

# オンライン調停教育支援システムの類似場面検索機能

## A Similar Disputation Scene Search on An Online Mediation Support System for Education

田中 貴紘      片上 大輔      新田 克己  
Takahiro Tanaka      Daisuke Katagami      Katsumi Nitta

東京工業大学大学院 知能システム科学専攻  
Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology

### 1. はじめに

近年インターネットでのオンライン取引に関する、多様なトラブルが生じてきている。特にネット通販やネットオークションなどにおけるトラブルは増加傾向にある。裁判による紛争解決方法はコストと時間がかかるため、このようなトラブルにおいて効果的な救済策を提供することは難しい。そこで迅速かつ低コストな ADR(Alternative Disputation Resolution)と呼ばれる調停や仲裁が注目されている。

ネットオークション等のオンライン取引で生じたトラブルの場合、当事者同士が遠隔地に住んでいることも多く、直接会って話し合うということが時間的・コスト的にも困難である。そのため、遠隔地同士を繋げるオンライン ADR が実験的に始まり、現在幾つかの ADR 機関によってメールを利用した調停が行われている[ECOM, Cyberpol]。しかし、何度もメールでのやり取りを行わなければならない、解決までに数ヶ月を要することもある。また、質の高い調停を行うために、調停を行う調停者の教育も必要となる。

そこで、われわれはオンライン調停教育支援システムの開発を目的とする研究を行っている[田中 03]。このシステムはオンライン論争環境を提供するだけでなく、事例ベースを利用して調停者の支援や教育を行う。システムによって蓄積されたデータを解析し教育に利用する、新たな教育モデルの提案をも含んでいる。本論文では、本システムの支援機能の一つである、類似場面検索の評価を、模擬調停記録を用いて行った。

本論文では、まず、第2章で ADR について述べ、3章でオンライン調停教育支援システムを紹介する。次に、4章は、今回行った模擬調停事例を用いたシステム評価実験について述べる。最後に、5章でまとめと今後の課題を述べる。

### 2. ADR について

#### 2.1. ADR とは

判決による紛争解決以外の紛争解決手段、例えば仲裁・調停などのことをまとめて ADR と呼ぶ[ADR 01]。ADR は裁判と比較して、迅速かつ低コストで紛争解決が可能であ

り、近年注目されている。調停は裁判の判決によらず、第三者の立会いのもと、当事者の話し合いで紛争を解決するものである。調停は、一般的なディベートや裁判と異なり、当事者同士が納得できれば、明らかになっていない事実や、白黒ははっきりしていない事実があったとしても、合意として終了する。また、調停では、当事者同士が言い合いをするのとは異なり、中立な第三者の立場である調停者に言い分を聞いてもらうというスタイルで進行する。調停の進行は、1.当事者双方から言い分を聞き、合意・不合意を確認・整理し、2.不合意点について、理由などを聞き話し合いを行う。最後に、3.調停者が調停案と呼ばれる妥協案を提案し、当事者がその案に合意できれば調停を終了する。

#### 2.2. ADR と教育

従来の法学教育では、解釈の暗記といった条文中心の教育であり、一部で、ゼミなどで模擬裁判・調停を行っている。そのため、組織的に模擬調停を行い、データの収集・解析を行っているところはない。調停を行う調停者の教育方法については、経済産業省を中心に議論されている[ADRT]。教育方法は、主にビデオによる学習や模擬事例を用いた学習が提案されている。アメリカでは、主に紛争解決の分野として、ロースクールにコースが設置されているが、わが国では教育体制は十分ではない。

#### 2.3. オンライン ADR の現状

ADR のニーズの増加を受け、オンライン ADR (ODR) が国内外で実験的に行われ始めている。一般的な ODR のシステムは、相談をメールで受け付け、当事者と調停者がメールのやり取りによって調停を行うものである。このようなシステムでは、解決までに何度もメールでのやり取りが必要となる。ある調停機関の実験では、解決までに平均 50 通前後のメールでのやり取りを必要とし、解決までに数ヶ月掛かっている。また、アメリカ政府の労働紛争を扱う機関には、実際に紛争解決にオンラインシステムを利用しているものがある[TAGS]。このシステムは、ブレインストーミングや Web チャットなどのツールを組み合わせただけのものであり、オンラインの論争環境としては、特に支援を行っておらず不十分と言える。

### 3. オンライン調停教育支援システム

本研究で開発しているシステムは、オンライン調停環境

の提供と、それに加えて「調停者の支援や教育」までを行うことを目的としている。本研究では、単に用意された調停問題を解くだけでなく、調停記録を解析・比較することによって、調停のスキルを身に付けさせるという、模擬事例を利用した教育モデルを提案している。事例を用いた教育の流れは以下の通りである。

- あらかじめ題材を教師が選定する
- 同一題材を元に、異なる学生が模擬調停を行い、調停記録を蓄積する
- 他の調停と比較することで、個々の調停を評価し、指導する
- 調停記録に対して、後で教師がコメントを付加し、学生が後からそれを参考にする
- エージェントが調停司会を代行し、調停内容を評価する

よって、本研究では、調停記録を事例ベースに蓄積し、事例ベースを利用したシステムを開発することで、模擬調停事例を利用した教育を実現する。さらに、その事例ベースを利用する調停者エージェントを開発し、調停者の司会の代行や、行われている論争内容を評価するといった、調停者支援を目標としている。

本章では、まずオンライン調停教育支援システムの概要と論争インタフェースについて述べる。そして、本システムによる事例ベースの構築方法とその利用方法について述べる。

### 3.1. オンライン調停教育支援システムの概要と構成

図1にオンライン調停教育支援システムの概要を示す。オンライン調停は、参加者が論争インタフェースを用いて、インターネットを介しサーバに接続することで行われる。調停は、調停の司会進行を行う調停者と、トラブルを抱えた当事者2人の3名で行われる。

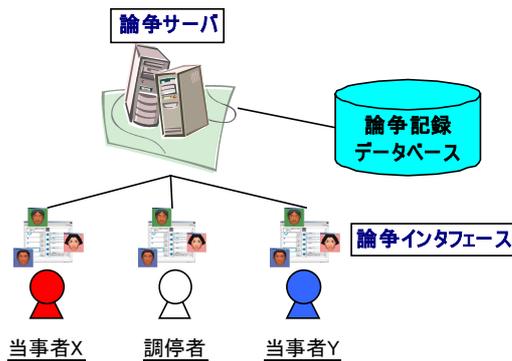


図1 システムの概要図

本システムの主な機能は次の2つである。

- オンライン論争環境の提供： オンラインで論争内容を交換し、論争状況（合意、決裂、未決着）を表示する。論争内容のダイアグラムによる表示を行い、その構造にしたがって論争グラフに変換する。アバタによる表情付けを行う。

- 事例ベースの構築： 論争の記録をXML文書として出力する。発言内容が含む話題（論点）を抽出し、ダイアグラム情報とともに記録に付加する。また、構築した事例ベースから、類似する事例・場面を検索する。

この機能を実現するための、サーバおよびクライアントの構成を図2に示す。図2の上半分はサーバの構成、下半分はクライアントの構成である。サーバは、オンライン調停を行う部屋全体と各参加者の管理を行う。発言は、サーバが参加者からの発言情報を受け取って、論争参加者全員にその情報を通知する。発言内容は、サーバにより文章に含まれる論点が認識され、XMLタグを用いてインデキシングされる。調停記録は、必要に応じて事例ベースへの格納、および、事例ベースからの読み込みが行われる。

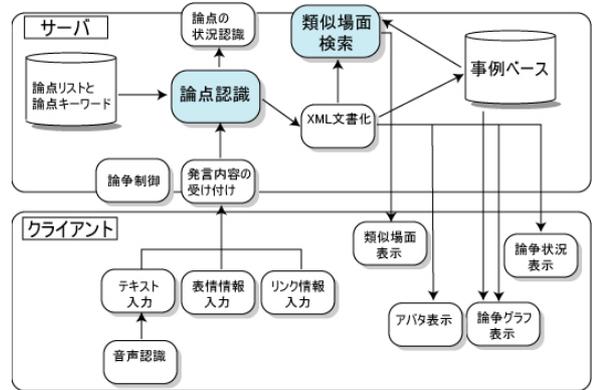


図2 システムの構成

### 3.2. 論争インタフェース

われわれの想定している模擬調停は、図3に示すような、論争インタフェースを用いて行われる。インタフェースはFlashで実装されており、一般的なWebブラウザ上で動作可能である。



図3 論争インタフェース

#### §1 ユーザによる入力

ユーザは、図3右下のサブウィンドウによって発言を行う。発言は、音声入力かキーボードによる直接入力で行い、その際、「リンクデータ」、「アバタの表情」を入力する。

リンクデータとは、「現在の発言と以前の発言との関係やタイプ」を表すデータである。リンクデータは、ある発言が「対象とする発言の ID」と「発言のタイプ（主張、合意、否定、補足、質問、回答、要求、拒否）」から成る。アバタの表情には Cool, Happy, Angry, Sad, Surprise の 5 種類がある。

法学部で実験されている多くの遠隔地模擬裁判では、TV カメラを使用して参加者の表情を伝送している。TV カメラは裁判の臨場感を保持するための有用なツールである。しかしながら、本研究ではあえて TV カメラを使う代わりにアバタを使用する。その理由は、オンライン調停の場合には、匿名性の保持が必要となることがあるからである。また将来的に、調停者をエージェントによって代行することを目標としているため、人間と入れ替わっても気付かれないようアバタを使用している。アバタを用いた場合でも、ノンバーバルな情報を相手に伝えることができ、論証や交渉を円滑に進めることができる[湯浅 03]。

## §2 システムの出力

システムは発言内容をテキストで表示し（図 3 左）、アバタが表情を変化させながら、音声による読み上げを行う。また、システムは論争状況をモニタし、リンクデータから発言を合意・決裂・未決着に振り分け表示する。さらに、論争の経緯を論争グラフとして視覚化してユーザに表示することができ、現在の論争状況と類似する類似場面検索機能を持つ。論争グラフと類似場面検索については次節で述べる。

### 3.3. 事例ベースの構築と利用

#### §1 取り扱う事例データの特徴

本研究で取り扱う事例データは、調停における論争内容を表す自然言語のテキストである（図 4）。テキスト自体は、口語で書かれていたり、指示語や省略、ノイズを含んでいたりする。

調停者の支援を行うためには、現在扱っているトラブルがどのような問題で、話し合うべきことは何かを判断できる必要がある。そこで調停記録を整理・分析し、トラブルが持つ特徴を抽出する。この特徴はトラブルの原因や事実、要求などであり、本研究ではこれらを論点と呼ぶ。論争テーマは、あらかじめ教材として用意される模擬調停事件であるので、教師が論点を列挙することができる。さらに紛争のタイプ・カテゴリごとに論点をまとめ、論点リストの作成を行う（これは Hypo の dimension と類似の概念である[Ashley 91]）。Hypo が裁判の結果である判例に出現する論点を対象としているのに対し、本研究では、論争推移における論点の出現順も対象としている点で異なる。

図 5 に模擬調停実験の論点リストの例を示す。各論点には論点 id が振られ、どちらに有利・不利かの情報が教師等の人手で付加されている。また、[f11]や[f14]は[f1]の具体例であり、このように論点間に階層関係が生じるものがある。

**発言者 = 山本**  
 市川さんにお聞きします。松田さんはノークレーム、ノーリターンとのこと、また事前に質問がなかったのはおかしいのではないかとのことですが、その点についてどうでしょうか？

**発言者 = 市川**  
 ノークレーム、ノーリターンについてですが、メーカーのホームページを参照して下さいとの記載があったので、こちらとしては、そのホームページ上の情報を得た上でノークレームノーリターンと考えております。実際に、この説明に偽りがないのでは構わないのですが、こちらで得られた情報と、実際の商品とを照らし合わせると疑問となる点が出てくるため、この場合、ノークレーム・ノーリターンに応じることが出来ない。

**発言者 = 山本**  
 事前にホームページなどを参照しておかしいと思ったことについて質問などはなかったのでしょうか？

図 4 調停における論争例

**X:Xに有利 Y:Yに有利**

- [f1]送られてきた商品に問題がある
- [f11] 広告と異なる商品が送られてきた [Y]
- [f14] 正規品ではなく特注品である [Y]
- [f2]説明に問題がある
- [f26] 説明画像が不鮮明 [Y]
- [f27] 説明は「詳細はメーカーホームページを参照」のみである [Y]
- [f3]ノークレームノーリターンの宣言をした
- [f35] ノークレームノーリターンと記載した [X]
- [f36] 1ヶ月以上経ってクレームを付けた [X]

図 5 模擬調停実験における論点リストの一部例

#### §2 事例のインデキシング

事例データの特徴を利用して、調停の各発言にインデキシングを行う。これは、各発言を 1 単位とし、リンクデータ用いてその発言のインデキシングを行う方法である。この方法は比較的容易なため、発言者が模擬調停の最中にインデキシングをすることが可能である。

#### §3 論点の自動認識

調停記録の全ての発言に、人手で論点を付加することは、システム管理者（または教師）にとって非常に負担が大きく、事例ベース構築のコストが大きくなる。そのため、論点を自動で認識し、付加する必要がある。

論点の認識は、まず自然文であるテキストを茶筌[Chasen]を用いて形態素解析を行う。次に、論点ごとに設定したキーワードを含むかどうかで論点を判定する。表 1 は論点を認識するための論点キーワードの例である。論点キーワードは、与えられた問題設定から、各論点に関する発言内に出現するであろう名詞を、キーワードとしてあらかじめ教師等の人間が人手で設定する。認識精度は論点キーワードに依存すると考えられるが、現在は人手で設定しているため、揺らぎが存在する。揺らぎがどの程度存在するかを検証は 4 章で行う。調停記録が蓄積される前にキーワードを設定するため、人手で設定を行って

るが、今後調停記録が蓄積されれば、TF/IDF 等の統計的手法によってキーワード設定を行うことも考えている。

表 1 論点キーワード例

論点 id	論点キーワード
f14	特注, 特注品, 試作, カスタマイズ
f16	画像, 塗装, 不鮮明
f17	参照, ホームページ, HP
f35	ノークレーム, ノーリターン

#### § 4 事例の記述法

論争終了後、システムは論争のログを XML 文章として出力する。システムが出力する調停記録の例を図 6 に示す。

1 回の発言内容には、発言 ID、発言者名、リンクデータ、表情値と入力されたそのままの自然言語が記述される。また、発言内容から認識した論点の ID や、発言時に明らかになっている事実データなども記述される。論争途中で、発言内容の論理構造を入力することはユーザにとって負担となるため、論理構造タグは論争終了後、必要に応じて人手で付加する。

```

<発言 id="11" 発言者="Yamamoto(1)" タイプ="質問" 表情="cool">
<論点>f35</論点>
<原文>市川さんにお聞きします。松田さんはノークレーム、ノーリターンとのこと、また事前に質問がなかったのはおかしいのではないかとのですが、その点についてどうでしょうか？</原文>
</発言>
<発言 id="12" 発言者="Ichikawa(2)" タイプ="回答" 表情="angry">
<論点>f35,f27,f28</論点>
<原文>ノークレーム、ノーリターンについてですが、メーカーのホームページを参照して下さいとの
  
```

図 6 XML 文書化された調停記録

#### § 5 論争グラフの表示

前述のように各発言を論点とリンク構造でインデキシングすることにより、調停記録は図 7 で示すような、論争グラフとして表現することができる。

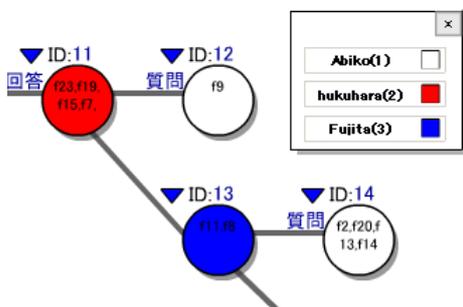


図 7 論争グラフの例

グラフは、発言者ごとに各発言が色分けされ、ID が発言順を表している。また、発言間の関係（質問や回答）や、発言内容が含む論点の ID が表示されている。調停記録の解析は、調停記録を論争グラフへと抽象化して、グラフ同士を比較することで行う。グラフ化することで、話題の推移の認識や類似場面の検索が容易となる。調停記録は Web

上で閲覧することができ、学習者は調停記録から論争グラフに適時変換することで、論争状況の把握が容易となる。

#### § 6 類似場面の検索

過去の調停記録のなかから、現在の論争状況に似た場面を検索する。検索は、事例データに付加されている論点情報やリンクデータを元に行われる。次節で類似場面検索の詳細について述べる。

模擬調停の途中で、事例ベースから過去の類似場面の検索をすることができれば、学生・調停者は、次にすべき発言の参考にすることができる。また、人間の調停者の代わりにエージェントを利用する際に、エージェントの発言内容として類似場面から発言を引用することができる。本検索機能は、現在は調停に参加する全ユーザが使用可能な機能となっている。当事者、調停者ともに検索結果を参考にすることができる。

#### 3.4. 類似場面検索

類似検索は、キーワード検索によって同じキーワードを持つ発言を検索するのではなく、幾つかの発言の「流れ」を比較し、類似する流れを持つ場面を検索するものである。

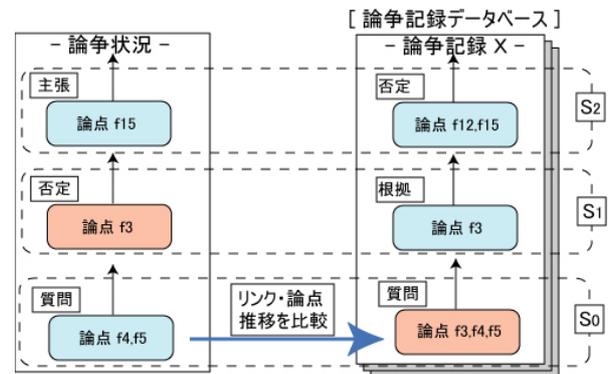


図 8 類似場面検索

同一の論点を含む事例を類似事例とし、過去の類似場面を検索する（図 8）。ユーザが検索対象とする発言 ID をクエリとして入力すると、システムはまず、発言内容に含まれるのと同種類の論点が出現した調停記録の中から、同じ論点を含む発言を検索する。次に、該当する発言が見つかったら、その発言のリンクデータを元に、幾つか過去の発言までリンクをさかのぼり、論点の推移を比較し類似点を計算する。そして類似点でランク付けた類似場面リストを作成する。

2 つの発言間の類似点は次のように計算される。

- **論点の類似点:** 2 つの発言が共有する論点数から、(a)により発言の持つ論点の類似点  $S_a$  を計算する。

$$(a) S_a = \frac{\text{共有する論点数} \times 2}{X\text{の論点数} + Y\text{の論点数}}$$

- **発言タイプの類似点:** 2 つの発言のタイプから、(b)により発言タイプの類似点を計算する。発言タイプは、主張、合意、否定、補足、質問、回答、要求、拒否の 8 種類であり、この発言タイプが一致すると、+0.5 ポイント、一致しない場合は 0 ポイントとする。発言グループとは、論争で使用される主な場面によって、発言タイプを 3 つのグループに分類したものである。相手 X の発言に対して Y が発言するタイプ「X←Y」には、

「合意・否定・拒否・質問・回答」, 自分の発言に対して発言するタイプ「X←X」には, 「根拠・理由・補足」, そして単独で存在する場合は多い発言タイプ「X」には, 「主張・要求」が属する. 発言グループが一致する場合は +0.5 ポイント, 「X」と「X←Y」または「X←X」の場合は +0.25 ポイント, 全く一致しない場合は 0 ポイントとする.

(b)  $S_b$  = 発言タイプの一致 + 発言グループの一致

- **発言者の役割の類似点:** 発言者が, 中立な立場である調停者か, 紛争当事者かの違いにより類似点を計算する. 発言者が共に調停者または当事者であった場合は +1 ポイント, 一致しなかった場合は 0 ポイントとする.

(c)  $S_c$  = 発言者の役割の一致

以上から, 2つの発言間の類似点  $S_i$  は, (d)のように計算される.  $S_0$  は, クエリとした発言と, その発言の持つ論点と同じ論点を持つ発言との類似度を示す.  $S_1$  は, それらの発言から1つリンクを遡った発言同士の類似度である.  $S_2$  は同様に2つ遡った発言同士の類似度である.

$$(d) S_i = Sa_i + Sb_i + Sc_i$$

そして, 場面間の類似点  $R_s$  を(e)によって計算する. size とは, リンクを辿って過去の発言を幾つまで判定に含めるかの値である. size = 3 であれば, 過去3つ分の発言が類似判定に含まれる. この場面間の類似点を元に, 類似場面検索を行う.

$$(e) R_s = \sum_{i=0}^{size-i} \frac{(size-i) \times S_i}{size} \quad (0 \leq i < size)$$

システムは, 検索された類似場面の最後の発言にリンクを張っている発言を, 回答候補としてユーザに提示する. そのため, 回答候補の数は, 検索結果よりも多くなる場合がある.



図 9 類似場面検索機能

類似場面検索は, 参加者がある発言に対する返答候補の検索を行ったり, 発見された類似場面をエージェントが

司会進行に組み込んだりすることに利用される. 図 9 は実際に類似場面検索を行っている例である. まず, 検索対象となる発言の ID を入力し検索を開始する. 検索結果が類似点の高い順に, リスト表示される (図 9 左上). リストから候補を選択すると, 類似判定の詳細情報が表示される (図 9 中央). ここには, 検索された発言の発言タイプ, 論点, 内容が表示され, 各発言間の類似点も表示される. 回答候補は図 9 右上に示す通り, 発言タイプと発言内容が表示される.

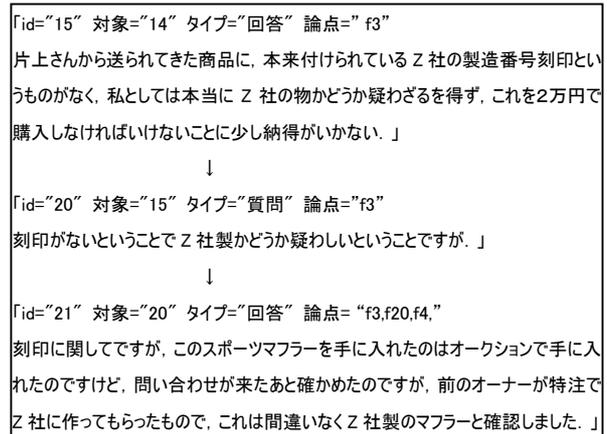


図 10 現在の論争状況の一部例

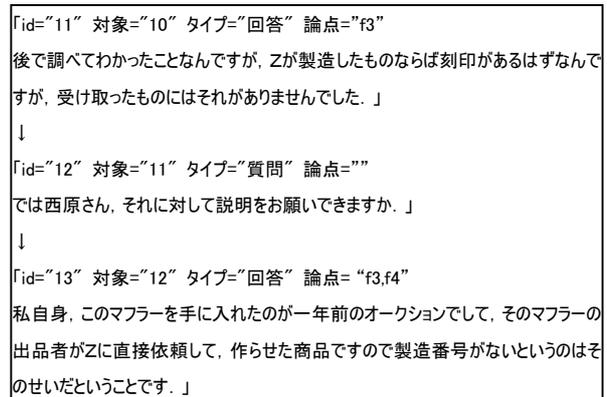


図 11 類似場面検索結果の一部例

次に, 実際の検索結果の例を示す. 図 10 は現在の論争状況の一部である. 発言 id=5 の発言を検索のクエリとして, 類似場面検索を行い, 得られた結果の一つが図 11 である. 幾つか認識された論点の違いはあるものの, その場面推移は類似していることがわかる. この場合に回答候補として提示されたものが図 12 である. この回答候補は現在の状況に適合しており, 再利用できることがわかる.

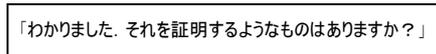


図 12 回答候補の一部例

#### 4. 類似場面検索の評価

本システムの支援機能の一つである類似場面検索は, 論争状況と類似した場面を事例ベースから検索し, ユーザに

回答候補を提示する。まず、模擬調停実験で収集した調停事例を用いて、論点自動認識の評価実験を行った。その後、以下の2つを目的とした、本システムの類似場面検索の評価を行った。

- 検索される場面の類似性は高いか
- 回答候補が実際の回答の参考となるか

本章では、まず本実験で使用する模擬調停記録について説明する。次に論点自動認識評価実験の実験方法と結果について述べる。次に類似場面検索の評価実験方法と実験結果について述べる。

#### 4.1. 実験で使用する模擬調停記録について

使用する模擬調停記録は、模擬調停実験により収集した調停記録を使用する。模擬調停実験は、ネットオークションで生じたトラブルの事例を用意し、調停者・落札者・出品者それぞれの役を被験者が担当し、調停を行った。用意した事例は、オークションで商品説明が不十分だったため、落札者から「商品に問題がある」としてクレームがついたというケースである。実験では、2種類の情報を被験者に与え調停を行った。一つは、調停を行う前に明らかになっている共通情報である。二つ目は、当事者どちらか一方しか知らない秘密情報である。調停では出品者と落札者の双方に弱みがあるが、それは相手に対して秘密になっている。調停の途中で、相手の秘密事項をうまく聞き出した側が有利になるという設定である。実験は、大学生33名を3人1組とし、11回の模擬調停を行った。論点総数は17個で、発言総数は754個、1事例あたりの発言数は平均69個であった。

この問題設定における論点リストの一部を図5に示す。被験者は、論点に関しては一切知らされていない。また、論争過程において秘密情報に属する論点を聞き出し、その論点を用いて他の論点を攻撃/補強する。つまり、論争で有利となる秘密論点を知り、有効活用することができれば、調停を自分に有利に進めることができる。

#### 4.2. 論点自動認識実験

調停記録には、調停者の司会や話題の推移によって、論争内に出現した論点の種類に偏りが生じる。当事者の一方のみが知る秘密論点に関しては、この傾向が特に強い。論点自動認識実験では、共通・秘密論点を共に含む事例の認識精度の評価を考え、模擬調停実験により収集した事例のうち、共通・秘密論点が共に出現し、17種類中15種類以上の論点を含む5事例(368発言)によって実験を行った。代名詞や省略内容からのみ判断できる論点は省く。評価は、人間が見て付加した論点と自動認識により付加した論点を比較し、適合率と再現率を計算した。また、論点キーワード設定による認識精度の揺らぎについても検証するため、3人の被験者が問題設定のみから作成した論点キーワードをそれぞれ用いて実験を行った。

表2に論点自動認識の評価実験結果を示す。有効対象発言数は151個であった。適合率は平均で52.6%であり、1つの発言に付加される論点の半数程度を抽出することが

できた。また、その半数に必須となる論点の6割以上が含まれていることになる。この程度の精度でも、論点を自動で付加することにより、「全く論点が付加されていない記録に人手で論点を付加する負担」から「付加された論点を必要に応じて修正する負担」へと、システム管理者が事例ベースを構築する負担を軽減することができる。適合率が低いことは、正しく論点が付加されている発言もあるが、発言内容に含まれると考えられる論点よりも、多くの論点を付加してしてしまった発言が多いとも言える。キーワードマッチングによる論点認識の場合、多義語として予期しない語が使われていたり、設定した論点キーワードが論点と無関係に使用されていたりすることが、論点認識精度が低下する原因となる。

表2 論点自動認識評価実験結果

	平均	被験者A	被験者B	被験者C
適合率	52.6%	55.9%	50.3%	51.5%
再現率	63.7%	66.7%	64.6%	59.8%

キーワード設定による精度差は、適合率で最大3.4%、再現率で4.0%となり、キーワード設定における認識精度の揺らぎは少ないと考えられる。揺らぎが少ない理由としては、問題設定からのキーワード作成は、与えられる情報が限られ、候補となる名詞が限られるためと考えられる。次節の類似場面検索実験では、被験者Aの論点キーワードを使用して、論点の自動認識を行った。

#### 4.3. 類似場面検索実験

##### §1 類似場面検索実験目的

本研究の類似場面検索では、複数発言のリンクを1場面として類似判定の対象としている。幾つの発言を1場面と考えるかが問題であるが、1, 2発言のリンクでは、相槌や訴訟指揮(指示など)が含まれる可能性が高いため論争推移を判断するには短く、3発言以上が望ましいと考えられる。また、論争の推移が、「主張→確認→反論」や「質問→指示→回答」といった、3発言を基本とした構造が多く観察された。よって本節では、類似場面検索の比較対象とする発言リンク数を3発言とした場合の検索精度の評価実験について述べる。

##### §2 類似場面検索実験方法

模擬調停実験により収集した全11事例を用いて類似場面検索を行った。各事例には、自動認識された論点が付加されている。今回は付加された論点の人手による修正やリンクデータの誤入力の修正を行っていない。実験方法は以下の手順で行った。

1. ある論点について最初に述べられている場面で、人間が見て、自然文として類似している場面を11事例の中から探し、類似場面集合Xを作成する：模擬調停実験の解析結果を参考に、A・Bの2種類の場面タイプについて検索を行った。場面タイプAは、比較的

どの事例にも出現する論点である共通論点が述べられている場面であり、場面タイプ B は、一部の事例に出現する論点である秘密論点が述べられている場面とした。検索対象とする論点は、論点リストの 3 つの論点カテゴリから、任意の論点をタイプ A・B それぞれ 2 つずつピックアップして決定し、4 つの類似場面集合を作成した。各類似場面集合は、教師の立場である 1 人の人間により選択された。場面タイプ A には、11 場面から成る集合と 4 場面から成る集合を含む、合計 15 場面が含まれ、場面タイプ B には、7 場面から成る集合と 6 場面からなる集合の、合計 13 場面が含まれる。事例によっては、1 事例中に 2 場面含むものも存在した。

2. 集合 X から 1 場面をピックアップして現在の論争状況とする
3. ピックアップした場面を検索のクエリとし、集合 X の他の場面を検索できるか類似場面検索を行う。類似点が閾値以上のものを選んで、再現率・適合率を計算: 1 つの場面をクエリとした場合、検索対象は、場面タイプ A では 10 場面と 3 場面、場面タイプ B では 6 場面と 5 場面となる。また類似点 Rs は、3 発言比較の場合 6.0 が最大値であるので、閾値を 3.0 とした。検索結果が正答かどうかの判断には、リンクデータや論点付加のノイズも考慮し、正答場面のサブセットも類似場面としている。例えば、発言 id8~10 から成る場面を正答とする。この場合、検索結果として、id8~10 を類似場面とするものとは別に、id7~9 や id9~11 を類似場面とする。本実験において類似しているかどうかの判断は、人間が自然文を読み、主観的に似ている場面を正解としている。この点に関しては、特に法律知識が必要ではなく、素人・専門家によって違いが生じることは少ないと考えられる。
4. 2 をピックアップする場面を変えて、集合 X の全場面に関して検索を繰り返す: 検索は、当事者・調停者どちらかの状況に特に限定して行っていない。本実験では、4 つの類似場面集合ごとに、上記 1~4 を繰り返して実験を行った。

場面タイプ A は、共通論点、つまり調停参加者全員が開始当初から知っている事実についてである。そのため、回答候補を得ること以外に、検索結果を論争推移の参考にしたり、事実間の有利・不利の関係を調べたりといった利用が考えられる。場面タイプ B の秘密論点は、当事者のみしか知らない事実である。よって、当事者以外の参加者は、その事実があることを知らず、調停開始時から反論を考えたり、論争推移の予測に組み込んだりすることはできない。そのため、その事実が述べられた際に、どう回答するか、どう調停を指揮するかを参考にするため検索を行うことが考えられる。

検索精度の評価は、適合率と再現率、そして利用可能な

回答候補が見つかるかを測るため、再利用率を設定し評価する。再利用率とは、提示される総回答数のうち、人間が見て利用できると判断した回答数の割合である。発言を再利用する場合、以下の段階が考えられる。

1. 発言内容をそのまま再利用できる
2. 主語の変換や接続詞の調整を行えば再利用できる
3. 発言内容の一部の文を再利用できる

今回の実験では、以上の 1, 2 を満たす回答候補を再利用可能な回答と判断している。

### §3 類似場面検索実験結果

表 3 に類似場面検索の評価実験結果を示す。ヒット数とは、システムが検索した場面数の総数である。正答数とは、検出した場面のうち、選択した類似場面集合に含まれた数である。出現数とは、選択した類似場面が、検索結果に出現した数である。また、回答候補数とは、システムがユーザに提示した回答候補の総数であり、有効回答数は、提示された回答候補のうち、利用できると判断された数である。

表 3 類似場面検索結果 (3 発言)

	場面タイプ A 平均 (11+4 場面)	場面タイプ B 平均 (7+6 場面)	全平均
ヒット数	13.4 個	6.3 個	9.8 個
正答数	5.9 個	3.1 個	4.5 個
出現数	3.6 個	2.3 個	3.0 個
適合率	52.6%	50.4%	51.5%
再現率	48.0%	41.4%	44.7%
回答候補数	18.6 個	9.0 個	13.8 個
有効回答数	6.8 個	6.0 個	6.4 個
再利用率	42.8%	66.3%	54.6%

適合率と再現率の計算方法は以下の通りである。

- ・ 適合率 =  $\frac{\text{正答数}}{\text{ヒット数}}$
- ・ 再現率(場面タイプ A) =  $\frac{\text{出現数}}{10}, \frac{\text{出現数}}{3}$
- ・ 再現率(場面タイプ B) =  $\frac{\text{出現数}}{6}, \frac{\text{出現数}}{5}$

### §4 類似場面検索の精度評価

表 3 より、場面タイプ A の場面検索の方が、場面タイプ B よりも高い適合率・再現率を示した。場面検索では、検索対象が、複数の発言から成る「場面」であるため、正答のサブセットを検出することが多く、ヒット数が増加することで精度が低下している。これは両場面検索において言えるが、場面タイプ B は A よりもサブセットが検索されにくかったため、ヒット数が A よりも少なくなっている。場面タイプ B のサブセットが少なくなる理由として、場面タイプ B が秘密論点で構成されているため、限定的な状況において出現することから、その場面の前後に

ノイズが少なく、また同じ論点が場面の前後に出現することが少ないからと考えられる。そのため、重要な論点が正しく付加されていない場合は、逆に検出することが難しくなる。今回の実験では、自動認識された論点の修正を行っておらず、論点の自動認識の精度が低いため、発言内容に適さない論点が付加されていることがあり、これもまた、類似場面検索の精度を低下させる原因となっている。また、発言のリンクデータは発言者によって入力されているため、誤ったデータが入力されている場合がある。この入力誤りが同様に精度を低下させる原因となっている。

類似場面検索の精度は、平均適合率で 51.5%、平均再現率で 44.7% という結果から、事例ベース内の類似場面の半数以上を検索することができ、また、検索結果の約半数は適切な類似場面である。よって、本機能により過去の調停記録から類似場面が検索できることがわかる。類似場面検索の精度を向上させるための方法として、類似判定対象の場面の修正が考えられる。論争推移の中には、相槌（「なるほど」「わかりました」）やオウム返し（「～ということですね」）、調停者の指示（「次に～さん述べてください」等）が含まれることが多い。これらの発言には論点が含まれていることが少なく、論争推移には影響しない。現段階の類似判定では、これらの発言が対象に含まれていると論争推移が捉えきれない場合がある。

#### §5 回答候補の再利用性

表 3 で示すように、平均回答候補数が 13 個であり、再利用率が 54.6% ということから、システムによって提示される回答候補のうち、半数以上をユーザが参考にすることができる。この結果から、本システムの類似場面検索機能によるユーザ支援が有効であることがわかる。

場面タイプ B の方が、A よりも提示される回答数が少ないにも関わらず再利用率が高いのは、これら論点を含む発言がなされたときに、これを補強または攻撃する論点が限られてくるために、回答候補の再利用性が高くなると考えられる。

今後調停エージェントによる調停司会の代行を考えた場合、回答候補をそのままエージェントの発言として流用するには、精度が低いと言える。回答候補の中から、論争状況に適合し、かつ論争を良い方向へ推移させる発言を選択したり、また当事者の感情を考慮し、宥める発言を選択したりするなどの機能が必要となる。また、半面教師としての教材になる発言の検索など、教育に有効な検索機能の拡張も今後検討していく予定である。

#### §7 類似場面検索の特性

本研究の類似場面検索がどのような場面の検索に向き・不向きがあるのかを述べる。

本研究の類似場面検索によって検索しやすい場面とは、対象となる場面に全て論点が付加されており、なおかつ出現論点に変化がある場合である。つまり、「それまで続いてきた論争に新たな話題が出現した場面」の検索に向いている。これは、論点の出現を類似判定に利用し、論争推移

が論点間の関係に大きく影響を受けているためである。このような話題転換に、発言のタイプや論点の使われ方が類似している。また、そういった論争推移の場面では、発言内容が似ていることが多く、システムから提示される回答候補が利用できる可能性が高い。

逆に、論点が付加されず重要でない発言を多くリンク内に含む場面の検索には不向きであり、また、検索対象とする発言数を減らすと、類似場面の検索精度が低下する。

類似場面検索の精度向上には、論点認識精度の向上以外にも、発言間のリンク誤りを訂正する機能や、重要度の低い発言をあらかじめ省いたリンクを使って検索する機能などが必要と考えられる。また、本検索機能を調停者の支援に特化した場合、どのような機能拡張が有効であるかを検討していく必要がある。

## 5. おわりに

オンライン論争環境を実装し、オンライン調停教育支援システムの支援機能の一つである、類似場面検索の評価実験を行い、本機能によって過去の調停記録から類似場面が検索できることが示された。また、ユーザは本支援機能により、現在の論争状況に適した複数の利用可能な回答候補を得ることができることを示した。本機能によって、ユーザは、論争中に回答に行き詰まった場合には、類似場面検索によって回答候補を調べ、論争を続けることができる。

### ◇ 参考文献 ◇

- [ADR 01] 特集・ADR の現状と理論－基本法制定に向けて、ジュリスト、No.1207、2001.
- [ADRT] 経済産業省 経済産業研究所 ADR 試行プログラム ミング、2003
- [Ashley 91] Ashley, K.D: Reasoning with cases and hypotheticals in HYPO, Int. J. ManMachine Studies 34, pp753-796, 1991
- [Chasen] 形態素解析システム 茶筌,  
<http://chasen.aist-nara.ac.jp/hiki/ChaSen/> .
- [Cyberpol] シロガネ・サイバーポール,  
<http://www.scyberpol.org/>
- [ECOM] 電子商取引推進協議会(ECOM)、ネットショッピング紛争相談室、<http://www.ecom.jp/adr/>
- [TAGS] Technology Assisted Group Solutions System,  
<http://tags.fmcs.gov/>
- [田中 03] 田中、安村、新田: 仲介エージェントを用いた論争支援システム、知識ベースシステム研究会、59th, 信学技報 vol.102 No.603 pp25-30 , Jan.2003.
- [Toulmin 84] Stephan Toulmin: An Introduction To Reasoning Second Edition, Macmillan Publishing Company, 1984.
- [湯浅 02] 湯浅、安村、新田: 交渉エージェントのための表情抽出アーキテクチャ、知識ベースシステム研究会、58th, Nov.2002.