

論文投稿のための国際会議マップ

A Map of Conferences for Paper Submission

篠田 孝祐
Kosuke Shinoda

産業技術総合研究所 知能システム研究部門
Institute of Intelligent System, National Institute of Advanced Industrial Science and Technology
kosuke.shinoda@aist.go.jp

佐藤 浩
Hiroshi Sato

防衛大学校 情報工学科
Dept. of Computer Science, National Defense Academy
hsato@nda.ac.jp

keywords: conference network, bipartite graph, conference ranking, DBLP

1. ま え が き

研究者にとって、自身の研究論文をどの国際会議に投稿するかを決めることは、研究のスケジュールを決める上でも、またその後の研究の方向性を決める上でも重要なプロセスである。国際会議はたくさんあるから、多くの方は何かしらの基準で国際会議の取捨選択を行っているのだろう。では、その際、候補となる国際会議をどのようにして選んでいるのか。おそらく、日々の研究活動のなかで、自分に関連する国際会議の存在を知り、その会議のメインピックや会議の重要性を認識し、そして自分の研究に適した会議を選んでのことと思う。その一方で、学生や若手の研究者にとっては、自分に関連する会議の存在を知る機会が限られているのではないだろうか。その理由として、国際会議に関する良質な情報を得ることが容易ではないことが挙げられる。そもそも情報を得る機会が、研究者個人が抱える社会的なネットワーク（所属先や共同研究相手など）に制約されているのも一因であろう。我々は、戦略的に国際会議に投稿していくためには、投稿する候補となる国際会議の立ち位置を把握し、関連する国際会議を俯瞰することが重要であると考えている。それにより、投稿論文の位置づけをより明確に意識することができるのではないだろうか。

本稿では、「国際会議に通すための英語論文執筆」の特集のひとつとして、国際会議の関係性を総体的に把握するために、国際会議マップの生成を試みる。以下、2章において国際会議マップの抽出方法について述べ、3章にて作成したマップをもとに、国際会議の関係を概観する。4章で、海外での同様の取り組みについて紹介し、最後にまとめを述べる。

2. 国際会議マップの作成

人工知能に関連する国際会議のマップを作るために、我々は DBLP^{*1}で公開されている論文メタデータを用いた。このメタデータは XML 形式で提供され、国際会議だけでなく、論文誌、ワークショップ、博士論文等も含まれている。国際会議の数が約 3700 件、国際会議の論文約 60 万本、著者約 40 万名のデータが登録されており、データマイニングや自然言語処理の研究で用いられることも多い、標準的なデータである^{*2}。

国際会議マップの作成には、国際会議の論文情報のみを対象とし、著者名と国際会議名を抽出した。著者および共著者名、ならびに国際会議名を用いて 2 部グラフを作成し、その 2 部グラフを用いて国際会議の類似性を計算した。その類似性によって国際会議マップ（ネットワーク図）を作成する。以下にその具体的な手順を示す。国際会議の集合を C 、著者の集合を A とする。

- (1) 著者（もしくは共著者） $a \in A$ による論文 p が、ある国際会議 $c \in C$ にて発表されていれば、著者 a は国際会議 c で一回発表したとする。発表回数が閾値 thr_{auth} 以下の著者、および発表された回数（=論文数）が閾値 thr_{conf} 以下の国際会議は除去する。
- (2) 各国際会議 $c \in C$ が、各著者 $a \in A$ によって何回発表されたかによって、国際会議ごとの著者ベクトル v_c を生成する。 v_c は $|A|$ 次元のベクトルである。
- (3) 2 つの国際会議 c_1 と c_2 の類似度を、ベクトルのコサイン類似度 $v_{c_1} \cdot v_{c_2} / |v_{c_1}| |v_{c_2}|$ で求める。
- (4) コサイン類似度が閾値 thr_{veccos} 以上の場合、2 つ

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

*2 ただし、登録されている国際会議のすべての情報が登録されているわけではない。例えば、コンピュータグラフィックスの分野で著名な国際会議 SIGGRAPH は、2004 年以降のデータがない。また、ドイツのデータベースの研究者によって作られているため、ドイツ国内の会議やデータベース系に偏っている傾向がある。

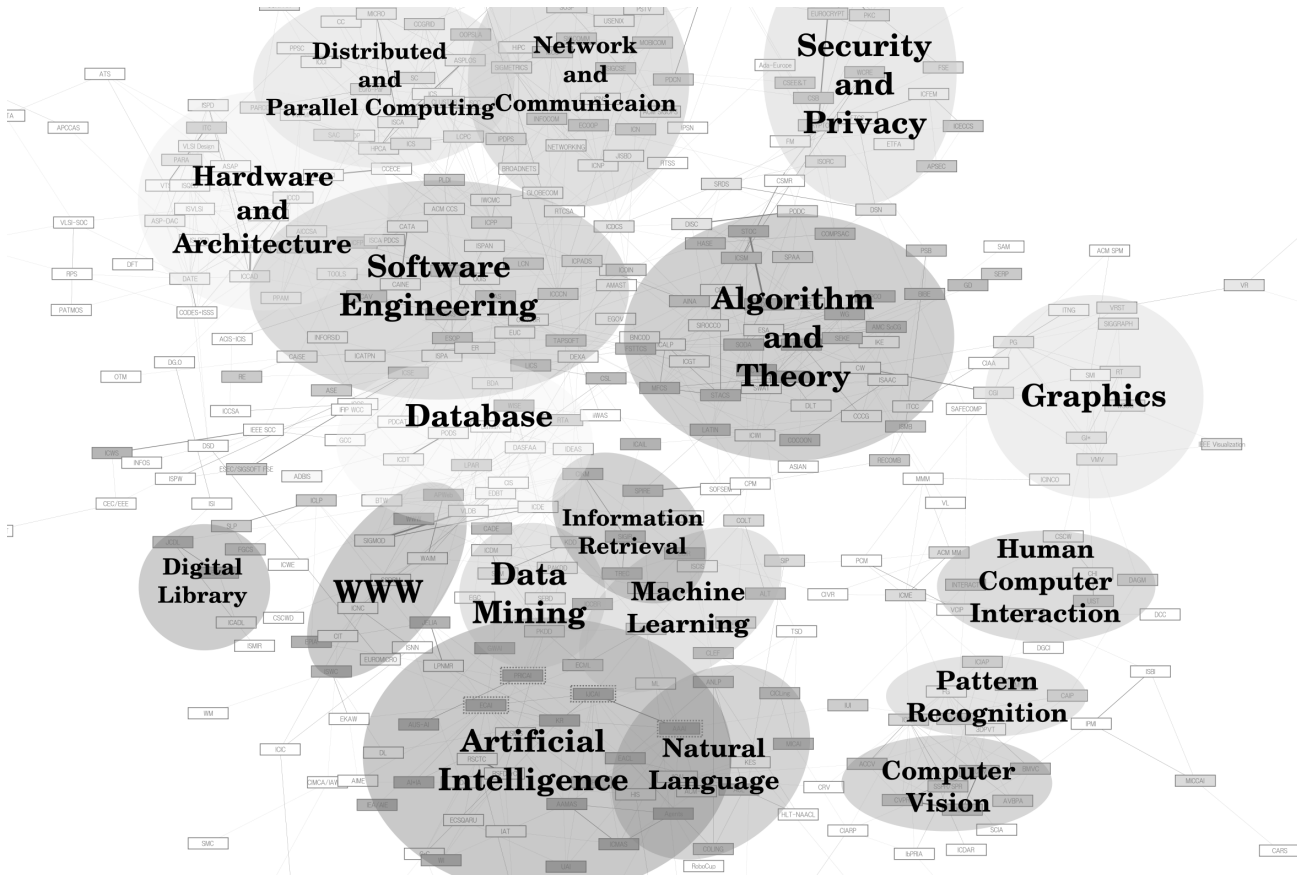


図 1 DBLP からの国際会議マップ (全体図)

の国際会議間にエッジを張り、ネットワークを構成する。

すなわち、本稿で生成する国際会議マップは、国際会議をノードとし、国際会議間の類似性をエッジとしたネットワーク図である。国際会議間の類似性は、採択された論文に同じ著者^{*3}がどの程度含まれているのかを基準に決定しており、エッジは方向なしである。本稿では、ネットワーク図を graphviz^{*4} を用いて描画している。ノードは、ばねモデルを用いて配置されており、その絶対的な位置に意味はなく、エッジの有無とそれによる距離の近さにのみ意味がある。

上記の処理は、2 部グラフからネットワーク図を構成する際に用いられる一般的な手法を用いているが [Zha 01, Wasserman 91]、以下のように若干の前処理を行っている。まず、DBLP に登録されているデータは、ドイツで作られているため、ドイツの国内会議が含まれる。したがって、ドイツ語名の会議等、地域性の高いと考えられる会議は削除した。次に、国際会議の中には、他の会議と合併するなど発展的に解消することがあり、そう

いったものは適宜、最終的な会議名に置き換えた^{*5}。どちらの作業もできるだけ客観的に行なっており、用いたデータと前処理に関する詳細の情報は、著者のウェブページ^{*6}に掲載しているので、必要であれば参照されたい。

3. 国際会議マップ

上記の手順で作成した国際会議マップが図 1 である^{*7}。全体を俯瞰しやすいように、トピックが近い国際会議が集まっている近辺にラベルを付与している。ラベルの名称ならびに分類については、DBLP や Libra Academic Search^{*8}で示されている分類を参考に行なった。また、分野に応じてノードのモノクロ階調を変えている。

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 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)、

*3 著者が同じかどうかは DBLP のデータに依存しており、同名の研究者がいた場合に厳密に別の著者になるかどうかは本稿では特段の処理をしない。統計的に見れば、仮に同名同名の著者が少数含まれていても会議の類似性が高いかどうかは相対的なものであるから、その結果に大きな影響はないと考えられる。

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

*5 例えば、Symposium on Logic Programming (SLP) などは、DBLP では <http://www.informatik.uni-trier.de/~ley/db/conf/slp/> をみると最終的に Joint International Conference and Symposium on Logic Programming (IJCSLP) という会議になっている

*6 <http://confmap.k4no.info>

*7 閾値は、図の見易さを考慮し、 $thr_{auth} = 15$, $thr_{conf} = 150$, $thr_{veccos} = 0.092$ とした。

*8 <http://libra.msra.cn/>

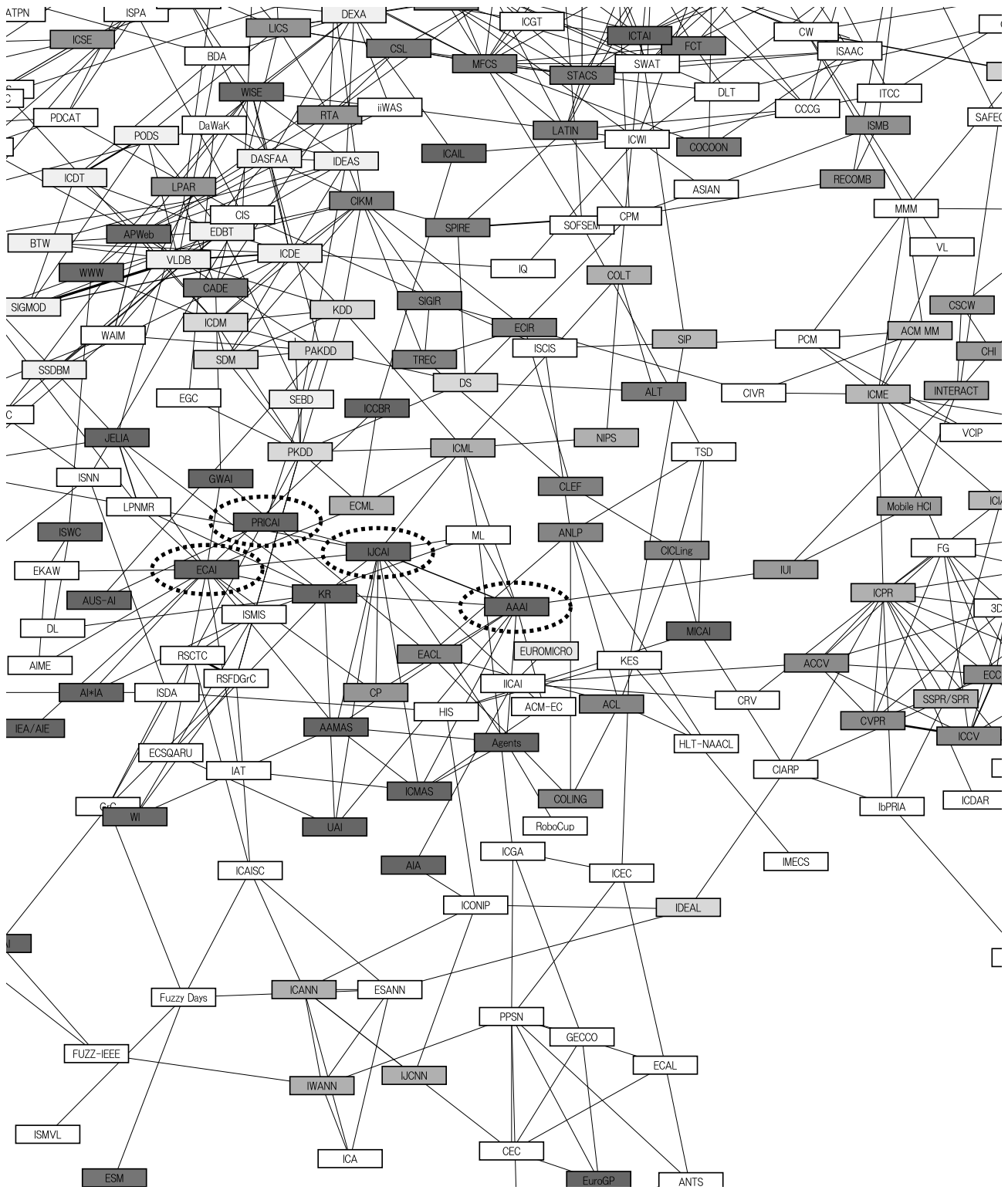


図 2 国際会議マップ (人工知能に関連する国際会議近辺の拡大図)

インタフェース (Human Computer Interaction) などの分野があることがわかる。機械学習 (Machine Learning) はほぼ人工知能と隣接しており、後述するように人工知能の分野の中には、知識表現やエージェントといったトピックが含まれる。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。

図 2 は、人工知能の近辺を拡大して表示したものである。人工知能 (AI) の主要国際会議である IJCAI*⁹ 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 が配置されている。図では、人工知能学会と直接関連の深いこれらの国際会議を楕円で囲っている。その近辺には、エージェント関係の国際会議 AAMAS、不確実性を扱う国際会議 UAI、知識表現の国際会議 KR 等、人工知能の中心に近いトピックの会議が見られる。

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。

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、論文を投稿する先を検討するには、有用な判断材料になると思う。読者自身が投稿している国際会議の周辺を確認して、今まで考えていなかった国際会議やその意外な関連に気づいてもらえれば幸いである。なお、DBLP のデータには含まれていない国際会議 (例えばユビキタス系やロボット系の国際会議の一部) もあることにご注意いただきたい。

参考までに表 1 に、図 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。自分に関連する分野で見逃しているものがないか確認していただければと思う。ここで挙げた会議は、主に Libra Academic Search 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。

4. 国外での取り組み

国際会議の位置づけに関する情報提供の試みは国外でも行われている。以下に、参考のためいくつかのサイトを挙げるので、論文投稿の判断材料として用いていただければと思う。

(1) Computer Science Conference Ranking*¹⁰

この Conference Ranking では、“Estimated Impact of Conference (EIC)” という指標に基づいて会議のランキングの提供をしている。ランキングのデータは、下記の 4 分類の分野の会議に対して提供されている。

- a Database / Knowledge and Data Management / Data Security / Web / Mining
- b Artificial Intelligence / Machine Learning
- c Architecture / Hardware / High-Performance Computing / Tools
- d Application / Education / Software / Theory / Communications

EIC を決定する要素としては、論文の引用頻度、査読者のレポートの質、会議による学生への参加費・旅費サポートなどを考慮に入れていると記述がある。質的な評価の方法に関しては明示されていないが、DBLP に掲載されている情報をもとに筆者やプログラム委員にメールを利用したインタビューを行っている。

(2) Australian Ranking of ICT Conferences*¹¹

主にオーストラリアの研究者が参加している ICT 関連の国際会議をランキングが掲載されている。このランキングは、国際会議参加者から会議に関するレポートを集め、20 名程度の委員によって 4 段階に評価している。ランキングのデータは、2005 年からほぼ毎年更新されている。

(3) Estimated impact of publication venues in Computer Science*¹²,

Citeseer*¹³ に登録されている論文情報をもとに、会議で採択された論文の引用頻度の平均を計算している。つまり、多くの研究者から読まれている国際会議であるかどうかを評価している。なお、国際会議に関する詳細情報は DBLP の情報を利用している。

(4) Libra Academic Search*¹⁴

各会議の引用頻度の平均で、会議のランキングの表示などを見ることが出来る。DBLP だけでなく、

*10 <http://www.cs-conference-ranking.org/conferencerrankings.html>

*11 [http://www.core.edu.au/rankings/Conference % 20Ranking % 20Main.html](http://www.core.edu.au/rankings/Conference%20Ranking%20Main.html)

*12 <http://citeseer.ist.psu.edu/impact.html>

*13 <http://citeseer.ist.psu.edu/>

*14 <http://libra.msra.cn/>

*9 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。

表 1 図 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, CILing, ...

ACM Digital Library^{*15} など複数のデジタルライブラリからデータの提供を受け、他の同様のサイトと比べて大量のデータを持っている点特徴である。また、検索した研究者と関係の深い会議や共著者を提示するなど、研究に係わる情報の検索ポータルとしての機能も備えている。

(5) confsearch^{*16}

国際会議ランキングを利用しながら、キーワードや研究者に関連する国際会議のリストを、投稿カレンダーとして表示できる。ランキングの手法に関する詳細は不明である。

このように、国際会議の重要度を決定する取り組みは Conference Ranking と呼ばれ、論文誌に対する Impact factor と同じく、会議の質を表すために提案されている指標である [Sidiropoulos 05, Zhuang 07, Nie 07]。海外でも、いかに国際的なインパクトを与えるために研究をアピールするかに腐心し、データに基づく客観的な指標による戦略的な論文投稿が行われているのではないだろうか。

5. ま と め

本稿では、DBLP のデータを用いて国際会議マップを生成する試みを行い、人工知能に関連する国際会議を取り巻く状況を概観した。また、海外での類似の取り組みについて紹介した。

本論文で作成したマップは、DBLP の全データを使い、できるだけシンプルに処理したが、さらに踏み込んで国際会議の投稿への意思決定支援としての利用を考えるならば、例えば、国際会議の重要度やトピックのトレンドを把握するなど重要であろう。国際会議に投稿し、研究をきちんとアピールするには、周りの状況をきちんと把握しておくことがその第一歩として不可欠ではないかと思う。本稿での試みで、その重要性を少しでも示せば幸いである。今後は、国際会議の投稿に役立つよう、

詳細なデータ分析を行なっていきたいと考えている。なお、本論文で、作成した国際会議マップはウェブページ (<http://confmap.k4no.info>) に掲載している。

◇ 参 考 文 献 ◇

- [Nie 07] Nie, Z., Ma, Y., Shi, S., Wen, J.-R., and Ma, W.-Y.: Web Object Retrieval, in *In Proceedings of the 16th international World Wide Web conference (WWW 2007)* (2007).
- [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).
- [Wasserman 91] Wasserman, S. and Iacobucci, D.: Statistical modeling of one-mode and two-mode networks: Simultaneous analysis of graphs and bipartite graphs, *British Journal of Mathematical and Statistical Psychology*, Vol. 44, pp. 13–44 (1991).
- [Zha 01] Zha, H., He, X., Ding, C. H. Q., Gu, M., and Simon, H. D.: Bipartite Graph Partitioning and Data Clustering, in *CIKM*, pp. 25–32 (2001).
- [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).

〔担当委員：×× 〕

19YY年MM月DD日 受理

—— 著 者 紹 介 ——

篠田 孝祐(正会員)

1999年名古屋工業大学 知能情報システム学科 卒業 2004年 北陸先端科学技術大学院大学 知能科学研究科 博士後期課程終了。博士(知能科学)、2005年 防衛大学校にて助手、2008年、産業技術総合研究所にて産総研特別研究員として、避難シミュレーションの研究に従事。マルチエージェントシステム、大規模社会シミュレーション、複雑系ネットワークに興味をもつ。人工知能学会、情報処理学会 学生会員

佐藤 浩(正会員)

1997年東京工業大学大学院総合理工学研究科知能科学専攻博士後期課程修了。博士(工学)。同年大阪府立大学総合

*15 <http://portal.acm.org/dl.cfm>

*16 <http://www.confsearch.org/>

科学部助手 . 1999 年防衛大学校情報工学科助手 . 2002 年同大学校同学科講師 , 2005 年同助教授 , 2007 年同准教授 , 現在に至る . 進化計算 , マルチエージェントなどの研究に従事 . 人工知能学会 , 日本知能情報フアジイ学会 , シミュレーション学会 , 進化経済学会各会員 . 2002 年ニューサウスウェールズ大学計算機科学科客員研究員 .

Abstract: