sidiropoulos 05] Sidiropoulos, A. and Manolopoulos, Y.: A New Perspective to Automatically Rank Scientific Conferences Using Digital Libraries, *Information Processing and Management*, Vol. 41, No. 2, pp. 289–312 (2005)

[Zhuang 07] Zhuang, Z., Elmacioglu, E., Lee, D., , and Giles, C. L.: Measuring Conference Quality by Mining Program Committee Characteristics, in *In Proceedings of The ACM-IEEE Joint Conference on Digital Libraries*, pp. 225–234 (2007)

[斎藤 80] 斎藤堯幸：多次元尺度構成法, 朝倉書店 (1980)

[篠田 08] 篠田 孝祐, 佐藤 浩：論文投稿のための国際会議マップ, 人工知能学会会誌, Vol. 23, No. 3, pp. 380–384 (2008)

[野呂 05] 野呂智哉, 根岸秀典, 徳田雄洋：グラフ的手法による国際会議プログラム情報の解析, 日本ソフトウェア科学会全国大会, pp. 1–8 (2005)