

音楽理論 GTTM に基づく 議論タイムスパン木の生成方式とその評価

Generating method and its evaluation of discussion time-span tree based on music theory GTTM

三浦 寛也*¹
Hiroya Miura

長尾 確*²
Katashi Nagao

平田 圭二*¹
Keiji Hirata

*¹ 公立はこだて未来大学
Future University Hakodate

*² 名古屋大学
Nagoya University

ディスカッションマイニングとは、会議における活動を複数メディアで記録し、そこから再利用可能な知識を抽出するための技術である。音楽理論とは、音の時系列を構文解析する技術である。本研究の目的は、音楽理論 GTTM の手法を用いて議論を「構文解析」することである。本稿では、議論タイムスパン木の生成方式について計算機上に実装する手法を提案し、プロトタイプシステムの有効性について検討する。

1. はじめに

ディスカッションマイニング[長尾 2003]とは、会議における活動を複数メディアで記録し、そこから再利用可能な知識を抽出するための技術である。音楽理論とは、音の時系列を構文解析する技術である。楽曲においては音イベントが、会議においては発言が時間の進行とともに発生しグループ(ゲシュタルト)を生成する点に着目すると、音楽理論の会議記録における時系列データの分析手法としての応用が考えられる。本研究の目的は、会議における議論を発言の時系列とみなし、音楽理論 GTTM(Generative Theory of Tonal Music) [Lerdahl 1983] の手法を用いて議論を「構文解析」することである。これにより、会議記録の各発言の重要度を階層的に表現する議論タイムスパン木の自動獲得を目指す。本稿では、議論タイムスパン木の生成方式について計算機上に実装する手法を提案し、プロトタイプシステムの有効性について検討する。

2. ディスカッションマイニングシステム

ディスカッションマイニング(DM)システム[土田 2010]では、会議記録から映像・音声情報やテキスト情報などの実世界情報を獲得し、それらから半自動的に構造化した会議コンテンツを作成することで、議論の内容を効率的に閲覧させる[Nagao 2004]。このシステムが支援している会議スタイルは、発表者がスライドを表示しながらその説明を行い、発表の途中あるいは終了のタイミングで会議参加者との質疑応答が始まるというものである。このシステムでは、会議参加者自ら議論の要素にタグ付けを行うことに関して若干のオーバーヘッドを強要するが、そのトレードオフとして、構造化された議論データが取得できる。

会議参加者は議論札と呼ばれる専用デバイスを用いることで、導入発言と継続発言の2つに発言タイプが分類される。先行する発言が無いものを導入発言と呼び、そうでないものを継続発言と呼ぶ。これを議論の構造化の主要な手がかりとしている。DM 木とは、このような各発言間の関係に基づいて構成された木構造である(図 2)。DM 木の根は導入発言である。ある1つの発言に対して、同時に複数の継続発言が付くと DM 木の分岐が増える。先行発言に継続発言が付く、さらにそれを先行発言として継続発言が付くと DM 木の枝が伸び、木が深くなる。

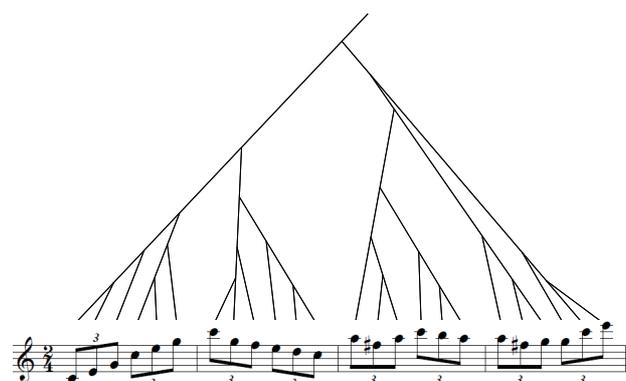


図 1. タイムスパン木

3. 音楽理論

楽曲を構文解析するとは、時間の進行に沿って生じる音イベント列をさまざまな時間長のレベルで分節し重要な音を発見することであり、その分節方法を記述したものが音楽理論である。中でも Fred Lerdahl と Ray Jackendoff により 1983 年に提案された Generative Theory of Tonal Music (GTTM) は現在最も正しい音楽理論の1つとされ[Lerdahl 1983]、音楽認知や音楽情報処理の多くの研究において参照され続けている。GTTM の分析では、隣り合った音イベントが生成する階層的なグループの情報と音イベントが生起している拍節位置の情報に基づいて、音イベントのグループ階層が生成するゲシュタルトを検出する。

GTTM の分析結果はタイムスパン木と呼ばれる木構造で表現される(図 1)。タイムスパン木は、時間的なまとまりを構成するもので、楽曲に含まれる音の相対的な重要度を表現する。また、GTTM において音楽を分析しタイムスパン木を生成するための規則は、グループ構造を獲得するための規則と、そのグループの中で重要な音はどれかを判定するための規則に大別される。例えば、発音時間や近い音どうしはグループを作りやすい、音価の大きい音は重要である等である。

4. 音楽理論の会議記録分析への応用

楽曲構造と会議構造を対比する。楽曲においては音イベント、会議においては発言がともに時間軸に沿ってゲシュタルトを生成する点に着目すると、音楽理論の会議記録分析への応用が考えられる。GTTM の楽曲分析のアプローチに基づき会議記録の分析への応用を行うと、発言の重要度を階層的に表現する木構造(議論タイムスパン木)の生成が考えられる。

議論タイムスパン木は DM 木の情報に基づき、以下 2 段階の処理によって生成する。(1)グルーピング獲得:DM 木に含まれるグループ(ゲシュタルト)の発見。(2)重要発言の選定:あるグループ全体の時間幅(タイムスパン)を代表する重要発言の選定。DM システムから、時間的な近さや発言順序、情報量の変化によって話題の変化や類似が判定できるだろう。また発言の重要度は、発話量や発話時間長、さらに発言の内容の意味を考慮し、賛同数、重要単語の頻出数から類推される。

図 2 の DM 木は、導入発言<1>から継続発言<2>、<3>が生じ、さらなる継続発言が生じていることを表しており、<1>→<2>→<5> と<1>→<3>→<4>の 2 つの仮想的な時間軸が存在していることが分かる。また、図 2 の議論タイムスパン木では、各時間軸ごとにタイムスパン木が作成され、最終的に 1 つのタイムスパン木に統合されることがわかる。各発言の重要度は高い方から順に<5>、<1>、<2>および<5>、<4>、<1>、<3>であることが分かる。

前章で述べた GTTM のタイムスパン木を生成する規則からの類推により、DM 記録分析のための GTTM 風規則を提案した[三浦 2013]。グルーピング獲得ルールには「発言間の間隔で境界が生じやすい」「発言者の順序の変化で境界が生じやすい」、重要発言の選定ルールには「発言時間の長い発言は重要である場合が多い」「重要発言は重要単語を含む場合が多い」などがある。

5. 議論タイムスパン木の生成アルゴリズム

5.1 ルールの実行管理

前節で述べた議論タイムスパン木生成に関するルールについて、ルールの競合と階層構造の獲得が主要な問題である。例えばルールの競合に関して、発話量は短いが多く数の賛同数を得られた発言の場合、「発言における賛同数の変化」「重要単語の初出箇所」の 2 つについて、両者を正しく境界判定することは難しい。また階層構造の獲得に関して、提案したルールは、局所的/大局的な観点からボトムアップ/トップダウンに生成するルールが混在する。このため、両者をどのように組み合わせる適切な階層構造を生成するかを判断することは難しい[浜中 2007]。

これらの問題に対処するため、局所的/大局的な処理を適切に組み合わせるアルゴリズムを構築した。また適切なルール実行管理としてルールを評価値として導入し、以下 2 項目を判定する。(1)発言間に生じる境界、(2)発言が内包する情報量。ルール実行の曖昧性をできるだけ排除するため、ルールが成立すれば値を 1、不成立ならば 0 に対応づける。ルールが成立する度合いが連続的な場合は重み付けを行う。その結果から、相対的なグループ発見を基にボトムアップに重要発言の選定を行い、以下のステップで議論タイムスパン木を生成する。

- (1)DM システムにより DM 木を取得する。
- (2)1 セクションを 1 つのグループにする。
- (3)局所的構造に関するグルーピング獲得ルールを適用する。
- (4)相対的な観点から高次の境界の強さを算出する。
- (5)最も強い境界でグループを 2 つに分類する。

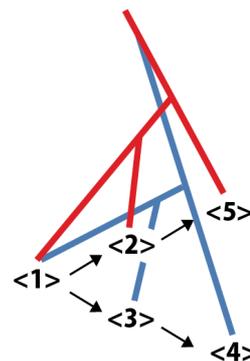


図 2. ディスカッションマイニング木と議論タイムスパン木

- (6)局所的境界がある場合、(3)(4)(5)を繰り返す。
- (7)重要発言の選定ルールを適用する。
- (8)グループ内での重要発言をボトムアップに選定する。

ここでは、ある導入発言から次の導入発言までの継続した発言群を 1 セクションと呼んでいる。階層的なグルーピング構造は、ボトムアップ処理により求めた局所的境界を用いて、トップダウンに獲得する必要がある。そのためには、グループ全体に局所的な構造に関するグルーピング獲得ルールを適用し、境界判定によって高次の境界の強さを算出する(3)(4)。この結果から最も強い境界でグループを 2 つに分類し、そのグループがその内部に局所的境界を含んでいる場合、この処理を繰り返す(5)(6)。このアプローチによって、局所的/大局的な階層構造を取得できると考える。また議論タイムスパン木は、上記の手順で得た局所的/大局的な階層構造と、重要発言の選定ルールをグループ全体に適用して得られる各発言の重要度合からボトムアップに生成する(7)(8)。

5.2 会議記録における重要単語の同定

重要発言の選定ルールの中には、発言内容の意味を考慮する必要がある。テキストデータにおける単語の重要度を求める方法として、一般に TF-IDF 法[天野 2007]が用いられるが、形式や内容が予め整えられているテキストとは性質が異なるため、会議記録の性質を考慮した重要単語の抽出手法が求められる。そこで 1 つの話題から複数の話題が派生する[松村 2009]という会議記録の性質から、TF-IDF 法の適用範囲について以下の 3 つの範囲を提案した。会議記録全体: 1 つの議題に対する議論全体。セクション: 導入発言から次の導入発言までの継続した発言群。仮想スレッド: 導入発言から各末端までの継続発言の連鎖であり、意味的な繋がりも考慮した仮想的な時間軸のまとまり。

DM システムで公開されている全 25 議論(1 議論あたりの平均議論時間: 1 時間 57 分、平均発言数: 65.5、平均セクション数: 13.4)を対象とし、適用範囲内に出現する全単語の TF-IDF 値を計算し、正解データの重要度を比較した。正解データは事前に準備しており、適合率 P(precision)と再現率 R(recall)を組み合わせた F 値を求めた(表 1)。

表 1. 各適用範囲における重要単語選定の F 値

適用範囲	会議全体	セクション	仮想スレッド
F 値	0.449	0.444	0.463

表 1 の結果から、範囲における F 値の差はあまりないことがわかった。しかし、会議記録内の複数のセクションに出現する場合、セクションのみならず会議全体の特徴となる単語として認識される場合がある。また仮想スレッドのみに制限した場合、範囲内に

表 2. 各発言要旨

<1>,O	危険ではない状況というのは、目と目があっている状況のことではないか.
1,0:33	
<2>,W	お互いに目が合っていないでも大丈夫. 目が合うというか, 認識できているかどうかだと思う.
0,0:30	
<3>,O	ずっと認識している必要はないが, 一度は相手が何処にいてどの方向に動いているか知る必要がある.
0,0:26	
<4>,W	その人が次にとる行動を予測するところまで考えないと「認識して回避する」と言えないのではないか.
0,0:34	
<5>,N	相手がこちらを認識していないときはその行動を予測できないと思うが, そこは従来研究に譲る.
1,0:40	
<6>,W	相手が人間だと認識したら, AT がやるべきは回避ではなくて人間に AT の存在を知らせることである.
0,0:32	
<7>,N	人間に乗り物の存在を気づいてもらえるクラクションなどの何らかのアクションをしなくてはならない.
2,1:05	
<8>,W	安全走行のためには, そういうことに気をつけることも必要だと思う.
0,0:16	

<1> → <2> → <3> → <4> → <5> → <6> → <7> → <8>

図 3. DM 木

出現しない単語があるため, セクション毎の適用が有効であると考えた. 議論タイムスパン木生成に関して, 重要単語を取り扱うルールでは, これらの結果を反映させ, 実装を行った.

5.3 ケーススタディ

本章で述べた議論タイムスパン木の生成アプローチに基づくケーススタディを行う. ここで, 議論セクションは話題の派生の仕方によって 3 タイプ(直線的, 途中から二股に分岐, 根元から分岐)に分類できる. 今回は, 直線的な議論セクションである例を対象としている. 例の発言要旨を表 2, DM 木を図 3 に示す. 表 2 の各発言要旨の左側は, 左上が発言番号(例:<1>), 右上が発言者(例:O), 左下が賛成ボタンの押下回数(例:1), 右下が発言に要した時間(例:0:33(33 秒の意))である. 発言者 O による導入発言<1>を聞いて, 発言者 W による継続発言<2>が生じ, さらなる継続発言が生じたことを表している.

発言番号<1>から<8>までの 1 セクションを 1 グループとする. このグループ全体に局所的な構造に関するルールを適用する. 各発言への適用結果から<4>-<5>の間に最も深い境界が生じ, それを境界とした<1>-<4>と<5>-<8>のサブグループが検出される. また<2>-<3>, <6>-<7>の間にも境界が生じる. この一連の処理をサブグループ内で繰り返すと, 最終的に<1>-<4>のグループでは, <1>-<2>, <3>-<4>と細かく分類され, 局所的/大局的な階層構造が得られる.

同様に重要発言の選定に関するルールをグループ全体に適用する. この処理によって各発言の重要度合が分かる. 以上より得られた局所的/大局的な階層構造と各発言の重要度合を基に, 議論タイムスパン木をボトムアップに生成する. <1>-<4>のグループにおいては, <1>-<2>と<3>-<4>のそれぞれで重要発言の選定を行う. この処理を繰り返し, 最終的には<1>-<4>と<5>-<8>でトーナメント式に得られた重要発言の比較を行い, このセクションでの最重要発言が決定する. このようにして議論タイムスパン木が得られる(図 4).

6. 実験と評価

前章で述べた議論タイムスパン木の生成アルゴリズムを自動化したプロトタイプシステムを構築した. 本章ではその有効性を評価するため, DM システムでの議論データを分析対象とし,

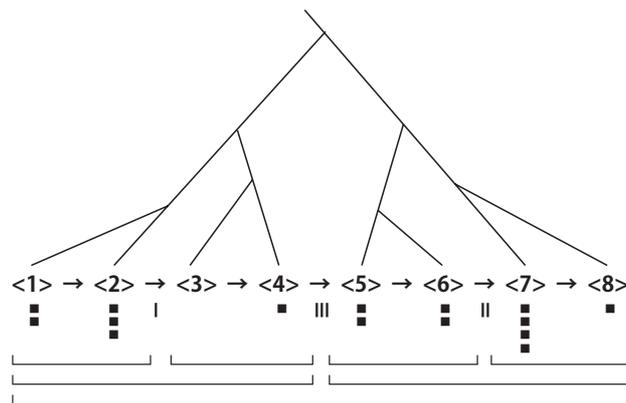


図 4. 各ルールの適用結果と得られる議論タイムスパン木

評価実験を行った. ここでは, グルーピング構造分析と重要発言選定の性能の評価を, 適合率 P と再現率 R を組み合わせた F 値で評価する. グルーピング獲得ではグループが所属する階層に関係なく, システム出力と正解データの両方に同じグループがあった場合を適合とした. 一方, 重要発言選定では, タイムスパン木の枝の交点システムの出力と正解データとともに同じ位置にある場合, 適合とした.

このような F 値による評価を行うため, 新たに評価用データを作成した. 評価用データは, DM システムで公開されているこれらの異なる議論タイプ(直線的, 途中から二股に分岐, 根元から分岐)各 10 件, 全 30 件の議論データでと, 手作業でグルーピング構造分析および重要発言選定を行った正解データである. 議論データは, 発言数や発表者数, 分岐数など議論構造の異なるセクションを選定している. また正解データは, GTTM や議論タイムスパン木を良く理解している本研究報告の筆者の 1 人が作成した. 事前実験により, 議論タイムスパン木の品質と要約としての妥当性がともに高いことが証明されたものを選定している. プロトタイプ出力のグルーピング構造獲得分析と重要発言選定の性能評価を表 3 に示す.

表 3. グルーピング獲得および重要発言選定に関する F 値

議論タイプ	グルーピング獲得	重要発言選定
直線的	0.62	0.58
根元から分岐	0.79	0.63
途中から分岐	0.76	0.63
全体平均	0.72	0.61

7. 考察

本章では, 評価結果から階層構造の獲得とルール競合の解消について考察する. まず階層構造の獲得について, 今回の評価データでは, 発言数の多い発言や直線的な議論セクションのデータに関して F 値が低いことが分かった. これはスレッド数の長い発言に対して, 大局的な階層構造の獲得が適切でないと推測できる. また全データの正解データとシステム出力に含まれるグループ数, グルーピング階層の数を比較するとグループ数とグルーピング階層の両方もが, 正解データに比べてシステムの出力の方の値が大きくなる傾向があった.

次にルール競合の解消について, 全 30 件でのグルーピング構造分析に関する F 値は 0.72, 重要発言選定に関する F 値は 0.61 であった. 本稿ではルールの重み付けによる優先順位の管理と, 評価値導入による境界および重要度合の選定を行ったが, グルーピング獲得が正確に行われていないことが評価結果

表 4. Q&A型議事録で想定される機能

Q&A型議事録想定機能		必要な要素技術			
		DMシステム	構造獲得	再構成	自然言語処理
検索	議論情報	○	-	-	-
	参加者情報	○	-	-	-
	決定事項	-	○	○	○
	今回の結論	-	○	-	-
	発散・収束	-	○	-	-
コンテンツ化	起承転結	-	-	○	-
	ストーリー化	-	-	○	○
	ハイライト	-	○	○	-
	意思決定プロセス	○	○	○	-
	トレンド	○	○	○	○
評価	フィードバック	-	○	○	-
	MVP	○	○	-	-
	課題に適切な結論か	○	○	-	○
	議論の妥協点	○	○	-	○
	議論の妥当性	○	○	○	-

から分かった。その理由として、今回の実験では、「発言者のパターンが変化した場所で境界が生じやすい」といった意味を考慮しない、形式的なルールを比重を重くしたためだと考えられる。この問題に対処するためには、発言内容を理解したルールの提案やルール優先順位の管理に関して新しい手法を提案する必要がある。

8. おわりに

本稿では、音楽理論の時系列データ分析への応用として、GTTM の楽曲分析アプローチに基づき、議論タイムスパン木の生成方式について述べた。議論タイムスパン木生成における計算機上への実装に対する問題は、階層構造の獲得とルール競合であった。これらの問題に対処するため、局所的/大局的な階層構造を獲得するためのアルゴリズム提案と適切なルール実行管理方法を考案し、プロトタイプシステムを作成した。このシステムの有用性を評価するため、正解データを作成し、実験によりグルーピング構造分析および重要発言選定の結果に関する適合率、再現率を評価した。今後は、外部からルールの優先順位や重み付けを管理できる外部パラメータの導入を検討している。

今後の展望として Q&A 型議事録の実現を考えている。Q&A 型議事録とは、会議記録データに対する検索エンジンのようなインタラクティブシステムのことであり、以下のような質問を受け付けることを想定している：「この話題はどういう結論だったのか？」「この結論に至ったプロセスを教えてください」「私は何を知らなかった上で次の会議に臨めばいいか？」。これらを実現するには、さまざまな機能が必要となる。そこで現在構想中の Q&A 型議事録について、アプリケーションで想定される機能を縦軸、さらにそれを実現するための要素技術を横軸とし、分類した(表 4)。

議論タイムスパン木の自動生成は会議の深層構造の分析を可能とするだけでなく、過去の会議コンテンツの柔軟な検索や加工、議論タイムスパン木に含まれる様々な情報をユーザの意図に沿った変換や抽象化、議論内容のコンテンツ化により、Q&A 型議事録への応用が期待できる。今後は、議論タイムスパン木に含まれる様々な情報を起承転結のような型に当てはめコンテンツとして提供するための木構造の生成を試みており、これにより会議記録の多種多様な再利用の実現を目指す。

謝辞

本研究は JSPS 科研費 23500145 の助成を受けたものです。

参考文献

- [長尾 2003] 長尾研究室: ディスカッションマイニングプロジェクト, <http://dm.nagao.nuie.nagoya-u.ac.jp/>
- [Lerdahl 1983] Lerdahl, F. and Jackendoff, R.: A Generative Theory of Tonal Music, The MIT Press (1983).
- [土田 2010] 土田貴裕, 大平茂輝, 長尾確, 対面式会議コンテンツの作成と議論中におけるメタデータの可視化, 情報処理学会論文誌, Vol.51, No.2, pp.404-416 (2010).
- [Nagao 2004] Nagao, K., Kaji, K., Yamamoto, D. and Tomobe, H., Discussion Mining: Annotation-Based Knowledge Discovery from Real World Activities, Proceedings of the Fifth Pacific-Rim Conference on Multimedia (PCM 2004), pp.522-531, (2004).
- [三浦 2013] 三浦寛也, 森理美, 長尾確, 平田圭二, 音楽理論 GTTM に基づく議論タイムスパン木の生成方式とその評価, (社) 情報処理学会 音楽情報科学研究会, 2013-MUS-100, No.2 (2013).
- [浜中 2007] 浜中雅俊, 平田圭二, 東条敏, 音楽理論 GTTM に基づくグルーピング構造獲得システム, 情報処理学会論文誌, Vol. 48, No. 1, pp. 284-299 (2007).
- [天野 2007] 天野真家, 石崎俊, 宇津呂武仁, 成田真澄, 福本淳一, IT Text 自然言語処理, オーム社(2007).
- [松村 2009] 松村真宏, 三浦麻子, 人文・社会科学のためのテキストマイニング, 誠信書房(2009).