

制約充足 / 最適化テクニックを用いた会議プログラム自動作成ツール

Development of a Tool for Automatically Generating a Conference Program

西村 直史
Naofumi Nishimura

大田 直樹
Naoki Ohta

櫻井 祐子
Yuko Sakurai

岩崎 敦
Atsushi Iwasaki

横尾 真
Makoto Yokoo

九州大学大学院システム情報科学府
Graduate School of ISEE, Kyushu University

We develop a tool for automatically generating a conference program. Most conference programs are manually generated based on some information about submitted papers. However, it is very difficult to determine a conference program in order to satisfy various constraint, condition, and criteria at a large-scale conference. Thus, we propose a tool for JSAI2008 conference program that based on search algorithms by utilizing techniques of constraint satisfaction/optimization problems, and we actually constructed JSAI2008 conference program with this tool. In addition, we collect and examine questionnaire for audience at the JSAI2008, and we considered and improve a tool. We experimentally constructed JSAI2009 conference program with this improved tool, and we considered it in comparison with the real program.

1. はじめに

大規模な学術会議では、複数の論文発表がセッションと呼ばれる単位にまとめられ、複数のセッションが並行して実施される。このような会議プログラムは通常、論文投稿時に著者により入力される登録情報（論文タイトル、キーワード等）に基づいて手作業で決定されていた。しかし、大規模な会議において、様々な制約条件 / 価値基準を満足するプログラムを作成することは非常に困難である。また、プログラム作成にかかる人的・時間的なコストも問題で、プログラム委員会には大変な負担となっていた。

この会議プログラム作成問題は、スケジューリング問題の一つと捉えることができる。スケジューリング問題は従来より盛んに研究されてきた分野であり [松井 99, Easton 01, Kaneko 99], 様々な手法が提案されている。また、会議プログラムの作成支援を対象とした研究はいくつか見られるが [小作 02], 会議プログラムの作成過程における一部の事務処理の自動化などを目的としたものが多く、会議プログラムの作成全体を取り扱ったものはほとんどなかった。

そこで本論文では、制約充足 / 最適化のテクニックを用いた会議プログラム自動作成ツールの開発を行う。このツールでは、会議プログラムの作成をセッションの構成と配置という2つの手順に分けて扱っている。それぞれの手順における評価基準として、セッション構成では、同じセッション内の論文同士の関連性が高くなることが望ましいとし、セッション配置では、同じ時間帯に実施されるセッションに同士の関連性が低くなることを望ましいとする。論文 / セッションの関連性を表すため、任意の2つの論文間の関連度を定義する。

2つの論文間の関連度は、発表申込み時の入力情報の1つである論文キーワードの類似性と著者間距離を用いて決定する。著者間距離は、人工知能学会全国大会の「大会支援システム」 [西村 04] の一部である Web 支援ツールにおける、研究者間の関連度を適用した。論文キーワードだけでなく、著者間距離も考慮することで、これまでの研究活動における研究者間の関連

性を評価可能となる。さらに、付加的な制約条件として、同一セッション内での同一所属の発表者数に上限を設定する。

こうした条件のもと、まず我々は JSAI2008 の大会プログラムを対象とした会議プログラム自動生成ツールを開発し [西村 08], 実際に大会プログラムを作成、採用された。本論文ではまず、この大会プログラムに関して、JSAI2008 の大会参加者に対して行ったアンケートの結果について報告し、考察を行う。

また、考察から得られた問題点等のうち、JSAI2008 で用いたツールによって得られる解が局所最適解であるという問題に対し、探索手法の改良を行う。具体的には、会議プログラム作成問題を整数計画問題として表現し、整数計画法を用いることで探索を行う手法を用いたツールを開発した。結果、大域的な最適解により近い解を得ることに成功した。

この改良されたツールを用いて、我々は JSAI2009 の大会プログラムを作成した。作成したプログラムは実際に採用されたプログラムと異なるが、その実際のプログラムと作成したプログラムを比較、考察した結果について報告する。

本論文では、まず会議プログラム作成問題の要件と定式化について述べ (2章)、次に JSAI2008 の大会プログラムについて述べる (3章)。また、改良したツールの最適化手法である整数計画法について述べ (4章)、そのツールを用いた JSAI2009 の大会プログラム作成 (5章)、及びそのプログラムと実際のプログラムとの比較と考察について述べ (6章)、最後に結論と今後の課題について述べる。

2. 会議プログラム作成問題

本章では、会議プログラム作成問題を制約最適化問題として取り扱うための定義を行う。

2.1 会議プログラムの作成要件

本節では、会議プログラムの作成要件について説明する。本研究では、会議プログラム作成問題をセッション構成、セッション配置という2つの段階に分けて取り扱う。

セッション構成 セッション構成では各セッションをどの論文で構成するかを決定する。複数の論文を1つのセッションとして、投稿論文をセッション単位に分割する。セッ

表 1: キーワード関連度 $f_k(i, j)$

		論文 j キーワード		
		第一希望	第二希望	第三希望
論文 i キー ワード	第一希望	+6	+4	+2
	第二希望	+4	+3	+1
	第三希望	+2	+1	+1

セッション構成において望ましい特徴として、以下の条件が考えられる。

- 評価基準：同一セッション内の論文が、同一もしくは類似した研究分野に分類される。
- 制約条件：同一セッション内の論文が同一の所属（大学の研究室 / 企業の部署）であることが少ない。

セッション配置 セッション配置では各セッションをどの時間帯、どの会場に配置するかを決定する。セッション配置では次の条件を満たすことが望ましいと考えられる。

- 評価基準：同一時間帯に配置されたセッション間における、研究分野の関連性が低い。
- 制約条件：同一時間帯に配置された異なるセッションに、同一の著者・共著者が存在しない。

2.2 問題の定式化

本節では、会議プログラム作成問題の定式化を行う。

2.2.1 作成に用いられる情報

会議プログラムを作成する際に与えられる情報として、論文の数 n 、セッションの数 m 、1セッション辺りの論文数 $d = \lceil n/m \rceil$ 、1つのセッションに許される同一所属論文数 l がある。

また、各論文には所属情報、著者情報、キーワード情報があり、任意の2つの論文 i, j に対し、所属の一致を示す $g(i, j)$ （一致なら1、そうで無ければ0）、著者の重複を示す $h(i, j)$ （重複があれば1、無ければ0）、及び次節で定義される関連度 $f(i, j)$ が与えられる。

2.2.2 評価関数

会議プログラムを作成する際に用いる、任意の2論文 i と j の関連度を示す評価関数 $f(i, j)$ を、論文に設定されたキーワードに関する関連度 $f_k(i, j)$ と著者間距離 $f_a(i, j)$ によって定義する。以下にその詳細を示す。

キーワード関連度 今回、各論文にはキーワード（希望セッション）として第三希望までの情報が与えられていた。そのため、任意の2論文 i, j のキーワード関連度 $f_k(i, j)$ を表1のように定義した。ここで、表中の数字はその行、列のキーワードが一致していたときに $f_k(i, j)$ に加算される値を表す。例えば、論文 i の第一希望と論文 j の第二希望が一致していた場合、 $f_k(i, j)$ に4が加算される。第一希望から第三希望までが同じ希望順位で全て一致した場合、 $f_k(i, j)$ は最大値10をとる。

著者間距離 著者間距離 $f_a(i, j)$ は文献 [松尾 05] の研究者間関連性 $dis(a, a')$ に基づく。 $dis(a, a') \in [0, 1]$ は関連性が全くない場合は0となり、関連性が高いほど1に近づく。2つの論文 i と j に関する著者間距離は、各論文の著者

集合 A_i と A_j において、任意の著者 $a \in A_i$ と $a' \in A_j$ の研究者間関連性 $dis(a, a')$ の最大値とする。

$$f_a(i, j) = \max_{a \in A_i, a' \in A_j} dis(a, a') \quad (1)$$

ここで、 $f(i, j)$ を

$$f(i, j) = f_k(i, j) + f_a(i, j) \quad (2)$$

とする。キーワード関連度 $f_k(i, j)$ は1刻みの不連続値、著者間距離 $f_a(i, j)$ は1未満に正規化された連続値であるので、この評価関数はまずキーワード関連度を基準とし、これが同値のときに著者間距離を用いて比較するものとなっている。

論文間の関連度を可視化する研究は、論文の本文全体を利用したもの [杉本 96] やアブストラクトのみを利用したもの [鈴木 07]、論文の引用関係を用いるもの [Small 73] など様々なものがあり、主に論文データベースにおける検索や分類を目的として研究されている。しかし、それらには事前処理や膨大なデータが必要となる場合が多いため、今回の評価関数では論文の登録情報から容易に算出可能なキーワードの類似性と、プログラムの作成対象とした人工知能学会の大会支援システムの提供する著者間距離データを用いる。

3. JSAI2008 プログラムについて

我々は、2章で定義したプログラム作成要件と評価関数と同様のものを用いて、会議プログラムを自動生成するシステムを開発し、JSAI2008の大会プログラムを作成した。このシステムでは、セッション構成・配置ともに、制約条件を満たす下で、評価基準を最適化するように貪欲法で初期解を生成し、2つの論文 / セッションの入れ替えによる反復改善を行い、解を求める。このシステムで作られた大会プログラムは若干の修正の後、実際のプログラムとして採用された。

また我々は、JSAI2008参加者に対し、この大会プログラムに関するアンケートを行なった。このアンケートでは開催中に紙ベースで行ったものと、終了後にWeb上で行われたものを併せ100件の回答を得た。

まず、本プログラムに関しての感想として、非常に満足から非常に不満までの5段階で評価してもらった。その結果、非常に満足 / 概ね満足という回答が7割、どちらとも言えないが2割、やや不満 / 非常に不満といった回答が1割という結果となった。次に、2章で定義した、プログラムの作成要件に関して、適切と思うかどうかを回答してもらったところ、8割5分の方から「適切である」との評価を頂いた。

アンケートの結果、プログラムに満足しているとの回答は7割であり概ね高評価であるが、その割合は作成要件を支持する割合よりも少ない。これは、作成要件を満たす手法にいくつかの問題点が存在するためと考えられる。そこで我々は、まず評価基準を最適化する手法について着目し、改良を行った。

4. 整数計画法による最適化

JSAI2008のプログラム作成では、最適化手法として貪欲法と反復改善法が用いられていた。これは、問題が大規模なことと条件を変えながらの試行を想定し、高速性を重視したためであるが、そのために解が局所最適解となっていた。

そこで、我々は会議プログラム作成問題を整数計画問題としてモデル化し、解を算出する。整数計画問題は数理計画法の一種で、整数の値をとる変数の集合が存在し、その変数が任意の

数の一次等式および一次不等式（制約条件）を満たす値をとる上で、ある1つの一次式（目的関数）を最小化もしくは最大化するような値を求める問題である。[Williams 99]。ここでは変数の値を0か1かに限定した0-1整数計画問題を取り扱う。整数計画問題としてモデル化し、解を求めることが出来れば、その解はその問題の大域的最適解となる。

4.1 整数計画問題としてのモデル化

この節では、セッション構成問題を整数計画問題として扱うためのモデル化を行う。

4.1.1 定数

初期入力として、2章で定義した論文数 n 、セッション数 m 、セッションあたり論文数 d 、所属の制限 l 、及び2論文間の関係 $f(i, j), g(i, j), h(i, j)$ が与えられる。

4.1.2 変数

制約条件や評価基準を表現するために、論文 i がセッション s_k にあることを表す変数 p_{is_k} を定義する。この値は、論文 i がセッション k にあれば1、なければ0をとるものとする。

4.1.3 制約

前節で定義した変数を用い、セッション構成の制約条件を一次式によって表現する。

論文は1つのセッションに所属する。全ての論文は、ある1つのセッションに所属する。すなわち、「ある論文 i に対し、全てのセッション $s_k (k = 1, 2, \dots, m)$ について p_{is_k} を考えたとき、 $p_{is_k} = 1$ となる k が唯一存在する」という条件が全ての論文 $i (i = 1, 2, \dots, n)$ において成り立てばよい。実際の制約式は以下のように表現する。

$$\sum_{k=1}^m p_{is_k} = 1 (\forall i \in \{1, 2, \dots, n\}) \quad (3)$$

よって、左辺の項数 m の制約式が n 個存在する。

1 セッションあたりの論文数の制限 ある1つのセッションに所属するセッションは d 個である。すなわち、「あるセッション s_k に対し、全ての論文 i について p_{is_k} を考えたとき、1であるものが d 個存在する」という条件が全てのセッション $s_k (k = 1, 2, \dots, m)$ において成り立てばよい。実際の制約式は以下のように表現する。

$$\sum_{i=1}^n p_{is_k} = d (\forall k \in \{1, 2, \dots, m\}) \quad (4)$$

よって、左辺の項数 n の制約式が m 個存在する。

1 セッション内の同一所属の論文数の制限 ある1つのセッション内の論文で、同一所属のものは l 個を上限とする。この制約は変数 p_{is_k} を用いると、以下のように表現出来る。

$$\sum_{j=1}^n \sum_{k=1}^m g(i, j) (p_{is_k} \wedge p_{js_k}) \leq l (\forall i \in \{1, 2, \dots, n\})$$

4.1.4 目的関数

セッション構成における評価基準は、同一セッション内の論文の関連度が高くなることである。同一セッション内の全ての組み合わせの2論文間の関連度 $f(i, j)$ の平均値を各セッションの評価値とし、全てのセッションの評価値の平均値をセッション構成の評価値とする。そして、このセッション構成の評価値

表 2: 整数計画法によるセッション構成の評価

探索手法		評価値	実行時間 (分)
JSAI2008 システム	10 分割 (平均)	3.32	
	240 論文	2.71	10
整数計画法	10 分割 (平均)	3.92	126
	240 論文	N/A	N/A

を最大化するように目的関数を設定する。これを2論文間の関連度 $f(i, j)$ と変数 p_{is_k} を用いて表すと以下ようになる。

$$\text{Maximize } \sum_{k=1}^m \sum_{i=1}^n \sum_{j=1}^n f(i, j) (p_{is_k} \wedge p_{js_k})$$

4.2 予備実験

前節で定義した整数計画問題が、実際の会議プログラム作成に使用出来るかどうかを確認するため、予備実験を行った。論文数を変えた問題を複数用意し解かせるため、論文数に対し実行時間が指数的に増大し、32 論文を越えたところで求解に数日を要するようになった。

実際の会議プログラムを作成するには、数百本の論文を一度に扱えなければならない。そこで我々は、第一希望キーワードの大分類によって論文を分割し、それらの小問題に対し整数計画法を適用することとした。会議プログラム作成に用いる評価関数 $f(i, j)$ は偏った疎な分布をとっており、大分類によって分割されたグループ間には関連性は殆ど見られなくなるため、問題を分割して最適化したものでも、全てを一度に最適化したものと遜色ない結果が期待できる。実験結果を表2に示す。

結果、整数計画法では評価値 3.92 と、JSAI2008 システムでの評価値 2.71 より大幅に良い評価値となった。また、JSAI2008 システムでも、分割した問題を解いたものは評価値が高くなったが、これは問題を分割することでデッドロックが起こりにくくなったためと考えられる。

5. JSAI2009 プログラム作成

本章では、改良したツールと JSAI2009 の投稿データを用いて、JSAI2009 の大会プログラムを作成した手法について説明する。

5.1 セッション構成

前章で定義した整数計画法を用いて、論文数 $n = 236$ 、セッション数 $m = 61$ 、1 セッションあたりの論文数 $d = 4$ 、また $l = 0$ 及び $l = 2$ の2通りの場合に関してセッション構成を行った。論文の分割は大分類に応じ、10 から 30 件程度の13 グループに分割し、それぞれ最適化を行った。

5.2 セッション配置

前節で行った2通りのセッション構成それぞれに対しセッション配置を行う。セッション配置においても、前章のセッション構成と同様に、 $h(i, j)$ を用いて著者の重複の制約条件を、同一時間帯のセッションの関連度を最小化するように目的関数を設定することで、整数計画問題としてモデル化できる。

しかし、セッション構成以上に求解が困難なため、今回はセッション構成で分割した各グループ（異なるグループは関連性が低い）を会場ごとに割当て、その中で各会場ごとに DP（動的計画法）を利用し順序を最適化する反復改善法を用いることとした。各会場について5回ほど最適化を行ったところ改善が終了した。

表 3: JSAI2009 プログラムの比較

プログラムの種類	セッション構成		セッション配置	
	評価値	制約違反	評価値	制約違反
実際のプログラム	5.67	7	1.22	30(4)
所属制約無し	6.72	8	0.60	0(0)
所属制約有り	6.43	0	0.74	0(0)

6. JSAI2009 プログラムの考察

前章では、4章の手法で改良したツールを用いて JSAI2009 の大会プログラムを自動作成したが、この自動作成したプログラムは実際に JSAI2009 において採用された大会プログラムとは異なる。そこで本章では、実際に採用された大会プログラムと、ツールを用いて自動作成した大会プログラムの比較、考察を行う。

6.1 セッション構成

まず、セッション構成について、実際のプログラムと本システムで作成されたプログラムを比較する。表 3 のセッション構成の項にその結果を示す。最初に、セッション構成における制約違反とは、所属が同一の論文の数が、同一のセッションの中で規定数（今回は 2）を越えている状態を表す。実際のプログラムではこの違反が 7 件起こっており、ほとんど考慮されていないことが分かる。

次にセッション構成の評価値とは同一セッション中の論文の関連度 $f(i, j)$ の平均値を表し、高いほうがより優れたセッション構成と言える。この値は、実際のプログラム 5.67 に対し所属制約無しでシステムが作成したプログラムは 6.72 と、1 ポイント以上高い結果となっている。所属制約有りのプログラムでも 6.43 となり、構成の自由度が低くなっているにも関わらず、実際のプログラムの値を大きく上回る結果となった。

6.2 セッション配置

次にセッション配置について実際のプログラムと本システムで作成されたプログラムを比較する。まずセッション配置における制約違反とは、同一時間帯に存在する異なるセッションに、同一の著者・共著者を含む論文が存在することである。本システムで作成したプログラムではこの制約違反がないように制約を設けているが、実際のプログラムでは同じ時間帯のオーガナイズド/近未来チャレンジセッションとの著者の重複が 10 件、同じ時間帯の一般口頭発表との著者の重複が 20 件の合計 30 件存在した。特に括弧内の数値（4 件）はそのうち発表順、時刻まで一致しているケースを示し、共著者は両方の発表に参加することが全く不可能となっている。

次に評価値を比較する。セッション配置における評価値は、同一時間帯に含まれるセッション間の関連度の平均値を示し、低いほうがより優れたセッション配置であると言える。実際のプログラムでは 1.22 となっているのに対し、本システムの所属制約無しの評価値が 0.60、所属制約有りが 0.74 と、実際のプログラムの値は大幅に高くなっている。これは比較的関連度の高い発表が同一の時間帯に存在することを示す。実際のプログラムでは大会参加者が興味があり聴講したいと思う発表を、全て網羅することが困難となっている可能性がある。

7. おわりに

本研究では、大規模な学術会議に対する会議プログラム自動作成ツールを開発した。また、本ツールを JSAI2009 の大会

プログラム作成に適用した結果について報告した。今回は残念ながら実際の会議プログラムとして採用されなかったが、開発した会議プログラム自動作成ツールにより、評価値ベースでは実際のものより優れたプログラムを作成することが出来た。

今後の課題としては、まず今回作成した JSAI2009 プログラムに関して行なった比較・考察の結果と、実際の大会参加者の感想や意識とが一致しているかの考察が挙げられる。また、評価関数の精度の向上のため、論文キーワードの関連性の詳細化若しくは別の評価基準を探ることや、現実的な時間でより大域的最適解に近い解を得ることの出来るアルゴリズムの開発等が挙げられる。

参考文献

- [Easton 01] Easton, K., Nemhauser, G., and Trick, M.: The traveling tournament problem description and benchmarks, in *Proceedings of Seventh International Conference on the Principles and Practice of Constraint Programming (CP'01)*, Vol. 2239, pp. 580–584 (2001)
- [Kaneko 99] Kaneko, K., Yoshikawa, M., and Nakakuki, Y.: Improving a Heuristic Repair Method for Large-Scale School Timetabling Problems, in *Proceedings of the 5th International Conference on the Principles and Practice of Constraint Programming (CP'99)*, Vol. 1713, pp. 275–288 (1999)
- [Small 73] Small, H.: Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents, *Journal of the American Society for Information Science*, Vol. 24, No. 4, pp. 28–31 (1973)
- [Williams 99] Williams, H. P.: *Model Buildings in Mathematical Programming*, Wiley-InterScience (1999)
- [小作 02] 小作 浩美, 内山 将夫, 村田 真樹, 内元 清貴, 井佐原 均: 自然言語処理技術を用いた大会プログラム作成支援について, *自然言語処理*, Vol. 9, No. 5, pp. 131–148 (2002)
- [松井 99] 松井 知己: スポーツのスケジューリング, *オペレーションズ・リサーチ*, Vol. 44, pp. 141–146 (1999)
- [松尾 05] 松尾 豊, 友部 博教, 橋田 浩一, 中島 秀之, 石塚 満: Web 上の情報から人間関係ネットワークの抽出, *人工知能学会論文誌*, Vol. 20, pp. 46–56 (2005)
- [杉本 96] 杉本 雅則, 小山 照夫, 堀 浩一, 大須賀 節雄, 絹川 博之, 間瀬 久雄: 文書間の関連性を可視化することによる文献検索システム, *情報処理学会研究報告. 自然言語処理研究会報告*, Vol. 96, No. 27, pp. 12–22 (1996)
- [西村 04] 西村 拓一, 濱崎 雅弘, 松尾 豊, 大向 一輝, 友部 博教, 武田 英明: 2003 年度人工知能学会全国大会支援統合システム, *人工知能学会誌*, Vol. 19, pp. 43–51 (2004)
- [西村 08] 西村 直史, 徳永 亮, 大田 直樹, 櫻井 祐子, 岩崎 敦, 横尾 真: 人工知能学会全国大会プログラム自動作成ツールの開発 (2008)
- [鈴木 07] 鈴木 雅人: 公開情報を用いた類似論文検索精度の向上 (2007)