

メディアデータを対象としたメタデータ抽出作用素と
その逆作用素による知識創造サイクルモデル

2019年 3月

岡田 龍太郎

メディアデータを対象としたメタデータ抽出作用素と
その逆作用素による知識創造サイクルモデル

岡田 龍太郎

システム情報工学研究科

筑波大学

2019年 3月

目次

目次.....	i
図目次.....	iv
表目次.....	vi
第 1 章 序論.....	1
1.1 論文の構成.....	6
第 2 章 知識創造サイクルモデルの提案.....	7
2.1 はじめに.....	7
2.2 関連研究.....	9
2.3 知識創造サイクルモデルの定義.....	10
2.3.1 知識創造と知識利活用の関係.....	10
2.3.2 知識創造と知識利活用の定義.....	10
2.3.3 知識創造サイクルモデルの定義.....	11
2.4 知識創造サイクルモデルの構成.....	13
2.4.1 センサーとアクチュエーター.....	13
2.4.2 知識創造と知識利活用の定式化.....	13
2.4.3 共通計量空間.....	15
2.4.4 特定分野に対する知識創造サイクルモデル.....	16
2.4.5 複合分野に対する知識創造サイクルモデル.....	17
2.5 まとめ.....	19
第 3 章 メディアコンテンツを対象とした統計的一般可逆作用素構成方式とその楽曲メディアコンテンツ生成への適用.....	20
3.1 はじめに.....	20
3.1.1 3 章の構成.....	22
3.2 関連研究.....	23
3.3 メディアデータを対象とした一般化逆作用素.....	25
3.3.1 Media-Lexicon Transformation Operator(\mathcal{ML}).....	26
3.3.2 Stochastic Generalized Inversed Media-Lexicon Transformation Operator($i\mathcal{ML}$).....	27
3.3.3 \mathcal{ML} - $i\mathcal{ML}$ 系の構成と検証.....	28
3.4 楽曲メディアコンテンツを対象とした統計的一般化逆作用素 $i\mathcal{ML}$ の構成.....	32
3.4.1 Hevner の研究.....	32
3.4.2 楽曲メディアコンテンツを対象としたメタデータ自動抽出方式(\mathcal{ML}).....	33
3.4.3 楽曲メディアコンテンツ自動構成方式($i\mathcal{ML}$).....	42
3.5 楽曲メディアコンテンツにおける \mathcal{ML} - $i\mathcal{ML}$ 系を用いた検証方式.....	49

3.5.1	\mathcal{ML} - $i\mathcal{ML}$ 系の保存性	49
3.5.2	\mathcal{ML} - $i\mathcal{ML}$ 系の検証方法.....	49
3.5.3	楽曲の表現力と分解能.....	50
3.6	任意の言葉の音相の印象に合致した楽曲の自動構成方式	52
3.6.1	音相理論.....	52
3.6.2	任意の言葉の音相の印象に合致したメタデータ自動抽出方式の概要.....	52
3.6.3	意味の数学モデルの概要	53
3.6.4	任意の言葉の音相の印象に合致した楽曲の自動構成方式.....	54
3.7	実験	57
3.7.1	実験システム	57
3.7.2	実験 1 (\mathcal{ML} - $i\mathcal{ML}$ の検証).....	57
3.7.3	実験 2 (楽曲生成および作曲システムの表現力の検討)	63
3.7.4	実験 3 (音相の印象に基づく作曲)	71
3.7.5	実験全体の考察.....	77
3.8	まとめ	78
第 4 章	会議の時系列構造分析によるトピック抽出および合意形成支援の研究.....	79
4.1	はじめに.....	79
4.1.1	4 章の構成.....	80
4.2	関連研究.....	81
4.2.1	会議の分析に関する研究.....	81
4.2.2	発散的思考と収束的思考に関する研究	81
4.3	会議の時系列解析による構造化および可視化手法	84
4.3.1	概要.....	84
4.3.2	セグメントベクトルの構成	85
4.3.3	各種分析.....	87
4.3.4	単一の会議を対象とした可視化手法.....	89
4.4	一連の会議のトピックの関係の可視化手法	91
4.4.1	概要.....	91
4.4.2	各種分析.....	91
4.4.3	複数会議を対象とした可視化.....	91
4.5	議論の発散/収束に着目した会議の時系列構造分析と状態遷移可視化手法.....	93
4.5.1	会議の状態を表す指標の定義.....	93
4.5.2	会議の進行評価指標および会議の状態遷移図の導出	95
4.6	実験	98
4.6.1	実験環境.....	98
4.6.2	実験 1:会議の構造可視化及びトピック抽出の検証.....	98

4.6.3	実験 2:会議の評価指標の検証	106
4.7	実システムへの展開.....	113
4.7.1	実装環境.....	113
4.7.2	展示会における来場者への聞き取り	115
4.8	まとめ	116
第 5 章	結論	117
謝辞	118
参考文献	120
論文目録	125

図目次

図 1-1 論文の構成.	6
図 2-1 人間の知識創造のベクトル空間モデルにおける表現.	10
図 2-2 知識創造と知識利活用の関係性.	11
図 2-3 知識創造サイクルモデルの概念図.	12
図 2-4 Specific-to-General 作用素と General-to-Specific 作用素.	14
図 2-5 特定分野に対する知識創造サイクルモデルにおける検証演算.	16
図 2-6 複合分野に対する知識創造サイクルモデル.	17
図 2-7 複合分野に対する知識創造サイクルモデルにおける検証演算.	18
図 3-1 ML と iML の概略図.	25
図 3-2 ML と iML に含まれる変換の関係.	29
図 3-3 Hevner による 8 つの印象語群.	32
図 3-4 楽曲からのメタデータ自動抽出方式の全体の流れ.	33
図 3-5 変換行列 T	41
図 3-6 楽曲メディアコンテンツ自動構成方式の流れ.	42
図 3-7 変換行列 T^\dagger	43
図 3-8 ハ長調 (C major) の長音階とハ短調 (C minor) の和声的短音階.	46
図 3-9 ハ長調 (C major) の 7th ダイアトニックコードとハ短調 (C minor) の和声的短音階における 7th ダイアトニックコード.	47
図 3-10 tc, oc, nc の和音との対応.	47
図 3-11 C major の和音の変形.	48
図 3-12 ML - iML 系の一般形と具体例の対応.	49
図 3-13 楽曲の ML - iML 系の検証のための実験の検証項目.	50
図 3-14 任意の言葉の音相の印象に合致した楽曲の自動構成方式の概要.	54
図 3-15 実験 1-4 の検証項目.	62
図 3-16 明るい印象から生成された楽曲.	64
図 3-17 暗い印象から生成された楽曲.	64
図 3-18 「ハンガリー舞曲 No.5」の冒頭 8 小節.	67
図 3-19 「ハンガリー舞曲 No.5」を入力として作曲された楽曲.	67
図 3-20 「家路」の冒頭 8 小節.	68
図 3-21 「家路」を入力として作曲された楽曲.	68
図 3-22 「イキイキ」から生成された楽曲.	74
図 3-23 「グズグズ」から生成された楽曲.	75
図 3-24 「ナマテー」から生成された楽曲.	76
図 4-1 セグメントベクトルの構成.	84

図 4-2 4.3.3 節に示す各種分析の流れ.	85
図 4-3 4.3.4 節に示す 4 種類の可視化とそれに用いるデータ.	85
図 4-4 会議の状態の定義.	95
図 4-5 会議の状態遷移図および進行評価指標導出方式の全体像.	95
図 4-6 会議 1 の会議アウトライン.	100
図 4-7 会議 1 の代表度グラフ.	101
図 4-8 実験結果：一連の会議におけるトピック間の関係の視覚化.	104
図 4-9 実験 2-1 結果：すべての会議の状態遷移図.	107
図 4-10 会議 ID1 の状態遷移図.	108
図 4-11 会議 ID10 の状態遷移図.	109
図 4-12 実装された会議室.	113
図 4-13 分析結果の表示例.	114
図 4-14 発話のタイムライン表示.	114

表目次

表 3-1 各 ML の性質.	31
表 3-2 Hevner による楽曲構造要素と楽曲印象語群の関係性.	33
表 3-3 楽曲の特徴値.	35
表 3-4 楽曲の特徴値と楽曲構造要素の対応関係.	37
表 3-5 拡張された各調の調性度(key).	39
表 3-6 歌曲の旋律音における各音程の出現頻度.	48
表 3-7 音相を表す表情語と表情属性.	53
表 3-8 実験 1 の検証項目と実験の対応.	58
表 3-9 実験 1-1：結果 1(入力と出力で一致した楽曲特徴値データの数).	59
表 3-10 実験 1-1：結果 2(分解能と最大の誤差).	59
表 3-11 実験 1-2：結果 1(楽曲構造要素の検証).	60
表 3-12 実験 2：結果 2(楽曲構造要素の分解能の検証).	60
表 3-13 実験 1-3 結果.	61
表 3-14 実験 1-4：実験に用いた楽曲.	61
表 3-15 実験 1-4：結果.	62
表 3-16 実験 2-1：生成された楽曲の楽曲構造要素ベクトルと楽曲特徴値の各値.	65
表 3-17 実験 2-2：「ハンガリー舞曲 NO.5」における抽出時と生成時の各値.	67
表 3-18 「家路」における抽出時と生成時の各値.	68
表 3-19 実験 2-3：実験に用いた楽曲.	69
表 3-20 実験 2-3 の結果.	70
表 3-21 実験 3：入力文字列から抽出された音相の印象語ベクトル.	73
表 3-22 「イキイキ」から生成された楽曲の各値.	74
表 3-23 「グズグズ」から生成された楽曲の各値.	75
表 3-24 「ナマテー」から生成された楽曲の各値.	76
表 4-1 実験 1-1 で用いる会議の概要.	99
表 4-2 実験結果：2つのセグメントにおけるトピック語と特徴語リスト.	102
表 4-3 実験結果：代表度を考慮した会議の重要語リスト.	103
表 4-4 実験 1-2：分析対象とする4つの会議の概要.	104
表 4-5 実験対象とした会議の概要.	106
表 4-6 実験対象とした各会議の参加人数.	106
表 4-7 会議の順調／不順の判定.	108
表 4-8 会議の評価指標一覧.	110
表 4-9 各会議の評定と会議評価指標の評価値.	110
表 4-10 会議の評定と評価指標の相関係数.	111

第 1 章 序論

掛谷は著書「学問とは何か」[1]の中で、学問を「予測する力を持つ体系的知識、およびその知識を得るための研究方法」と簡潔に定義した。予測が成り立つということは、すなわち、その体系的知識が現実の現象と合致していると思なしてよいことを意味していると考えられる。逆に言えば、予測した結果が現実にかかる事象と合致していないということは、その学問による知識は不十分であると言える。ニュートンの運動方程式に代表されるような古典物理学の知識は、まさしくこの予測の能力を備えていたために、輝かしい応用例を多数生み出すとともに、学問の中心あるいは規範として長く参考にされてきた。

しかし、現在我々が学問的知識として参照している知識のすべてが、予測を可能にするほど十分な知識であるとはいえない。例えば、気象学の分野ではバタフライエフェクトという用語があるが、これは気象における現象はあまりに多くの要素が相互に影響し合って発生するために、ほんの僅かな初期値の違いによって、実際に起きる現象が大きく変わってしまうこと、すなわち予測が困難であることを示している。

相互に影響する要素が多数あるために、結果を予測することが困難な事象をカオスと呼ぶが、ハラリは著書[2]の中で、そのカオスにはまた1次のカオスと2次のカオスという分類があることを示した。気象のような現象は、確かに予測が困難ではあるものの、その入力情報や、現象を表現するモデルを精緻化していけば、予測は現実近づいていく。これは1次のカオスに相当する。しかし、経済学分野や社会学分野での現象のように、予測そのものが現実のプレイヤーの行動に変化を起こしてしまう場合には、その予測はより困難になる。このような、現状に対応して振る舞いを変化させる要素を含む系を2次のカオスと呼ぶ。振り返って考えてみると、ニュートンの運動方程式が予測を実現することが出来たのは、当時対象としていた物理現象の要素は、現実に対して対応しようとするプレイヤーではないことに一因があった。

デカルトが方法叙説の中で示した「分析と総合」、あるいはその後に還元主義と呼ばれようになった手法は、科学的発見のための手法の基礎となっている。これは、世界を要素に細かく分割しその要素を詳しく調べることでその性質を明らかにし(分析)、そうして明らかになった要素を組み合わせることで世界全体を知ろうというものである(総合)。還元主義的な立場における実験とは、ある変数を操作したときの影響を見るために、その他の要因を無視できるように系を切り取り、その系の中で試行し現実にかかる事象を観察する行為と言える。

しかし、仮に切り取った実験系の中で得られた知識が正しいとしても、それを総合して生まれた知識が正しいこと、すなわち予測可能性を持つことを保証するには、切り取られた系同士の相互作用が無視できること、そしてその結果としての「部分の総和が全体に一致する」という条件が必要である。これは大変厳しい条件であるにも関わらず、しばしば無視される。長沼はその著書[3]において、この条件が無視されがちな原因は、まさに先述のニュ

ートンの運動方程式が、当時のヨーロッパの学者の最大級の関心事であった天体の動きを見事に説明出来てしまったことであると指摘している。ここで着目していた天体というのは、太陽系における太陽とその惑星群である。太陽系において、太陽はその他の惑星に比べて圧倒的に大きな質量を持つために、太陽と惑星の間の引力に比べると、惑星同士の間の引力は無視できるほど小さくなる。そのため、分析対象の系を「太陽と各惑星」という小さな系に分割して、それらの関係を明らかにしたうえで、それらを総合することで、太陽系全体の運動を知ることが出来たのである。運動方程式が規定しているのは二つの物体の間に発生する力であり、もし、対等に近い質量を持つ物体が3つあったとすれば、その運動を記述することは三体問題と呼ばれる問題となり、代数的な解法は存在しなくなってしまう。太陽系のような「分析と総合」に都合の良い条件を持つ事象はこの世の全ての事象の中では例外的であると考えられるが、先述の通り、この運動を明らかにすることが当時のヨーロッパの学者にとって最大級の関心事だったこともあって、それを実現したニュートンの運動方程式は「この世の全ての物体の動きを記述できる」とさえ思われてきた。しかしこれは過大評価と言わざるを得ない。

ニュートン以後の科学的発見は、長くこの分析と総合を前提とする手法によって行われてきた。しかしこうした方法では解決が難しい問題も当然ながら多数存在した。そうした問題は、1980年代ごろから「複雑系」として広く語られ始めた。しかし、もともとほとんどの事象は複雑系であり、一部の特殊なケースのみ分析と総合による手法が有効であったと言うべきであろう。

長沼はさらに、大学等の学術活動を行う組織がそれぞれの専門分野に細かく分かれていることも、「部分の総和が全体に一致する」という条件を前提にしていると指摘している。「部分の総和が全体に一致する」という条件があれば、細かく分かれた専門分野の中で、その分野の専門家が見出した知識を、他の分野の専門家が組み合わせて利用しても問題ないことが保証される。そのため、各専門家はその専門における知識を生み出してさえいれば、学術活動全体に貢献できることになる。しかしもしこの条件が成立しないのであれば、一人の人間あるいは課題を共有する小規模のチームが、対象としている課題の解決に必要な様々な分野の知識を一度に持たない限り、実際の課題の解決には役に立たないことになる。近年、「学際」を謳う大学組織が多数生まれて来ているのも、上記のような問題意識が広まってきているからであると思われる。

こうした現状を踏まえ、本論文では、ある知識によって構成される系、あるいは知識を組み合わせて構成される系が、現実を正しく予測できるのかを検証可能にするためのモデルとして、知識創造サイクルモデルを提案する。知識創造サイクルモデルは、知識創造の作用素と知識利活用の作用素を組み合わせて構成されるモデルである。知識創造とは、現実の事象から知識を抽出することを指す。知識利活用とは、知識を用いて現実に影響を及ぼすことを指す。知識創造は抽象化に相当し、知識利活用は具体化に相当する。構成された知識創造サイクルモデルが現実の現象を正しく予測できない場合は、知識を更新する、あるいは知

識の組み合わせ方を変更することによって、予測可能なモデルを模索することを促すことになる。

これまで筆者らは、本モデルにおける知識創造に相当する研究を行ってきた。特に、メディアデータを対象として、そのメディアデータから人間が受ける印象を「印象語」と呼ばれる言葉のメタデータとして抽出する枠組を Media-Lexicon Transformation Operator (\mathcal{ML}) [4] と定義しており、この \mathcal{ML} の枠組みにおける具体的な事例として、様々な種類のメディアに対する印象語の抽出システムを構築してきた。例えば、音楽心理学者 Hevner の研究 [5] [6] [7] を用いた、音楽の印象を言葉のメタデータに変化する研究 [8]、カラーイメージスケール [9] を用いた、色の印象を言葉に変換する研究 [4] [10] などである。これらは、専門家が調査して得た知見を元にした知識創造の作用素である。

筆者らは、今回さらに、この印象を表すメタデータから、その印象を表すメディアデータを生成することに挑戦した。これは知識創造サイクルモデルにおいては、知識利活用に相当する。本論文の 3 章で述べるのは、音楽の印象メタデータからそれを表す楽曲を生成する研究である。

知識利活用のための作用素を作るのは、一般に、知識創造の作用素を作るのに比べて難しい。なぜなら、抽象化とは情報を捨象することであるのに対し、具体化とは情報を付加することだからである。情報を付加する場合、その方法は一意に定まらない。したがって、知識利活用によって現実の事象を作り出す場合には、何か制約を加える必要がある。この部分は対象とする事象に依存するため、一般的な方法を示すことは難しい。その実現の例については 3 章で述べる。

データを生成するには、データがどのような規則に従って生成されるのかという制約条件が必要である。例えば、あるデータ群があったとして、そのデータ群の性質を知るために平均値を求めるという計算をしたとすると、求めた平均値のデータにおいては、個々のデータの持つ情報は失われている。したがって、そこから元のデータを再現することはできない。これは、抽象化によって情報を捨象していることと言える。しかし、ここで例えばそのデータ群の個々のデータの生成規則がある分布に従うという仮定があり、その仮定が正しいとするならば、分布を表現するのに必要なパラメータさえ保持していれば、元の現象をある程度再現できる。例えば、ある現象から観察されるデータが正規分布に従うのであれば、平均値と標準偏差を保持していれば、元の現象を再現できるといった具合である。これは抽象化されたデータを再び具体化することであり、その処理においてはデータの情報量が増加している。しかしこの処理によっては、元のデータそのものを再現できるわけではない。元のデータと生成されたデータの間で保持されているのは、生成したデータから再び平均値と標準偏差を抽出したときの、その値である。

本論文では、3 章において、楽曲におけるデータ生成の実現例を示す。本研究で示す手法で生成される楽曲は、先述の正規分布の例のように、統計的な情報を保持したうえで、ある分布の上に存在する点としての存在になる。ここで問題となるのは、本当に統計的な情報を

保持できているのかということと、統計的でない、個々のデータが持つ情報が失われているにも関わらず、そのデータ群を同一の現象として扱ってよいのかということである。これらのことについても3章で詳しく述べる。なお、上記のような性質があることを考慮して、このメディアデータ生成の枠組みを統計的一般化逆作用素“Stochastic Generalized Inverse Media-lexicon Transformation Operator”(“iML”)と、「統計的」という言葉を含めて名付けている。

また、データ生成が可能であるということは、既存の知識によって明らかになっていることを、現実の現象として出現させ、それを検証するというループを構成できることを意味する。これは機械学習の文脈では強化学習を可能にする条件であるということも出来る。2017年に、DeepMindによって開発されたAlphaZero [11]という強化学習を用いた一つのプログラムが、それまで囲碁、将棋、チェスにおいてチャンピオンだったプログラムに勝利した。AlphaZeroは、ゲームのルールを与えるだけで、そのゲームにおける対戦AIを作り出す。そうして作られたAI同士を対決させることで、より強いAIを生み出していく。こうした自己対戦を用いた強化学習の手法によって、棋譜を学習データとして与える既存のアプローチを用いたプログラムに勝つことが出来たのは、これらのゲームにおけるデータを生み出す行為、すなわち「手」が離散的かつ完全に記述可能であったことが要因として大きいと思われる。現実の現象をモデル化する場合、その多くは連続的である。例えば、楽曲を生成しようと考えた時、音の長さの取り得る値というのは連続的であるため、音符一つでさえ無限の選択肢が生まれてしまう。このような条件で楽曲を生成しようと思うのであれば、プログラマが音の長さの取り得るパターンを記述せざるを得ず、生成されるデータはその想定範囲を超えることが出来ない。それに対して、囲碁等における手は離散的であり、可能な手はルールの範囲で与えられているため、データ生成に必要な条件は完全に記述することができる。AlphaZeroによる輝かしい成果は、現実と比べて遜色のないデータ生成系を含むモデルの強さの一例と考えることができる。

機械学習の文脈に限らなくても、データ生成系を持つことで検証のためのフィードバックループを構成できることは、既存の知識の更新のために非常に役に立つ。本論文で示す知識利活用の実現例である、印象を表す楽曲メディアデータとしての生成システムは、それ単体では自動作曲システムとなる。しかしこのシステムを印象抽出システムの逆演算として構築したことによって、印象抽出システムの検証システムとしての意味も持つ。この検証システムによって、印象抽出システムが正しく機能する範囲を限定することができる。例えば、Hevnerは人に様々な楽曲を聴かせて質問紙調査をすることによって楽曲の構造と印象の対応付けを行っていたのだが、そこで用いられていた楽曲はクラシックの曲に限定されていた。したがって、その研究を参考にした楽曲からの印象メタデータ抽出方式は、クラシックの曲以外に適用したときの有効性は未知数とすべきであった。しかし印象抽出のシステムを構築する段階ではどのような音楽に対しても適用できると想定していた。ここで楽曲生成システムを構成することで、生成された楽曲を人間が聴いて印象を判断できるようにな

り、実際にはすべての楽曲ではなくある特定のジャンルにおいてしか機能していないことを発見した。これは、あくまで楽曲自動生成システムを印象抽出システムの逆演算として構成しているからこそ可能なことである。理論的には正しい楽曲を生成しているにも関わらず、それを人間が有用性がないと判断することは、元になっている知識が不完全であることを示している。

ここで重要となるのは、メディアデータからの印象の抽出に用いる方式と、その逆である印象からメディアデータを生成する方式とを、互いに逆演算となるように構成することである。抽出と生成における演算が逆演算となっていることによって、生成されたメディアデータから再び印象を抽出した際に、メディアデータの生成の元となった印象と同じデータになれば、これらの演算は閉じた系において整合性を持っていることが言える。このことは、抽出と生成の二つの作用素を、本質的には一つの知識から構成しているからこそ可能になる。閉じた系の中での正しさが保証できるからこそ、知識そのものを検証することに繋がるのである。

また、4章においては、会議の効率化及び合意形成支援の研究について述べる。この研究においては、会議の概要を振り返るための可視化や、会議が順調に進行しているのかを判断するための指標を提案し、参加者に対するフィードバックを可能にする。本論文ではまだ実現していないが、これを会議の進行中に示すことによって、ファシリテーションを支援することを想定している。そうすると、進行中の会議に影響を与えるため、参加者の行動が変化する。これは、意思決定を行うプレイヤーである人間を含むフィードバックループを含んだ知識創造サイクルモデルを構成することを目指している。

1.1 論文の構成

論文の構成を以下に述べる．論文の構成の概要図を図 1-1 に示す．

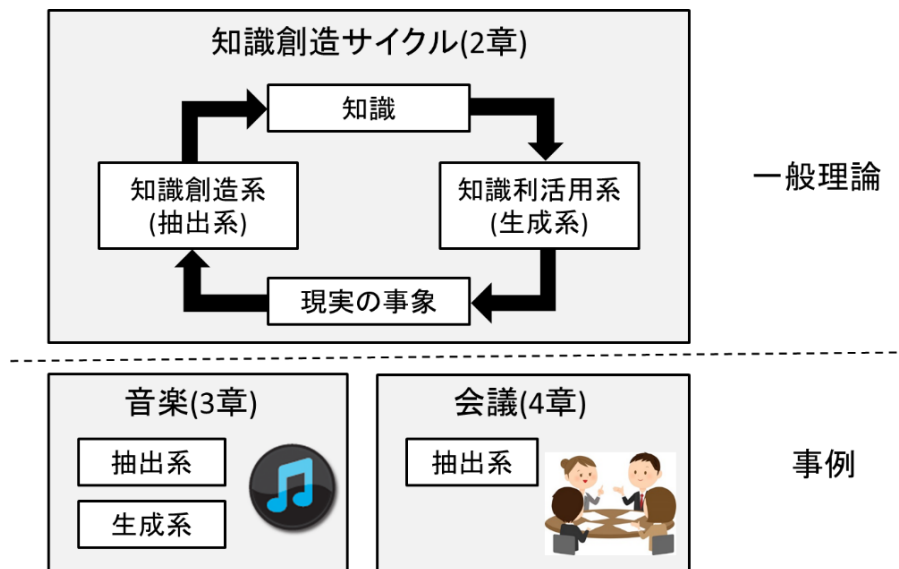


図 1-1 論文の構成.

2 章では，提案モデルである，知識創造サイクルモデルの定義と構成方法について述べる．知識創造サイクルモデルは，知識創造系と知識利活用系を組み合わせでサイクルにすることで知識の再創造を促すモデルである．知識創造系は現実の事象から知識を抽出する抽出系に，知識利活用系は知識を現実の事象として生成する生成系に相当する．これは以下の3章4章で述べる事例の一般理論に当たる．

3 章では，メディアコンテンツを対象とした統計的一般可逆作用素構成方式とその楽曲メディアコンテンツ生成への適用について述べる．これは，知識創造サイクルモデルの実現例として，音楽の分野を対象とした，抽出系と生成系を備えたシステムとなる．特に，メディアデータを対象とした知識創造サイクルモデルの実現例を $ML-iML$ 系と定義し，また，メディアデータを生成する作用素を統計的一般化逆作用素“Stochastic Generalized Inverse Media-lexicon Transformation Operator”(“ iML ”)として定義する．

4 章では，会議の時系列構造分析によるトピック抽出および合意形成支援の研究について述べる．これは，知識創造サイクルモデルにおける，抽出系の実現例となる．

5 章で本論文をまとめる．

第2章 知識創造サイクルモデルの提案

2.1 はじめに

近年、コンピュータを用いた新たな価値創造の枠組みとして、ビッグデータを対象とした研究が盛んに行われている。ビッグデータ研究の持つ特徴は、現実の事象から、人間には気付かないような関係性を発見するところにある。すなわち、我々人間が仮説を立てて、その仮説を検証するためのデータを集めて検証するといったような方法ではなく、まずデータがあり、そこにある関係性をコンピュータに可能な限り探させるという方法を採用。しかし、これはそうして見つかる関係性に一般性があるかどうかは、入力したデータに依存するところが多い。結果として、そうして見つかる法則や理論は、物理科学の法則のような常に成り立つようなものばかりではなく、ある限定された環境下でのみ成り立つような法則も多数含まれる。機械学習の文脈で言えば、これは過学習に相当する。そして交差検定などの検証手法は、その法則や理論に一般性があるか、すなわち、現実の問題に適用したときにも有効であるのかを示す手法であると整理できる。しかし一般的に言って、ある法則が見出された時に、それが成り立つ前提条件は非明示的であることが多く、それを明示化することは想定外を想定するようなことであって本質的な困難を伴う。既存の分野の学術研究においても、ある分野の研究者が専門家としての立場で明示した前提については知ることができるが、その研究者にとって想定外だったこと、あるいは自明すぎて明示しなかったことなどを知ることは同様に困難である。

本章では、この非明示的な前提条件を明示化するために、見出された法則や理論を検証するための枠組みである「知識創造サイクルモデル」を提案する。知識創造サイクルモデルは、知識創造の作用素と知識利活用の作用素を組み合わせることで構成されるモデルである。知識創造とは、現実の事象から知識を抽出することを指す。知識利活用とは、知識を用いて現実に影響を及ぼすことを指す。知識創造は抽象化に相当し、知識利活用は具体化に相当する。本研究では知識創造と知識利活用に互いに逆の演算と見なし、それら両方を備えることによって知識の検証が可能となることを示す。知識創造サイクルモデルは、現実の現象から知識を見出し、それを現実に作用させることで、想定したことと現実に起こる現象との差を検証可能にし、より精緻化された知識の創造を促すモデルである。

提案モデルである知識創造サイクルモデルは、意味の数学モデルを用いた意味的連想検索システム [12] に発想の基礎を置いている。このシステムでは、様々な分野のメディアデータを言葉という共通のメタデータに変換したうえで、言葉の意味的な類似性を計量するモデルである意味の数学モデルを用いてメタレベルで連結することで、異なるメディアデータを相互運用することを可能にしていた。

筆者はこのシステムに、メタデータからメディアデータを生成する方式を追加して連結することに挑戦した。その成果は3章で述べている。この生成の方式を追加することで、現実世界に新たなメディアデータを生み出すことが可能になった。これは、知識創造サイクル

モデルにおいては、データの利活用に相当する。そして、生成したデータから再度メタデータを抽出することによって、生成したデータを検証することが可能であることを示した。そして、抽出系と生成系を一つの知識から構成することによって、互いを検証するサイクルが構成できることを示した。

本章で提案する知識創造サイクルモデルは、上記の意味の数学モデルを用いた意味的連想検索システムをさらに抽象化したうえで、現実における影響を加味したモデルと言える。知識利活用は現実に影響を与えることを指す。新たに導入したこの知識利活用の概念によって、現実に影響を与え、その影響を再度観察することで、サイクルが構成できる。

本章の内容の大部分は [13] にて報告している。また、知識創造サイクルモデルは抽象的なモデルであるため、具体的な実装は、扱う対象を明確にしたうえで構成する必要がある。その実現例である楽曲メディアコンテンツを対象とした事例は 3 章に示す。3 章で示す事例が本章で示す内容の実現例になっているという関係性についての議論は、[14] にて報告している。

本章の構成は以下の通りである。2.2 節で関連研究について述べる。2.3 節で提案方式である知識創造サイクルモデルを定義する。2.4 節で、知識創造サイクルモデルの構成方法を示す。2.5 節で本章をまとめる。

2.2 関連研究

本章の研究は、2.1 節で述べた通り、意味の数学モデルを用いた意味的連想検索システム [12] が発想のベースになっている。我々はこれまでに、言葉と言葉の関係の計量による検索機構として、意味の数学モデルによる意味的連想検索システムを提案している。これは、単語群を文脈として解釈する機構により、言葉と言葉、あるいは言葉とメディアデータ、ドキュメント間の相関を文脈に応じて動的に計算することを可能とするモデルである。意味の数学モデルは、潜在的意味インデキシング(LSI) [15] などに代表される、ベクトル空間モデル [16] を応用したモデルの一つである。また、様々なメディアデータを言葉という共通のメタデータとして表現する枠組みである、Media-lexicon Transformation Operator (\mathcal{ML}) [4] の研究を行ってきた。それぞれの \mathcal{ML} は、対象となるメディアコンテンツに対応する心理学などの専門分野における研究成果を用いることにより構成されている。異種の分野のメディアデータに対して \mathcal{ML} をそれぞれ構成し、それらを意味の数学モデルを組み合わせることで、異種のメディアデータをメタレベルで連結することで、相互運用可能なシステムとなる。意味の数学モデルにおける言葉と言葉の関係を計量する空間を生成する際には、“Longman Dictionary of Contemporary English” [17] という英英辞典を利用していた。

本研究においても、異なる分野における知識を相互連結するための、共通計量空間が必要となる。意味の数学モデルもその一つとして利用可能である。一方、この共通計量空間を生成する方法は、その他にも考えられる。WordNet [18] のような言葉の概念の階層構造を規定した概念辞書を用いても、言葉との言葉の類似度を計量することが可能であると考えられる。

また、ニューラルネットワークを用いて単語をベクトルで表現することで、ベクトルの演算によって意味的な関係を計量可能なモデルである word2vec [19] [20] [21] を用いることでも、共通計量空間を構成することができる。

本研究全体の関連研究としては、フィジカル環境を丸ごとサイバー空間に構築することでシミュレーションや分析を可能にする、デジタル・ツイン [22] [23] の研究が挙げられる。

2.3 知識創造サイクルモデルの定義

2.3.1 知識創造と知識利活用の関係

本研究では、人間が知識を創造する仕組みをベクトル空間モデルを使って表現する。人間の知識創造のベクトル空間モデルにおける表現を図 2-1 に示す。

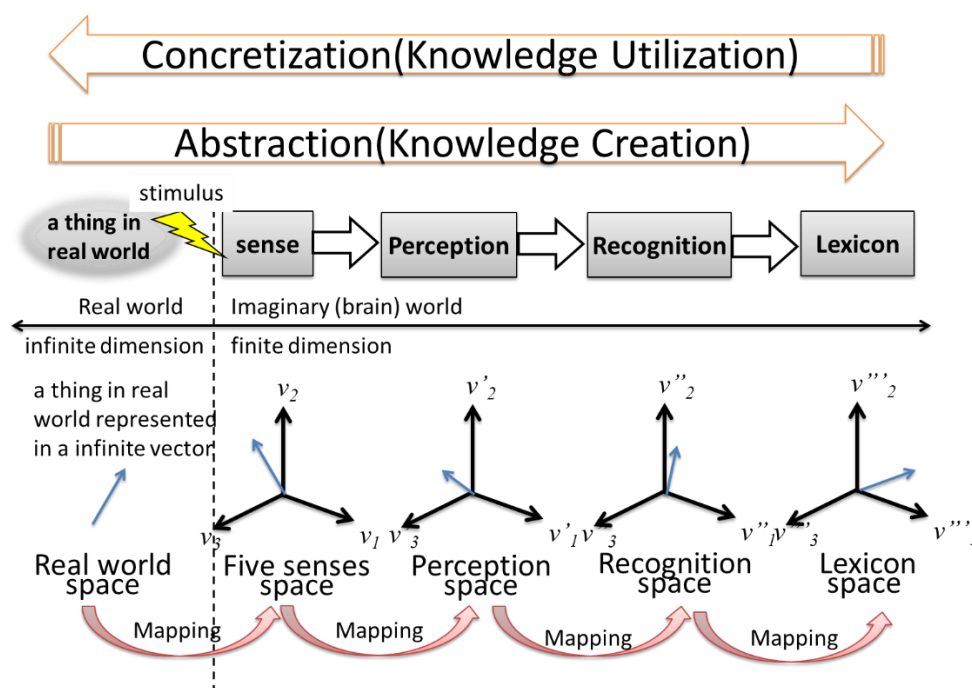


図 2-1 人間の知識創造のベクトル空間モデルにおける表現。

まず、このモデルでは、現実存在するなんらかの事象は、観測される以前には無限の次元のデータとして想定される。これは、ある事象に対しては、人間が観察項目を決める、あるいは観察の粒度を細かくしていくことで、無限のデータが取り出し得ることを意味している。人間は、触覚、味覚、聴覚、視力、嗅覚の五感の感覚を使ってこれらのデータを観察する。観察された事象は、五感を軸とする空間にマッピングされる。これは脳内で感覚(sense)となる。なお実際には、例えば視覚は三色で感じるというように、一つの感覚器で一次元のデータを取得するわけではないので、五感を軸とする空間は五次元よりも多い次元の空間になる。そしてそれらをよりメタな特徴を扱う空間である、認知(perception)の空間、認識(recognition)の空間へと写像していき、最終的に言語(Lexicon)の空間に写像する。この言語の空間で表されたデータであれば、我々は記録することや他者に伝達することが可能になる。大枠として、知識創造は抽象化に相当し、知識利活用は具体化に相当する。

2.3.2 知識創造と知識利活用の定義

本節では、提案モデルにおける知識創造と知識利活用の定義を示す。提案するモデルにおいては、写像は行列で表現する。異なる空間の軸の間の関係を表した行列で写像を表現す

ることによって、ベクトルと行列の計算によって異種の空間上のデータを写像することができる。提案するモデルでは、知識創造は相関行列によって実現される。以上の過程を知識創造と定義する。

続いて提案モデルにおける知識利活用を定義する。知識利活用は、知識創造の逆演算として定義される。それは人間が知識を使って現実の世界に事象を生み出すこと、つまり現実に影響を及ぼすことを意味する。

知識創造(Knowledge Creation)と知識利活用(Knowledge Utilization)の関係を図 2 に示す。知識創造は抽象化(Abstraction)と関係の発見(Finding Relation)に分かれる。知識利活用は、関係の利用(Using Relation)と具体化(Concretization)に分かる。それぞれ、抽象化と具体化が、関係の発見と関係の利用が逆作用素として対応している。

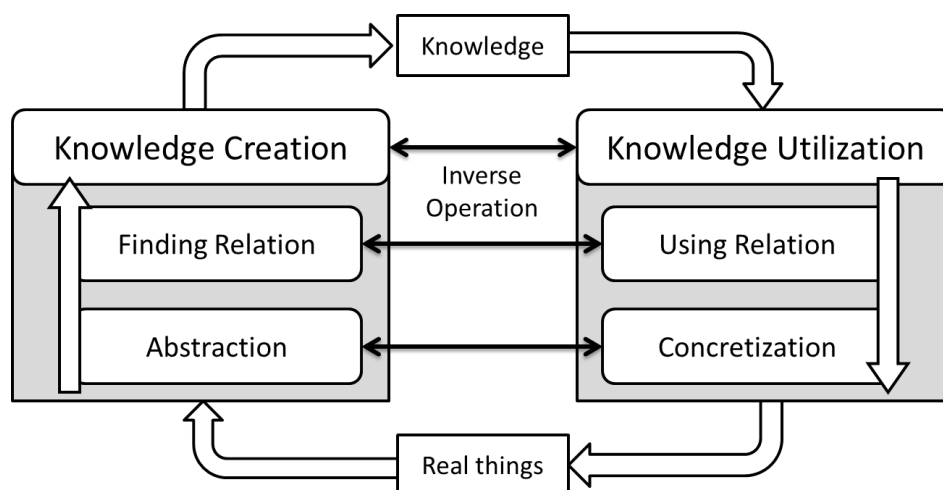


図 2-2 知識創造と知識利活用の関係性.

このように知識創造と知識利活用の演算を揃えることによって、サイクルを作ることができる。ここで、知識を利活用して現実作用させた場合には、想定通りにならないことが発生する。逆に言えば、利活用していない知識は、想定通りになるのかがまだ検証されていない知識である。したがって、より正しい知識を得るためには、知識の利活用をして現実作用させて、現実の変化を観察して、さらに知識を再創造するというサイクルを作ることが重要になる。このサイクルを知識創造サイクルモデルと名付ける。知識創造サイクルモデルの詳細については 2.3.3 節で述べる。

2.3.3 知識創造サイクルモデルの定義

本節では、提案モデルである、知識創造サイクルモデルについて述べる。これは、前節で定義した知識創造と知識利活用を組み合わせるサイクルの形にしたものである。知識創造とは、現実の事象を観察しそこから知識を生み出すことであり、知識利活用とは、知識を用いて現実の世界に何か影響を与えることである。知識創造サイクルモデルの概念図を図 2-3 に示す。

通常、知識を利活用する際には、現実の世界に影響を与えることで、現実の事象に変化を起こそうとする。この際に、達成したかった現実の事象と、実際に起こる現実の事象には多かれ少なかれ誤差が生じる。誤差が大きいことは、知識が不完全であることを意味する。その場合は、誤差を検証し、知識を再創造することで、知識を更新する。

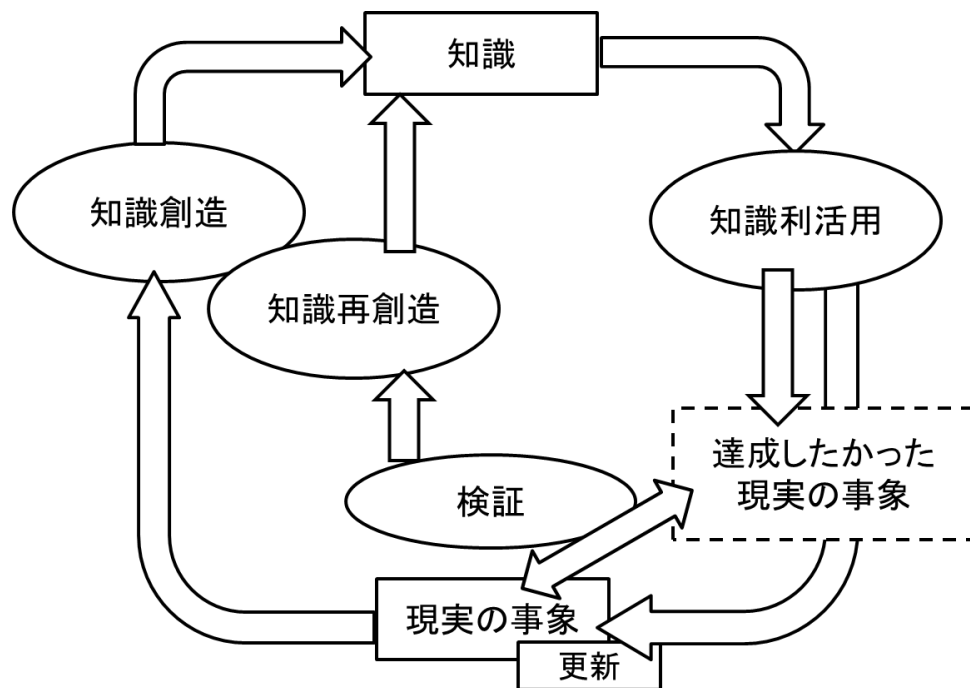


図 2-3 知識創造サイクルモデルの概念図.

2.4 知識創造サイクルモデルの構成

本節では、知識創造と知識利活用を組み合わせることで知識の再創造を行うためのモデルである知識創造サイクルモデルの構成方法を提案する。

2.4.1 センサーとアクチュエーター

現実の事象からデータを取得する装置を一般にセンサーと呼び、目的を実現するために現実に影響を与える装置を一般にアクチュエーターと呼ぶ。知識創造サイクルモデルを構成するためには、センサーとアクチュエーターをもって現実を経由するサイクルが成立していなくてはならない。現実を経由するサイクルとは、すなわち、アクチュエーターを用いて現実に影響を与えた結果をセンサーで取り込み直して、知識に問題がないか検証するプロセスのことである。

着目した特定の分野が、センサーとアクチュエーターの両方を備えている場合は、その分野のみで知識創造サイクルモデルを構成できる。センサーとアクチュエーターを備えた分野を「操作可能(operable)」と呼ぶことにする。しかし、分野によっては、センサーはあるがアクチュエーターは無い、すなわちデータの取得は出来るが、データから得られた知識から目的を現実反映させる手段を持たない場合がある。その場合は、目的を達成するためには、操作可能な他の分野の知識と連動することによって、現実の世界において間接的に影響を与える必要がある。センサーしか持たない分野を「操作不能(inoperable)」と呼ぶことにする。

特定分野で閉じた形で構成される知識創造サイクルモデルを「特定分野に対する知識創造サイクルモデル」、複数の分野を組み合わせることで構成される知識創造サイクルモデルを「複合分野に対する知識創造サイクルモデル」と呼ぶことにする。

2.4.2 知識創造と知識利活用の定式化

本節では、知識創造と知識利活用の構成方法を示す。知識創造は、センサーによる現実世界の事象からの特定分野の特徴ベクトルの抽出と、そのベクトルの Specific-to-General 作用素による一般分野の特徴への変換の二段階からなる。Specific-to-General 作用素は、Specific なデータ、すなわち特定の分野におけるローカルなデータを、General なデータ、すなわち言語などの広く汎用的に用いられる一般的な分野のデータに変換する。知識利活用はその逆演算であり、General-to-Specific 作用素による一般分野の特徴から特定分野の特徴ベクトルへの変換と、アクチュエーターによるその現実世界への反映からなる。

Specific-to-General 作用素と General-to-Specific 作用素のそれぞれを行列として構成する。Specific-to-General 作用素にあたる行列 T は、特定の特徴と一般的な特徴の間の相関関係から構成する。General-to-Specific 作用素にあたる行列 T^* は、行列 T のムーア・ペンローズ一般化逆行列を計算することによって構成される。したがって、2つの行列は数学的に

互いに対となる。これらの関係を図 2-4 に示す。この図はまた、それ自体が特定分野に対する知識創造サイクルモデル」の概要を示した図となっている。

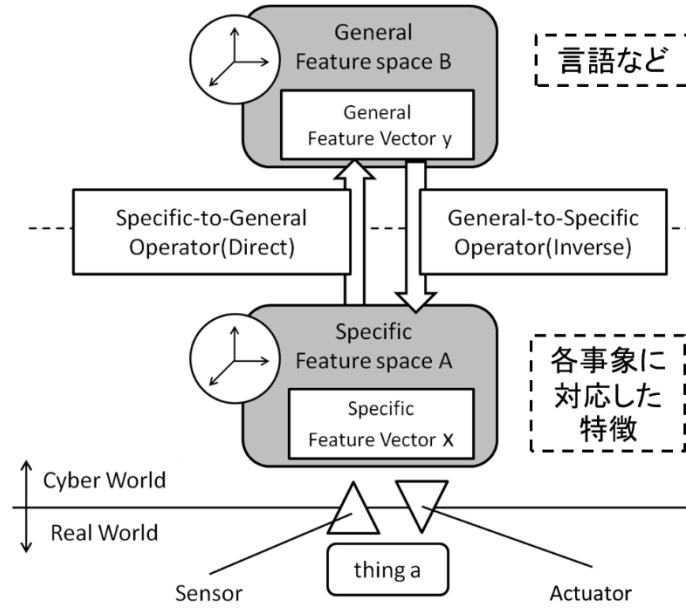


図 2-4 Specific-to-General 作用素と General-to-Specific 作用素.

2.4.2.1 Specific-to-General 作用素

本節では、Specific-to-General 作用素の構成方法について述べる。

§ Specific-to-General 変換行列 T の構成

n 個の要素を持つ特定分野の特徴集合 S_1 を定義する。

$$S_1 = \{fs_i | 1 \leq i \leq n\}. \quad (2-1)$$

m 個の要素を持つ一般分野の特徴集合 S_2 を定義する。

$$S_2 = \{fg_j | 1 \leq j \leq m\}. \quad (2-2)$$

変換行列 T は、特定分野の特徴と一般の特徴との間の相関関係から構成される。 T は $m \times n$ 行列である。

$$T = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mn} \end{bmatrix}, \quad w_{ij} = \text{correlation}(fs_i, fg_j). \quad (2-3)$$

§ 変換行列 T による写像

特定分野の特徴ベクトルを n 次元ベクトル \mathbf{x} で表現する。ベクトルの各要素の添え字は、 S_1 の要素である特徴 fs_i の添え字と対応している。すなわち、ベクトルの各要素は、その特徴についての重みとして表現される。

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T. \quad (2-4)$$

一般分野の特徴ベクトルを、 m 次元ベクトル \mathbf{y} で表現する。特定分野の特徴ベクトルと

同様に、一般分野の特徴ベクトルの各要素は、一般分野の特徴についての重みとして表現される。

$$\mathbf{y} = (y_1, y_2, \dots, y_m)^T. \quad (2-5)$$

特定分野の空間から一般分野の空間への写像は、特定分野の特徴量ベクトル \mathbf{x} と一般分野の特徴量ベクトル \mathbf{y} および変換行列 T を用いて次式のように表される。

$$\mathbf{y} = T\mathbf{x}. \quad (2-6)$$

2.4.2.2 General-to-Specific 作用素(一般化逆作用素)

本節では、General-to-Specific 作用素の構築方法について述べる。これは、Specific-to-General 演算子の逆演算に相当する

§ Specific-to-General 変換行列 T の構成

T のムーア・ペンローズ一般化逆行列 T^\dagger を計算する。 T^\dagger は $n \times m$ 行列となる。

$$T^\dagger = \begin{bmatrix} w'_{11} & \dots & w'_{1n} \\ \vdots & \ddots & \vdots \\ w'_{m1} & \dots & w'_{mn} \end{bmatrix}. \quad (2-7)$$

§ 変換行列 T^\dagger による写像

一般分野の空間から特定分野の空間への写像は、特定分野の特徴量ベクトル \mathbf{x} と一般分野の特徴量ベクトル \mathbf{y} および変換行列 T^\dagger を用いて次式のように表される。

$$\mathbf{x} = T^\dagger \mathbf{y}. \quad (2-8)$$

§ T と T^\dagger による再変換

T^\dagger は T のムーア・ペンローズ一般化逆行列である。この場合、これらの行列は以下の特徴を持つ。

$$\begin{aligned} T^\dagger T \mathbf{x} &= \mathbf{x} \quad (n > m) \quad \dots (a), \\ T T^\dagger \mathbf{y} &= \mathbf{y} \quad (m > n) \quad \dots (b). \end{aligned} \quad (2-9)$$

式(a)と(b)は、ベクトルに対して順演算と逆演算をすることで、元のベクトルに戻るか条件について述べていることになる。多くの場合、特定分野の特徴の数 n は一般的な特徴の数 m よりも大きい。したがって、多くの場合には(b)の性質が成り立つ。これは、 T と T^\dagger による二つの変換による再変換によって、情報が保持されるケースと保持されないケースがあることを示している。情報が保持されないケースについては、保持されている分だけで問題ないのか、そうではないのかを、実際に知識を利活用する際に十分であるかによって判断することになる。本章では2.4.4.1節と2.4.5.1節にてその検証の方式を示すとともに、3章では、楽曲という具体的な事例において検証を行い、情報の保存性について詳しく論じる。

2.4.3 共通計量空間

2.4.1節で述べた通り、センサーとアクチュエーターの両方を備えていない操作不能は分

野では、単独では知識創造サイクルモデルを構成することができない。こうした分野を含む知識創造サイクルモデルを構成するには、二つの条件をそろえる必要がある。一つ目は、操作可能かつ現実世界でその分野の事象に影響を与えることができる別の分野を用意することである。もう一つは、その二つの分野において、General なデータを扱う空間を共通化することによって、データを相互に変換可能にすることである。この複数の分野を扱う際の共通の空間を「共通計量空間」と呼ぶことにする。共通計量空間を構成することによって、異種の分野のデータを相互に接続・変換することが可能になり、複合分野に対する知識創造サイクルモデルを構成可能になる。

2.4.4 特定分野に対する知識創造サイクルモデル

一つの特定期分野が、センサーとアクチュエーター、および変換行列 T と T^\dagger の全てを備えている時、その分野において閉じた形の知識創造サイクルモデルが構成できる。その全体像は、図 2-4 に示したものと同様である。以下に、本モデルを用いて知識を検証する手法について示す。

2.4.4.1 特定分野に対する知識創造サイクルモデルにおける検証演算

特定分野に対する知識創造サイクルモデルを用いて、知識を検証する手法について示す。ユーザーは達成したい目標を一般分野の特徴空間におけるベクトル y として表現する。それを T^\dagger を用いて特定分野の特徴ベクトル x に変換する。そして、アクチュエーターを用いてそれを現実の事象として発生させる。その事象からさらにセンサーを用いて特徴ベクトル x' を取得する。この流れを図 2-5 に示す。

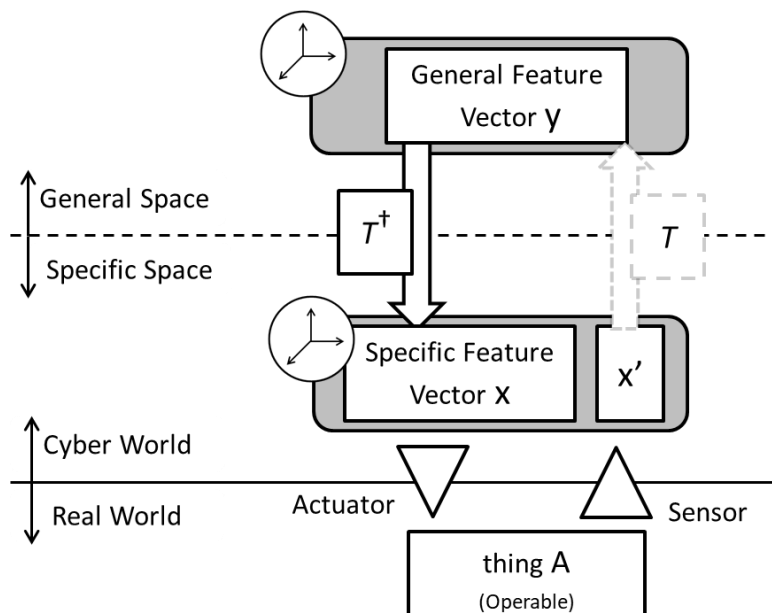


図 2-5 特定分野に対する知識創造サイクルモデルにおける検証演算。

この一連の変換によって構成される x と x' は同じ分野の事象を表現する同じ形式のベク

トルであるため、直接の比較が可能である。この検証の関数を f_{ver} とし、以下に示す。 $sim(,)$ は、2つの任意のベクトル間の類似度を意味し、 ε は任意に設定する閾値を表す。

$$f_{ver}(\mathbf{x}, \mathbf{x}') \begin{cases} true, & sim(\mathbf{x}, \mathbf{x}') > \varepsilon, \\ false, & otherwise. \end{cases} \quad (2-10)$$

類似度は高いほどモデルの性能が高いことを意味するが、目的に応じて必要な性能は変化するため、この ε は目的に応じて変更されるべき値である。

また、この検証によって、モデルの性能が不十分であると判断した場合には、変換行列 T を用いて知識の再創造を行うことが推奨される。

2.4.5 複合分野に対する知識創造サイクルモデル

2.4.1 節で述べた通り、センサーとアクチュエーターを兼ね備えていない操作不能な分野を含む知識創造サイクルモデルを構成する場合には、操作可能かつ現実世界でその分野の事象に影響を与えることができる別の分野と、その二つの分野を相互に接続・変換することが可能な共通計量空間を用意する必要がある。これらを組み合わせて構成される複合分野に対する知識創造サイクルモデルを図 2-6 に示す。

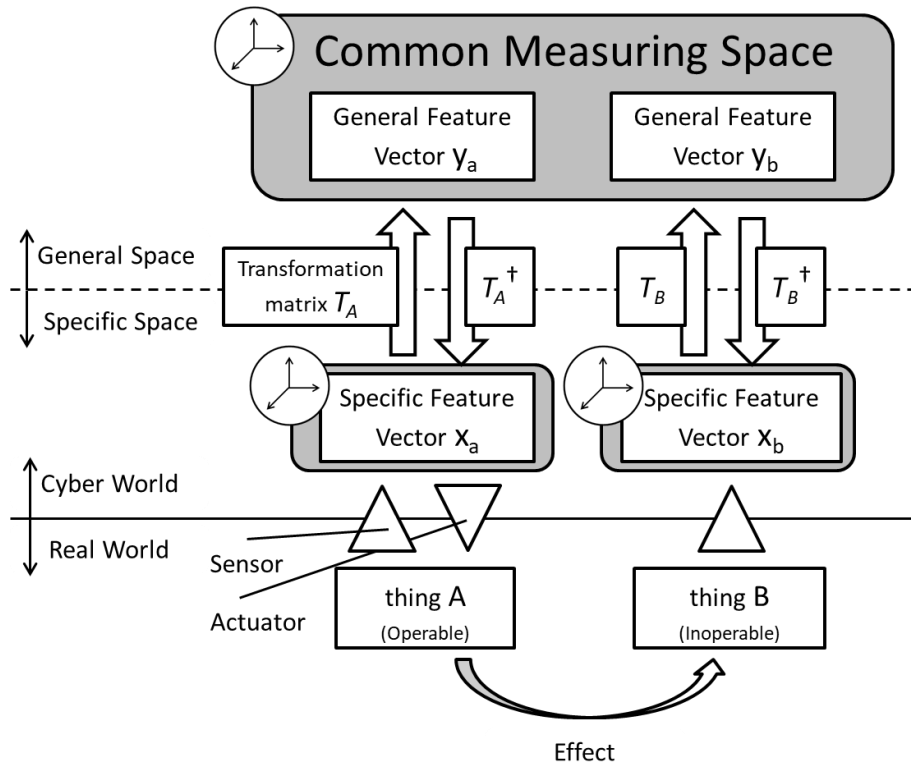


図 2-6 複合分野に対する知識創造サイクルモデル。

2.4.5.1 複合分野に対する知識創造サイクルモデルにおける検証演算

特定分野に対する知識創造サイクルモデルと同様に、複合分野に対する知識創造サイク

ルモデルにも検証演算を定義する．検証の流れを図 2-7 に示す．検証は，二つの変換のルートに分かれて最後に合流するという形になる．

一つ目のルートについて述べる．操作可能な分野を分野 A，操作不能な分野を分野 B とし，この二つの分野を接続したとする．ユーザーは達成したい目標を，操作可能な分野である分野 A おける一般空間におけるベクトル y_a として表現する．これをまず，共通計量空間を用いて，分野 B の一般空間におけるベクトル y_b に変換する．そしてそれを T_B^+ を用いて分野 B における特定分野のベクトル x_b に変換する．この x_b は，既存の知識で再現できる範囲のシミュレーション結果と捉えることができる．

二つ目のルートについて述べる．先ほど定めた y_a を T_A^+ を用いて特定分野の特徴ベクトル x_a に変換する．そして，アクチュエーターを用いてそれを現実の事象 A として発生させる．その事象 A から，別の分野の事象 B に何らかの影響が発生する．事象 B から，今度は別の分野におけるセンサーを用いて特徴ベクトル x'_b を取得する．この x'_b は実際に知識を利活用した際の現実の事象を表していると捉えることができる．

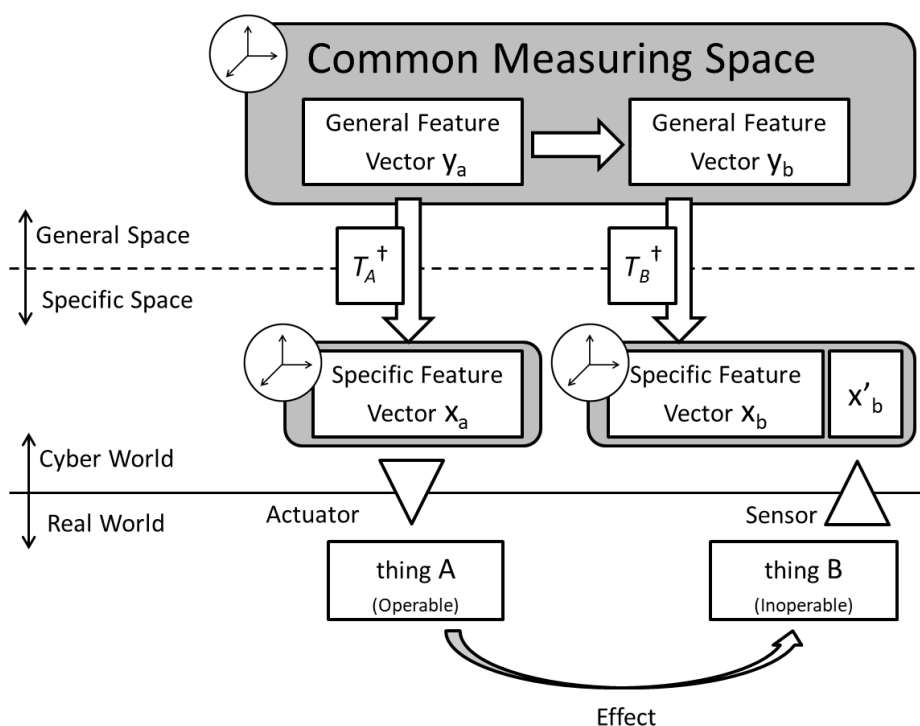


図 2-7 複合分野に対する知識創造サイクルモデルにおける検証演算．

この x_b と x'_b は，特定分野に対する知識創造サイクルモデルと同様に，同じものを表すベクトルとなるため，比較による検証が可能である．検証の演算は式(2-10)に示す関数 f_{ver} を用いて， $f_{ver}(x_b, x'_b)$ と表現される．この場合も特定分野に対する知識創造サイクルモデルと同様に，精度が不足している場合は，知識の再創造をすることが推奨される．

2.5 まとめ

本章では、知識創造と知識利活用を組み合わせることで知識の再創造を行うためのモデルである知識創造サイクルモデルの定義と構成方法を示した。その中で、知識創造と知識利活用をベクトル空間モデルの応用として定義した。知識創造サイクルモデルには、「特定分野に対する知識創造サイクルモデル」と「複合分野に対する知識創造サイクルモデル」の二つがあることを示し、それぞれにおいて、現在の知識が十分であることを検証するための演算を構成できることを示した。

提案モデルである知識創造サイクルモデルは、意味の数学モデルを用いた意味的連想検索システムに発想の基礎を置いている。このシステムでは、様々な分野のメディアデータを言葉という共通のメタデータに変換したうえで、言葉の意味的な類似性を計量するモデルである意味の数学モデルを用いてメタレベルで連結することで、異なるメディアデータを相互運用することを可能にしていた。

本章で提案した知識創造サイクルモデルは、上記の意味の数学モデルを用いた意味的連想検索システムをさらに抽象化したうえで、現実における影響を加味したモデルと言える。扱う対象をメディアデータから知識一般に拡張し、知識創造と知識利活用の関係について定義した。知識利活用は現実に影響を与えることを指す。新たに導入したこの知識利活用の概念によって、現実に影響を与え、その影響を再度観察することで、サイクルが構成できることを示した。具体的には、知識創造を行う作用素と知識利活用を行う作用素の両方を揃えることで、現実が発生した事象と、既存の知識との差異を検証できることを示した。検証によって、既存の知識が不十分であることが分かれば、知識の再創造を促すことができる。以上の流れによって、知識創造サイクルモデルが、知識を検証し再創造を促すモデルであることを示した。

知識創造サイクルモデルは抽象的なモデルであり、具体的な知識を対象とした構成方法までは定めていない。本論文では、3章において、メディアデータを対象とした知識創造サイクルモデルの実現例として *ML-iML* 系の構成方法を示し、さらにその具体的な適用例として、楽曲メディアデータを対象とした知識創造サイクルモデルの例を示している。3章では、音楽という特定分野に対する知識創造サイクルモデルを構成するとともに、3.6節では、複合分野に対する知識創造サイクルモデルの事例である、任意の言葉の音相の印象に合致した楽曲の自動構成方式も示している。ただしこちらは、まだ完全にはサイクルが閉じておらず、その実現は今後の課題となっている。4章では、会議の中で話された言語情報を対象とした、知識創造にあたる研究を示している。こちら、まだ実現してはいないが知識利活用によるフィードバックを含めたサイクルを構成し、現実との整合性を重視したシステムを作ることを目標としている。

第3章 メディアコンテンツを対象とした統計的一般可逆作用素構成方式とその楽曲メディアコンテンツ生成への適用

3.1 はじめに

近年、コンピュータが人間の生活する環境に遍在するようになり、コンピュータとのコミュニケーションにおける人間の感性が着目されつつある。これまでのコンピュータと人間の論理情報の伝達だけでは、その操作による人間の負荷が大きくなっており、人間の感性や直感に合致した、ユーザーへの負担が少ないコミュニケーションメディアの実現が重要な課題となってきている。コンピュータと人間がより密接に関わるようになっていくにつれて、人間の感性に基づく情報の獲得方式やコンピュータによる感性の表現方式は一層重要度を増していくことが予想される。

我々のコミュニケーションにおいて、互いの感情を正確に理解することは重要であり、言葉のほか、顔の表情、ジェスチャなど非言語行動が様々な情報を相手に効果的に伝えるメディアとして重要な役割を担っている。このような感性的な非言語行動にあたるマルチメディアコンテンツが人間とコンピュータの間におけるコミュニケーションメディアとして導入されれば、人間の感性や直感に合致した、ユーザーへの負担が少ないインターフェイスの実現が可能となると考えられる。

我々はこれまでに、様々なメディアを対象として、そのメディアコンテンツが持つ印象や感性的な情報を、言葉のメタデータとして自動抽出する枠組みである Media-lexicon Transformation Operator(\mathcal{ML}) [4]の研究を行ってきた。

それぞれの \mathcal{ML} は、対象となるメディアコンテンツに対応する心理学などの専門分野における研究成果を用いることにより構成されている。“ \mathcal{ML} ”によって抽出される、メディアコンテンツの持つ印象や感性的な情報を、以下では“印象メタデータ”と呼ぶ。印象メタデータは、単語とその重みの組によって構成されている。

これまでの“ \mathcal{ML} ”の実現例としては、楽曲を対象としたもの [24] [25]、画像の色を対象としたもの [10]、“音相”と呼ばれる言葉の発音情報を対象としたもの [26]などがある。

本研究では、上記のようなメディアコンテンツから印象メタデータを自動抽出する手法とは逆に、印象メタデータを入力として、メディアコンテンツを自動生成する新たな枠組みである、メディアコンテンツを対象とした統計的一般化逆作用素“Stochastic Generalized Inverse Media-lexicon Transformation Operator”(“ $i\mathcal{ML}$ ”)を定義する。統計的一般化逆作用素 $i\mathcal{ML}$ は、 \mathcal{ML} の逆演算を一般化したものである。

そもそも、メディアコンテンツから印象メタデータを自動抽出する方式の逆演算によって印象語を入力としたコンテンツ自動生成システムを構成する際には、印象語を表す次元数とメディアコンテンツの特徴を表す次元数は異なることがほとんどであり、その次元数

の不一致から、逆作用素を構成した際に、不確定要素が発生する。

この不確定要素を、一般化作用素を生成する際に発生する不良設定問題と呼ぶ。そこで、不確定要素となった部分を対象とするメディアに関連する統計情報やそのメディアを対象とした研究成果・理論を導入することにより、対象メディア依存の制約条件を設けることが出来る。その制約条件によって、不確定要素を最小限に抑えながら、印象を表したメディアコンテンツを生成することが可能となる。メディアコンテンツ生成の過程に表れる不良設定問題を対象とするメディアに関連する統計情報を導入することによって解決することを本稿では統計的(“stochastic”)と名付けている。

本章では、 iML の適用事例として、楽曲メディアコンテンツ生成を対象とする。本方式の内容の大部分は [8]にて報告されている。本方式は、音楽心理学者 Hevner の研究を用いて、楽曲メディアに対する ML と、その逆演算である iML を構成する。Hevner の研究 [5] [6] [7]では、楽曲構造要素として調性(key)・テンポ(tempo)・音高(pitch)・リズム(rhythm)・和声(harmony)・旋律(melody)の6つを挙げており、この6つの楽曲構造要素と8つの印象語群によって表現される印象との相関関係を調べ、表にまとめられている。Hevner の研究は楽曲の構造要素と、それに対して人間の抱く感情との対応関係を調べたものである。楽曲メディアに関する ML と、その逆演算として構成された iML を組み合わせることで、入力した楽曲メディアコンテンツの持つ印象の抽出、さらに印象を表す印象語からそれに合致した楽曲メディアコンテンツとして構成する双方向のシステムが実現できる。

ML と iML を組み合わせることで、お互いを検証する系である ML - iML 系を構成することができる。この ML - iML 系の構成方法を示すとともに、それによってどのような検証が可能になるのかについても示す。そして実験によって、実際に検証を行う。

本研究における ML - iML 系は、2章で提案した知識創造サイクルモデルのメディアデータにおける具体的な実現例である。また、楽曲を生成するシステムは、 iML の具体的な実現例である。

さらに、構成した楽曲自動生成システムを、別のメディアと連結する事例を示す。本方式の内容は [27]にて報告されている。我々はこれまでに、言葉と言葉の関係の計量による検索機構として、意味の数学モデルによる意味的連想検索を提案している [12]。これは、単語群を文脈として解釈する機構により、言葉と言葉、あるいは言葉とメディアデータ、ドキュメント間の相関を文脈に応じて動的に計算することを可能とするモデルである。また、メディアデータから人間の心理的要素が反映されたメタデータを抽出する方式として、「音相理論 [28]」と呼ばれる理論を用いて、日本語の音の印象に合致したメタデータを抽出する方式 [26]を実現している。

本研究では、この音相理論を用いた言葉の音の印象を抽出するシステムと、楽曲自動生成システムを意味の数学モデルに代表される意味的連想検索機構を用いてメタレベルで連結することにより、日本語の音の印象を表す楽曲を生成するシステムを構築する。本方式は、任意の日本語をコンテキストとした楽曲の自動構成を可能とする。これにより利用者の意

図する印象をより直感的に反映させた楽曲の構成が実現され、音楽を通じた情報の新しい価値を生み出すことが可能になる。さらに、未知語や擬音語など、普段我々がその言葉を持つ印象を意識しない言葉でさえ、その印象を表した楽曲を構成することができるので、人間の感性に新たな発見を見出す手助けになると考えられる。こうした連結方式を実現することは、独立に実装されたメディア情報の相互運用性を実現し、既存のデータベース群の利用価値を高め、新たな価値の創造が可能になると考えられる。

3.1.1 3章の構成

本章の構成は以下の通りである。3.2節では、関連研究を示す。3.3節では、提案方式である iML および $ML-iML$ 系の定義について述べる。3.4節では、 iML の具体的な事例として、楽曲メディアコンテンツを対象とした統計的一般化逆作用素 iML の構成方式について述べる。これは楽曲自動性システムとなる。3.5節では、楽曲メディアコンテンツにおける $ML-iML$ 系を用いた検証方法について述べる。3.6節では、3.4節で構成した楽曲自動生成システムと、日本語の音相の印象を抽出するシステムを組み合わせ、任意の日本語の入力語の音相の印象を表現する楽曲を生成するシステムの構成について述べる。3.7節では、実験を行い、3.4節、3.5節、3.6節のそれぞれの内容について検証を行う。3.7節で本章をまとめる。

3.2 関連研究

本研究は、メディアコンテンツを対象とした統計的一般化逆作用素”Stochastic Generalized Inverse Media-lexicon Transformation Operator”(“ iML ”)構成方式を示しているが、本研究で対象メディアとしている楽曲に焦点を絞ると、自動作曲システムの一つと位置付けることができる。

自動作曲の現状については、文献 [29]が明るい。

自動作曲のはじまりは、1957 年に発表された「イリアック組曲」 [30]であると言われている。これは、音高のマルコフ連鎖と乱数生成によって音高列を決定するモジュールと生成結果が妥当であるかを検証するモジュールで構成されるものである。妥当性の判断は音高同士の不協和回避ルールというルールベース的なアプローチで行われる。

現在、自動作曲システムについては様々なアプローチが提案されている。音楽には和声法や対位法などの方法論や理論が存在することから、これらをルールとして計算機上に実装することによって作曲システムを構築する、ルールに基づくアプローチが存在する。例えば、旋律生成システムの Ebcioğlu の研究 [31]が、それにあたる。

また、遺伝的アルゴリズムを用いたアプローチも提案されている [32] [33]。実際にユーザーに聴かせたときの評価を用いることで、自然かつ嗜好に合った旋律を得ていくというものである。これらのアプローチの欠点としては、ユーザーの負担が大きいことが問題として挙げられている。

一方で、既存の楽曲の断片を事例として部分的に用いて、組み合わせて用いる手法 [34]も提案されている。

さらに、近年では確率モデルを用いたアプローチも提案されている。例えば、白井らの研究 [35]では、旋律の音の遷移の偏りを階層 Pitman-Yor 言語モデルを用いて学習し、そのモデルから旋律を生成している。また、深山らの研究 [36]では、日本語歌詞から、音楽要素の組み合わせによる制約条件の下で、動的計画法の音高系列の探索問題として作曲を定式化することで歌詞の韻律に合った旋律を生成することを実現している。

近年、ユーザー視点に立ち、ユーザーとインタラクションを考え、作曲の創作支援を考える実用的な研究も増えてきている。例えば、Tsuchiya らの研究 [37]では、旋律の大まかな音の流れを曲線で入力させることにより、旋律を作曲させる手法がある。この研究の場合は、ユーザー意図をどのように伝えるかというユーザーインターフェイスを研究の主眼としている。また、北原ら [38]は、生成結果をユーザーがどう修正するかということに着目し、システムが生成した楽曲をシステムの支援を受けながら編集できる human-in-the-loop 型の自動作曲というものを提案している。

本研究で提案する我々の手法は、上記の手法とは異なり、そのメディアコンテンツを持つ印象や感性的な情報を、言葉のメタデータとして自動抽出する枠組みである ML の逆作用素 iML を構成することによって自動作曲を行うシステムである。本方式では、音楽心理

学者 Hevner の研究 [5] [6] [7]を用いる。Hevner の研究は、楽曲構造要素というメディアコンテンツから抽出しうる特徴量と人間の抱く感情の関係を心理学的に調査した研究である。また、Hevner は、クラシックのピアノ曲に対して調査を行っているため、本システムの適用範囲はクラシックのピアノ曲に限定される。

本研究では、音楽心理学者 Hevner の研究を元として、統計的一般化逆作用素 $i\mathcal{ML}$ を構成する。 $i\mathcal{ML}$ を用いることで、印象メタデータを入力とし、その印象に対応する楽曲構造要素を持つ楽曲を生成することで、入力した印象を表現するクラシックピアノ曲を生成する自動作曲システムとなる。

3.3 メディアデータを対象とした一般化逆作用素

本節では、対象とする任意のメディアデータから人間が受ける印象を抽出し、言葉によって表現されるメタデータを自動的に抽出する方式である、Media lexicon Transformation Operator(\mathcal{ML})[5]と、提案方式であり、 \mathcal{ML} の逆演算である、印象を表現したメタデータからメディアデータを生成する統計的一般化逆作用素 Stochastic Generalized Inversed Media-Lexicon Transformation Operator($i\mathcal{ML}$)の性質と構成方式を述べる。この時、 \mathcal{ML} と $i\mathcal{ML}$ は対になるよう構成し、これらを組み合わせて \mathcal{ML} - $i\mathcal{ML}$ 系を構成する。 \mathcal{ML} は、メディアデータからの特徴の抽出と、特徴量を印象語群のベクトルへと変換する線形作用素によって構成されている。 \mathcal{ML} の逆演算においては、まず順演算(\mathcal{ML} による変換)で用いた線形作用素の逆作用素を用いて印象メタデータをメディアデータの特徴量に変換する。そこから先は不良設定問題となるため、他の知識を制約条件として導入する事により要素の配置を行い、メディアデータを生成する。ここで制約条件として各種メディアに対応する統計的な研究成果を用いることにより、メディアデータはメタデータの情報(特徴や印象)を保ちはするが、一意には定まらず、毎回違う物が生成される。以上を実現する逆作用素を、統計的一般化逆作用素“Stochastic Generalized Inversed Media-Lexicon Transformation Operator”(“ $i\mathcal{ML}$ ”)と呼ぶ事とする。 \mathcal{ML} と $i\mathcal{ML}$ の概略図 3-1 をに示す。

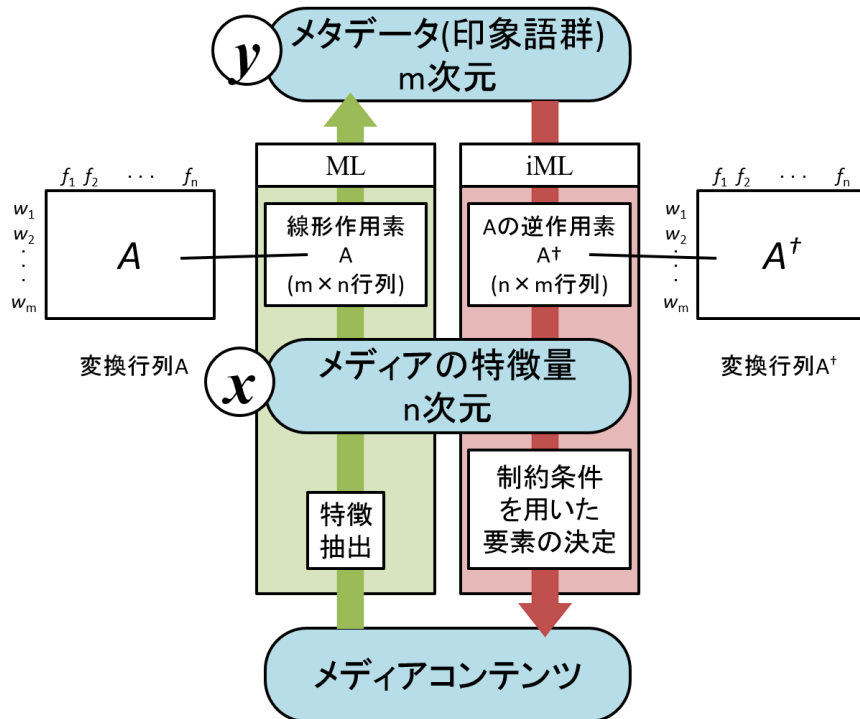


図 3-1 \mathcal{ML} と $i\mathcal{ML}$ の概略図.

ここで、 $i\mathcal{ML}$ は \mathcal{ML} の逆演算であるため、その性質を利用してお互いの演算を検証する事が出来る。例として、印象メタデータから $i\mathcal{ML}$ によってメディアデータを構成し、メ

ディアデータから \mathcal{ML} を用いて印象抽出を行い、生成元と等しい印象メタデータが得られた時、構成の演算を抽出の演算で検証し、正しい事を示した事となる。このようにして、順演算と逆演算を組み合わせ、お互いの検証を行う系を \mathcal{ML} - $i\mathcal{ML}$ 系と呼ぶ事とする。 \mathcal{ML} - $i\mathcal{ML}$ 系は閉じた系となり、その中に保存される情報を明示する事が出来る。3.3.1 節では \mathcal{ML} について概説し、3.3.2 節では、 $i\mathcal{ML}$ の性質と構成方式を述べる。3.3.3 節では \mathcal{ML} - $i\mathcal{ML}$ 系の構成方法と、その性質について述べる。

3.3.1 Media-Lexicon Transformation Operator(\mathcal{ML})

Media-lexicon Transformation Operator(\mathcal{ML})は、対象とするメディアデータに関する分野の専門家による研究や評論、統計などを用いることにより、人間がそのメディアデータから受ける印象を表す単語の抽出を実現する機構である。 \mathcal{ML} はもともと、言葉同士の相関を計量する機構とセットで考案されたものであり、様々な種類のメディアデータを言葉という統一的なメタデータで表現し、さらにそれらのメタデータ同士の関係を計量する機構と組み合わせて、メディアデータの分野をまたいだ統一的な操作を行う事を目指している。 \mathcal{ML} の概要を図 1 に示す。 \mathcal{ML} は一般に次のように表される。

$$\mathcal{ML}(Md): Md \mapsto Ws. \quad (3-1)$$

(Md : メディアデータ, Ws : (重み付き) 単語群)

以下に、 \mathcal{ML} の一般的な構成方法について示す。

(1) 専門家の研究による変換行列

A の作成メディアデータから抽出可能な特徴量(画像であれば色彩情報など、楽曲であればテンポ・音程など)と印象語の関係を表す研究(カラーイメージスケール[8], Hevner の研究[1]~[3]など)を用いて、変換行列 A を作成する。変換行列 A は、特徴量と印象語の関係を表現したものとなり、特徴を m 個、印象語を n 個とする時、 $m \times n$ 行列となる。ここで、 $w_1 \sim w_m$ は印象語、 $f_1 \sim f_m$ は特徴として、変換行列 A を図 3-1 のように作成する。

(2) メディアデータから特徴量ベクトル \mathbf{x} を抽出

対象のメディアデータを解析し、メディアデータの印象を決定する特徴量を抽出する。ある特徴量ベクトルを \mathbf{x}_a とし、以下のように示す。

$$\mathbf{x}_a = (f_{1a}, f_{2a}, \dots, f_{na})^T. \quad (3-2)$$

\mathbf{x} は印象語 $f_1 \sim f_n$ にそれぞれ重みを付けた n 次元のベクトルとなる。

(3) 変換行列 A と特徴量ベクトル \mathbf{x}_a から印象語ベクトルを抽出

抽出した特徴量ベクトルと変換行列 A により、メタデータ \mathbf{y}_a を抽出する。

$$\begin{aligned} \mathbf{y}_a &= A\mathbf{x}_a \\ &= (w_{1a}, w_{2a}, \dots, w_{na})^T. \end{aligned} \quad (3-3)$$

\mathbf{y} は印象語 $w_1 \sim w_m$ にそれぞれ重みを付けた m 次元のベクトルとなる。

3.3.2 Stochastic Generalized Inversed Media-Lexicon Transformation Operator($i\mathcal{ML}$)

本節では提案方式である、 \mathcal{ML} の一般化逆作用素である Stochastic Generalized Inversed Media-Lexicon Transformation Operator($i\mathcal{ML}$)の構成方法について述べる。専門家により選ばれた印象語群とその重みを入力として \mathcal{ML} の逆演算を行うことによってメディアデータの特徴量を決定し、この特徴量に基づいてメディアデータを自動的に構成することを考える。本研究では、この逆演算を行う作用素を逆作用素と呼ぶこととする。 \mathcal{ML} とその逆作用素は、対をなす存在であると言える。 \mathcal{ML} の逆作用素を、Stochastic Generalized Inversed Media-Lexicon Transformation Operator($i\mathcal{ML}$)としてここでは定義する。

$i\mathcal{ML}$ は一般に次のように表される。

$$i\mathcal{ML}(Ws): Ws \mapsto Md \quad (3-4)$$

(Md : メディアコンテンツ, Ws : (重み付き) 印象語群)

逆作用素は、2.1 節に示した \mathcal{ML} の逆演算により構成される。 $i\mathcal{ML}$ によるメディアデータ生成の概要を図 3 に示す。 $i\mathcal{ML}$ は、印象語から特徴量への変換と、特徴量からメディアデータへ変換という 2 段階で構成される。印象語から特徴量への変換は、 \mathcal{ML} における変換行列 A の一般化逆行列を構成する事で実現する。特徴量からメディアデータへの変換は、 \mathcal{ML} における特徴抽出の逆演算に対応するが、その演算は一般に不良設定問題となる。ここで、対象とするメディアにおける統計情報を新たに制約条件として導入し、不良設定問題を収束させる。用いる統計情報は目的に応じて変更可能である。

\mathcal{ML} の逆演算である $i\mathcal{ML}$ を構成することにより、印象メタデータからメディアコンテンツの生成を実現する。

3.3.2.1 $i\mathcal{ML}$ を用いたメディアデータの構成方法

$i\mathcal{ML}$ を用いたメディアデータの構成方法の手順について詳細を述べる。

(1) 逆変換行列 A^\dagger の生成

図 2 に示した変換行列 A から、ムーア・ペンローズ一般化逆行列を計算し、逆変換行列 A^\dagger を導出する。ここで、 A を $m \times n$ 行列とすると、 $m > n$ の場合と $m < n$ の場合で逆演算の性質が異なる。性質の詳細については 3.3.3.2 節にて述べる。逆変換行列 A^\dagger は、図 3-1 中に示されている。

(2) \mathbf{y} からの特徴量ベクトル \mathbf{x} の抽出

ある印象語のベクトルである \mathbf{y}_b と、 A^\dagger を用いて、特徴量ベクトル \mathbf{x}_b を導出する。これは 3.3.1 節の式(3-3)の逆演算に相当する。

$$\begin{aligned} \mathbf{x}_b &= A^\dagger \mathbf{y}_b \\ &= (f_{1b}, f_{2b}, \dots, f_{nb})^T. \end{aligned} \quad (3-5)$$

ここでも $m > n$ の場合と $m < n$ の場合で性質が異なる。性質の詳細については 3.3.3.2 節にて述べる。

(3) 特徴量ベクトルと統計情報に基づくメディアデータの要素の構成

式(3-5)により得られた特徴量に基づき、メディアデータの構成要素（楽曲であればテンポや音符の配置など、言葉の音であれば子音と母音の組み合わせやその並びなど）を決定していく。しかしながら、この問題は一般に不良設定問題となる。なぜなら、元は異なるメディアデータであっても、ほぼ同様のメタデータが得られることはあり得るからである。そのためここでは、得られた特徴群に加えて、メディアデータごとの統計的知識に基づいて適切な制約条件を与えることにより、自動生成を可能な限り適切に終了させる。これらの制約条件は、対象とするメディアデータ毎にそれぞれ個別に考案する必要がある。

特徴量からメディアコンテンツへの変換は、 \mathcal{ML} における特徴抽出の逆演算に対応するが、その演算は一般に不良設定問題となる。一般化逆行列のような一意な逆演算のみでメディアコンテンツを生成できればよいのだが、後半の特徴量からメディアコンテンツへの変換については自由度が大きく、一意なコンテンツを作成することができない。そこで、対象とするメディアにおける統計情報やそのメディアを対象とする研究成果・理論を新たに制約条件として導入することで、自由度を抑え、メディアコンテンツの生成を可能とする。

特徴量からメディアコンテンツを生成する段階において、ルールに基づくアプローチである、例えば旋律生成システムの Ebcioğlu の研究 [31]などを応用することは可能である。

$i\mathcal{ML}$ による印象メタデータから特徴量への変換によって、印象に合致した特徴量を抽出することができ、その特徴量から、そのメディア依存のルールなどによって、メディアコンテンツの要素を確定し、メディアコンテンツを構成することができる。本方式によって、印象メタデータからその印象に合致するようなメディアコンテンツを構成可能となる。

3.3.3 \mathcal{ML} – $i\mathcal{ML}$ 系の構成と検証

\mathcal{ML} と $i\mathcal{ML}$ を組み合わせて \mathcal{ML} – $i\mathcal{ML}$ 系を構成する。 \mathcal{ML} と $i\mathcal{ML}$ による変換をお互いの検証に用いることで、閉じた系内での保存性を保証するモデルとなる。本節では、 \mathcal{ML} – $i\mathcal{ML}$ 系を用いた検証方法について示す。

3.3.3.1 \mathcal{ML} – $i\mathcal{ML}$ 系の定義

\mathcal{ML} と $i\mathcal{ML}$ をそれぞれ 2 段階の操作として表現すると以下の様になる。

$$\begin{aligned} \mathcal{ML}: \\ MF(Md): Md \mapsto Fs, \\ FL(Fs): Fs \mapsto Ws. \end{aligned} \quad (3-6)$$

$$\begin{aligned} i\mathcal{ML}: \\ iFL(Ws): Ws \mapsto Fs, \\ iMF(Fs): Fs \mapsto Md. \end{aligned} \quad (3-7)$$

(Md : メディアデータ, Fs : 特徴セット, Ws : (重み付き) 単語群)

ここで, それぞれ MF にとって iMF が逆演算に, FL と iFL が逆演算に相当する. これらの演算全てを備えた系を \mathcal{ML} - $i\mathcal{ML}$ 系と定義する. これらが逆演算であるということとは, 一般に以下 (a) ~ (d) の性質を満たすことが期待される. また, これらの関係を図 3-2 に示す.

$$\begin{aligned} iMF(MF(Md)) &= Md, & \dots (a) \\ MF(iMF(Fs)) &= Fs, & \dots (b) \\ iFL(FL(Fs)) &= Fs, & \dots (c) \\ FL(iFL(Ws)) &= Ws. & \dots (d) \end{aligned} \quad (3-8)$$

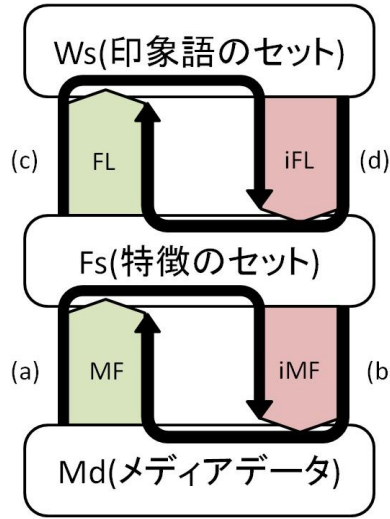


図 3-2 \mathcal{ML} と $i\mathcal{ML}$ に含まれる変換の関係.

しかしながら \mathcal{ML} と $i\mathcal{ML}$ においては, これらすべてを実現することはできない. \mathcal{ML} と $i\mathcal{ML}$ で実現されるのは, (a) または (b), そして (c) または (d) の性質のみとなる. (a) と (b) については, 通常想定されるシステムにおいては (b) が成り立つ. これは, メディアデータの持つ情報量が一般に可変であり, 無限に大きな情報量を持ち得るのに比べて, 抽出によって得られる特徴量は決まった情報量を持つからである. (b) が満たされるという事は, 特徴量から生成したメディアデータが同じ特徴量を持つという事である. (c) が成り立つという事は, 特徴量と印象の変換において, 特徴量の方が不変であるという事であり, (d) が成り立つ場合は印象の方が不変であるという事である. (c) が満たされる時の \mathcal{ML} - $i\mathcal{ML}$ 系を, “特徴保存の \mathcal{ML} - $i\mathcal{ML}$ 系”, (d) が満たされる時を, “印象保存の \mathcal{ML} - $i\mathcal{ML}$ 系”と呼ぶことと

する。(c)または(d)のどちらの性質が満たされるかの条件については、次節で詳細を述べる。

したがって、本研究で目指す $i\mathcal{ML}$ の構築とは、(b)および(c)または(d)の性質を満たす逆作用素を構築する事となる。本稿における適用事例として3節以降で実現する $i\mathcal{ML}$ は、(c)の性質を満たすため、本稿では特徴保存の \mathcal{ML} - $i\mathcal{ML}$ 系を実現する事が目標となっている。

3.3.3.2 \mathcal{ML} - $i\mathcal{ML}$ 系の保存性

変換行列 A から、ムーア・ペンローズ一般化逆行列を計算し、逆変換行列 A^\dagger を導出する。この時、 A^\dagger は広義可逆元となり、以下の性質を持つ。

$$AA^\dagger A = A. \quad (3-9)$$

$$A^\dagger AA^\dagger = A^\dagger. \quad (3-10)$$

また、 AA^\dagger 、 $A^\dagger A$ はエルミート行列となり、以下の性質を持つ。なお、行列 A の随伴行列を A^* と表す。

$$(AA^\dagger)^* = AA^\dagger. \quad (3-11)$$

$$(A^\dagger A)^* = A^\dagger A. \quad (3-12)$$

ここで、 A を $m \times n$ 行列とすると、 $m > n$ の場合と $m < n$ の場合で逆演算の性質が異なる。 $m > n$ かつ、各列が線形独立であるとき、 A^*A は正則となり、逆元を持つ。この時 A^\dagger は以下のように表現される。

$$A^\dagger = (A^*A)^{-1}A^*. \quad (3-13)$$

これを变形すると、 A^\dagger は A の左逆元となる事が分かる。

$$A^\dagger A = (A^*A)^{-1}A^*A = I_n. \quad (3-14)$$

反対に $m < n$ かつ、各行が線形独立であるとき、 AA^* が正則となり、 A^\dagger は A の右逆元となる。

$$AA^\dagger = AA^*(AA^*)^{-1} = I_m. \quad (3-15)$$

ここで、印象語のベクトルを \mathbf{y} 、特徴量のベクトルを \mathbf{x} とする。行列 A が $m > n$ のとき(行列 A^\dagger が $n < m$ のとき)、以下が成り立つ。

$$\mathbf{x} = A^\dagger \mathbf{y} = A^\dagger A \mathbf{x}. \quad (3-16)$$

すなわち、 $\mathbf{x} \rightarrow \mathbf{y} \rightarrow \mathbf{x}$ と変換した場合は、元のベクトルと等しくなる。 $\mathbf{y} \rightarrow \mathbf{x} \rightarrow \mathbf{y}$ と変換した場合は、一般には元に等しくなる事が保証されない。また、 A が $m < n$ のとき、以下が成り立つ。

$$\mathbf{y} = A \mathbf{x} = AA^\dagger \mathbf{y}. \quad (3-17)$$

したがってこの条件では $\mathbf{y} \rightarrow \mathbf{x} \rightarrow \mathbf{y}$ と変換した場合は元のベクトルと等しくなる。

これらのことをまとめると、メタデータのベクトルの次元を n 、メディアデータの特徴量ベクトルの次元を m としたとき、 $m > n$ のとき“特徴保存の \mathcal{ML} - $i\mathcal{ML}$ 系”が実現され、 $m < n$ のとき“印象保存の \mathcal{ML} - $i\mathcal{ML}$ 系”が実現されることが分かる。

これまでの研究で実現されてきたいくつかの \mathcal{ML} について、 n と m の関係を表 3-1

に示す.

表 3-1 各 \mathcal{ML} の性質.

\mathcal{ML}	n(特徴)	m(印象)	性質
楽曲 (Hevner)	6	8	特徴保存
カラーイメージスケール[8]	180	130	印象保存
音相理論[7]	78	20	印象保存

3.4 楽曲メディアコンテンツを対象とした統計的一般化逆作用素 iML の構成

本節では、3.3 節で示した iML の具体的事例として、印象メタデータから楽曲メディアコンテンツを自動生成する方式を示す。

3.4.1 節では、対象とする楽曲コンテンツに関する分野の専門家による研究として用いる、Hevner の研究について述べる。3.4.2 節では、3.4.1 節で述べた Hevner の研究を用いて楽曲からメタデータを自動抽出する方式、つまり、楽曲コンテンツを対象とした ML について述べる。さらに、3.4.3 節で、印象メタデータから楽曲メディアコンテンツを生成する方式、つまり、本研究の提案方式である、楽曲メディアコンテンツ自動生成方式(iML)について示す。

3.4.1 Hevner の研究

Hevner の研究 [5] [6] [7]では、楽曲メディアにおける特徴量である楽曲構造要素として調性(key)・テンポ(tempo)・音高(pitch)・リズム(rhythm)・和声(harmony)・旋律(melody)の6つを挙げている。Hevner は、この6つの楽曲構造要素と8つの印象語群によって表現される印象との相関関係を調べ、表にまとめた(表 3-2)。8つの印象語群は、印象語間で類似性があるものをまとめて一つの印象語群を作り、さらに印象語群間で類似性があるものを隣接するよう、円形に配置している(図 3-3)。

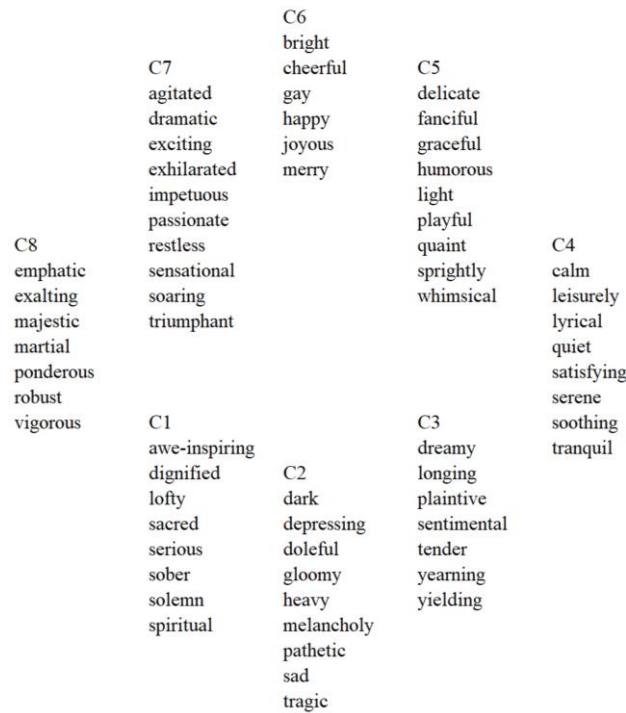


図 3-3 Hevner による 8 つの印象語群.

表 3-2 Hevner による楽曲構造要素と楽曲印象語群の関係性.

印象語群名 構造要素名	c1	c2	c3	c4	c5	c6	c7	c8
key	長調 4	短調 12	短調 20	長調 3	長調 21	長調 24	---	---
tempo	遅い 14	遅い 12	遅い 16	遅い 20	速い 6	速い 20	速い 21	速い 6
pitch	低い 10	低い 19	高い 6	高い 8	高い 16	高い 6	低い 9	低い 13
rhythm	固定 18	固定 3	流動 9	流動 2	固定 8	流動 10	固定 2	固定 10
hamony	単純 3	複雑 7	単純 4	単純 10	単純 12	単純 16	複雑 14	複雑 8
melody	上昇 4	---	---	上昇 3	下降 3	---	下降 7	下降 8

3.4.2 楽曲メディアコンテンツを対象としたメタデータ自動抽出方式(\mathcal{ML})

前節で述べた Hevner の研究を用いて楽曲からメタデータを自動抽出する方式について述べる. 詳細は [24] に述べられている. この方式では, Hevner が調べた楽曲の印象と言葉の関係性をコンピュータを使って自動で計量できるように, Standard MIDI ファイルから印象語群を計量するまでの手順を具体的に定めている.

3.4.2.1 全体の流れ

楽曲からのメタデータ自動抽出方式を以下の 4 つの Step で説明する. また, 全体の流れを図 3-4 に示す.

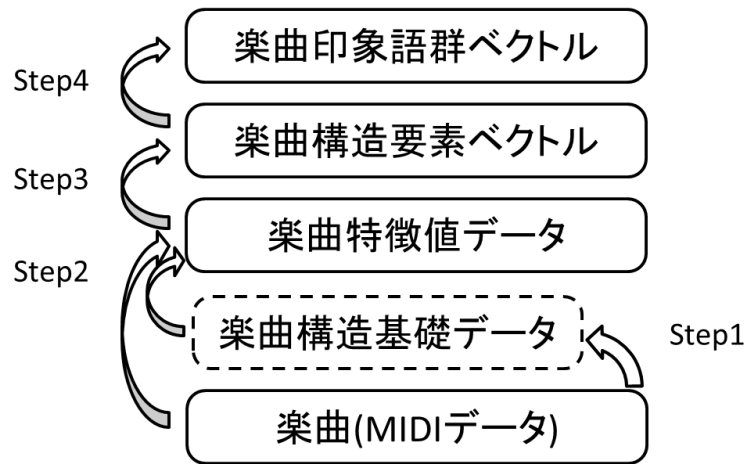


図 3-4 楽曲からのメタデータ自動抽出方式の全体の流れ.

Step1: 楽曲構造基礎データの抽出

MIDI データから, 楽音に相当するデータを抽出し, 主旋律と伴奏 (和音) のデータを定義する. 詳細は 3.4.2.2 節で述べる.

Step2: 楽曲特徴値データの抽出

MIDI データの情報と Step1 で抽出した楽曲構造基礎データの値をもとにして, 楽曲特徴値データを抽出する. 詳細は 3.4.2.3 節で述べる.

Step3: 楽曲の生成・出力

Step2 で抽出した楽曲構造要素ベクトルから、楽曲構造要素ベクトルを抽出する。詳細は 3.4.2.4 節で述べる。

Step4: 楽曲印象語群ベクトルの抽出

Step3 で抽出した楽曲構造要素ベクトルに、変換行列 T を作用させ、楽曲印象語群ベクトルを抽出する。詳細は 3.4.2.5 節で述べる。

3.4.2.2 楽曲構造基礎データの抽出 (Step1 の詳細)

§ 想定する楽曲データ

想定する楽曲データは SMF(Standard MIDI File) とし、チャンネル 1 に主旋律データを持ち、チャンネル 2 に伴奏データ（主旋律以外すべて）を持つとする。音色（楽器）はピアノとする。

§ 楽音データ Note

楽曲データに含まれる楽音数を notenum とする。楽音データ $Note_a(a = 1, 2, \dots, \text{notenum})$ を以下に示す。

$$Note_a = (npitch_a, nvel_a, nsdel_a, nedel_a, ndur_a). \quad (3-18)$$

Note の各要素は、それぞれ音程番号(pitch)、音の強さ(vel)、発音開始デルタタイム値(sdel)、発音終了デルタタイム値(edel)、楽音の長さ(dur)を意味する。また、Note は nsdel 値により昇順にソートしておく。

§ 主旋律データ Mnote

Note のうち、チャンネル 1 番にある楽音(主旋律)を抽出し、主旋律データとする。主旋律楽音数を melnum とする。主旋律楽音データ $Mnote_b(b = 1, 2, \dots, \text{melnum})$ を次に示す。

$$Mnote_b = (mnpitch_b, mnvel_b, mnsdel_b, mnedel_b, mndur_b). \quad (3-19)$$

各要素の意味は Note の場合と同様である。また、Mnote は mnsdel により昇順にソートしておく。

§ 伴奏データ Cnote

Note のうち、チャンネル 2 番にある楽音（伴奏）を抽出し、伴奏データとする。伴奏楽音数を chonum とする。主旋律楽音データ $Cnote_c(c = 1, 2, \dots, \text{chonum})$ $Cnote_c(c = 1, 2, \dots, \text{chonum})$ を次に示す。

$$Cnote_c = (cnpitch_c, cnvel_c, cnsdel_c, cnedel_c, cndur_c). \quad (3-20)$$

各要素の意味は Note の場合と同様である。また、Cnote は cnsdel により昇順にソートしておく。

§ 和音データ Chord

CNote(チャンネル 2 番のデータ) のなかで、同時に発音している楽音の集合を和音と

し，楽曲データに含まれる和音数を $chnum$ とする．和音データ $Chord_d (d = 1, 2, \dots, chnum)$ を次に示す．

$$Chord_d = (chname_d, chsdel_d, chedel_d, chdur_d). \quad (3-21)$$

Chord の各要素は，それぞれコード名，和音の開始デルタタイム値，和音の終了デルタタイム値，和音の長さである．また，Chord は $chsdel$ 値により昇順にソートしておく．なお，文献では上記の通りの方針となっているが，このままでは発音開始タイミングは同じだが発音終了タイミングが違う楽音をどう Chord にまとめるかということが不明確であるため，本研究では発音開始タイミングが同じものを一つの Chord としてまとめ，その長さ $chdur$ は，その中の楽音の長さ $ndur$ がいちばん大きいものを採用することとした

3.4.2.3 楽曲特徴値データの抽出 (Step2 の詳細)

MIDI データおよび 3.2.2 で抽出した楽曲構造基礎データから，楽曲特徴値データを抽出する．なお，文献にはこの他に演奏時間データ pt が定義されているが，これらは次に変換する楽曲構造要素ベクトルの計算には使われないため，ここでは取り上げないこととする．表 3 に，最終的に抽出する楽曲の各特徴値の意味を示す．

表 3-3 楽曲の特徴値.

楽曲の特徴値のラベル	内容
kn	楽曲の特徴
mt	拍子
tem	テンポ
mnap	旋律の音高の平均
unac	四分音符以上の長さの和音の総演奏時間
ac	四分音符未満の長さの和音の総演奏時間
tc	三和音の総演奏時間
oc	三和音以外の和音の総演奏時間
nc	三和音未満の和音の総演奏時間
um	上昇する旋律音の総演奏時間
dm	下降する旋律音の総演奏時間
lm	水平な旋律音の総演奏時間

§ 調名データ kn

SMF のメタイベントに含まれる調データから，楽曲の調を抽出する．調名データ kn は，調性を表す文字列により，次のように定義される．

$$kn = (\text{調名}). \quad (3-22)$$

§ 拍子 mt

SMF のメタイベントに含まれるメーターから、拍子を抽出する。拍子には分子と分母がある。 mt は次のように定義される。

$$mt = n/m = (\text{メーター}). \quad (3-23)$$

§ テンポデータ tem

SMF のメタイベントに含まれるセット・テンポデータから、1 分間に演奏される四分音符の数 tem を抽出する。テンポデータの値をそのまま tem の値とする。

$$tem = (\text{テンポ}). \quad (3-24)$$

§ 平均音高値 $mnap$

楽音の音高値の平均を $mnap$ とし、次に示す。

$$\begin{aligned} mnap &= \frac{mnap_{chord} + mnap_{melody}}{2}, \\ mnap_{chord} &= \frac{\sum_{i=1}^{chonum}(cndur_i)(cnpitch_i)}{\sum_{i=1}^{chonum}(cndur_i)}, \\ mnap_{melody} &= \frac{\sum_{i=1}^{melnum}(mndur_i)(mnpitch_i)}{\sum_{i=1}^{melnum}(mndur_i)}. \end{aligned} \quad (3-25)$$

ここでは、主旋律と伴奏の音高平均をまず算出し、その平均を平均音高値 $mnap$ としている。文献では全楽音データの平均であったが、このように変更した。理由は、伴奏に音が多い場合に、平均が下に引っ張られるため、同じ音高値でも主旋律の高さが高くなってしまい、違った印象を与えてしまっていたからである。

§ 固定/流動リズム $unac/ac$

文献を参考に、固定的なリズムとは、和音が四分音符以上の長さの楽音で構成されていることとする。また、流動的なリズムとは、和音を構成する楽音中に四分音符未満の楽音が含まれている事とする。固定的なリズムの和音の長さの和を固定リズム値とし、流動的なリズムの和音の長さの和を流動リズム値とする。ここで、固定的リズムの Chord を示す指標を Cchord、流動的なリズムの場合を Achord とし、固定/流動リズム $unac/ac$ を次に示す。

$$\begin{aligned} unac &= \sum_{i \in Cchord} (chdur_i), \\ ac &= \sum_{i \in Achord} (chdur_i). \end{aligned} \quad (3-26)$$

§ 単純/複雑/ノーハーモニー値 $tc/oc/nc$

主要三和音により構成される Chord を示す指標を Tchord、それ以外の和音により構成

される Chord を示す指標を Ochord とする．しかし「それ以外の和音」の定義が文献には示されていないため，新たに定義した．Chord の構成音数が 2 以下の場合を表す指標を Nchord とし，Tchord でも Nchord でもないものを Ochord とすることとした．単純/複雑/ノーハーモニー値 tc/oc/nc を次のように定義する．

$$\begin{aligned}
tc &= \sum_{i \in Tchord} (chdur_i), \\
nc &= \sum_{i \in Nchord} (chdur_i), \\
oc &= \sum_{i=1}^{chnum} (chdur_i) - (tc + nc).
\end{aligned} \tag{3-27}$$

§ 上昇/下降/水平メロディー値 um/dm/lm

Mnote において，次に発音される音との音高差が正，負，0 であるとき，それぞれ上昇，下降，水平メロディーであるとする．ここで，上昇，下降，水平メロディーの Mnote を表す指標を，それぞれ Umnote，Dmnote，Nmnote，とし，上昇/下降/水平メロディー値を次のように定義する．

$$\begin{aligned}
um &= \sum_{i \in Umnote} (mndur_i), \\
dm &= \sum_{i \in Dmnote} (mndur_i), \\
lm &= \sum_{i \in Nmnote} (mndur_i).
\end{aligned} \tag{3-28}$$

なお，この際 Mnote の最後の Note には次の Note が無いため，最後の Note は無視する事とする．

3.4.2.4 楽曲構造要素ベクトルの抽出 (Step3 の詳細)

前節で得られた楽曲特徴値データを用いて，楽曲構造要素量ベクトルを求める．楽曲特徴値と楽曲構造要素の対応関係を表 3-4 に示す．

表 3-4 楽曲の特徴値と楽曲構造要素の対応関係.

楽曲特徴値		楽曲構造要素
kn		key
tem		tempo
mnap		pitch
unac, ac		rhythm
tc, oc, nc		harmony
um, dm, lm		melody

また、それらの特徴量により楽曲を特徴づけしたベクトル f を生成する。 f を次のように定義する。

$$f = (\text{key}, \text{tempo}, \text{pitch}, \text{rhythm}, \text{harmony})^T. \quad (3-29)$$

f の各要素は数値であるが、それぞれ別のものを表している。そこで、各特徴量が $-1 \sim 1$ の値をとるように正規化する。なお、key と pitch と melody については文献から変更を加えた。各特徴量を以下に定義する。

§ 調性度 key

key は、楽曲データの調名データ kn および、調性度の表 3-5 を利用として決定する。この表は、梅本・中島の研究[9] による、各調の長調らしさ、短調らしさを表す値を $-1 \sim 1$ に正規化して作成された、文献 [24] に示されている表を拡張したものである。元々の表では二つ以上の調に同一の調整度が割り当てられている場合があり、そのままでは逆変換の際に調が一意に復元できなくなってしまうため、全ての調が異なる調性度を持つように修正を加えた。具体的には、二つ以上の調に同一の調性度が割り当てられている場合には、前後の調の持つ調性度の間隔を同一の調性度に割り当てられている調の数で等分割し、分割したそれぞれの点に一つの調をランダムに選んで対応させた。元の調性度とそれを拡張した調性度の表を表 3-5 に示す。この表を用いて、kn から key の値を決定する。

§ テンポ tempo

tempo は、楽曲データのテンポデータ tem を用いて次のように定義する。

$$\text{tempo} = \begin{cases} \frac{\text{tem} - \text{tstnd}}{\text{tmax} - \text{tstnd}} & \text{tem} - \text{tstnd} \geq 0, \\ \frac{\text{tem} - \text{tstnd}}{\text{tstnd} - \text{tmin}} & \text{tem} - \text{tstnd} < 0, \end{cases} \quad (3-30)$$

ここで、tmax, tmin, tstnd は、それぞれ M.M(メルツェル・メトロノーム) の最大値 (=184)、最小値 (=40)、基準値 (=88) である。

§ 音高 pitch

pitch は、楽曲データの平均音高値 mnap を用いて次のように定義する。

$$\text{pitch} = \begin{cases} \frac{\text{mnap} - \text{pstnd}}{\text{pmax} - \text{pstnd}} & \text{mnap} - \text{pstnd} \geq 0, \\ \frac{\text{mnap} - \text{pstnd}}{\text{pstnd} - \text{pmin}} & \text{mnap} - \text{pstnd} < 0. \end{cases} \quad (3-31)$$

ここで、pmax, pmin, pstnd は、それぞれ mnap が最大の時、最小の時、標準の時の SMF の音程番号であるが、これは楽器に依存する値である。今回はピアノ曲を対象としているため、ピアノ曲において採り得る値の最大最小を考慮して決定する。ピアノの最高音に当たる SMF の音程番号は 108、最低音は 21 であるが、今回はそれより範囲を狭めて、pmax(= 84) (最高音から 2 オクターブ下のド)、pmin(= 45) (最低音から 2 オクターブ上のラ)、pstnd(= 64) (中央ミ) とした。

表 3-5 拡張された各調の調性度(key).

調	調性度(key)	Key の範囲		
G	1	0.96875	～	1.00000
C	0.9375	0.89583	～	0.96875
A	0.854167	0.81250	～	0.89583
D	0.770833	0.72917	～	0.81250
F	0.6875	0.65625	～	0.72917
B	0.625	0.56250	～	0.65625
F#	0.5	0.46875	～	0.56250
E	0.4375	0.40625	～	0.46875
E ♭	0.375	0.34375	～	0.40625
A ♭	0.3125	0.21875	～	0.34375
B ♭	0.125	0.06255	～	0.21875
D ♭	0.0001	0.00000	～	0.06255
Am	-0.0001	-0.10422	～	0.00000
Gm	-0.20833	-0.31250	～	-0.10422
G#m	-0.41667	-0.48958	～	-0.31250
Bm	-0.5625	-0.59375	～	-0.48958
Em	-0.625	-0.65625	～	-0.59375
F#m	-0.6875	-0.69492	～	-0.65625
Dm	-0.70233	-0.74100	～	-0.69492
Fm	-0.77967	-0.80650	～	-0.74100
C#m	-0.83333	-0.87500	～	-0.80650
B ♭ m	-0.91667	-0.95833	～	-0.87500
Cm	-1	-1.00000	～	-0.95833

§ リズム rhythm

rhythm は、楽曲データの固定/流動リズム値 $unac/ac$ を用いて次のように定義する.

$$rhythm = \frac{unac - ac}{unac + ac}. \quad (3-32)$$

§ ハーモニー harmony

harmony は、楽曲データの単純/複雑/ノーマーモニー値 $tc/oc/nc/$ を用いて、次のように定義される.

$$harmony = \frac{tc - oc}{tc + oc + nc}. \quad (3-33)$$

§ メロディー melody

$melody$ は, 楽曲データの上昇/下降/水平メロディー値 $um/dm/lm$ を用いて次のように定義される.

$$melody = \frac{um - dm}{um + dm + lm}. \quad (3-34)$$

ただしここで, 楽曲を構成することを考慮し, 変更を加える. $melody$ は上式より, $-1 \sim 1$ の値を採るが, $melody$ がその上限値または下限値であるという事は極めて特殊な状態である. 例えば $melody$ が 1 であるということは旋律音が全て上昇列で構成されているという事になるが, それではすぐ使用可能な音域の上限に達してしまう.

ここで, メロディーの生成に用いる制約に依存して, $melody$ の値の採り得る範囲が変化することが言える. 例えばメロディー生成の制約として, “長音階または和声的短音階の構成音のみ用いる”, “音程は最大 8 度 (1 オクターブまで)” の二つを与えたとすると, $melody$ が最大になるのは, 上昇は全て 2 度音程で, 下降は全て 8 度音程になる時である. この時, 8 度音程は 2 度音程 7 つ分であるので, $um : dm = 7 : 1$ の状態, つまり, $um = 0.875, dm = 0.125, lm = 0$ の時 $melody$ は最大値を取り $melody = 0.75$ となる. 下降についても同様なので, 上記の制約における $melody$ の採り得る値の範囲は, $-0.75 \sim 0.75$ となる. このことを踏まえ, 制約によって決まる $melody$ の上限下限値 (上記の例では 0.75) を $limit$ として定義する. $limit$ の算出は, 統計的な研究などから音程の出現確率を与えた上で, 上昇音と下降音の生起確率の比 ($um : dm$) が最大, かつ, 音程の移動量の期待値が 0 となる場合の, $melody$ の値として行う.

上記を考慮し式(3-34)で $melody$ を求めた後, $limit$ を用いて $melody$ を以下の式に従って正規化する. 正規化後の値を $melody'$ とする.

$$melody' = \begin{cases} 1 & melody > limit, \\ -1 & melody < -limit, \\ \frac{1}{limit} melody & otherwise \end{cases} \quad (3-35)$$

これは $melody$ のドメインの一部を拡大し -1 から 1 に写像することに相当する. $melody$ の絶対値が $limit$ 以上だった時の情報は失われてしまうが, 先に触れたように $melody$ の絶対値が大きい楽曲は稀であるため問題となるケースは少ない.

実際に用いる $limit$ の値については, メロディー生成の制約条件に依存するため, 楽曲の生成・出力(3.4.3.4 節)時に述べる.

3.4.2.5 楽曲印象語群ベクトルの抽出 (Step4 の詳細)

3.2.4 節で抽出した f と, 変換行列 T (図 9) を用いて, 楽曲印象語群ベクトル c を定義する. c は次のように定義される. $C_1 \sim C_8$ はそれぞれ, Hevner が定めた楽曲印象語群に対応する.

$$c = (C_1, C_2, \dots, C_8)^T. \quad (3-36)$$

§ 変換行列 T の生成

変換行列 T は, Hevner が調べた, 各印象語群に対する楽曲構造要素の相対重要性の表 (2) に正負の符号を付けて, さらに行列全体を 2 ノルム正規化することで得る. 正負の符号を付けただけの行列を T' とすると, T は次の式で表される.

$$T = \frac{T'}{\|T'\|}. \quad (3-37)$$

楽曲印象語群ベクトルへの変換変換行列 T と f を用いて, c を以下の式で求める.

$$c = Tf. \quad (3-38)$$

	key	tempo	pitch	rhythm	harmony	melody
c1	0.070	-0.244	-0.174	0.314	0.052	0.070
c2	-0.349	-0.209	-0.331	0.052	-0.122	0.000
c3	-0.209	-0.279	0.105	-0.157	0.070	0.000
c4	0.052	-0.349	0.139	-0.035	0.174	0.052
c5	0.366	0.105	0.279	-0.139	0.209	-0.052
c6	0.418	0.349	0.105	-0.174	0.279	0.000
c7	0.000	0.366	-0.157	0.035	-0.244	-0.122
c8	0.000	0.105	-0.227	0.174	-0.139	-0.139

図 3-5 変換行列 T .

3.4.3 楽曲メディアコンテンツ自動構成方式(i \mathcal{ML})

本節では、印象メタデータから楽曲メディアコンテンツを構成する方式の各機能の詳細について示す。

3.4.3.1 全体の流れ

楽曲印象語群ベクトルから楽曲を生成する流れを以下の3つのStepとして示す。また、その流れを図3-6に示す。

Step1: 楽曲構造要素ベクトルの出力

楽曲印象語群ベクトルに変換行列 T を作用させることにより、楽曲構造要素ベクトルを求める。また、その際にベクトルの正規化を行う。詳細は3.4.3.2節で述べる。

Step2: 楽曲特徴値の出力

Step1 で求めた楽曲構造要素ベクトルの値をもとにして、生成する楽曲の特徴値を求める。詳細は3.4.3.3節で述べる。

Step3: 楽曲の生成・出力

Step2 で求めた楽曲構造要素ベクトルから、楽曲構造要素ベクトルを抽出する。詳細は3.4.3.4節で述べる。

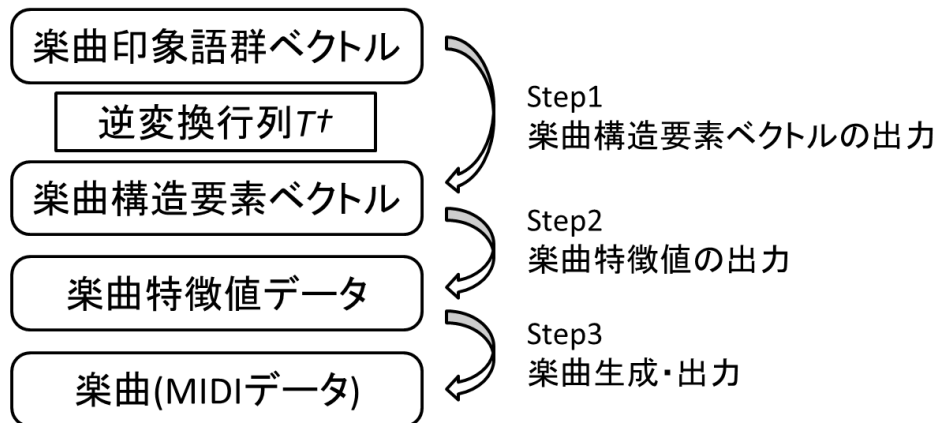


図 3-6 楽曲メディアコンテンツ自動構成方式の流れ。

3.4.3.2 楽曲構造要素ベクトル出力(Step1 の詳細)

本節では楽曲印象語群ベクトルを楽曲構造要素ベクトルに変換する方式について示す。

§ 逆変換行列 T^\dagger の生成

3.3 節で示した方式を用いることによって、図3-5に示した変換行列 T から、ムーア・ペンローズ一般化逆行列を計算し、逆変換行列 T^\dagger を導出する。具体的には図3-6のようになる。これは印象メタデータから楽曲構造要素に変換する行列となる。ここで、 T は 8×6 行列と、列より行の方が多く、各列は線形独立である。そのため、次式より、 T^\dagger は T の左逆元となる事が分かる。

	c1	c2	c3	c4	c5	c6	c7	c8
key	-0.048	3.466	-4.947	6.726	3.141	-0.775	5.398	-3.403
tempo	0.063	-2.550	1.882	-4.760	-2.146	1.279	-2.820	1.519
pitch	-0.155	-4.226	2.028	-2.954	-0.631	-1.560	-2.465	1.209
rhythm	2.096	-4.914	3.016	-5.756	-2.188	-0.303	-4.799	3.909
harmony	1.295	-3.138	5.905	-7.057	-2.625	3.301	-7.006	5.116
melody	0.376	1.602	-5.154	2.235	-1.780	-0.027	1.956	-7.191

図 3-7 変換行列 T^\dagger .

$$T^\dagger T = (T^* T)^{-1} T^* T = I_6. \quad (3-39)$$

楽曲構造要素ベクトルを楽曲印象語群ベクトルに変換し，さらに楽曲構造要素ベクトルに変換した場合は元のベクトルに戻る．しかし，その逆は元に戻ることは保証されない．

$$\mathbf{f} = T^\dagger \mathbf{c} = T^\dagger T \mathbf{f}. \quad (3-40)$$

§ 楽曲印象語群ベクトルの構成

楽曲印象語群ベクトル \mathbf{c} を構成する． \mathbf{c} は式(3-36)に定義されている．ここで，要素である C1～C8 をそれぞれ -1 ～ 1 として設定し， \mathbf{c} を構成する．これが印象メタデータに相当する．

§ 楽曲構造要素ベクトルの構成

次式のように，楽曲印象語群ベクトルに対して変換行列 T^\dagger を作用させることにより 6 つの楽曲構造要素への重みを表す楽曲構造要素ベクトル \mathbf{f} を構成する．

$$\begin{aligned} \mathbf{f} &= (\text{key}, \text{tempo}, \text{pitch}, \text{rhythm}, \text{harmony}, \text{melody})^T, \\ &= T^\dagger \mathbf{c}. \end{aligned} \quad (3-41)$$

§ 楽曲構造要素ベクトルの正規化

楽曲構造要素ベクトルの各要素は，楽曲特徴値との関係上，各要素に対応する性質が最大である状態を 1，最小である状態を -1 とし，-1 ～ 1 の値を持つように定義されている．しかし入力する楽曲印象語群ベクトルの値によってはその範囲に収まらなくなるため，正規化を行う．正規化後の楽曲構造要素ベクトルを \mathbf{f}' とし，以下の式で表す．

$$\mathbf{f}' = \begin{cases} \mathbf{f} & \|\mathbf{f}\|_\infty < 1 \\ \frac{\mathbf{f}}{\|\mathbf{f}\|_\infty} & \text{otherwise.} \end{cases} \quad (3-42)$$

3.4.3.3 楽曲特徴値出力 (Step2 の詳細)

出力された楽曲構造要素ベクトルの値をもとに，表 3-3 に示した，生成する楽曲の特徴値 $kn, tem, mnap, unac, ac, tc, oc, um, dm, lm$ を決定する．これらを，抽出の逆になるように定義していく．この定義は，心理学の研究成果や楽典を参考に構成している．これによって，自然な楽曲メディアコンテンツ生成を可能とする．

§ 調性 kn の決定

調性度 key の値から、生成する楽曲の調 kn を定める。 kn は印象抽出の際に作成した拡張された調性度の表 3-5 に基づいて、調性度の値 key に対応する調とする..

§ テンポ tem の決定

式(3-30)を変形し、生成する楽曲のテンポを決定する.

$$tem = \begin{cases} tstnd + (tmax - tstnd)tempo & (tempo \geq 0), \\ tstnd + (tstnd - tmin)tempo & (tempo < 0). \end{cases} \quad (3-43)$$

ただし, $tmax$, $tmin$, $tstnd$ は、抽出と同じとし、それぞれ 180, 40, 88 とした.

§ 音高 $mnap$ の決定

式(3-31)を変形し、生成する楽曲の旋律音の高さの平均 $mnap$ を定める.

$$mnap = \begin{cases} pstnd + (pmax - pstnd)pitch & (pitch \geq 0), \\ pstnd + (pstnd - pmin)pitch & (pitch < 0). \end{cases} \quad (3-44)$$

ただし, $pmax$, $pmin$, $pstnd$ は抽出と同じとし、それぞれ 84,45,64 とした.

§ リズムの特徴 $unac$, ac の決定

式((3-32)) を変形し、 $unac$ と ac を求める. ここで、 $rhythm$ は $unac$ と ac の比率さえ保持していれば同一の値になるため、 $unac$ と ac の関係を比率と捉えることとし、仮に $unac + ac = 1$ とすると $unac$, ac を一意に定めることができる.

$$\begin{aligned} unac &= \frac{1 + rhythm}{2}, \\ ac &= \frac{1 - rhythm}{2}. \end{aligned} \quad (3-45)$$

§ 和声の特徴 tc , oc , nc の決定

式(3-33) を変形し、 tc, oc, nc を求める. ここで、 $harmony$ は tc, oc, nc の三つの値の比率さえ保持していれば同一の値になるため、 tc, oc, nc の関係を比率と捉えることとし、リズムの場合と同様に $tc + oc + nc = 1$ とすると、 tc , oc は以下のように表現出来る.

$$\begin{aligned} tc &= \frac{1 - (nc - harmony)}{2}, \\ oc &= \frac{1 - (nc + harmony)}{2}. \end{aligned} \quad (3-46)$$

この場合、 nc を決めない限り tc, oc を一意に決定する事は出来ない. また、 $harmony$ の大きさによって nc の採り得る値が変化する. ここで、 $|harmony| < 1$ であるため、この条件と上式より以下の関係が導かれる.

$$0 \leq nc \leq |harmony|. \quad (3-47)$$

nc の値がこの条件を満たさない場合、 tc, oc, nc から $harmony$ を導出する際に値を保持す

る事が出来なくなる．よって nc の値は，この範囲における任意の値を取ることが出来る．本研究では，生成される楽曲にバリエーションを持たせる目的から，取り得る値の範囲の中心を中心とした正規分布から確率的に決定することとした． nc の値を，取り得る値の範囲の中心を中心とした正規分布から確率的に決定する． nc を決定する事が出来れば， tc, oc は一意に決定する事が出来る．

§ 旋律の特徴 um, dm, lm の決定

まず 3.4.2.4 節の melody の生成の項で述べた limit を用いて melody を以下の式に従って正規化する．正規化後の値を $melody'$ とする．

$$melody' = limit \cdot melody. \quad (3-48)$$

これは melody のドメインを狭い範囲に写像し直すフィルタに相当する．実際に用いる limit の値については，旋律生成の制約条件に依存するため，楽曲の生成・出力(3.4.3.4 節)時に述べる．

式(3-34)を変形し， um, dm, lm を定める．ここで，melody は um, dm, lm の三つの値の比率さえ保持していれば同一の値になるため， um, dm, lm の関係を比率と捉えることとし，仮に $um + dm + lm = 1$ とすることで， um, dm を一意に定めることができる．

$$\begin{aligned} um &= \frac{1 - (lm - melody')}{2}, \\ dm &= \frac{1 - (lm + melody')}{2}. \end{aligned} \quad (3-49)$$

和声の場合と同様に，上式中の水平旋律音 lm を決定する．ここで， $|melody'| \leq 1$ であるため，この条件と上式より以下の関係が導かれる．

$$0 \leq lm \leq |melody'|. \quad (3-50)$$

よって lm の値は，この範囲における任意の値を取ることが出来る．予備実験により，本研究では lm の値を，取り得る値の範囲の中心を中心とした正規分布から確率的に決定することとした．それによって um, dm を一意に決定する事が出来る．

3.4.3.4 楽曲生成・出力 (Step3 の詳細)

楽曲の特徴値に沿って楽曲を生成する．以下に楽曲データの生成手順を示す．

§ 生成する楽曲の概要

生成する楽曲は有限の小節数を持ち，調性を持ち，旋律と伴奏を持つとする．途中で調やテンポ，拍子の変更は含まれない．また，音色に当たる楽器はピアノを想定する．旋律および伴奏に用いる楽音の音価は四分音符と八分音符に限定する．ただし，最後の一小節における旋律の楽音の音価は一小節分とする．

§ テンポ

tem の値をそのまま楽曲の M.M(メルツェル・メトロノーム)におけるテンポの値とす

る. M.M は一般的なメトロノームにおけるテンポを示している.

§ 小節数

楽曲の小節数を size とする. 本研究における実験システムでは size は 8 とした.

§ 拍子

楽曲の拍子を mt を分子分母に分け n/m とする. 本研究における実験システムでは, $4/4$ とした.

§ 和音による伴奏

決定した特徴値のうち, kn, unac, ac, tc, oc, nc を用いて, 和音を構成する音を以下の手順により生成する.

- (1) 曲中で使用する和音を kn によって決定する. 伴奏に用いる和音はその調のダイアトニックコードとする. なお, 短調の曲に関しては和声的短音階に準じる. ハ長調の長音階とハ短調の和声的短音階を図 3-8 に示す.
- (2) unac は四分音符以上の音価を持つ楽音の総演奏時間を表わし, ac は四分音符未満の音価を持つ楽音の総演奏時間を表わすが, ここでは, unac を四分音符の演奏時間割合, ac を八分音符の演奏時間割合と限定して割り当てる. size と unac と ac の値から四分音符の和音と八分音符の和音の個数・配列を決定する. unac と ac の値を, それぞれ四分音符と八分音符の和音の演奏時間の割合に対応させる. 四分音符と八分音符の割合が決まったら, 音価の配列として表現し, 配列をランダムにシャッフルする.
- (3) 作曲に用いる和音は, 7th を含むダイアトニックコードの I~VII の和音を用いる. ハ長調の 7th ダイアトニックコードとハ短調の和声的短音階における 7th ダイアトニックコードを図 3-9 に示す. なお, ここで短調の I の和音のみ, ダイアトニックコードの 7th が長七度であるところを短七度としている. 和声法を参考にし, コード進行はトニック・サブドミナント・ドミナントの関係を考慮して決定する. 一小節に入る和音の種類としては, I~VII のどれか一種類だけとし, 最後の一小節は必ず I 度の和音とする.
- (4) tc, oc, nc の値をもとに和音を割り当てる. それぞれ, nc には「二和音(root と 3rd)」, tc には「三和音(二和音+5th)」, 「四和音(三和音+7th)」を割り当てる. この対応関係を図 3-10 に示す.

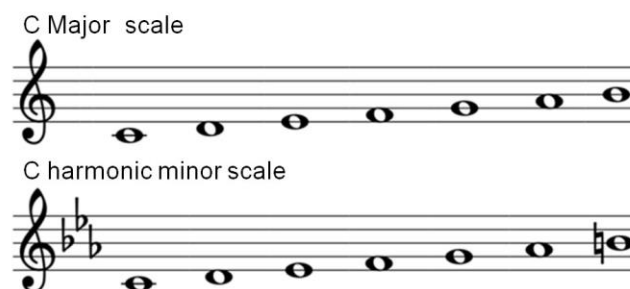


図 3-8 ハ長調 (C major) の長音階とハ短調 (C minor) の和声的短音階.

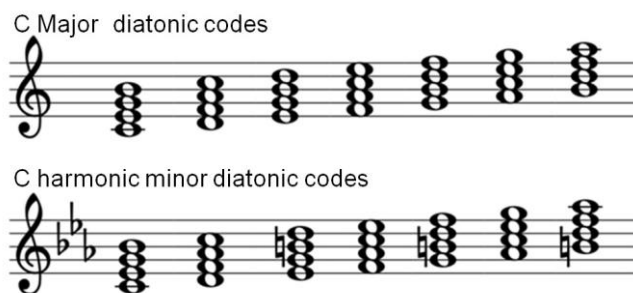


図 3-9 ハ長調 (C major) の 7th ダイアトニックコードとハ短調 (C minor) の和声的短音階における 7th ダイアトニックコード.

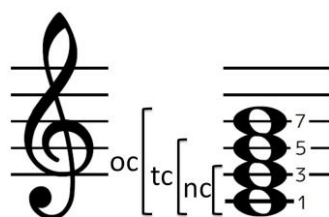


図 3-10 tc, oc, nc の和音との対応.

§ 旋律の生成

特徴値のうち kn, um, dm, lm を用いて旋律音を以下の手順により生成する.

- (1) kn より使用する音階を決定する. 長調ならば長音階, 短調ならば短音階 (和声的短音階) とする. 旋律に使用する音はその調の音階に沿ったものとする.
- (2) um, dm, lm の割合に応じて上昇音, 下降音, 水平音の割合を決定する.
- (3) size と um, dm, lm から, 上昇音, 下降音, 水平音それぞれに属する四分音符・八分音符の割合をランダムに決定し, 音価の配列を作る.
- (4) 配列をシャッフルし, 上で定めた上昇音, 下降音, 水平音の順序を決定する.
- (5) 曲の最後の音が調の基音となるようにし, 旋律の上下幅はそこから上下 1 オクターブの, 最大 2 オクターブとする. また, Hevner の研究では旋律音の長さやリズムには言及していないので, 旋律の楽音の長さは原則として四分音符と八分音符のみとするが, 曲が終わった感じを出すため, 最後の一小節は主音とし, 一小節分の長さとする.
- (6) 旋律を具体的に決定する. 最後の音をその調の主音に固定して後ろから順に決定していく. 上昇音と下降音における音程の変化量は, 統計的な情報を用いて決定する. 統計情報として文献 [39]における, 隣接する二つの旋律音の音程とその出現する割合についての研究を用いる. この研究では, 歌曲の旋律について, 1 度から 8 度までの音程の出現頻度を調べている (表 3-6). この統計情報を制約条件として旋律を確率的に決定していく. この際, 3.4.2.4 節の式(3-35)および 3.4.3.3 節の式(3-50)における limit は, 0.5538 となる.

表 3-6 歌曲の旋律音における各音程の出現頻度.

音程		出現頻度(%)
1 度	完全 1 度	26
2 度	短 2 度	17.5
	長 2 度	24
3 度	短 3 度	10
	長 3 度	6
4 度	完全 4 度	9
	増 4 度	0.5
5 度	減 5 度	0.5
	完全 5 度	3
6 度	短 6 度	1
	長 6 度	1
7 度	短 7 度	0.5
	長 7 度	0
8 度	完全 8 度	1

§ 音高の修正

音の高さの平均値が $mnap$ に等しくなるように伴奏と旋律の高さを調節する. 伴奏や旋律をシフトまたは変形して, 音高平均値が $mnap$ が等しくなるよう近づけて行く. まず, 旋律は伴奏より上に来るように一オクターブずつ平行移動する. 次に, 伴奏の和音(楽音のセット)を一つずつ一オクターブ分シフトして, 平均値を $mnap$ に近づけていく. その処理では一つの変更における $mnap$ の変化量が大きくて計算が収束しない場合は, さらに和音の最低音を固定したまま, その他の音を一つずつ一オクターブ分シフトし, 音高を微調節する. C メジャーコードの変形の例を図 3-11 に示す.

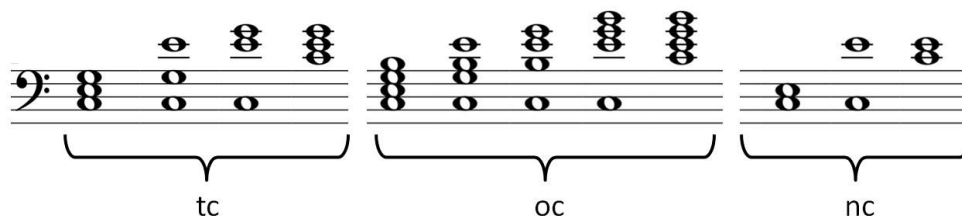


図 3-11 C major の和音の変形.

3.5 楽曲メディアコンテンツにおける ML - iML 系を用いた検証方式

これまでに、楽曲を対象とした ML と iML の作成方法を示した。 ML と iML を組み合わせることで ML - iML 系を作る事が出来る。本章では、 ML - iML 系が持つ保存性と、その検証方法について示す。

3.5.1 ML - iML 系の保存性

3.3.3 節で示した ML - iML 系の一般形と、3.4 節で構成した具体的な ML - iML 系の対応関係を図 3-12 に示す。Md は楽曲データ (MIDI) に対応する。Fs は楽曲構造要素ベクトルに対応する。Ws は楽曲印象語群ベクトルに対応する。楽曲特徴値データは、Md から Fs に変換する際の間データセットと捉えられる。楽曲への適用例には、それぞれのデータの持つ情報量も示した。また、それぞれの変換において、情報量が増加するものについては「増」、減少するものについては「減」と示した。今回の例では、Fs にあたる 6 次元の楽曲構造要素ベクトルにおいて情報量が一番少なくなるため、2 章で示した通り、“特徴保存の ML - iML 系”となることが期待される。

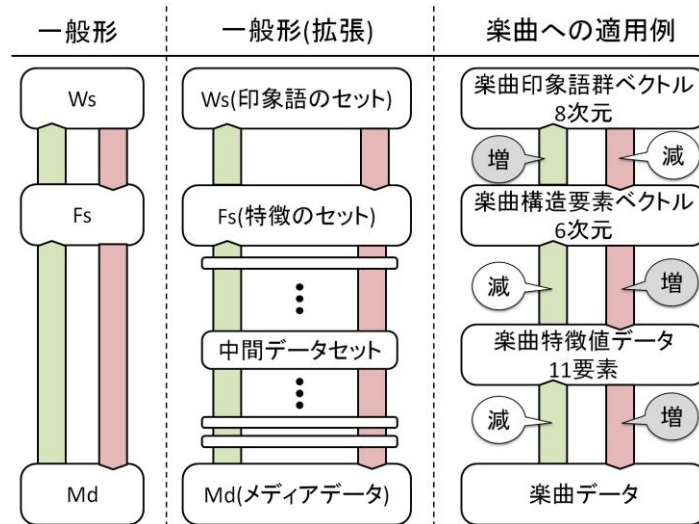


図 3-12 ML - iML 系の一般形と具体例の対応。

3.5.2 ML - iML 系の検証方法

節 3. で構成した一般可逆作用素 iML と ML によってつくられる ML - iML 系が、式(3-8)のうち以下の性質を満たす事を確認する。

$$MF(iMF(Fs)) = Fs, \quad \dots (b)$$

$$iFL(FL(Fs)) = Fs, \quad \dots (c)$$

これを満たす時、この系は特徴保存の ML - iML 系とも言える。

ML - iML 系は、あるメディアに対して、言葉とメディアデータを相互に変換する枠組み

である．そのため，ユーザーとしては，言葉を入力として楽曲を作る，あるいは，楽曲を入力として言葉のメタデータを出力する，といった機能が期待される．しかし，そうしたアプリケーション全体としての評価をするだけでは，検証としては不十分と考えられる．なぜなら，情報量が保存されない処理を元々含んでいるため，本来評価が完全なものにならないことが明白であり、アプリケーション全体を検証するだけでは、それぞれの処理が最終的な結果に及ぼしている影響を知ることが出来ないからである．そのため、検証項目を可能な限り小さな単位に分けることで、それぞれの処理の全体への影響を検証する。 ML - iML 系の保存性についての検証項目を図 3-13 に示す．検証項目は 6 個となる．

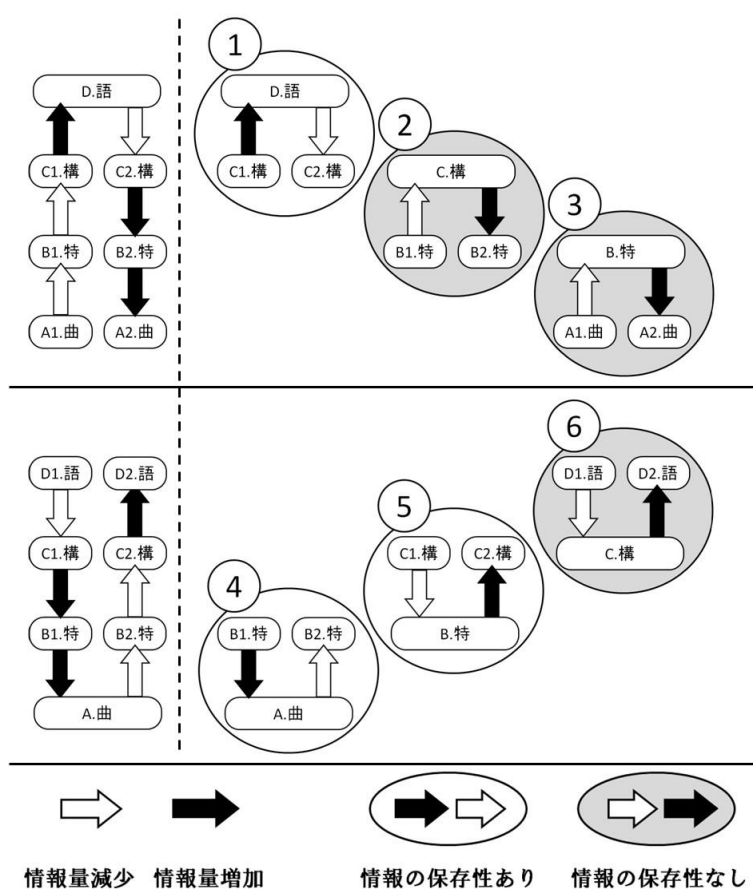


図 3-13 楽曲の ML - iML 系の検証のための実験の検証項目．

3.5.3 楽曲の表現力と分解能

楽曲構造要素ベクトルの harmony や melody の値や，そこから変換される oc,um といった値は連続値であり，音符数や小節数は離散値であるため，楽曲を生成する際には端数の情報が失われてしまう．ここでは，楽曲の仕様を決めた場合に表現出来る情報の最小値を分解能として定式化する．実験では，これらの分解能以下の値の差については，情報が失わ

れてはいるものの想定内の事とし、誤りではないとみなすこととする。

3.5.3.1 楽曲特徴値の分解能

一小節に一種類の和音のみを用いるとした時、tc,oc,nc の分解能 δ_{tcocnc} は以下の式で表される。

$$\delta_{tcocnc} = \frac{1}{size \cdot \frac{n}{m}} \quad (3-51)$$

音符の音価（長さ）を四分音符または八分音符のみとした時、unac,ac の分解能 δ_{unacac} は以下の式で表される。

$$\delta_{unacac} = \frac{1}{4 \cdot size \cdot \frac{n}{m}} \quad (3-52)$$

音符の音価（長さ）を四分音符または八分音符のみとし、曲の最後の旋律音は一小節分とした時、um,dm,lm の分解能 $\delta_{umdm lm}$ は以下の式で表される。

$$\delta_{umdm lm} = \frac{1}{4 \cdot (size - 1) \cdot \frac{n}{m}} \quad (3-53)$$

これらを踏まえた上で、今回の実験では曲の生成がおよびそこからの抽出が出来れば良いと考え、簡易的な曲を作成することとし、楽曲の小節数 size を 8、拍子 n/m を 4/4 とした。これらの値は任意に変更可能である。

3.5.3.2 楽曲構造要素の分解能

key の分解能 δ_{key} は、表 5 における、同一の調に割り振られた、抽出時と構成時の値の最大の差であり、以下の式で表される。

$$\delta_{key} = 0.10417. \quad (3-54)$$

tempo の分解能 δ_{tempo} は、楽曲特徴値 tem を実数の状態から四捨五入して整数にする際に切り捨てられる量の最大値を tempo に直したものである。式を以下に示す。四捨五入の最大丸め誤差が 0.5 であり、また、tempo がプラスの時とマイナスの時でマイナスの時の方が、tem の変更に対する tempo の影響が大きくなるため、以下の場合に最大になる。

$$\delta_{tempo} = \frac{0.5}{tstand - tmin} = \frac{0.5}{88 - 40} = 0.01042. \quad (3-55)$$

pitch の分解能 δ_{pitch} は、 δ_{tempo} と同様に、楽曲特徴値 mnap を実数の状態から四捨五入して整数にする際に切り捨てられる量の最大値を pitch に直したものである。式を以下に示す。

$$\delta_{pitch} = \frac{0.5}{pstand - pmin} = \frac{0.5}{64 - 45} = 0.02632. \quad (3-56)$$

3.6 任意の言葉の音相の印象に合致した楽曲の自動構成方式

本節では、3.4 節で構成した作曲システムと、言葉の音声情報の印象を抽出する研究を接続することで、任意の入力語の音声情報の印象に合致した楽曲の自動構成方式を示す。言葉の音声情報を用いる研究として、音相理論を用いる。また、二つのシステムを接続するためには、それぞれの分野において印象語とその重みとして表現されたメタデータの類似度を計量する機構が必要になる。この機構として、言葉同士の類似度を計量する機構である、意味の数学モデル [12]を用いる。

3.6.1 音相理論

本節では音相理論の概略について述べる。音相理論は、木通がまとめた日本語の音声情報が持つ印象についての理論である [28]。木通はそれぞれの言葉が持つ音の構造の違いによって生まれる表情を音相と呼んだ。音相理論では、音相を捉えるための単位として、それぞれの音の最小単位である音素と、それらの音素を発音するために使われる調音器官を表す調音点、および発音の方法にあたる調音法が挙げられている。さらに、各音相基がどのような印象を表すときに使われているかを調査することにより、その音素の勁性(強さ)と輝性(明るさ)を明らかにした。音相理論において、音相の単位として挙げられたこれらの要素は音相基と呼ばれており、40 の音相基が定義されている。そして単独の音相基、または 2 つの音相基の組み合わせと、音相理論では“表情”と呼ばれる、音相基から生じる印象の相関関係、および各表情とそれらに合致する印象語群についてまとめられている。

3.6.2 任意の言葉の音相の印象に合致したメタデータ自動抽出方式の概要

本節では、音相理論に基づいて、言葉の音を持つ印象を表す言葉をメタデータとして自動抽出するための方式を示す。本方式は具体的に構成された ML の一つである。詳細は [26]に示されている。

(1) 音相基の抽出

初期パラメータとして、入力語から 40 種の音相基を表現するのに必要十分な 29 の要素からなる初期ベクトル \mathbf{v}_b を抽出し、音相基と表情の関係を用いて変換行列 T_1 を生成し、 T_1 を用いて 40 種の音相基を表す 40 次元のベクトル \mathbf{v}_k へと変換する。

$$\mathbf{v}_k = T_1 \mathbf{v}_b \quad (3-57)$$

(2) \mathbf{v}_k の拡張

それぞれの表情が有効であるかを判断する尺度として各音相基の標準使用率が示されているので、これを用いて \mathbf{v}_k を補正する。さらに、2 つの音相基の組み合わせによりさらに生じる 38 種の表情の判定を行い、 \mathbf{v}_k を 78 次元のベクトル \mathbf{v}_f に拡張する。

(3) 印象語群とその重みの出力

それぞれの表情とそれらから抽出される印象語群の関係から変換行列 T_2 を作成し、以

下により 40 の印象語より構成される, 2 語を 1 組とした 20 の表情語群を表現している 20 次元のベクトル \mathbf{v}_c に変換する.

$$\mathbf{v}_c = T_2 \mathbf{v}_f \quad (3-58)$$

この \mathbf{v}_c が, 本方式により抽出される, 言葉の音の印象を表す重み付き印象語群によるメタデータである.

表 3-7 音相を表す表情語と表情属性.

表情語群	表情属性	訳語
A	シンプルな, 明白さ	plain, obvious
B	躍動感, 進歩的	vibrant, advance
C	新鮮さ, 新奇さ	fresh, unprecedented
D	動的, 活性的	dynamic, active
E	派手さ, 賑やかさ	florid, bustle
F	軽やかさ, 軽快感	light, trippingly
G	若さ, 澆刺さ	young, effervescent
H	現代的, 都会的	modern, urban
I	明るさ, 開放的	bright, open-minded
J	合理的, 現実的	reasonable, real
K	個性的, 特殊的	individual, special
L	強さ, 鋭さ	powerful, sharp
M	適応性, 庶民的	adaptable, popular
N	清らかさ, 爽やかさ	pure, brisk
O	健康的, 清潔感	healthy, clean
P	暖かさ, 安らぎ	warm, comfortable
Q	安定感, 信頼感	stable, confidence
R	高級感, 充実感	expensive, fulfil
S	高尚な, 優雅さ	profound, elegant
T	静的, 非活性的	static, inactive

3.6.3 意味の数学モデルの概要

本節では, 言葉と言葉の類似度を計量することができるモデルである意味の数学モデルの概要を示す. 詳細は文献 [12], [4] に述べられている.

(1) メタデータ空間 MDS の設定

メタデータ空間 MDS と呼ばれる, 検索対象となるメディアデータをベクトルで表現したデータにマッピングするための正規直交空間(以下, MDS)を設定する.

(2) メディアデータのメタデータを MDS へ写像

設定された MDS へ、メディアデータのメタデータをベクトル化し写像する。これにより、検索対象データのメタデータが同じメタデータ空間上に配置されることになり、検索対象データ間の意味的な関係を空間上での距離として計算することが可能となる。

(3) MDS の部分空間(意味空間)の選択

検索者は与える文脈を複数の単語を用いて表現する。検索者が与える単語の集合をコンテキストと呼ぶ。このコンテキストを用いて MDS に各コンテキストに対応するベクトルを写像する。これらのベクトルは、MDS において合成され、意味重心を表すベクトルが生成される。意味重心から各軸への射影値を相関とし、閾値を超えた相関値(以下、重み)を持つ軸からなる部分空間(以下、意味空間)が選択される。

(4) MDS の部分空間(意味空間)における相関の定量化

選択された意味空間において、メディアデータベクトルのノルムを検索語列との相関として計量する。これにより、与えられたコンテキストと各メディアデータとの相関の強さを定量化している。この意味空間における検索結果は、各メディアデータを相関の強さについてソートしたリストとして与えられる。

3.6.4 任意の言葉の音相の印象に合致した楽曲の自動構成方式

本節では、任意の言葉の音声情報からその印象に合致した楽曲を自動生成する方式について示す。本方式の概要は以下の通りである。全体の概要図を図 3-14 に示す。

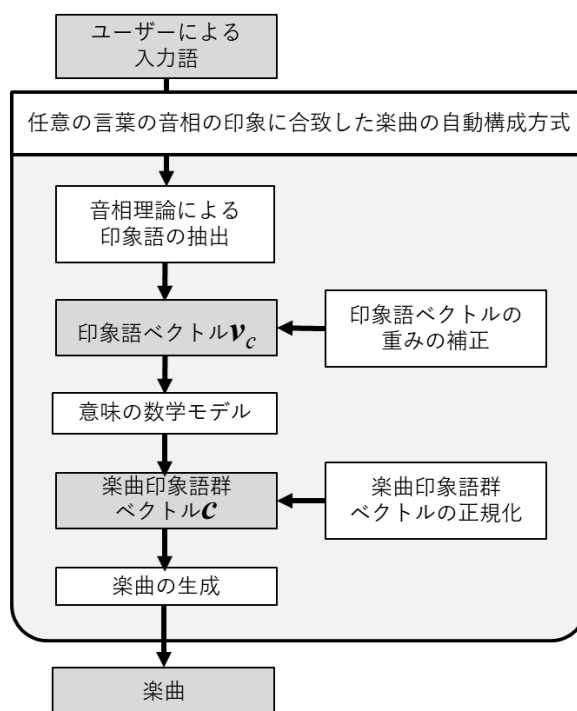


図 3-14 任意の言葉の音相の印象に合致した楽曲の自動構成方式の概要。

(1) 音相理論による印象語の抽出

3.6.2 節に示した，対象となる語の音相に合致した印象語の抽出方式を用いて，ユーザーから入力された任意の言葉の音相から，印象語ベクトル \mathbf{v}_c を抽出する．

(2) 音相の印象語ベクトル \mathbf{v}_c の重みの補正

(1)により得られた音相の印象語ベクトル \mathbf{v}_c を，より入力語の印象を的確に反映できるように補正する．詳細は 3.6.4.1 節に示す．

(3) 音相の印象語ベクトルと各楽曲印象語群との相関の計量

3.6.3 節に示した意味の数学モデルを用いて，(2)で得られた印象語ベクトル \mathbf{v}_c と Hevner の指定する 8 つの各楽曲印象語群との相関を計量し，楽曲印象語群ベクトル \mathbf{c} を構成する．

(4) 楽曲印象語群ベクトルの正規化

楽曲印象語群ベクトル \mathbf{c} の取る値は不定であるため，これを正規化し，最小の要素が-1，最大の要素が1になるように正規化する．詳細は 3.6.4.2 節に示す．

(5) 楽曲の構成

(4)で得られた楽曲印象語群ベクトル \mathbf{c}' から，3.4 節に示した楽曲自動生成方式を用いて楽曲を構成する．

3.6.4.1 音相を表す印象語の重みの補正

音相を表すとして \mathbf{v}_c に選ばれた語は，肯定的な意味を表す印象語のみで構成されているため，Hevner の指定する 8 つの印象語群との相関を計量した際に，特定の印象が強調されやすい傾向があり，出力される楽曲において入力語毎の印象の違いを正確に反映できていない可能性がある．そこで，重みの小さい語をマイナスの印象として見ることにより，印象語ベクトル \mathbf{v}_c を Hevner の指定する 8 つの印象語群に変換した際に強調されにくい印象を反映することができると考えられる．ここでは，予備実験に基づき，音相から選ばれた印象語の重みを，最小を-1，最大を1になるよう写像する．印象語ベクトル \mathbf{v}_c の要素を v_{ci} ，写像後の要素を v'_{ci} としたとき，上記の写像を以下の式に示す．ただし， $\max(\mathbf{v}_c)$ ， $\min(\mathbf{v}_c)$ は，それぞれ \mathbf{v}_c の要素のうち最大，最小のものを表す．

$$v'_{ci} = \frac{2v_{ci} - (\max(\mathbf{v}_c) + \min(\mathbf{v}_c))}{\max(\mathbf{v}_c) + \min(\mathbf{v}_c)}. \quad (3-59)$$

さらに，多くの印象語が入力された場合，関連の低い語がノイズとなり，入力語ごとの印象の差を弱めると考えられるので，相関値の低い要素の重みを 0 とする．ここでは， ε_m から ε_p までの値を持つ要素の重みを 0 とする．式を以下に示す．

$$v_{ci} = \begin{cases} 0 & (\varepsilon_m \leq v_{ci} \leq \varepsilon_p), \\ v_{ci} & (\varepsilon_m > v_{ci}, v_{ci} > \varepsilon_p). \end{cases} \quad (3-60)$$

具体的な ε_m と ε_p の値は実験の際に決める．これらの補正方式は，この方法に限ったものではない．

3.6.4.2 楽曲印象語群ベクトルの正規化

音相を表す印象語群ベクトル \mathbf{v}_c から楽曲印象語群ベクトル \mathbf{c} への変換は、二つの異なるメディアの接続にあたり、その値の正規化方法には解釈に応じた様々な方法があり得る。ここでは、より多様な楽曲が生成されることが良いと仮定し、楽曲印象語群ベクトル \mathbf{c} の値に差が付きやすい方式として、 \mathbf{c} の最小の要素が-1，最大の要素が1 になるように \mathbf{c} を正規化する。正規化後の楽曲印象語群ベクトルを \mathbf{c}' ，その \mathbf{c}' の各要素を $c'_i (i = 1, 2, \dots, 8)$ としたとき、正規化後の c'_i を以下の式で表す。

$$c'_i = \frac{2c_i - (\max(c) + \min(c))}{\max(c) - \min(c)}. \quad (3-61)$$

この正規化方式は、上記の解釈に応じて設定されたものであり、この方式に限定されるものではない。

3.7 実験

本研究で示した方式のシステムを構築し、検証実験を行う。大きく分けて三種類の実験を行う。実験 1 では、 $ML-iML$ 系が実現されているかどうかの検証を行う。実験 2 では、提案システムにより楽曲が実際に生成されていることと、作曲システムとしての表現力が十分であるかについて検討する。3.6 節で構成した、任意の日本語を入力として、その音相の印象から楽曲を生成し、その検証を行う。

3.7.1 実験システム

実験システムを、Python 言語及び Perl 言語を用いて実装した。本システムは、WWW ブラウザをユーザーインターフェイスとし、任意の楽曲印象語群ベクトルを入力として、楽曲を生成し、提示する。楽曲印象語群ベクトルの入力には、テキストファイルで事前に用意して選択項目から選んで与えることと、各語群 C1~C8 のそれぞれについて、ブラウザ上から-1 から 1 の値を選んで入力とする事が出来る。さらに、3.6 節で構成した、任意の日本語を入力として、その音相の印象から楽曲を生成するシステムに対しては、任意のひらがなまたはカタカナの文字列を入力として与えることができる。楽曲は Standard MIDI File の形式で出力される。

3.7.2 実験 1 ($ML-iML$ の検証)

3.5 節で示した検証方法に従って、3.4 節で構成した楽曲を対象とした印象メタデータ抽出方式と印象メタデータから楽曲を生成する方式が $ML-iML$ 系を構成できているかの検証の実験を行う。

3.7.2.1 実験 1 の目的

統計的一般化逆作用素が理論上正しく構成されていることを示す。図 3-13 における 6 つの検証項目と実験の対応を表 3-8 に示す。この中で、理論上保存性を持つ検証項目は 1, 4, 5 である。これらが保存されていることが、特徴保存の $ML-iML$ 系が理論上正しく構成されていることを意味する。これらは数値実験によって示す。その他の項目については理論上は保存性を持たないが、「指定した印象を表現することが出来るシステム」としての実効性を示すために被験者実験を行う(実験 1-4)。また、一般化逆行列を用いている項目 6 については、限定的な範囲での保存性を持つと考えられるため、その検証を数値実験によって行う。なお、検証項目 1 は行列計算をするだけであり、図 3-13 における C1 と C2 が等しくなる事は自明である。誤差の生じる余地も計算機による丸め誤差しかないため、実験結果は省略する。

表 3-8 実験 1 の検証項目と実験の対応.

検証項目	理論上の保存性	数値実験			ユーザー実験
		完全な保存性	誤差を含めた保存性	相対的な保存性	
1	○	●説明のみ			
2	×				●実験 1-4-2
3	×				●実験 1-4-1
4	○(誤差あり)		●実験 1-1		
5	○(誤差あり)		●実験 1-2		
6	×(部分的な保存性あり)			●実験 1-3	●実験 1-4-3

実験 1-1 では、検証項目 4 について検証する。検証項目 4 は理論上保存性を持つが、楽曲を作成する際に楽曲の長さや音符の音価が離散化されるため、必ず誤差が発生する。その誤差が理論上想定される範囲内であることを確認する。実験 1-2 では、検証項目 5 について検証する。検証項目 5 は理論上保存性を持つが、こちらでも一部の要素に誤差が発生する。実験 1-1 と同様に、その誤差が理論上想定される範囲内であることを確認する。実験 1-3 では、検証項目 6 について検証する。検証項目 6 は理論上保存性を持たないが、一般化逆行列を用いた変換を行っているため、部分的な保存性を有していると考えられる。その保存性について検証する。

実験 1-4 では、検証項目 2,3,6 について検証する。これらの検証項目は保存性を持たない。これらはユーザー実験によりシステムの実用性を確認する。

3.7.2.2 実験 1-1:検証項目 4 特→曲→特

§ 実験の目的と方法

検証項目 4 について実験を行う。「楽曲特徴値データ→ 楽曲 → 楽曲特徴値データ」と変換を行う。そして、入力と出力の間の保存性について検証を行う。楽曲特徴値データを採り得る値の範囲内でランダムに 100 個作成し入力とする。入力データと出力データの値が同じ場合を正解とする。分解能により必ず誤差が生じる値については、差が分解能未満であれば正解と見なす。

§ 実験結果 1-1

表 3-9 に、入力と出力で一致した楽曲特徴値データのそれぞれの正解数を示す。全て正解となった。表 3-10 に、誤差が生じる値についての分解能と、100 回のうちの最大の誤差を示す。どれも想定以内になっている。以上から、理論上想定した結果が得られ、検証項目 4 は保存性を持つと言える。

表 3-9 実験 1-1：結果 1(入力と出力で一致した楽曲特徴値データの数).

楽曲特徴値のラベル	同じになった数 / 試行回数
kn	100 / 100
mt	100 / 100
tem	100 / 100
mnap	100 / 100
unac	100 / 100
ac	100 / 100
tc	100 / 100
oc	100 / 100
nc	100 / 100
um	100 / 100
dm	100 / 100
lm	100 / 100

表 3-10 実験 1-1：結果 2(分解能と最大の誤差).

分解能		最大の誤差	
δ_{unacac}	0.03125	unac	0.01554
		ac	0.01554
δ_{tcocnc}	0.12500	tc	0.06143
		oc	0.06182
		nc	0.11444
δ_{umdmlm}	0.01786	um	0.00884
		dm	0.00892
		lm	0.01693

3.7.2.3 実験 1-2:検証項目 5 構→特→構

§ 実験の目的と方法

検証項目 5 について実験を行う。「楽曲構造要素ベクトル → 楽曲特徴値データ → 楽曲構造要素ベクトル」と変換を行う。そして、入力と出力の間の保存性について検証を行う。楽曲構造要素ベクトルを採り得る値の範囲でランダムに 100 個作成し入力とする。入力データと出力データの値が同じ場合を正解とする。分解能により必ず誤差が生じる要素については、差が分解能未満であれば正解と見なす。理論上は誤差が生じない要素についても計算機による丸め誤差が発生するため、その場合は差が十分小さければ正解とする。十分小さい誤差を分解能= 0.00001 として定義する。

§ 実験結果 1-2

実験結果を表 3-11, 表 3-12 に示す. 表 3-11 では, 楽曲構造要素ベクトルのそれぞれの値が保存性を持つかどうかを示した. その結果, 楽曲構造要素ベクトルのすべての値について, 保存性がある事が示された. 表 3-12 では各パラメータの分解能と, 100 件のうち, そのパラメータにおいて発生した最大誤差を示した. その結果, どのパラメータも範囲内に収まっている事が示された. 以上から, 理論上想定した結果が得られ, 検証項目 5 は保存性を持つと言える. 実験 1 と 2 の結果, およびに検証項目 1 により, 特徴保存の \mathcal{ML} - $i\mathcal{ML}$ 系が構成出来ていることが確認できた.

表 3-11 実験 1-2: 結果 1(楽曲構造要素の検証).

楽曲構造要素のラベル	数 / 試行回数
key	100 / 100
tempo	100 / 100
pitch	100 / 100
rhythm	100 / 100
harmony	100 / 100
melody	100 / 100

表 3-12 実験 2: 結果 2(楽曲構造要素の分解能の検証).

分解能		最大の誤差	
δkey	0.10417	key	0.10415
$\delta tempo$	0.01042	tempo	0.01005
$\delta pitch$	0.02632	pitch	0.02536
その他	0.00001	rhythm	0.00000
	0.00001	harmony	0.00000
	0.00001	melody	0.00000

3.7.2.4 実験 1-3: 検証項目 6 語→構→語

§ 実験の目的と方法

検証項目 6 について実験を行う. 「楽曲印象語群ベクトル → 楽曲構造要素ベクトル → 楽曲印象語群ベクトル」と変換を行う. そして, 入力と出力の間の保存性について検証を行う. 検証項目 6 は理論上は保存性を持たないが, 部分的な保存性を持つと考えられる. 入力と出力に対してベクトルの類似度計算を行い, 有している保存性を明らかにする. 楽曲印象語群ベクトルを各要素の値が-1~1 になるようにランダムに 100 個生成する. さらに上記の変換を行う. そして出力データ一つずつから, 入力データ 100 個について類似度計算をし, 元になった入力データとの類似度が最大であったとき, それを正解とする. ベクトルの類似度計算としては, コサイン尺度とユークリッド距離の二つを用いる.

§ 実験結果 1-3

実験の結果を表 3-13 に示す。コサイン尺度では全ての変換後のベクトルにおいて、自身の変換前のベクトルが他のベクトルに比べて最大の類似度があるという結果となった。ユークリッド距離ではそれよりは大きく正解率が低い。これは、変換によって、各ベクトルの相対的な関係のみ保存していることを示していると考えられる。

表 3-13 実験 1-3 結果.

コサイン尺度	ユークリッド距離
100 / 100	58 / 100

3.7.2.5 実験 1-4:検証項目 2,3,6 被験者実験

§ 実験の目的と方法

検証項目 2,3,6 について、被験者実験を行う。実験番号と検証項目の対応は以下の通りである。

- 実験 1-4-1:検証項目 3
- 実験 1-4-2:検証項目 2
- 実験 1-4-3:検証項目 6

実験に用いる、変換の元となる楽曲を表 3-14 に示す。全 10 曲とする。これらを 2 曲ずつセットとする。実験 4 で比較に用いる楽曲の生成の流れを図 3-15 に示す。表 3-14 に示す楽曲を、変換の元（図 3-15 における A1）として利用する。被験者は、曲名を伏せられた中で、検証項目における変換元にあたる楽曲を 2 曲聴き、さらに、変換後にあたる楽曲のうちランダムに選ばれた楽曲 1 つを聴き、変換元にあたる 2 曲のうちどちらが変換後の楽曲に似ているかを答える。変換元の楽曲を当てられれば正解とする。被験者は 10 人とする。

表 3-14 実験 1-4：実験に用いた楽曲.

番号	曲名（作曲者）
1	ハンガリー舞曲 5 番（ブラームス）
2	前奏曲 Op.28-7(ショパン)
3	ラ・カンパネラ（リスト）
4	カノン（パッヘルベル）
5	ソナタ「月光」第一楽章（ベートーベン）
6	新世界より「家路」（ドヴォルザーク）
7	交響曲第三番第三楽章（ブラームス）
8	菩提樹（シューベルト）
9	ウィリアム・テル序曲（ロッシーニ）
10	埴生の宿（イングランド民謡）

上正しく構成されていることを示した。ただし、数値実験によって示した系の正当性に比べて、システムの実効性についてはあくまで従来通りの被験者実験に依っている。被験者実験の結果は、個々人の感性や音楽的知識の有無によるばらつきの他に、メディアデータ生成の制約条件として用いる統計情報に依存すると考えられる。メディアデータ抽出の元となる楽曲が、用いた統計情報の分布に含まれる場合、生成される楽曲が抽出元の楽曲の持つ特徴をより厳密に表現可能となる。

3.7.3 実験 2 (楽曲生成および作曲システムの表現力の検討)

実験 1 では、 ML - iML 系が理論上正しく生成されていることを確認した。ただし、作曲システムが総体として有用なものであるかはそれだけでは示されていない。実験 2 では、提案システムにより楽曲が実際に生成されていることと、作曲システムとしての表現力が十分であるかについて検討する。

実験 2-1 では、 iML を実現したシステムによって実際に楽曲が構成される事を実証する。楽曲印象語群ベクトルを入力として、楽曲を構成する例を示す。

実験 2-2 では、既存の楽曲から ML を用いて印象を抽出し、その印象を iML の入力として楽曲を生成することで、逆演算において保持される情報と保持されない情報について比較検討する。具体的には、抽出時と生成時に現れる楽曲構造要素ベクトル、楽曲特徴値、楽曲をそれぞれ比較する。

実験 2-3 では、実験システムによって構成された楽曲が、指定した印象を判別可能であるほどの表現力を有しているかを検証する。方法としては、二つの異なる楽曲を元に ML を用いて楽曲印象語群ベクトルを抽出し、片方の楽曲印象語群ベクトルから iML を用いて楽曲を生成し、元になった楽曲を当てる被験者実験を行う。

3.7.3.1 実験 2-1 (楽曲印象語群からの作曲)

§ 実験の目的と方法

楽曲印象語群ベクトルを指定し、実験システムを用いて楽曲を 2 曲生成する。今回、印象メタデータとして入力とした楽曲印象語群ベクトルは、明るい曲とその反対の暗い曲を示すことを想定したものである。明るい曲を表現する楽曲印象語群ベクトルの要素は、C6(“Happy, Bright”など)と C5(“Graceful, Humorous”など)の値を“1”とし、C2(“Sad, Dark”など)を“-1”，C3(“Dreamy, Tender”など)を“-0.5”とし、残りは“0”とした。反対の暗い曲については、明るい曲の楽曲印象語群ベクトルの要素の符号をそれぞれ反転し、C6 および C5 を“-1”，C2 を“1”，C3 を“0.5”とした。

§ 実験結果 2-1

楽曲生成の中間データである楽曲構造要素ベクトルと楽曲特徴値の各値を表 3-16 に示す．生成された楽曲の楽譜を図 3-16，図 3-17 に示す．



図 3-16 明るい印象から生成された楽曲．



図 3-17 暗い印象から生成された楽曲．

図 3-16 および図 3-17 から，実際に楽曲が生成されていることが確認できる．また，楽曲の仕様として指定した制約を逸脱しておらず，想定通りの楽曲生成が行われていることが確認できた．表 3-16 の楽曲構造要素ベクトルを比較すると，各構造要素についての絶対値が等しく，符号は逆になっていることが分かる．

表 3-16 実験 2-1：生成された楽曲の楽曲構造要素ベクトルと楽曲特徴値の各値.

楽曲構造要素ベクトル			楽曲特徴値		
	明るい曲	暗い曲		明るい曲	暗い曲
key	1.0000	-1.0000	kn	G	Cm
tempo	0.5398	-0.5398	mt	4/4	4/4
pitch	0.7432	-0.7432	tem	140	62
rhythm	0.6669	-0.6669	mnap	79	50
harmony	0.6271	-0.6271	unac	0.8335	0.1665
melody	-0.6056	0.6056	ac	0.1665	0.8335
			tc	0.7127	0.0354
			oc	0.0856	0.6625
			nc	0.2017	0.3021
			um	0.3008	0.6273
			dm	0.5848	0.3433
			lm	0.1143	0.0295

明るい曲を作ることと想定した曲(以下, 明るい曲)では, 音高 mnap が大きく, 暗い曲を作ることと想定した曲(以下, 暗い曲)では mnap は小さい. 楽曲においてもそれが反映されており, 明るい曲の楽音は全体的に高い音で構成されており, 反対に暗い曲では低い音で構成されている.

二つの曲の楽曲特徴値を比べると, 調性 kn は明るい曲では長調のト長調になっており, 暗い曲では短調のハ短調になっている. テンポ tem は, 標準である 88 から, 明るい曲では 140 と速く, 暗い曲では 62 と遅くなっている. これらの楽曲特徴量は, 入力の影響メタデータに対して想定した通りとなっている.

「明るい」「暗い」という印象とは直接関係しないと思われる, その他の特徴について述べる. rhythm に対応する ac と unac は, この二曲ではちょうど逆になっている. それを反映して, 明るい曲では伴奏に四分音符が多く, 暗い曲では八分音符が多い. 和声 harmony に関する特徴値 tc, oc, nc について見ていくと, 絶対値は異なるものの, 三和音に当たる tc と四和音に当たる oc の値の大小関係が逆転している. 楽曲でもそれが反映され, 明るい曲は三和音が多く, 暗い曲では四和音が多い. 同様に, 旋律 melody に関する特徴値 um, dm, lm についても, 旋律の上昇に相当する um と下降に相当する dm の割合が逆転している.

以上の結果から, 実験によって, 入力した印象メタデータを反映した楽曲が生成できていると言える. しかしながら, 本方式においては, ML と iML が正しく逆の演算になっていることが, 印象メタデータと楽曲の印象が正しく対応しているということによって示されることが重要である. この点について, 実験 2, 実験 3 にてより深く検討していくこと

とする。

3.7.3.2 実験 2-2 (既存楽曲から抽出したメタデータからの作曲)

§ 実験の目的と方法

既存の楽曲から ML を用いて印象を抽出し、その印象を iML の入力として楽曲を生成することで、逆演算において保持される情報と保持されない情報について比較検討する。具体的には、抽出時と生成時に現れる楽曲構造要素ベクトル、楽曲特徴値、楽曲をそれぞれ比較する。実験に用いる既存の楽曲として、「ハンガリー舞曲 No.5」と「家路」の2曲について、印象抽出と楽曲生成を行った。

§ 実験結果 2-2

印象抽出の元になった楽曲と、抽出された印象メタデータから生成された楽曲の例を示す。「ハンガリー舞曲 No.5」の冒頭8小節の楽譜を図 3-18 に、生成された楽曲を図 3-19 に示す。「家路」の冒頭8小節の楽譜を図 3-20 に、生成された楽曲を図 3-21 に示す。またこの二つの曲の、印象抽出時と楽曲生成時の楽曲特徴値、楽曲構造要素ベクトルをそれぞれ表 3-17、表 3-18 に示す。

抽出と生成の変換を経る中で、楽曲の情報として保持されているものと、保持されていない物に分けて述べる。二つの楽曲について、楽曲構造要素ベクトルは全て保持されている。僅かな差は丸め誤差によるものと思われる。楽曲特徴値の中で情報が保持されているのは、調(kn)、テンポ(tem)、音高(mnap)、伴奏の四分音符の演奏時間割合(unac)、伴奏の八分音符の演奏時間割合(ac)となる。

保持されていないのは、拍子、和声に関する特徴値(tc, oc, nc)、旋律に関する特徴値(um, dm, lm)となる。この中で、生成の度に値が変化するものに"*"を付けて示した。

楽曲特徴値に着目すると、harmony に関する特徴値である tc, oc, nc と、旋律に関する特徴値である um, dm, lm は、原曲と生成時で大きく異なっている。このようになる理由としては、1つの楽曲構造要素が3つの楽曲特徴値によって決まる関係にあるため、逆演算の際の値の決定に自由度が含まれてしまうからである。この自由度による特徴値の違いがあっても、楽曲の印象が保持されているかについての被験者調査を実験3にて示す。

表 3-17 実験 2-2 : 「ハンガリー舞曲 NO.5」における抽出時と生成時の各値.

楽曲構造要素ベクトル			楽曲特徴値		
	原曲	生成時	原曲	生成時	
key	-0.6875	-0.6874	kn	F#m	F#m
tempo	0.4583	0.4583	mt	2/4	4/4
pitch	-0.3684	-0.3684	tem	132	132
rhythm	-0.8654	-0.8655	mnap	57	57
harmony	0.1217	0.1217	unac	0.0673	0.0673
melody	-0.0870	-0.0870	ac	0.9327	0.9327
			tc	0.1217	*0.5186
			oc	0.0000	*0.3970
			nc	0.8783	*0.0844
			um	0.4463	*0.2951
			dm	0.4463	*0.3359
			lm	0.0667	*0.3690



図 3-18 「ハンガリー舞曲 No.5」の冒頭 8 小節.



図 3-19 「ハンガリー舞曲 No.5」を入力として作曲された楽曲.

表 3-18 「家路」における抽出時と生成時の各値.

楽曲構造要素ベクトル			楽曲特徴値		
	原曲	生成時		原曲	生成時
key	0.9375	0.9375	kn	C	C
tempo	-0.8750	-0.8750	mt	4/4	4/4
pitch	0.3000	0.3000	tem	46	46
rhythm	0.2942	0.2941	mnap	70	70
harmony	-0.1955	-0.1955	unac	0.6471	0.6471
melody	-0.0018	-0.0018	ac	0.3529	0.3529
			tc	0.3152	*0.1522
			oc	0.5107	*0.3478
			nc	0.1741	*0.5000
			um	0.4463	*0.2496
			dm	0.4472	*0.2504
			lm	0.1065	*0.5000



図 3-20 「家路」の冒頭 8 小節.



図 3-21 「家路」を入力として作曲された楽曲.

3.7.3.3 実験 2-3（作曲システムの表現力の検証）

§ 実験方法

システムによって構成された楽曲が、指定した印象を表現出来ているのかを被験者実験によって検証する。二つの異なる楽曲から ML を用いて楽曲印象語群ベクトルを抽出し、片方の楽曲印象語群ベクトルを iML の入力として楽曲を生成し、元になった楽曲を当てる被験者実験を行う。楽曲の印象を判断する際には、楽曲が長調であるか短調であるかが最も大きな影響を及ぼすと考えられるため、その一点のみの評価とならないように、長調の曲同士または短調の曲同士の比較においても楽曲の印象が判断できるかを検証する。印象を抽出する既存の楽曲を 12 曲用意し、「長調-短調」の組を 2 問、「長調-長調」の組を 2 問、「短調-短調」の組を 2 問、計 6 問の実験とする。実験に用いる、印象抽出の元となる楽曲を表 3-19 に示す。この表には、あわせて、これらの楽曲が長調であるか短調であるかを示している。またこれらの楽曲は、曲の冒頭から、転調やテンポの大幅な変化などの大きな曲調の変化を含まない範囲を著者が任意の小節数分抜粋したものである。抜粋した小節数も同表に示している。さらに、参考のため、各楽曲から抽出した印象メタデータである楽曲印象語群ベクトルの要素のうち、最大のものからと三番目に大きいものまでを同表に示している。なお今回のケースでは、これら三番目までの楽曲印象語群ベクトルの要素の値はすべて正となっている。

表 3-19 実験 2-3：実験に用いた楽曲。

ID	曲名(作者)	調の 長短	小節数	楽曲印象語群ベクトルの 最大～第三要素
1	ワルツ第 7 番(ショパン)	短調	32	C7, C2, C8
2	カノン(パッヘルベル)	長調	10	C1, C5, C8
3	ラ・カンパネラ(リスト)	短調	21	C3, C6, C4
4	前奏曲 Op.28-7(ショパン)	長調	16	C1, C5, C4
5	ウィリアム・テル序曲(ロッシーニ)	長調	48	C6, C7, C5
6	新世界より「家路」(ドヴォルザーク)	長調	15	C4, C5, C1
7	埴生の宿「イングランド民謡」	長調	16	C6, C5, C4
8	ソナタ「ワルトシュタイン」第一楽章(ベートーベン)	長調	13	C6, C7, C8
9	ソナタ「月光」第一楽章(ベートーベン)	短調	16	C2, C1, C8
10	飛翔(シューマン)	短調	16	C7, C2, C3
11	ハンガリー舞曲 5 番(ブラームス)	短調	34	C2, C7, C3
12	前奏曲 Op.28-4(ショパン)	短調	26	C3, C2, C4

被験者は、印象の抽出元にあたる楽曲を 2 曲聴き、続けて、自動生成によって得られた楽曲のうち一曲のみを聴き、始めに聴いた二曲のうちのどちらにより似ているかを答える。

選んだ曲が生成元であれば正解とする。被験者は10人とした。

また、生成された楽曲の音楽的・芸術的評価についても、自由記述欄を設けて、被験者に記入してもらった。

§ 実験結果 2-3

実験結果を表 3-20 に示す。最も高い正解率は 100%、最も低い正解率は 70%、平均正解率は 90%と、全体的に高い正解率となった。実験結果から、作曲システムが、楽曲印象語群として表わされた楽曲の印象を、楽曲として表現する能力を持っていると言える。長調同士、短調同士の楽曲においても高い正解率となっており、楽曲の印象を楽曲の調の長短以外の要素によっても表現できていると言える。

表 3-20 実験 2-3 の結果。

実験セット	曲の長短	正答率
1 と 2	長-短	80%
3 と 4		70%
5 と 6	長-長	90%
7 と 8		100%
9 と 10	短-短	100%
11 と 12		100%
		平均正答率 90%

楽曲印象語群ベクトルに着目すると、ID7 と ID8 の実験セットにおいては、どちらも最大要素が C6 となっており、印象メタデータが比較的類似しているケースと言えるが、これも正解率は 100%となっている。ここから、楽曲印象語群ベクトルの各要素は、特に大きい値を持つ要素だけではなく、各要素が複合的に影響して楽曲の印象を決めていると言える。

ID3「ラ・カンパネラ」と ID4「前奏曲 Op28-7」の実験セットが比較的低い正解率となった原因としては、「ラ・カンパネラ」の抜粋した部分のメロディーのほとんどが 16 分音符で構成されているにも関わらず、現状の楽曲生成システムでは 4 分音符と 8 分音符のみを用いて楽曲を生成しているため、16 分音符で構成されるメロディーを再現していないことなどが考えられる。これは作曲システムの制約を変更することで改善が可能であると考えられる。

被験者による楽曲の音楽的・芸術的評価としては、1.「メロディーが掴みにくい。和声改善すればメロディーがより際立つのではないか」、2.「曲調が淡々としている」、3.「楽曲のフレーズが途中から始まっているように感じた」、といった意見が挙がった。一方 4.「響きがおかしいとは感じなかった」という意見もあった。1.「メロディーが掴みにくい。和声改善すればメロディーがより際立つのではないか」については、アヴェイラブル・ノート

スケールを考慮した制約を加えることで改善が可能であると考えられる。また、本システムにおいては音量が考慮されていないため、メロディー部分の音量を大きくする制約を加えることや、音符の小節内での位置における音量を考慮した制約を加えることによっても改善が可能であると考えられる。2.「曲調が淡々としている」については、そもそも本システムは単一の印象を表す楽曲を生成することを目的としているため一概に悪い結果とは言えないが、本システムを拡張し起伏のある楽曲を構成する方法としては、複数の印象を入力として与え、それらの印象の遷移を連続的に表現するような楽曲を生成するシステムを構築することが考えられる。3.「楽曲のフレーズが途中から始まっているように感じた」については、曲の末尾と同様に曲の冒頭に制約を加えることで改善が可能であると考えられる。これらは今後の課題とする。

3.7.3.4 実験2 考察

実験 2-1 により、実験システムが提案方式を想定通り実装できており、メディアコンテンツ生成において生じる不良設定問題を解決して楽曲が生成できる事が確認できた。実験 2-2 により、 ML による印象抽出と iML による楽曲生成の過程において、保持される情報と保持されない情報があることが確認できた。実験 3 により、変換の過程で保持されない情報があるにもかかわらず、生成した楽曲が印象を判別するのに十分な情報を保持していることが確認できた。以上により、楽曲を対象とした iML が実現されていること、およびその有効性を示した。

今回の被験者実験では高い正解率を得ることが出来たが、被験者実験の結果は、個々人の感性や音楽的知識の有無によるばらつきの他に、実験に用いる抽出元の楽曲のタイプに依存すると考えられる。今回実験に用いた楽曲は、Hevner の研究において対象とされていた、クラシックのピアノ曲に限定している。それ以外の形式の楽曲においては、それぞれ新たな ML の構成及びに、メディアコンテンツ生成における統計情報の選択が必要となると考えられる。ジャンル別、時代別、作曲家別などの ML 及び iML を構成していく事は今後の課題である。

3.7.4 実験3 (音相の印象に基づく作曲)

3.6 節で構成した、任意の日本語を入力として、その音相の印象から楽曲を生成し、その検証を行う。なお、現段階では生成が可能であることと簡単な考察を示すことに留まる。システムの定量的な検証、ユーザー実験に基づく有効性の検証、それに基づく提案方式の改善は今後の課題である。

3.7.4.1 実験システムの設定

§ 意味の数学モデルの構成条件

本システムで用いる意味の数学モデルは入力に応じた計量空間を生み出すシステムであ

るため、その構成条件について述べる。意味の数学モデルの基本構成は 3.6.3 節に示す通りである。メタデータ空間 MDS の設定については、“Longman Dictionary of Contemporary English” [17] という英英辞典を利用した。同辞書は約 2,000 語の基本語を用いて約 56,000 語の見出し語を説明している。この基本語を特徴とみなし、見出し語の説明で肯定的に使われている語を“1”，否定的に使われている語を“-1”，説明に使われていない語を“0”として 2. 節(1)におけるデータ行列 M を作成した。これにより、約 2,000 次元の正規直交行列であるメタデータ空間 MDS が生成された。MDS では、約 2 の 2,000 乗通りの意味の様相を表現可能である。

§ 印象語の重みの補正における閾値の設定

3.6.4.1 節における、音相の印象を表す語群ベクトルを補正する式における閾値 ε_m と ε_p を、それぞれ $\varepsilon_m = -0.75$, $\varepsilon_p = 0.25$ とする。これは予備実験により、音相理論における輝性の値が高い場合に長調の楽曲が、低い場合に短調の楽曲が生成されやすい値を選んだものである。

3.7.4.2 実験 3 任意の入力語からの楽曲生成実験

§ 実験の目的と方法

日本語の音声を表現する文字列として、任意のカタカナまたはひらがなのみで構成された文字列を入力し、楽曲が生成されることを確認し、その印象を表現できているか検討する。

入力として、文献 [28] において“活動，活性”に関する語として分類されており、任意の言葉の音の印象抽出における輝性についてプラスの印象を持つ“イキイキ”という言葉と，“非活性，非行動性，否定，抗争”に関する語として分類されており、任意の言葉の音の印象抽出における輝性についてマイナスの印象を持つ“グズグズ”という言葉を用いて楽曲を構成する。さらに、特に意味を持たない“ナマテー”という言葉を用いて楽曲を構成する。そして、生成された楽曲およびその特徴値等について考察する。

§ 実験結果

各入力文字語から抽出された音相の印象語ベクトルを表 3-21 に示す。生成された楽曲の楽曲印象語群ベクトル，楽曲構造要素ベクトル，楽曲特徴値の各データを示す。表 3-22 にイキイキの各データを，表 3-23 にグズグズの各データを，表 3-24 にナマテーの各データを示す。また，生成された楽曲を示す。図 3-22 にイキイキの楽曲を，図 3-23 にグズグズの楽曲を，図 3-24 にナマテーの楽曲を示す。

表 3-21 実験 3：入力文字列から抽出された音相の印象語ベクトル.

イキイキ			グズグズ			ナマテー	
special	1.000000		profound	1.000000		popular	1.000000
individual	1.000000		elegant	1.000000		adaptable	1.000000
reasonable	0.609081		vibrant	-0.760714		sharp	-0.758073
real	0.609081		advance	-0.760714		powerful	-0.758073
young	0.550038		healthy	-0.803994		healthy	-0.783986
effervescent	0.550038		clean	-0.803994		clean	-0.783986
modern	0.396003		popular	-0.807572		vibrant	-0.785311
urban	0.396003		adaptable	-0.807572		advance	-0.785311
confidence	-0.942733		young	-0.867924		pure	-0.949109
stable	-0.942733		effervescent	-0.867924		brisk	-0.949109
inactive	-1.000000		light	-0.869603		confidence	-1.000000
static	-1.000000		trippingly	-0.869603		stable	-1.000000
			modern	-0.915269			
			urban	-0.915269			
			brisk	-0.927465			
			pure	-0.927465			
			open-minded	-1.000000			
			bright	-1.000000			

表 3-22 「イキイキ」から生成された楽曲の各値.

楽曲印象語群ベクトル		楽曲構造要素ベクトル		楽曲特徴値	
C1	-0.3910	key	0.3519	kn	Eb
C2	0.0549	tempo	1.0000	mt	4/4
C3	-0.3478	pitch	-0.2551	tem	184
C4	-1.0000	rhythm	0.1297	mnap	59
C5	0.4894	harmony	-0.2919	unac	0.5648
C6	1.0000	melody	-0.2201	ac	0.4352
C7	0.6674			tc	0.1041
C8	0.8768			oc	0.3959
				nc	0.5000
				um	0.2190
				dm	0.3223
				lm	0.4587



図 3-22 「イキイキ」から生成された楽曲.

“イキイキ”を入力として作曲された楽曲は、長調の曲となっており、これは想定に合致している。その他特徴としては、楽曲構造要素の tempo が最大値となっており、想定される範囲で最速のテンポの楽曲が生成されている。非常に速い印象を与える曲であり、これは入力語の印象を適切に反映できていると考えられる。

表 3-23 「グズグズ」から生成された楽曲の各値.

楽曲印象語群ベクトル		楽曲構造要素ベクトル		楽曲特徴値	
C1	-0.0041	key	-1.0000	kn	Cm
C2	1.0000	tempo	-0.7857	mt	4/4
C3	0.2063	pitch	-0.0260	tem	50
C4	0.1771	rhythm	-0.4975	mnap	64
C5	-1.0000	harmony	-0.1794	unac	0.2513
C6	-0.7610	melody	0.1360	ac	0.7487
C7	-0.6536			tc	0.1603
C8	-0.6172			oc	0.3397
				nc	0.5000
				um	0.5232
				dm	0.4595
				lm	0.0173



図 3-23 「グズグズ」から生成された楽曲.

“グズグズ”を入力として作曲された楽曲は、短調の曲となっており、これは想定に合致している。その他特徴としては、テンポも遅く、緩慢な印象を与える曲であり、これは入力語の印象を適切に反映できていると考えられる。

表 3-24 「ナマテー」から生成された楽曲の各値.

楽曲印象語群ベクトル		楽曲構造要素ベクトル		楽曲特徴値	
C1	-0.2694	key	-0.1291	kn	Gm
C2	1.0000	tempo	-0.7454	mt	4/4
C3	-0.1452	pitch	0.6529	tem	52
C4	0.1839	rhythm	-0.5130	mnap	77
C5	0.4040	harmony	0.5062	unac	0.2435
C6	0.3695	melody	0.2326	ac	0.7565
C7	-1.0000			tc	0.5333
C8	-0.9522			oc	0.0271
				nc	0.4395
				um	0.4402
				dm	0.3311
				lm	0.2287



図 3-24 「ナマテー」から生成された楽曲.

“ナマテー”を入力として作曲された楽曲は、短調の、遅いテンポの、音高の高い曲となった。音相の印象を表すベクトルを見ると、popular(庶民的), adaptable(適応性)のみが正の値となっており、その他は全て負の値となっている。もともと意味のない語を入力としているため、この楽曲が入力語の印象を的確に表しているかを判断することはできないが、音相の印象を表すベクトルとの比較では、庶民的であることと、テンポが遅いことには関連があると分析できる。

3.7.4.3 実験3の考察

実験3で示した限りにおいては、入力した文字列の音相の印象を表す楽曲を構成できていると考えられる。しかし、定量的な検証やユーザーによる評価などは行っていないため、このシステムの実効性についてはまだ検討段階にある。今後の研究によって検証および改善を行っていく予定である。

3.7.5 実験全体の考察

実験1では、 $ML-iML$ 系が理論上正しく構成されていることを確認した。また、理論上は情報が保持されない変換についても、ユーザーにとって十分な情報が保存されていることを確認した。

実験2では、システム全体を通しての有効性を確認した。楽曲が提案方式で定めた通りに生成されていることを確認した。また、ユーザー実験によって、システムの有用性を確認した。

実験1と2により、楽曲メディアコンテンツを対象とした統計的一般可逆作用素、および楽曲メディアコンテンツを対象とした印象語抽出方式と組み合わせた $ML-iML$ 系が構成されていることを確認した。

実験3では、実験1と2で確認した自動作曲システムと、日本語の入力語からその音相を表現する印象語を抽出するシステムを接続することで、任意の日本語の入力語の音相の印象に合致した楽曲を生成するシステムが正しく構成されていることを確認した。ただし現段階では定量的な分析やユーザーによる評価は行っていない。それらを行ってシステムを更新・改善するのは今後の課題である。

3.8 まとめ

本章では、メディアコンテンツから、印象語としてあらわされたメタデータを自動抽出する ML の逆演算である、印象メタデータを入力として、メディアコンテンツを自動生成する新たな枠組みである、メディアコンテンツを対象とした統計的一般化逆作用素”Stochastic Generalized Inverse Media-lexicon Transformation Operator”(“ iML ”)を定義した。さらに、 ML と iML を組み合わせて検証を行う枠組みである ML - iML 系の枠組みを提案した。また、 iML の具体的な実現例として、楽曲メディアコンテンツを対象として、印象メタデータから楽曲メディアコンテンツを生成する方式について示した。上記の提案方式をシステムとして実装し、実験を行い、 ML - iML 系が実現されていること、およびに、提案システムの楽曲生成システムとしての有効性を示した。さらに、異種メディア間の連結の事例として、音相理論と上記システムを組み合わせることで、任意の日本語の入力語の音相の印象に合致した楽曲を生成するシステムを構築した。このシステムについても実験を行い、印象に合った楽曲が生成されることを示した。

楽曲を対象とした統計的一般化逆作用素 iML の構成によって、印象を表す言葉からの楽曲メディアコンテンツの生成が可能となった。今後、他のメディアを対象とした iML を構築することによって、楽曲のみならず、画像、動画などの任意の種類のメディアコンテンツを対象としたメディアコンテンツ自動生成方式が構成できると考えられる。

本方式を応用することで、例えばWeb上の楽曲コンテンツおよびSNS等のデータから、楽曲の特徴と印象の相関が求められるのであれば、楽曲構造要素と印象との関係を表す ML をWeb上のデータから作り出せる可能性がある。そうなれば、本研究で提案する統計的一般化逆作用素 iML の構成を行うことにより、データドリブンな計算機による自動作曲方式の実現の可能性がある。

また、音相との連結事例では、未知語でも入力として用いることができるので、我々が印象を知らない新しい言葉の印象をコンピュータが教えてくれるシステムを構築したと考えることもできる。本研究はまだプロトタイプではあるものの、こうした観点でのシステムは、連結の事例の数だけ考えられるため、大きな発展性を持つものと予想される。

今後の課題として、Web上のデータを利用した ML 、 iML の構成方式、個人化方式、および、ジャンル別、時代別、作曲家別などに対応した、様々な音楽メディアに対する新たな ML 及び iML の構成方式が挙げられる。また、さまざまなメディアとの連結事例を構築していくことも課題として挙げられる。

第4章 会議の時系列構造分析によるトピック抽出および合意形成支援の研究

4.1 はじめに

日々の生産活動において、コミュニケーションをする、対話をするということは、情報を交換したり、新たな発想を生んだりする上で非常に重要な役割を担っている。企業などの組織において、コミュニケーション、対話をする場として会議がある。業務時間の中で会議の占める割合は大きい。NTT データ経営研究所の調査結果 [40]によると、会社で開催される会議等が全体業務に占める割合の平均は 15.4%にのぼる。さらに、これらの会議等について問題を抱えている点として、「無駄な会議が多い」、「会議の時間が長い」、「会議の頻度が多い」という回答が上位を占めることが分かっている。これらのことから、会議等、コミュニケーションや対話に費やす時間が全体業務の中で占める割合が高いにも関わらず、これらの効率化がうまく行われていないという問題点が浮かびあがる。

会議の効率化といえば、会議資料の事前の共有や時間厳守、進行役などの役割を明確にするなどの会議の形式に対する改善は行われているが、会議の対話の中の発言などの内容や流れに関して分析をすることによって効率化を図るところまでは行われていないのが現状である。

会議を分析する方法として、これまでエスノグラフィ的な手法 [41]が主に取り組みされてきた。これらの手法は実際の会議の現場に立ち会い、分析者が気になった動作や発言を会議中に観察し続けるものである。これらの手法は、1回の会議での分析コストが高く、継続的に分析を行うことは難しい。また、人間による観察から傾向を導き出すという定性的な研究にとどまっている。

マイクの性能の向上や廉価化により音声データの取得が容易になった。そして音声認識技術の向上により、音声データをテキストデータに変換することが実現しつつある。このテキストデータについてテキストマイニングすることができれば、実際対話で発言された意味を加味した分析が可能となり、会議の効率化のための一助となりうる。ここでいう会議の効率化とは、会議の時短や、より生産的で有意義な会議にすること、さらには、そもそも会議を開催すべきか判断し無駄な会議を排除することなどを指している。

会議の効率化のために必要な方策として、会議のファシリテーションや振り返りが重要であると考え、会議中の対話における最終的な合意事項や結論だけでなく、そこに至ったプロセスを知ることが会議の効率向上につながると考える。

本章では、2章で示した知識創造サイクルモデルのうち知識創造の実現例として、会議で話されている内容から、会議のトピックや、会議が円滑に進行したのかを判断するための指標を抽出する方式を示す。これらの情報を会議の参加者にフィードバックすることで会議の進行を補佐することを目指す。

本章で述べる方式は大きく二つに分かれる。一つ目は、会議の中でどのような議論が行われていたのかを、出現した単語などを用いて可視化する研究である。この研究の中の小分類として、単一の会議を対象としたものと、同一プロジェクト内の複数の会議を対象にしたものがある。これらの研究は [42] [43] [44] で報告している。二つ目は、会議における議論の状態を定義したうえで、その状態の遷移から会議が円滑に進行したのかを判断する指標を構築する研究である。この研究は [45] で報告している。これらの研究は、分析の基礎となる箇所では多くの共通部分を持っている。

4.1.1 4章の構成

本章の構成は以下の通りである。4.2 節では関連研究について述べる。4.3 節では、会議の時系列解析による構造化および可視化手法について述べる。4.4 節では、一連の会議のトピックの関係の可視化手法について述べる。4.5 節では、4.5 議論の発散/収束に着目した会議の時系列構造分析と状態遷移可視化手法について述べる。4.6 節では実験として、可視化手法についての事例を示すとともに、会議の進行を評価する指標の妥当性について検証する。さらに、4.7 節で、会議の進行を評価する手法のハードウェアを含めた実システムのプロトタイプの実装例を示す。4.8 節でまとめと今後の課題を述べる。

4.2 関連研究

本章では関連研究について述べる．4.2.1 節では一般的な会議の分析に関する関連研究について述べる．4.2.2 節では，発散的思考と収束的思考に関する研究について述べる．

4.2.1 会議の分析に関する研究

対話を印象付ける要因は様々である．話している内容はもちろん，声の質，声の大きさ，強いては参加しているメンバーの属性，発言者の振る舞い，聴取者の振る舞いなどが考えられる．

前記のエスノグラフィ的な手法 [41] は，話している内容はもちろんのこと，参加しているメンバーの属性や会議中の振る舞いについて総合的に観察し，対話の評価を行う．これらから導き出された結果は，定性的な評価であり，観察者の経験によるものが大きい．

それに対して，定量的な対話の評価としては，非言語情報に着目した分析手法の研究が盛んに行われている(例えば文献 [46] [47] [48])．このような手法では，発言と沈黙のタイミング，抑揚，声の高さ・大きさ，速度などをデータから読み取り，対話の性質を考察するものである．対話についてこれらの非言語情報に着目した分析の研究が先行して行われた背景として，音声データのままで取り出しやすい特徴であることがあげられる．つまり，認識技術を導入することなく，周波数解析のみで取得できる数値データのため，容易に導入しやすい．

本稿で示す手法は，言語情報に着目した分析手法の一つである．これらの手法は，音声データとして得られた対話データをさらに音声認識技術によって内容をテキストデータに変換したのち，テキストマイニングによって対話の特に発言された内容に着目して性質を導出するものである．

定性評価，定量評価，非言語情報による評価，言語情報による評価とあげてきたが，これらはそれぞれを補完する関係にある．例えば，非言語情報による評価の一つである声の抑揚による評価と言語情報による評価の一つであるトピックの変遷とを組み合わせ，どのトピックのところで声の抑揚が大きくついているかによって，主要トピックを探す手がかりとなるかもしれない．

今後，対話分析の研究も言語情報による評価手法が主流となると考えられる．音声認識の向上による対話のテキストデータ化の進展が主な理由である．これにより，内容や意味にまで踏み込んだ深い分析，評価が可能になる．さらに，テキストデータ化された対話の蓄積が対話ビッグデータとなり，会議一つ一つだけでなく，複数の会議を横断した会議の進展や効率を振り返ることが可能となる．

4.2.2 発散的思考と収束的思考に関する研究

人間の思考プロセスに発散と収束という種別があることを初めて示したのは心理学者の

J.P. Guilford [49]である。Guilford は知能の因子を表現するモデルとして知能構造モデルを提案した。このモデルでは、人間の知能の因子は「領域(content)」「所産(product)」「働き(operation)」の三つの組み合わせで表現される。そのうちの「働き」を大きく「思考能力」と「記憶」に分け、さらに「思考能力」を「認知」と「生産的能力」と「評価力」の三つに分けた。そして「生産的能力」をさらに二つに分け、「収束的思考」と「発散的思考」があるとした。ここで収束的思考とは既知の情報から論理的に一つの解へ至るような思考を指し、発散的思考とは多くの解決策を発想するような思考を指している。

Guilford の知能構造モデルにおける発散的思考と収束的思考は個人の知能に関する分類であった。会議における集団の状態としての発散と収束の段階があることを示したものとしては、A.F.Osborn の開発したブレインストーミング [50]が有名である。ブレインストーミングは会議形式で行われる発想支援のための手法である。以下の 4 つの原則を守ってアイデアを出し合うことによって、お互いの意見による発想の誘発を狙う。1.他人の出した意見を批判しない、2.自由奔放なアイデアを歓迎する、3.質より量を重視する、4.既存のアイデアに便乗してそれらを繋げたりより良いものにするを推奨する、となっている。これはまさに発散的思考に対応すると思われるが、特にこの手法のポイントは、発散的思考をする際に同時に収束的思考をするとそれが阻害されてしまうので、アイデアを出す場面では発散的思考に専念し、その後は収束的思考にあたる評価・批評をする、というように、発散と収束を会議の状態として明確に分離している点である。

収束的思考のための手法としては、川喜田二郎の開発した KJ 法 [51] が有名である。KJ 法もまた会議あるいはワークショップ形式で行われる手法で、創造的問題解決のための手法とされる。KJ 法では、まずブレインストーミングを行ってアイデアを出し、それらをカードとして記録する。そしてそのカードをグルーピングして空間的に配置し、グループ同士の関係やグループ内のアイデア同士の関係を明らかにしていく。このように、アイデアを統合して把握していくプロセスによって新しい知見を生み出す手法である。KJ 法においても発散と収束の段階を意識的に分離していることが見て取れる。

こうした研究を受け、効率的な会議のやり方として、発散と収束の違いを意識してファシリテーションを行うことが重要であるということはいしばしば指摘されている(例えば [52])。

しかし実際の会議の多くは発散と収束を明確に分けて行われることは稀で、実態としては発散と収束を小規模に繰り返しつつ行われているものと思われる。このことは非言語情報から発散と収束のセグメントを判別する市野らの研究 [53]でも指摘されている。また、この論文の中では、従来の会議支援システムが発散と収束のどちらかに特化したものが多く、もしこうした手法およびシステムを適用するのであれば、発散と収束のセグメントを判別する必要があることも指摘されている。

会議における発散と収束を具体的に判定するための指標も様々なものが考案されている。上記の市野の研究 [53]では、会議の時系列分析における非言語情報の特徴として、発話頻

度、オーバーラップ発話時間割合、発話長、交替潜時、話者移行パターン別の頻度の5つを挙げ、これらと発散と収束の会議セグメントを結びつけて、教師あり学習を行って会議が発散フーズにあるか収束セグメントにあるかを判別している。

また富山らの研究 [54] では、議論における各発話を分析単位として、言語情報と非言語情報の両方を用いて発散／収束を判別している。ここでは言語情報としては各発話における新規登場単語の数を発散の指標として対応させており、これは本研究の見解とも一致している。他の言語情報としては直前の発話との単語の一致数を指標としている。非言語情報としては、発話長、オーバーラップ時間、発話の音量、頭部の回転量を用いて、これらの指標と発散／収束の関係を判定するモデルを学習によって構築している。

本研究でも、富山らの研究と同じく、新規出現単語の数を指標の一つである「鮮度」として用いている。ただし、富山らが発言の一つ一つに対して発散／収束の判定を行っているのに対して、本研究ではある時間帯に相当するセグメントを対象としており、個人の発言ではなく会議の場そのものについての判定を行っている。

4.5 節で示す手法では、さらに、着目しているセグメントの全体から見た際の重要度にあたる「代表度」を定義している。そして、「鮮度」と「代表度」を組み合わせることで、さらに議論の性質を表す状態として会議の四状態を定義している。

4.3 会議の時系列解析による構造化および可視化手法

本節では会議中の対話プロセスを構造化して可視化する手法について述べる。本研究の内容は [42] [43] で報告されている。本研究では、会議中の発言を文字起こしして得たテキストデータを対象としている。会議を時系列に沿ってセグメントに分割したうえで、各セグメント間の類似度からセグメントをクラスタリングしてグループとしてまとめ、そのグループを会議の主要な話題のまとまりとして抽出している。そのうえでセグメントを時系列に並べることで、会議のどのようなところでどのような話題が話されていたのかを可視化する。その中では、同一グループ内でよく出てくる語をグループのトピックを表す語として表示し、各セグメントにおける特徴的な語をその時点で話されていたことを示す語として表示している。また、会議全体から見た各セグメントの類似度の高さを、そのセグメントの重要性を表す指標として「代表度」と定義した。これによって、いつどこで重要な話をしていたのかを可視化して提示している。

4.3.1 概要

提案方式のうち、4.3.2 節で示すセグメントベクトルの構成までの流れを図 4-1 に示す。この部分は、囲碁の分析の共通部分となる。入力音声から書き起こされたテキストデータである。まず、テキストデータを任意の数 n のセグメントに分割し、共通のプロセスとして各セグメントをベクトルに変換する。ベクトルの各要素は、4.3.2.3 節に示す各単語の Tf-iSF の重みとなる。

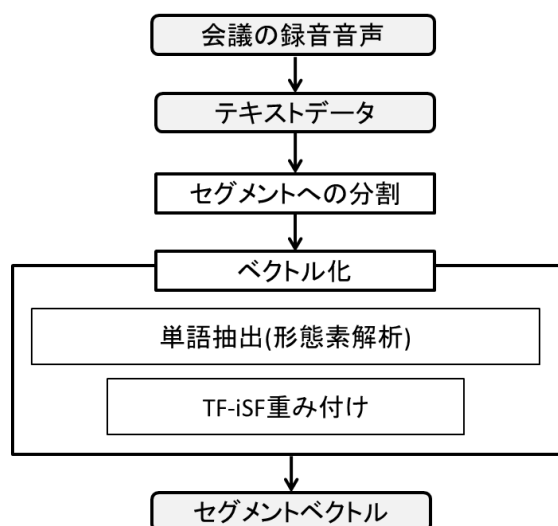


図 4-1 セグメントベクトルの構成。

次に、4.3.3 節で示す、各種分析の流れを図 4-2 に示す。

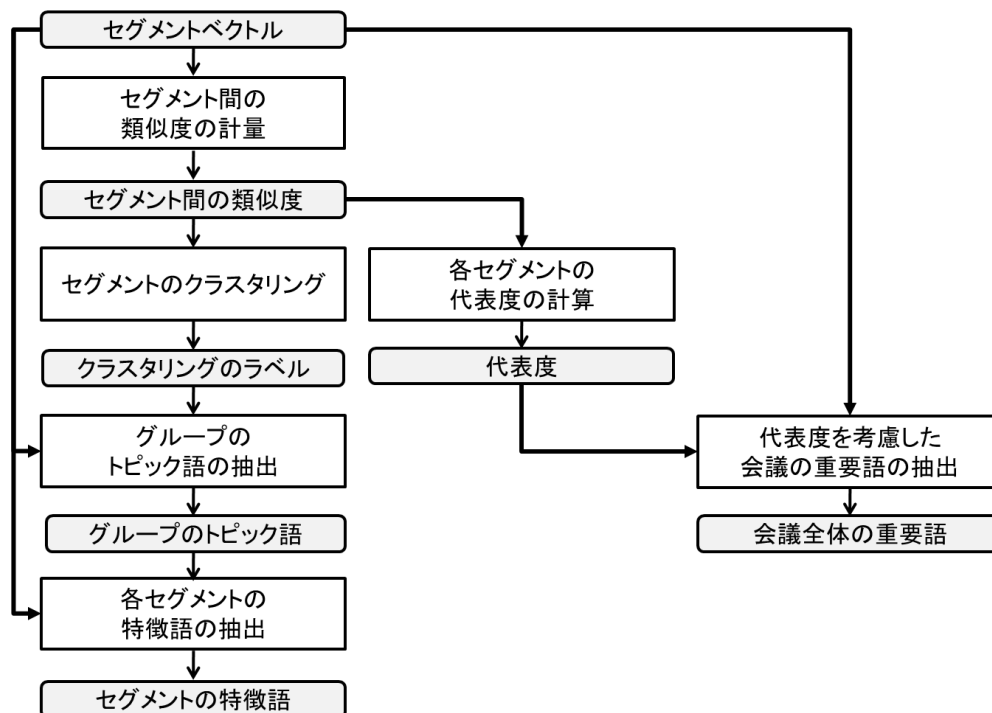


図 4-2 4.3.3 節に示す各種分析の流れ.

これらの分析によって得られたデータを用いて、4.3.4 節において可視化を 4 種類定義する．それぞれの可視化とそれに用いるデータを図 4-3 に示す．

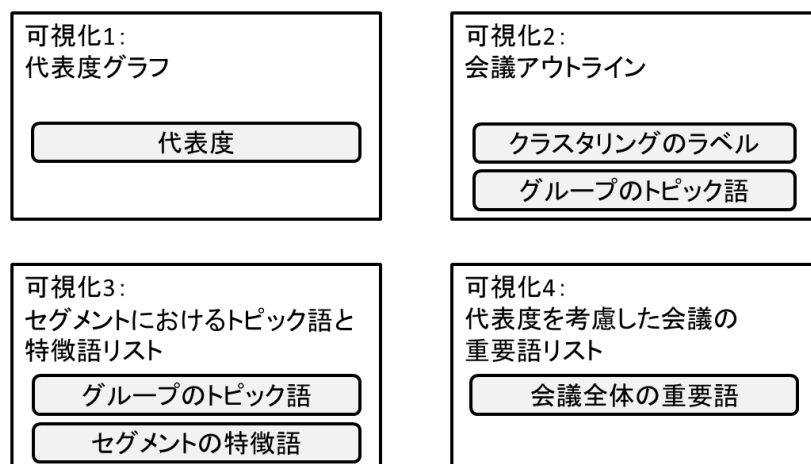


図 4-3 4.3.4 節に示す 4 種類の可視化とそれに用いるデータ.

4.3.2 セグメントベクトルの構成

4.3.2.1 入力データ

本方式で入力とするテキストデータの形式は「話者：発話内容」の繰り返しになっているが、本研究では話者情報は考慮に入れていないため、発話内容のみの一続きのテキストデータとして処理している．ただし、話者の交代が起きるところでは、文が区切れたものと見

なしている。

4.3.2.2 セグメントへの分割

テキストの文字数を m とする。本システムでは以下のステップでテキスト全体を n 個のセグメントへ分割する。

- (1) $l = m/n$ となる l を算出する。 l はセグメントの文字数の基礎となる値である
- (2) テキスト全体を文に分割する
- (3) 前から順に文を結合していき、結合してできた文字列が l を超えた時に、そこまでの文字列を一つのセグメントとして構成する。これを最後の文まで繰り返す

なお、ここで、セグメントへの分割は上記の手法に限定するものではなく、TextTiling [55]等のテキストセグメンテーション手法を用いることも可能である。また、本手法は時系列に着目した手法であるので、文字数ではなく実際の時間経過に基づいた分割を行えるのであればその方が望ましいが、そのためにはタイムスタンプ付きの発言データが必要になる。本手法では文字数を疑似的な時間経過と見なして分析を行っている。

4.3.2.3 ベクトル化

各セグメントをベクトル化する。この処理は形態素解析用いた単語抽出と、TF-iSF(term frequency-inversed Segment frequency)重み付けの二つのステップから構成される。TF-iSF重み付けとは、文書検索等で広く用いられる TF-IDF を応用したもので、ここでは各セグメントを TF-IDF における文書と見なしている。

§ 単語抽出(形態素解析)

形態素解析を行い、セグメント毎にテキストから単語を抽出する。形態素解析器としては MeCab [56]を用いた。抽出する単語の品詞は名詞のみとした。

セグメント数を n とする。文書全体に現れる単語の数を k とする。この時、セグメント集合 S とテキストで現れる全ての単語の集合を W とすると、 S と W は以下の式で表される。

$$S = \{s_1, s_2, \dots, s_n\}, \quad (4-1)$$

s_i : セグメント i .

$$W = \{w_1, w_2, \dots, w_k\}, \quad (4-2)$$

w_i : 単語 i .

セグメント s_i における単語 w_j の出現回数 $x_{i,j}$ を表す関数を $word_count(s_i, w_j)$ とする。この $x_{i,j}$ を、各セグメントを行に、単語を列に対応させて並べて、行列 \mathbf{X} を構成する。 \mathbf{X} は各セグメントにおける単語の出現回数を表す行列となる。

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,k} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,k} \end{pmatrix} \in \mathbf{R}^{n \times k}, \quad (4-3)$$

$$x_{i,j} = \text{word_count}(s_i, w_j).$$

§ TF-ISF 重み付け

$tfisf(x_{i,j})$ はベクトル \mathbf{x}_i をセグメント s_i に対応するセグメントベクトルとした場合の、単語 j に対応する要素の $tfisf$ 値を求める関数である。

$$tfisf(x_{i,j}) = tfisf_{i,j} = tf_{i,j} \cdot isf_i,$$

$$tf_{i,j} = \frac{x_{i,j}}{\sum_{r=1}^n x_{r,j}},$$

$$isf_i = \log \frac{n}{|\{s: s \ni w_i\}|}. \quad (4-4)$$

$a_{i,j}$ をセグメント j に現れる単語 i の数とする。 $\sum_k a_{i,j}$ は全てのセグメントに現れる

単語の合計である。 $|A|$ を集合 A の要素の数とする。 $|\{p: t_i \in p\}|$ は単語 t_i が含まれるセグメントの数である。この $tfisf$ を使って行列 \mathbf{X} を正規化する。 \mathbf{X}' は \mathbf{X} の各要素を $tfisf$ によって正規化した行列を表す。

$$\mathbf{X}' = \begin{pmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_n \end{pmatrix} = \begin{pmatrix} x'_{1,1} & \cdots & x'_{1,k} \\ \vdots & \ddots & \vdots \\ x'_{n,1} & \cdots & x'_{n,k} \end{pmatrix} \in \mathbf{R}^{n \times k},$$

$$x'_{i,j} = tfisf(x_{i,j}). \quad (4-5)$$

4.3.3 各種分析

4.3.3.1 セグメント間類似度の算出

行列 \mathbf{X}' からセグメント同士の類似度行列 \mathbf{Q} を算出する。

$$\mathbf{Q} = (q_{i,j}) = \begin{pmatrix} q_{1,1} & \cdots & q_{1,n} \\ \vdots & \ddots & \vdots \\ q_{n,1} & \cdots & q_{n,n} \end{pmatrix} \in \mathbf{R}^{n \times n},$$

$$q_{i,j} = \cos(\mathbf{x}'_i, \mathbf{x}'_j). \quad (4-6)$$

このとき、対角成分 $q_{i,i}$ は 1 となる。この対角成分は全ての列(行)で同一となるので、以下の計算では無視することとする。

4.3.3.2 代表度の算出

ある時間帯に対応するセグメントにおいて、会議全体から見てどれだけ主要な話題が話されていたかを表す指標として「代表度を定義する。セグメント同士の類似度を計量し、他のセグメントからの類似度の総和が大きいものを高い代表度として表現する。この指標を

用いることで、会議を振り返る際に、どのセグメントで主要な話題が話されていたのかを把握することが出来る。

セグメント間の関係を表す $n \times n$ 相関行列から各セグメントの重要度を計算する。行列の各要素の値は、2つのセグメントのコサイン類似度である。行列の1列は、あるセグメントと他の各セグメントとの間の類似度を表している。ここで、列の値の合計値を計算したとすると、大きな合計値を持つセグメントは、会議の他のセグメントと大きな関連があり、中心的なトピックであると言える。すべてのセグメントの列の合計を計算し、それらを無限大ノルムで正規化する。正規化された値をセグメントの代表度とする。正規化により、最も重要度の高いセグメントの重要度が1となり、会議の重要度が相対的に表現される。

行列 Q の j 列を対角成分を無視して足し合わせた値を b_j とする。 b_j は以下の式で求められる。また、それを並べてまとめてベクトル \mathbf{b} とする。

$$b_j = \sum_{i=1}^n q_{i,j} - 1. \quad (4-7)$$

$$\mathbf{b} = (b_1, b_2, \dots, b_n). \quad (4-8)$$

\mathbf{b} を無限大ノルム正規化して \mathbf{b}' を得る。 \mathbf{b}' は各セグメントの代表度を表す。

$$b'_i = \frac{b_i}{\|\mathbf{b}\|_{\infty}}. \quad (4-9)$$

$$\mathbf{b}' = (b'_1, b'_2, \dots, b'_n). \quad (4-10)$$

4.3.3.3 セグメントのクラスタリング

セグメント間の類似度に基づいてセグメントをクラスタリングする。各ベクトルをグラフ上のノードと見なす。ノード間の接続に関する閾値を ε とする。ノードの各ペアは、そのペアの類似度が ε よりも大きいときにノードを繋ぐエッジを獲得する。最終的に、エッジで接続されたノードのグループをクラスタとする。二つ以上のノードがエッジで接続された場合のみをクラスタとし、エッジの接続されていないノードはクラスタには属さない。クラスタリングされたセグメントをグループあるいはトピックのグループと呼ぶ。

本手法との特徴として、必ずしもクラスタに含まれないノードが発生する。これは、会議の分析という目的からして、周囲の話題から浮いていたセグメントを発見するという目的に沿っている。

クラスタリング結果は閾値 ε に依存する。ここでは ε を自動推定する。その基準は、クラスタ数を最大にすることである。 ε が1の時、エッジは一つも発生せず、したがってクラスタも発生しない。 ε が0のとき、すべてのノードが含まれる単一のクラスタとなる。したがって、このどちらもクラスタ数が最大になる場合ではなく、クラスタ数が最大になる ε の値はその中間にある。 ε を1から0に動かしていくことで、クラスタ数が最大になる値を見つけることができる。実装上は、ノード間の類似度を降順に ε の値として割り当てていくことで、エッジが増加する区切りとなる ε をを見つけることができる。クラスタ数を最大に

することで目指しているのは、周囲の話題と関連の薄かったセグメントを見つけることと、出来る限り多くのセグメントを大きなトピックの一部として位置づけることのバランスを取ることである。

4.3.3.4 グループのトピック語の抽出

トピック語として、セグメントベクトルから、同じグループに属するセグメントに共通して存在する単語を選ぶ。これらの単語の中で、より多くのセグメントで現れる単語から順に選択することによってトピック語として単語を抽出する。選ばれた単語は、少なくとも2つ以上の段階で現れる単語である。トピック語の数は任意に設定可能である。

4.3.3.5 各セグメントの特徴語の抽出

各セグメントの特徴語を抽出する。対象のセグメントがグループに属していない場合は、セグメントベクトルに対応する単語を、ベクトルの要素の降順で任意の語数選ぶことで抽出する。対象のセグメントがグループに属している場合は、そのセグメント属するグループのトピック語として抽出された語に対応する要素をセグメントベクトルから除外し、残った要素の降順で任意の語数選ぶことで抽出する。こうして選ばれた語を各セグメントの特徴語とする。

4.3.3.6 代表度を考慮した会議の重要語の抽出

ここまでの手法は、会議をセグメントに分割したうえでそのセグメントに着目して分析する手法であったが、本節で示す手法は会議全体の重要語を抽出する手法である。各セグメントベクトルにそのセグメントの代表度の値を掛け、すべてのセグメントベクトルを足し合わせることで、会議全体の単語の重要性を表すベクトルを構成する。ベクトルにより大きな重みが割り当てられている単語は、その会議において重要な単語と見なされる。ベクトルの各要素に対応する語をその重みの降順で出力することで、会議全体の重要語を表現する。

会議を n 個のセグメントに分割したとする。 k 番目のセグメントに対応するセグメントベクトルを \mathbf{x}_k とする。 k 番目のセグメントに対応する代表度を y_k とする。このとき、会議全体の単語の重要性を表すベクトル \mathbf{z} を以下の式で表す。

$$\mathbf{z} = \sum_{i=1}^n y_i \mathbf{x}_i. \quad (4-11)$$

\mathbf{z} の次元は \mathbf{x}_k と同じであり、会議全体で分析対象となった単語の数である。 \mathbf{z} から要素を任意の数だけ降順に選ぶことで、その要素に対応する単語を重要語として取り出すことができる。

4.3.4 単一の会議を対象とした可視化手法

前節までに述べた分析手法を組み合わせることで、4つの可視化手法を以下に提案する。

4.3.4.1 可視化 1：代表度グラフ

代表度グラフは各セグメントの代表度を時系列順に表示する可視化である。代表度の高いセグメントを優先的に確認することで、会議の中心的なトピックを効率的に把握できる。逆に、代表度の低いセグメントを分析することによって、会議の潜在的な問題を分析することもできる。

4.3.4.2 可視化 2：会議アウトライン

会議アウトラインは、会議の概要として各セグメントでどのような会話が行われていたのかを確認するための可視化である。会議アウトラインは、二つの可視化の組み合わせになっている。一つ目は、各セグメントを時系列順に配置した図である。各セグメントは、4.3.3.3 節に示したセグメントのクラスタリングによって形成されたグループに従って色分けする。二つ目は、4.3.3.4 節に示したグループのトピック語の抽出で得られたトピック語の一覧表である。この二つの可視化において、グループの色付けを統一することによって、時系列上のどのセグメントでどのようなトピックが話されているのかを、大まかにつかむことができる。

4.3.4.3 可視化 3：セグメントにおけるトピック語と特徴語リスト

セグメントにおけるトピック語と特徴語リストは、任意のセグメントを選択した際の、そのセグメントが属するグループのトピック語(4.3.3.4 節にて取得)と、そのセグメントの特徴語(4.3.3.5 節にて取得)の可視化である。セグメントがいずれかのグループに属する場合、システムはそのグループのトピック語とセグメントの特徴語の両方を表示する。セグメントがどのグループにも属していない場合、システムはセグメントの特徴語のみを表示する。ユーザーはこれらの語を見ることでセグメントの内容をより詳細に確認できる。

4.3.4.4 可視化 4：代表度を考慮した会議の重要語リスト

代表度を考慮した会議の重要語リストは、4.3.3.6 節で示した方法で求めた会議全体の単語の重要性を表すベクトル z から、任意の語数を降順で選んで表示する可視化である。上記 3 つの手法がセグメントを基準として可視化するのとは違って、会議全体における重要語を確認することができる。

4.4 一連の会議のトピックの関係の可視化手法

本節では、同一プロジェクト内の一連の会議におけるトピック間の関係を可視化する手法について述べる。本研究の内容は [44] で報告されている。4.3 節までで述べてきたのは単一の会議に対する分析であったが、多くの組織においては、会議はテーマを継承しつつも繰り返し開催されるものであり、それらの会議の関連性を把握することの需要は大きい。本研究では、あるプロジェクトにおける一連の会議を対象として、それらの会議の中でのテーマの変遷を追うための、トピックの関係の可視化手法を提案する。

4.4.1 概要

本方式は多くの構成要素を 4.3 節に示す単一の会議の分析手法と共有している。セグメントベクトルの構成(4.3.2 節)、セグメント間の類似度の算出(4.3.2 節)、セグメントのクラスタリング(4.3.3.3 節)、グループのトピック語の抽出(4.3.3.4 節)を行い、グループの情報とトピック語を得る。これらの処理を分析対象とする全ての会議に対して行う。その上で、以下の処理によって、各グループが、隣り合う会議におけるグループとどれぐらいの関連の大きさを持っているかを可視化する。

4.4.2 各種分析

4.4.2.1 グループベクトルの構成

会議におけるグループ G_i を表すベクトルであるグループベクトル \mathbf{g}_i を定義する。グループ G_i に属するセグメントベクトルの平均ベクトルを \mathbf{g}_i とする。 G_i に属するセグメントベクトルは $\mathbf{x} \in G_i$ である。 \mathbf{g}_i は次式で表される。 $|G_i|$ は集合 G_i の要素数を表す。

$$\mathbf{g}_i = \frac{\sum_{\mathbf{x} \in G_i} \mathbf{x}}{|G_i|}. \quad (4-12)$$

4.4.2.2 隣り合う会議のグループ間の類似度の算出

プロジェクト内に複数の会議があるとする。ある会議を M_i とし、次の会議を M_{i+1} とする。 M_i に属するグループベクトルと M_{i+1} に属するグループベクトル間の類似度のすべての組合せを計算する。類似度の計算にはコサイン尺度を使用する。この際、異なる会議のグループベクトルは異なる単語セットを持っている。したがって 2 つの会議に対応する 2 つのグループを表す単語セットの和集合に基づいてグループベクトルを再構築する。自身の会議には存在しなかった単語に対応するベクトルの要素は 0 となる。

会議 M_i のグループ a と会議 M_{i+1} のグループ b との類似度 $z_{i,a,b}$ は、次式のように定義される。関数 \cos_sim は、コサイン測度による類似度計算を表す。

$$z_{i,a,b} = \cos_sim(\mathbf{g}_{i,a}, \mathbf{g}_{i+1,b}). \quad (4-13)$$

4.4.3 複数会議を対象とした可視化

前節までに述べた分析手法を組み合わせることで、1つの可視化手法を以下に提案する。

4.4.3.1 可視化：トピックの類似度に基づく複数会議のグラフネットワーク

会議におけるグループ間の関係を時系列に基づくグラフネットワークとして可視化する。まず、システムは各会議を個別に可視化する。各会議のグループをいくつかのトピック語によって表示する。このグループは、最終的なアウトプットとなるグラフネットワークではノードに対応する。次に、入力とした会議を時系列に基づいて左から右に配置する。最後に、システムは4.4.2.2節(4-13)式の z の値に基づいてノード間にエッジを描画する。 $0 < z \leq 1/3$ の場合は線を描画しない。 $1/3 < z \leq 2/3$ の場合は細い線を描画する。 $2/3 < z \leq 1$ の場合は太線を描画する。線を引くためのこれらの基準は予備実験により定めた仮のものである。本研究では、実験結果と見比べながら、3つのケースがほぼ等しく見えるように基準を設定した。今後の研究では、デフォルトの描画を示した後には閾値をインタラクティブに変更しつつ描画することを可能にすることを目指す。

4.5 議論の発散/収束に着目した会議の時系列構造分析と状態遷移可視化手法

本節では、テキストデータを分析することで会議の状態遷移を可視化する会議の時系列構造分析方式を示す。本節の内容は [45] にて報告している。本方式では、会議における「発散」と「収束」の状態を表現する指標として「鮮度」と「代表度」の二つを定義する。それぞれ、会議をいくつかのセグメントに区切ったうえで、「鮮度」はあるセグメントの中で新しい語がどれだけ話されたかを表し、「代表度」は全てのセグメントから見たあるセグメントの類似度の総和、つまり全体から見たそのセグメント関連の大きさを表す。さらに、これら二つの指標を軸として平面座標に各セグメントをプロットすることで、会議の状態遷移を可視化する。状態の種類としては、鮮度の大小、代表度の大小の組み合わせに対応する状態名として「探索」「深掘り」「合意形成」「停滞」の四状態を定義する。

これらの指標や状態の区分を用いて分析することで、対話の中でどのあたりで議論が停滞したか、どのあたりで議論がまとまっていったかを客観的につかむことができる。そして、会議がどのようなプロセスを構造として持っているのかを明らかにし、振り返りやファシリテーションに活かすことで、会議の効率向上に繋げていくことが出来ると考える。

4.5.1 会議の状態を表す指標の定義

本節では、分析結果となる会議の状態を定義する。この定義は大きく分けて二段階で行われる。まず、議論が発散しているか収束しているかを表現する指標として、「鮮度」と「代表度」の二つを定義する。さらに、これら二つの指標を軸にして平面座標を構成することで、議論の状態を表現する。この平面座標を、鮮度の大小、代表度の大小をそれぞれ組み合わせることで、「探索」「深掘り」「合意形成」「停滞」の四つの状態を定義する。

4.5.1.1 「鮮度」と代表度」の定義

会議を分析するにあたって、議論の発散と収束という観点から分析を行うことは有力な候補である。本稿では発散と収束を判別するための定量的な指標として「鮮度」と「代表度」の二つを定義する。

本稿では会議の時系列分析を行うために、会議全体の発話テキストデータを一定時間ごとにセグメントに分割したものを基本分析単位とする。今回定義する指標はこれらセグメントを評価する指標として用いられる。これによって会議のどの時間帯がどのような状態であったのかを表現する。

§ 鮮度

一つ目の発散/収束の定義は「鮮度」である。これは、ある時間帯に対応するセグメントが、どれだけ新しい単語が話されていたかを表す指標である。発散は豊富なアイデアを提案するフェーズと位置付けられるため、そこまでに出来ていない単語の数を新規アイデ

ィアと見なし、発散の度合いとして用いる。

§ 代表度

二つ目の発散／収束の指標は「代表度」である。これは、ある時間帯に対応するセグメントにおいて、会議全体から見てどれだけ主要な話題が話されていたかを表す指標である。導出方法は既に 4.3.3.2 節に示している。各セグメントに出現する単語からセグメント同士の類似度を計量し、他のセグメントからの類似度の総和が大きいものを高い代表度として表現する。代表度は話題に着目した際の発散／収束の指標となっている。発散／収束という区分において、豊富なアイディアを発案して話題のバリエーションを増やしていくことを発散、そうして出て来たアイディアを統合あるいは取捨選択して一つの結論を導いていくことを収束であるとする、話題という観点からは、全体から見て関連度の高い話題について話しているセグメントが収束に相当する。この指標を用いることで、会議を振り返る際に、どのセグメントで主要な話題が話されていたのかを把握することが出来る。

4.5.1.2 議論の性質を表す会議の四状態の定義

鮮度は単語に着目した発散／収束の指標であり、代表度は話題に着目した発散／収束の指標である。議論において、新しい語が出てくるとは新しいアイディアが出ていることに相当すると思われるが、例えば議論の本題とは全く関係ない雑談になってしまっている場合でも鮮度は高くなると思われる。同様に、新しい語が出てこないことが必ずしも議論がまとまっていることに繋がっているとは限らず、それはただ議論に進展がないだけかもしれない。こうした状況に対して、鮮度と代表度の二つの指標を組み合わせることで、人間が会議を振り返る際に有益な判断のできるような指標である、会議の状態を定義できる。

ここで、鮮度の高低を「新鮮⇔常套」、代表度の高低を「中心⇔周辺」と呼ぶことにする。この鮮度と代表度を軸として、平面座標を構成する。このとき、二つの指標のそれぞれ二種類の状態の組み合わせによって、以下の四種類の状態を定義する。

- 話題：周辺 かつ 語：常套 → 停滞
- 話題：周辺 かつ 語：新鮮 → 探索
- 話題：中心 かつ 語：新鮮 → 深掘り
- 話題：中心 かつ 語：常套 → 合意形成

これらの会議の四状態を図 4-4 に示す。これら四つの会議の状態を発散的思考と収束的思考という観点から捉え、まずバラエティ豊かなアイディアを出してテーマを探索「探索」、主要テーマについてのアイディアを深めていく「深掘り」、そしてそれらを統合・選別して結論に導く「合意形成」と進行していくのが会議の進行として良い流れであると考えられる。したがって、分析にあたっては、右下の「探索」から左上の「合意形成」に全体として向かうこと、あるいは、その両者の間を行き来すること、そしてその際にできれば右上の「深掘り」を通ることが望ましいと考えられる。分析においてはこの流れを標準として、それに沿っているか、あるいは逸脱しているかということに注目して評価する。

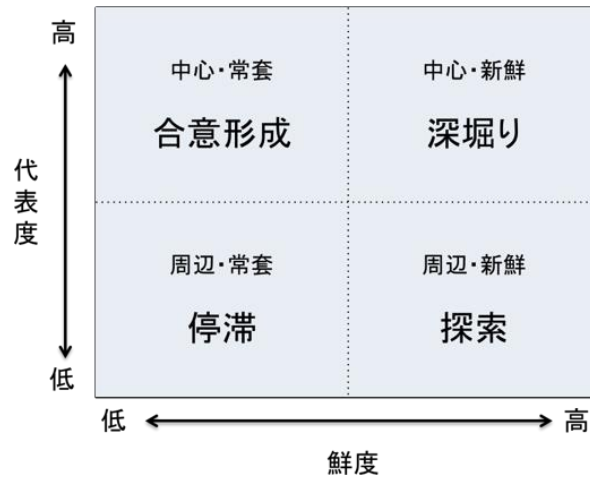


図 4-4 会議の状態の定義.

4.5.2 会議の進行評価指標および会議の状態遷移図の導出

本章では 4.5.1 節で定義した指標および会議の状態遷移図を導出する過程の定式化を行う．提案方式の全体像を図 4-5 に示す．セグメントベクトルの構成，および，代表度の算出は 4.3 節で示した手法と共通している．4.5.2.2 節ではセグメントベクトルから鮮度を導出する方式について述べる．4.5.2.4 節では，鮮度と代表度を組み合わせて会議の状態遷移を図示する方式について述べる．4.5.2.5 節では，状態遷移図から，会議の進行を評価する指標の値を導出する方法について述べる．

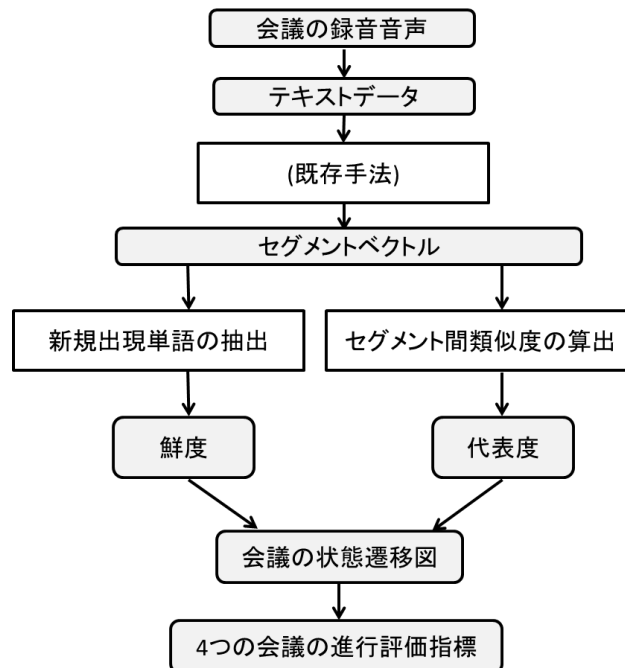


図 4-5 会議の状態遷移図および進行評価指標導出方式の全体像.

4.5.2.1 セグメントベクトルの構成

本方式では、分析の基礎となるセグメントベクトルを、4.3.2 節で示した方式を用いて構成する。入力とするテキストデータの形式も 4.3.2 節と同様で、「話者：発話内容」の繰り返しとなっている。

4.5.2.2 鮮度の出力

セグメントベクトルを前から順にみていき、注目しているセグメントにおいて初めて出現した単語の数を数えていくことで、鮮度を算出する。

§ 新規出現単語の抽出

セグメント s_i における新規出現単語の数を c_i とする。また、それを並べてまとめてベクトル \mathbf{c} とする。

$$\mathbf{c} = (c_1, c_2, \dots, c_n) \quad (4-14)$$

§ 鮮度の算出

\mathbf{c} を無限大ノルム正規化して \mathbf{c}' を得る。 \mathbf{c}' は各セグメントの鮮度を表す。

$$c'_i = \frac{c_i}{\|\mathbf{c}\|_\infty},$$
$$\mathbf{c}' = (c'_1, c'_2, \dots, c'_n). \quad (4-15)$$

4.5.2.3 代表度の出力

セグメントベクトル同士の類似度を計量することによって代表度を算出する。詳細は 4.3.3.2 節に示した通りである。

4.5.2.4 会議の状態遷移の可視化

鮮度と代表度を軸として用いることで平面座標を構成し、各セグメントを時系列順にプロットすることで会議の状態遷移を表現する。

§ 会議の状態を表す平面上へのプロット

鮮度を横軸、代表度を縦軸とする。セグメント i における座標上の点を \mathbf{r}_i とする。また、それを並べてまとめてベクトル \mathbf{r} とする。

$$r_i = (c'_i, b'_i),$$
$$\mathbf{r} = (r_1, r_2, \dots, r_n). \quad (4-16)$$

可視化を行う際には、 \mathbf{r} の要素を、順番が分かるようにプロットする。

本手法における定義では、プロットする点の座標の x 軸と y 軸の取り得る最小値は 0、最大値は 1 となる。また、代表度と鮮度の両方を無限大ノルムで正規化しているため、少なくとも一つのセグメントは x 軸または y 軸において最大値である 1 を持つ。それに対して他のセグメントの x 軸または y 軸の値は、正規化する前の値のばらつきの大きさに応じて、座

標上に広く分布することもあるれば偏って分布することもある。このような状況では、プロットされた点の絶対的な位置、例えばグラフの中心である(0.5,0.5)から見てどの方向にあるか、といった基準が会議の性質を表しているかについては疑問が残る。ただしこの条件でも、二点間の相対的な位置関係については、その二点の相対的な性質を表していることは問題なく言えると考えられる。したがってこれ以降は、主に二点間の相対的な位置関係を基準に議論の性質を取り扱い、全体から見たある点の性質を検討する場合は、その他の点の重心を基準にして検討することとする。

4.5.2.5 会議の評価指標

4.5.1.2 節での議論に基づいて、各セグメントに対応する座標が、右下の「探索」から左上の「合意形成」に全体として向かうこと、あるいは、その両者の間を行き来すること、そしてその際できれば右上の「深掘り」を通ることが望ましい会議の進行であると仮定して分析を行う。ここでは、会議の進行の評価指標として以下の $I_1 \sim I_4$ の四つを定義する。

§ 指標 I_1 ：会議が全体として探索から合意形成に向かうことを示す指標

会議が全体として探索から合意形成に向かうことを示す指標として、4.5.2.4 節で定義した座標の右下から左上に向かうことを、始点と終点を用いて表現する。始点から終点を繋ぐ線分をベクトルと見なし、そのベクトルと、座標の左上方向を示す単位ベクトルである $(-1,1)$ との内積によってこの指標 I_1 を表現する。以下に式を示す。

$$I_1 = (c'_n - c'_1) \times (-1) + (b'_n - b'_1) \times 1. \quad (4-17)$$

§ 指標 I_2 ：最終的に合意に達したことを示す指標

最終的に合意に達したことを示す指標として、終点の位置を用いる。全ての点の重心を基準点として、そこから終点を繋ぐ線分をベクトルと見なし、そのベクトルと、座標の左上方向を示す単位ベクトルである $(-1,1)$ との内積によってこの指標 I_2 を表現する。

$$I_2 = \left(c'_n - \frac{\sum_i^n c'_i}{n} \right) \times (-1) + \left(b'_n - \frac{\sum_i^n b'_i}{n} \right) \times 1. \quad (4-18)$$

§ 指標 I_3 ：会議全体の議論の振れ幅の大きさを示す指標

会議全体の議論の振れ幅の大きさを示す指標として、座標上の移動量の総和を用いる。この指標を I_3 とする。以下に式を示す。

$$I_3 = \sum_{i=1}^{n-1} \sqrt{(c'_{i+1} - c'_i)^2 + (b'_{i+1} - b'_i)^2}. \quad (4-19)$$

§ 指標 I_4 ：開始時と終了時の議論の振れ幅の大きさを示す指標

開始時と終了時の議論の振れ幅の大きさを示す指標として、始点と終点の距離を用いる。この指標 I_4 とする。以下に式を示す。

$$I_4 = \sqrt{(c'_n - c'_1)^2 + (b'_n - b'_1)^2}. \quad (4-20)$$

4.6 実験

本節では、大きく分けて二つの実験を行う。4.6.1 節では、実験環境について述べる。4.6.2 節では、4.3 節および 4.4 節で提案した会議の構造化可視化およびトピック抽出の手法の例を示す。4.6.3 節の実験 2 では、4.5.2 節で提案した会議の状態遷移図の可視化および評価指標の検証を行う。

4.6.1 実験環境

4.6.1.1 実験で用いる会議ログデータ

実験で入力として用いた会議ログデータは、実際に行われた会議を録音機で録音し、得られた音声データを人手によって書き起こすことで得られたテキストデータである。ここで行われた会議は、異なる企業の構成員同士が集まって、企業間をまたいだ協業の可能性を探ることを目的とした会議であり、会議のタイプとしてはいわゆる企画会議と呼ばれるような、新たなアイデアを出してそこから方針をまとめていくような会議である。一度に複数のテーブルで会議が行われ、また、日をまたいで行われた。メンバーは全体ではおおむね固定だが日によってはテーブルでの入れ替わりがあった。計 5 日間で一回平均 3 テーブル、延べ 15 回分の会議のログデータを入手した。

実験 1 では、これらのうち一つの会議をピックアップし、会議の構造とトピックを可視化する。実験 2 では、これらの全てのデータを使い、提案した会議の評価指標について検討する。詳細については各実験にて再度述べる。

4.6.1.2 実装

提案手法を実装し、会議分析システムを構築した。実験システムは R と Python によって実装されている。日本語の形態素解析には MeCab [56]を用いている。形態素解析を行う際には、ひらがな一文字、カタカナ一文字などの単語はストップワードとして除去している。

会議のログデータをセグメントに分割する際には、分割数は全ての実験で 20 とした。

4.6.2 実験 1:会議の構造可視化及びトピック抽出の検証

本節では 4.3 節および 4.4 節で提案した会議の構造可視化およびトピック抽出を例示し、検証を行う。実験 1-1 で、単一の会議に対する分析を、実験 1-2 で複数の会議に対する分析を行う。

4.6.2.1 実験 1-1 会議の構造とトピックの可視化(単一会議の分析)

§ 実験の目的と方法

実験 1-1 では、4 つの可視化を行うことで、会議の構造とトピックを分析する。可視化を行うとともに、それが会議の実際の状況を反映しているかを考察する。

一つ目の可視化は 4.3.4.2 節で示した、会議アウトラインである。これは、会議を 20 の

セグメントに分けた上で、それらのセグメントをグループにクラスタリングすることで、どこどこのセグメントで似たような話題が話されていたかを可視化するものである。グループ内のセグメントに共通する語であるトピック語をあわせて表示することで、会議全体で何が話されたのかを大まかに把握することができる。また、このクラスタリングにおいては、すべてのセグメントをグループに含めるわけではない。グループに含まれなかったセグメントは他のセグメントとの関連が弱いことを意味する。そのため会議のトピックに寄与しなかったセグメントとして分析することが可能である。なお、グループのトピック語は、二つ以上のセグメントに出現した語を、出現したセグメント数が多い順に表示している。表示数の最大数は 20 としているが、今回は全てのグループで抽出されたトピック語は 20 未満であった。

二つ目の可視化は、4.3.4.1 節で示した、代表度のグラフである。各セグメントの他のセグメントから見た関連の大きさの総和である、あるセグメントにおける代表度を、時系列に沿って並べることでグラフとして可視化する。このグラフによって、会議中のどこで重要なことが話されていたのか、あるいはどこで重要でないことが話されていたのかを分析することができる。重要なセグメントを見ることで、会議の主な論点を理解することができる。さらに、重要性の低い段階に焦点を当てることで、潜在的な問題を見つけ、会議のプロセスを妨げる理由を見つけることができる。

三つ目の可視化は、4.3.4.3 節で示した、トピック語とセグメントの特徴語リストである。ある一つのセグメントを選ぶと、選んだセグメントがグループに属している場合は、そのグループのトピック語とそのセグメントにおける特徴語を表示し、グループに属していない場合はセグメントにおける特徴語のみを表示する。これまでに示した二つの可視化とあわせてもちいることで、詳しく内容を知りたいセグメントにフォーカスして分析することができる。セグメントの特徴語は、TF-idf 値の高い順に表示して、表示する単語数は 15 とした。

四つ目の可視化は、4.3.4.4 節で示した、代表度を考慮した会議の重要語リストである。会議全体における、代表度を考慮した重要度の高い語を表示する。表示する単語数は 30 とした。

分析対象とするのは、実際に行われた一つの会議のログデータである。表 4-1 にその会議の概要として、各会議の文字数、文の数、メンバーの数、および会議の所要時間を示す。

表 4-1 実験 1-1 で用いる会議の概要。

	会議 1
文字数	15,553
文の数	323
メンバーの数	6
会議の所要時間	2 時間 18 分

§ 実験 1-1：実験結果

会議 1 の分析結果を示す．図 4-6 に，会議 1 の会議アウトラインを示す．会議アウトラインは，各セグメントがどのグループに属しているかを示す図と，各グループのトピック語の一覧からなる．トピック語の一覧表の中の「****」は，会社名，個人名など，開示できない単語である．トピック語の表示数は，最大を 20 としたが，選択されたトピック語の数は最大でグループ 1 の 16 語であった．

結果を見ると，3 つのグループが形成されている．グループ 1 は最もセグメント数が多いグループである．セグメント 6 から 12 は継続的にグループ 1 に属している．セグメント 1 もグループ 1 に属している．セグメント 3～4 およびセグメント 20 はグループ 2 に属している．セグメント 13 および 18 はグループ 3 に属している．セグメント 2, 5, 14～17, および 19 はどのグループにも属していない．どのグループにも属していないセグメントで，議論が混乱していたと考えることができる．

segment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
group:	1	2	2			1	1	1	1	1	1	1	3					3		2

グループ 1 の トピック語	グループ 2 の トピック語	グループ 3 の トピック語
地域	室	時間
位置	空間	健康
****	オフィス	ログ
生活	登録	行動
エリア	ICT	漫画
属性	SNS	能力
物件	会議	会社
特性	提供	
趣味	嗜好	
価値観	趣味	
健康	デバイス	
繁華街	顧客	
住居	口座	
売買	予約	
富裕		
予測		

図 4-6 会議 1 の会議アウトライン．

グループ 1 には，「地域」，「位置」，またはそれらに関連する単語がいくつかある．グループ 1 では，主に地域と人々の属性との関係を発見することについて話されていることが

分かる。つまり、グループ1は人々の住む場所と属性についてのトピックである。

グループ2には、「オフィス」、「趣味」、「SNS」、「アカウント」などの単語がある。これから、主に人間の活動から得られるデータについて話していることが分かる。つまり、グループ2は人間の活動データについてのトピックである。

グループ3には、「漫画」、「健康」、「ログ」などの単語が含まれている。これらから、主に人間の活動の記録について話されていることが分かる。特に、「コミック」および「健康」に関するアプリケーションのログについて話されている。つまり、グループ3は漫画および健康のアプリケーションのデータのトピックである。

この結果、会議1の会話の流れを捉えることができる。最初に、参加者は人の領域と属性のトピックについて話し始めた。次に、人間活動データのトピックについて話した。会議の中盤では、人々の領域と属性のトピックについて集中的に話した。その後、議論は一時期手に混乱した。最後に、再び人間活動データの話題について話した、

続いて図4-7に、会議1の代表度のグラフを示す。

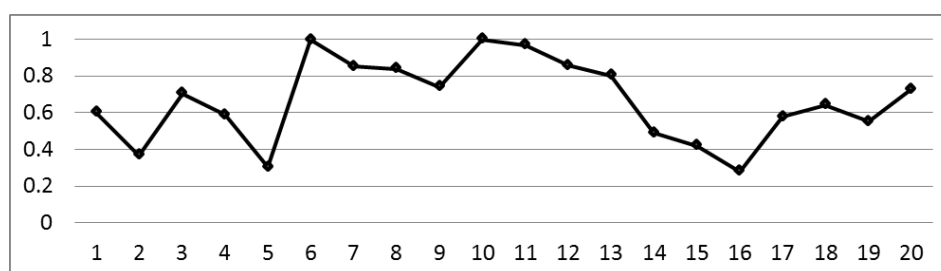


図4-7 会議1の代表度グラフ。

このグラフを見ると、最も重要なセグメントはセグメント6で、最も重要でないセグメントはセグメント16であることが分かる。また、第3セグメントから第13セグメントまでが良い議論が行われた区間であったことも分かる。第5セグメントや第14セグメントから第16セグメントまではあまり結果的にあまり重要でないことが話されていたことが分かる。第5セグメントおよび第14～16セグメントは、ミーティングアウトラインにおいても、どのグループにも属していない。これらの結果は正しく連動していることが分かる。

三つ目の可視化として、トピック語とセグメントの特徴語を示す。表4-2に、会議1のセグメント2つのセグメントにおける可視化の例を示す。代表度のグラフから、セグメント6が最も代表度の高いセグメントで、セグメント16が最も代表度の低いセグメントであることが示されているため、この二つのセグメントについて可視化した結果を示す。セグメント16はどのグループにも属していないため、トピック語は存在せず、特徴語のみを表示している。特徴語の表示数は、セグメントベクトルの重みの上位15語とした。

セグメント6の特徴語を見ると、「顧客」「像」「学生」「予測」、「ユーザー」、「クライアント」などの語が現れている。トピック1が人々の住む場所と属性についてのトピックであるとする、その具体的な属性を分析することについての話をしていることが推察できる。

また、セグメント 16 の代表度が低い原因について、実際の会議のログを確認した。するとセグメント 14 から、遅れて参加した人が自己紹介をしていて、話題が一度リセットされていたことが分かった。セグメント 16 の代表度が低いもその影響と考えられる。

表 4-2 実験結果：2つのセグメントにおけるトピック語と特徴語リスト。

セグメント 6		セグメント 16
グループ 1 の トピック語	セグメント 6 の 特徴語	セグメント 16 の 特徴語
地域	顧客	ソフト
位置	像	バイオ
****	ユーザー	辞書
生活	学生	アルゴリズム
エリア	漫画	駅
属性	嗜好	駅名
物件	toB	処理
特性	toC	表記
趣味	ワード	万
価値観	意識	揺れ
健康	意味合い	分析
繁華街	違い	全部
住居	一般	****
売買	完全	****
富裕	区分	でため
予測		

4 つ目の可視化として、会議 1 における代表度を考慮した会議の重要語リストを示す。比較のため、代表度の値を全てのセグメントで一定にした場合の結果をあわせて示す。表 4-3 に分析結果を示す。代表度が一定の場合をケース A、提案方式をケース B とする。ここでは、それぞれ上位 30 語を出力した。

ケース A とケース B を比較すると、ランキングが変化している。ケース B すなわち提案方式では、特に、「人」、「ライフスタイル」、「好み」という単語のランクが上昇した。逆に、「健康」、「コミック」、「ログ」という単語はランクが下がった。会議全体としては、人々の属性を推定することについて幅広い話題を出して話し合っていると考えられるため、代表度、すなわち他のセグメントから関連の大きさを利用して全体からの重要語を推定する際に、一般的な用語である「人」、「ライフスタイル」、「好み」の重要度が増し、その個々の事態を示す「健康」、「コミック」といった言葉の重要度が相対的に減じるのは妥当な結果であると思われる。

表 4-3 実験結果：代表度を考慮した会議の重要語リスト.

ケース A 代表度を考慮しない		ケース B 代表度を考慮	
単語	重要度	単語	重要度
情報	1.00000	情報	1.00000
地域	0.85105	地域	0.80194
健康	0.69718	価値観	0.68347
広告	0.65734	人々	0.60358
価値観	0.62379	広告	0.58841
空間	0.59027	登録	0.56466
漫画	0.58711	ライフスタイル	0.55159
分析	0.56772	健康	0.54032
位置	0.53856	空間	0.51476
登録	0.53539	位置	0.47640
親	0.53329	嗜好	0.47411
ログ	0.52738	漫画	0.47233
地域性	0.52549	趣味	0.46867
物件	0.51197	特性	0.43701
生活	0.50204	物件	0.42491
特性	0.50030	非常	0.42473
人々	0.49808	属性	0.41400
****	0.48161	ログ	0.41330
ライフスタイル	0.47877	地域性	0.40967
子ども	0.46971	エリア	0.40387
趣味	0.46476	生活	0.40176
データ	0.45279	保有	0.39665
オフィス	0.45034	****	0.38955
保有	0.44833	分析	0.37541
エリア	0.44813	この辺	0.36698
属性	0.43965	オフィス	0.36372
嗜好	0.43419	子ども	0.35489
行動	0.42332	傾向	0.35114
この辺	0.41729	繁華街	0.34905
傾向	0.41728	顧客	0.34423

4.6.2.2 実験 1-2 類似度に基づくグラフネットワーク(複数会議の分析)

実験 1-2 では、4.4 節に示した、同一プロジェクトに属する複数の会議の分析および可視化手法を、実際の一連の会議に適用しする。

§ 実験の目的と方法

4 つの会議を入力として分析する。これらの会議は、1 つのプロジェクト内の一連の会議である。これらの会議は異なる日に行われた。会議のメンバーはほとんど同じだったが、完全に同じではなかった。表 4-4 に、この 4 つの会議の概要として、各会議の文字数と合計時間を示す。

表 4-4 実験 1-2：分析対象とする 4 つの会議の概要。

	会議 1	会議 2	会議 3	会議 4
文字数	13,826	20,106	16,748	12,622
所要時間	2 時間 18 分	2 時間 02 分	1 時間 55 分	1 時間 44 分

§ 実験結果 1-2

提案手法による可視化の結果を図 4-8 に示す。

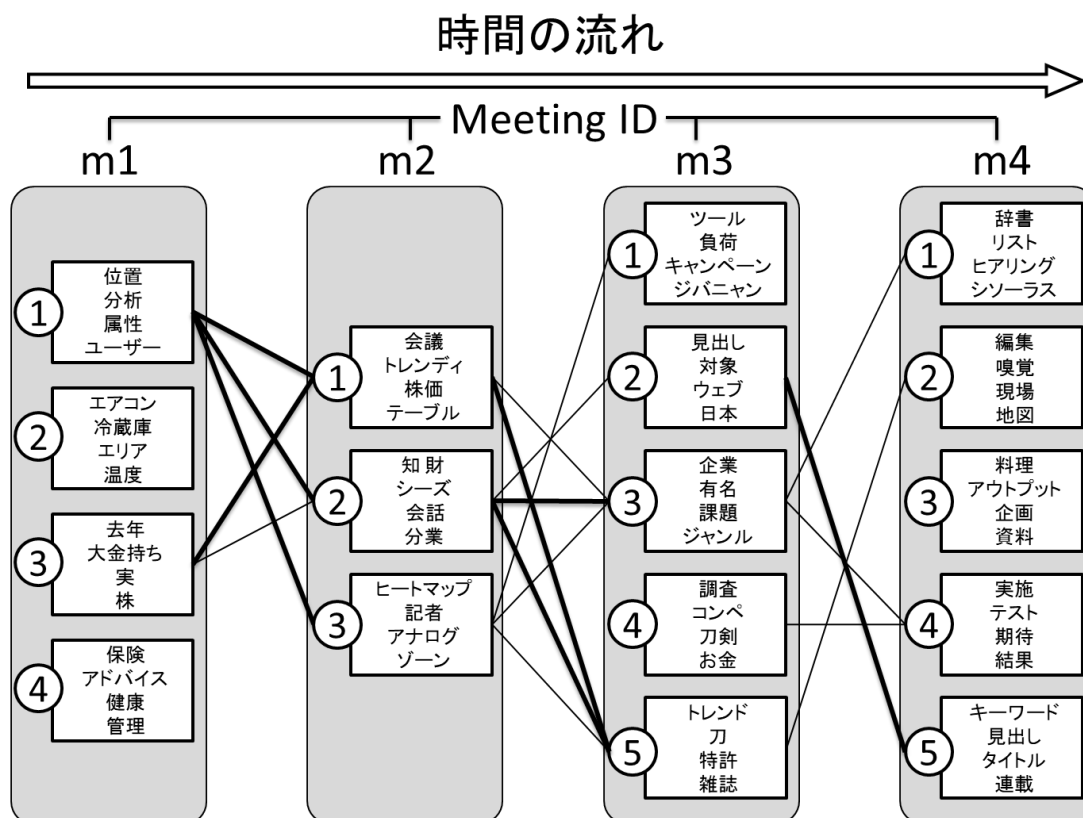


図 4-8 実験結果：一連の会議におけるトピック間の関係の視覚化。

この可視化では、各会議の各グループはノードとして表示される。各会議に必ずしも同数のグループがあるとは限らない。この図では、各会議のグループの各トピックを表すため

の4つの単語を表示している。これらの4つの単語は、各グループの上位4つまでのトピック語である。表示される話題語の数は4つに限ったものではなく、目的に応じて調整する必要がある。最後に、グループ間の類似性に基づいてエッジを描画する。なお、本実験における会議1は、前節において分析した会議と同一であるが、ストップワードの設定など実験の設定の違いによって、生成されるグループやトピックが異なっている。

結果を考察する。図2から、次の会議の他のトピックと関係のないトピックがあることが分かる。この場合、会議1のグループ2と4、および会議3のグループ1がそれにあたる。これらのトピック語は、実際には他のミーティングのトピック語としては現れていない。これらのトピックは他のトピックとの関係が弱いことが分かる。この部分は少なくとも2つの解釈が可能である。この部分が役に立たなかったのか、あるいは単に無視されたのか。現状ではこの部分がそのどちらかなのかを判断することはできない。いずれにせよ、この視覚化はプロジェクトのトピックを振り返るのに用いることは可能である。

次に、エッジに焦点を当てる。太字のエッジで接続された一对のノードを見ると、同じ単語あるいはほぼ同一の単語が現れる複数の例を見つけることができる。例えば、会議1のグループ3と会議2のグループ1の「株」と「株価」、会議3のグループ2と会議4のグループ5の「見出し」などである。これは、エッジを見ていくことで、会議をまたがって継続しているトピックを知ることが出来ることを示していると考えられる。また、エッジをたどることで隠れたトピックのつながりを発見する手助けとして用いることもできると考えられる。

また、接続されているエッジの数を確認することで、ノードの重要性を知ることが出来る。例えば、エッジが多いノードを重要だと考えることができる。これはプロジェクトの振り返りにも役に立つ。重要と思われるノードを詳細に分析することで、会議の改善に役立つ知見を発見することに繋がると考えられる。

4.6.2.3 実験1全体の考察

以上で示したきた結果から、提案方式は、会議の概要を把握したり、振り返りを助けるという観点からは有用であると考えられる。しかし、現段階では定量的な評価やユーザーによる評価実験などは行えていないため、さらなる改善の余地があると思われる。本研究は、4.7節にも示す通り、実システムとしての実装を見据えた研究であるため、利用者にとって何が必要なのかを検討しながらシステムを発展させていくべきであると考え。あるいは、利用者が可視化についての要望をフィードバックとして与える方式を合わせて検討していくべきであると考え。

4.6.3 実験 2:会議の評価指標の検証

4.5 節で示した方式を実装し、実際の会議ログデータを対象として実験を行う。実験 2-1 では、定性的評価として、各会議について、会議の状態遷移を平面座標にプロットし、その推移を観察し、考察を行う。実験 2-2 では、定量的評価として、各会議に対してアンケートを用いて評点を付け、それらと 4.5.2.5 節で定義した会議の評価指標が一致しているかを確認する。

4.6.3.1 実験 2-1 定性的評価

実験 2-1 では、定性的評価として、15 回分の会議について会議の状態遷移を平面座標にプロットし、その推移を観察し、考察を行う。

§ 実験の目的と方法

実験で入力として用いる 15 回分の会議は、4.6.1.1 節の「実験で用いる会議ログデータ」で述べたものである。延べ 15 回の会議は、計 5 日間で一回平均 3 テーブルに分かれて行われた。一テーブル当たりの参加人数は、平均 6.4 人、最低 2 人、最大 11 人であった。イベント開催日とその中で行われた会議の ID、そして会議の所要時間を表 4-5 に示す。各会議の参加人数を表 4-6 に示す。

表 4-5 実験対象とした会議の概要。

開催日	会議 ID	所要時間
第 1 回	①,②,③	111 分
第 2 回	④,⑤	103 分
第 3 回	⑥,⑦,⑧	110 分
第 4 回	⑨,⑩,⑪,⑫	97 分
第 5 回	⑬,⑭,⑮	95 分

表 4-6 実験対象とした各会議の参加人数。

会議 ID	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮
人数	6	11	7	10	8	5	4	6	7	6	2	8	8	2	6

本分析手法では会議を任意の数のセグメントに分割する必要がある。実験ではセグメント数 n を 20 とした。

§ 実験 2-1 結果

提案システムを用いて会議を分析した結果を図 4-9 に示す。図中のノードはセグメントに対応しており、最初のフェーズを黒丸(●)で、最後のフェーズを白丸(○)で囲んで強調している。また、遷移を表現するために、ノードとノードの間を線でつないでいる。なお、ラベルの Freshness は鮮度、Representativeness は代表度をそれぞれ表す。

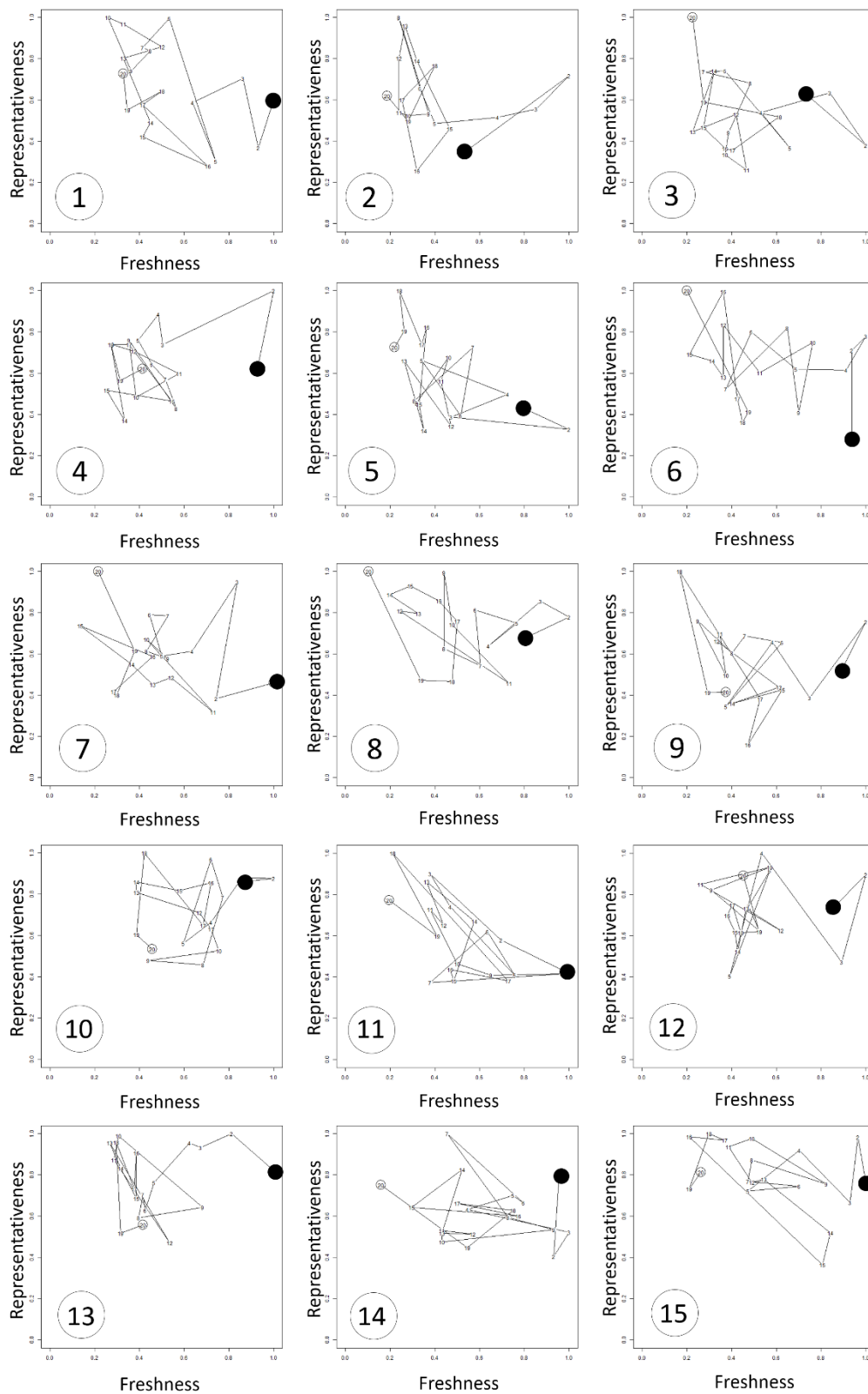


図 4-9 実験 2-1 結果：すべての会議の状態遷移図。

今回用いた指標においては、議論が順調に行われた場合には右下の「探索」から左上の「合意形成」に向かうと考えられる。ここでは、始点と終点の位置関係を見て、その向きが左上に向かっていているものをこの条件に合致するものとして、各会議を会議が順調に進行したかどうかを峻別する。表 4-7 にその結果を示す。

表 4-7 会議の順調／不順の判定.

	会議 ID	計
順調な進行	①,②,③,⑤,⑥,⑦,⑧,⑪,⑫,⑮	10
不順な進行	④,⑨,⑩,⑬,⑭	5

表 4-7 から分かるように、15 回行われた会議のうち、10 回が順調に進行し、5 回が順調に進行していないという結果になった。

会議が順調に進行した例として、会議 ID1 の会議を取りあげる。会議 ID1 の状態遷移図を図 4-10 に示す。この例では、第 1 セグメントが最も鮮度が高く、そこから最終的には代表度が高く鮮度が低い左上方向へと移動している。途中で折り返しもあるが、大きな流れとしては、左上方向と右下方向、つまり合意形成と探索を繰り返しながら会議が進行している。実際の会議でも、この日はワークショップ型の会議が行われており、アイデアを出す段階やまとめる段階を意識したファシリテーションが行われていた。同日に行われた ID1,2,3 の会議はどれも右下から左上へと順調に進行している。

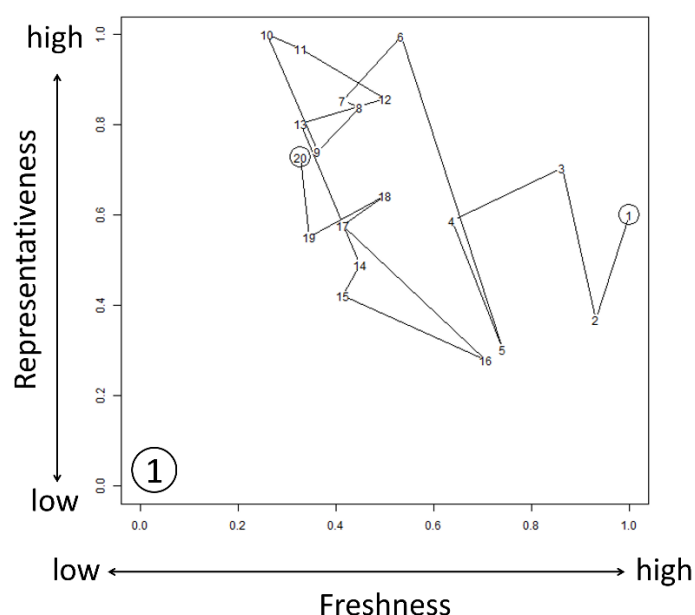


図 4-10 会議 ID1 の状態遷移図.

それに対して、会議が順調に進行しなかった例として、会議 ID10 の会議を詳しく取り上げる。会議 ID10 の状態遷移図を図 4-11 に示す。この会議では、主要メンバーの一部が開始時間からは参加することが出来なかったため、メンバーが揃うまで情報共有をしていた。代表度が最大になっている第 18 セグメントにメンバーが揃い、それまで話したことを再度

説明したため、そこで代表度が高く、鮮度が低くなった。そしてその後少しそれまでと違う話をしたが、その分はまともらずに時間切れとなった、と解釈出来る。

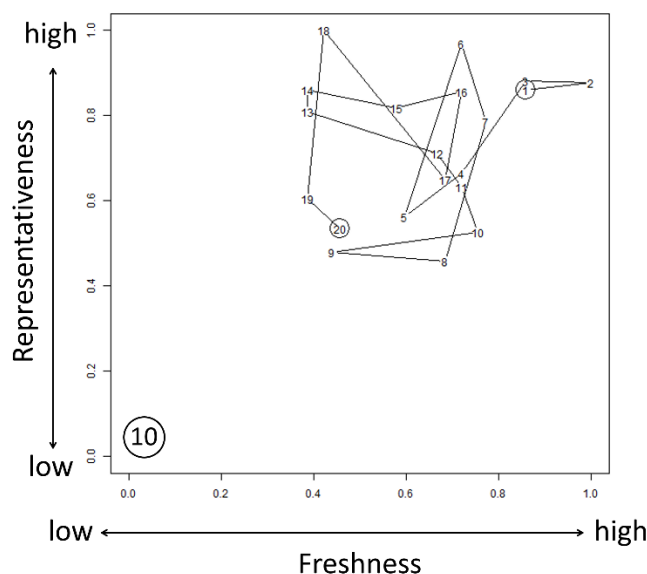


図 4-11 会議 ID10 の状態遷移図.

上記のように、分析結果である状態遷移図は会議の実際の状況を概ね反映しており、したがって状態遷移図を見ることで会議が順調であったかどうかについて把握することが可能である。

§ 実験 2-1 考察

実験 2-1 では、状態遷移図と会議のログデータを照らし合わせる形で検証を行い、会議の実際の状況が状態遷移図にある程度反映されており、状態遷移図から会議の状況を把握することが可能であることを示した。

4.6.3.2 実験 2-2 定量的評価

§ 実験の目的と方法

実験 2-2 では、実験 2-1 で評価対象とした 15 回の会議のそれぞれについて、アンケートによって会議の進行に関する評価を行い、4.5.2.5 節で定義した会議の評価指標との相関をみることでそれらの評価指標の有効性を確認する。アンケートには、当該事業の統括および会議のファシリテーションを担当した 1 名が回答した。回答者が参照できる資料としては文字起こしのテキストを用意した。状態遷移図は提示していない。

アンケートでは、15 回の各会議について「会議の進行が順調に行われたか」「合意形成がなされたか」という項目について、5(良い)から 1(悪い)までの 5 段階で評価をつけてもらった。さらに、実験 1 でも取り上げた会議 ID1 と 10 の会議について、「会議が順調にしたか、あるいはしなかったのかをどのような点から判断したか」「合意形成されている、あるいはされていないことをどのような点から判断したか」という項目について、自由記述によ

って回答してもらった。

4.5.2.5 節で定義した会議の評価指標の一覧を表 4-8 に示す。

表 4-8 会議の評価指標一覧。

I_1	会議が全体として探索から合意形成に向かうことを示す指標
I_2	最終的に合意に達したことを示す指標
I_3	会議全体の議論の振れ幅の大きさを示す指標
I_4	開始時と終了時の議論の振れ幅の大きさを示す指標

§ 実験 2-2 結果

実験結果として、各会議のアンケートによる評定と会議評価指標による評価値を表 4-9 に示す。評定及び評価値は、それぞれ値が大きいほど良い評価であることを表している。評定の数値をそのまま点数として用いて指標の評価値との相関係数を計量する。また、進行の評価と合意の評価のそれぞれについて、4 つの指標との相関係数を計量した。結果を表 4-10 に示す。

表 4-9 各会議の評定と会議評価指標の評価値。

ID	進行 評価	合意 評価	I_1	I_2	I_3	I_4
1	4	3	0.80	0.26	4.58	0.68
2	4	4	0.62	0.20	4.67	0.44
3	4	2	0.86	0.70	4.77	0.62
4	4	2	0.50	0.02	4.39	0.52
5	2	1	0.88	0.37	4.99	0.65
6	2	4	1.45	0.71	5.36	1.03
7	3	4	1.32	0.68	4.60	0.95
8	5	3	1.02	0.67	4.74	0.77
9	3	2	0.42	-0.01	5.57	0.52
10	4	2	0.08	-0.01	4.26	0.52
11	4	5	1.16	0.52	6.14	0.88
12	2	1	0.56	0.24	5.24	0.43
13	3	3	0.32	-0.16	4.48	0.64
14	5	4	0.76	0.60	5.67	0.80
15	2	2	0.78	0.32	4.90	0.74
平均	3.40	2.80	0.77	0.34	4.96	0.68

表 4-10 から、進行が順調であることと $I_1 \sim I_4$ の指標にはほとんど相関がないこと、合意がなされたこととは正の相関があることが読み取れる。実験 1 で評価基準とした、右下の

「探索」から左上の「合意形成」に向かうことを表現した指標である I_1 は、合意形成がなされたという評価とは 0.52 とやや高い正の相関があった。他の指標でもそれほど大きな差はなかったが、一番相関係数が大きかったのは始点と終点の距離を表す I_4 における 0.63 であった。これらの結果から、座標上の進行方向や振れ幅の大きさは会議において合意がなされたかどうかを判断する指標として利用可能であると考えられる。また、進行が順調であるかを判断する指標としては適当ではないと考えられる。

表 4-10 会議の評定と評価指標の相関係数.

	I_1	I_2	I_3	I_4
相関係数-進行	0.13	0.12	-0.08	-0.05
相関係数-合意	0.52	0.40	0.34	0.63

§ 実験 2-2 考察

本研究では、会議が順調に進行していることを示す指標を 4 つ定義し、アンケートによる評定との相関を計量したが、順調に進行しているという評価とこれらの指標との間には相関はほとんど見られなかった。それに対して、合意形成がなされたことについては相関が見られた。この要因について、アンケートの自由記述を元に考察する。

ID1 の会議において、回答者は進行の評定を 4、合意の評定を 3 としている。進行の評定の理由として、「主催者が司会進行やファシリテーションを行い順調に進行した」ことを挙げており、これは進行に対する相関係数が一番大きかった I_1 が 0.80 と平均の 0.77 と同等かやや高かったことと合致している。また、合意の評定の理由として「議論が中座しなかった」という意味では合意形成がなされたが、発想の拡散という意味ではなされていない」としている。これは一長一短の評価と捉えられ、各指標が概ね平均に近い値を示していることと合致している。したがって、ID1 の会議においては、各評価指標は概ね状況を正しく表現していると考えられる。

ID10 の会議において、回答者は進行の評定を 4、合意の評定を 2 としている。進行の評定の理由として、「出席者の欠席があったが、出席者にはファシリテーションの経験者が多く、豊かな発想が得られた」こと、「トピックは散逸気味で収束には至らなかった」ことを挙げている。これに対して指標 I_1 の値は 0.08 と全会議中最低であり、回答者の見解とは合致していない。また、合意の評定の理由としては、「本筋から外部環境の話題に逸れ過ぎたきらいがある」ことを挙げている。これは各指標が平均より低い数値を示していることと合致している。したがって、ID10 の会議においては、各評価指標は合意形成については状況を正しく表現できているが、会議の順調な進行については状況を正しく表現できていないと考えられる。

会議の順調な進行を表現できていないという事態が生じた原因としては、本手法では会議が順調に進行するということを、最終的に合意形成がなされることに主眼を置いて設計していたのに対して、実際の会議はそのような会議ばかりではなく、合意形成に至らずとも、

例えば豊かな発想が得られたことをもって順調に進行したと判断されることがあることが挙げられる。

また、今回分析対象とした各会議は、毎回司会やファシリテーターが用意されていたこともあり、進行が著しく順調ではない会議はそもそも存在しなかったと考えられる。したがって、相対的には評価の差はついているもの、些細な差を強調しすぎている可能性があり、より進行の順調でない会議を混ぜて分析すれば、より高い相関係数が得られる可能性があると考えられる。

4.6.3.3 実験 2 全体の考察

本実験では、会議の状態遷移を平面座標にプロットし、時間経過にともなってその座標の右下から左上に推移することを会議が順調に進行することであると仮定して指標の作成及び評価実験を行った。実験 2-1 では、座標上の動きが会議の議論の動きを概ね反映していることを確認した。実験 2-2 では、4.5.2.5 節で定義した会議の評価指標が「会議が順調に進行しているか」「合意形成がなされたか」の二点と相関があるかについて定量的評価を行った。その結果として、作成した指標はどれも合意形成については相関が見られたが、会議が順調に進行したかについては相関が見られなかった。したがって本手法で定義した指標は、最終的に合意形成に至るタイプの会議において、合意形成がなされたかの判定に用いることができることが分かった。他のタイプの会議を評価するには別の指標を用意する必要があると考えられる。

今後は会議のログデータを蓄積する際に、アンケート等によって、行った会議に対して様々な指標で評価を付けた形でデータを蓄積していくことで、会議のタイプ毎にどのような遷移のパターンが目標としている形なのかということを学習し、会議の評価を洗練させたり、あるいはファシリテーションに活かしていくことが可能であると考えている。

4.7 実システムへの展開

本研究で提案している分析システムを、国際大学 GLOCOM と株式会社イトーキの共同で、ハードウェアを含んだ実システムとして実装中である。以下にその概要を示す。

4.7.1 実装環境

会議室の現場の機器の開発およびセッティングは株式会社イトーキが担当した。音声入力には、会議用の音声録音マイクを用いた。この機器 1 台で話者を 8 名程度まで識別できる。音声のテキスト化には音声認識技術を用いており、リアルタイムにテキスト化を行うことが出来る。分析結果の表示にはタブレット端末を用いた。表示するプログラム等もイトーキが担当した。国際大学 GLOCOM は本論文に係る分析部分を担当した。イトーキと GLOCOM 間のシステムを API によって連携する形で全体のシステムを実装した。

実装された会場室の写真を図 4-12 に示す。マイクは天井に設置されている。奥のタブレット型端末に分析結果が表示されている。



図 4-12 実装された会議室。

この会場で話された言葉を入力として分析システムが出力したデータをタブレット端末上で可視化した例を図 4-13 に示す。なお、この画面が作成された時点では「鮮度」の事を「発散度」と呼んでいたため、現在の分析システムとはその点で表示が異なる。

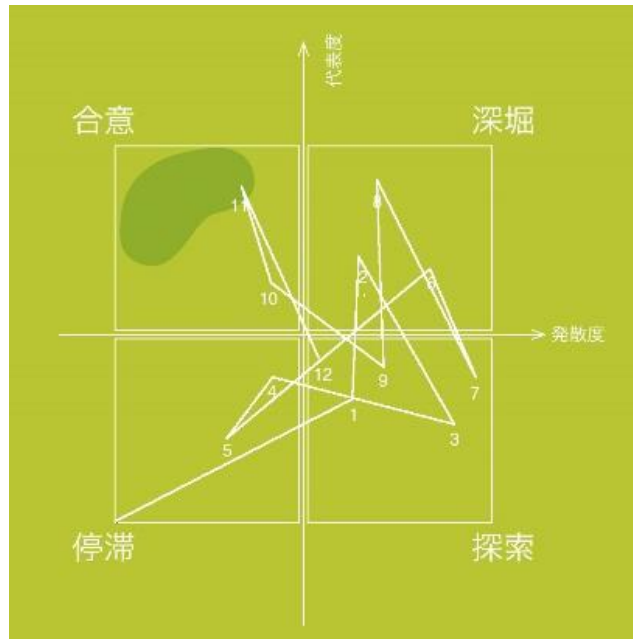


図 4-13 分析結果の表示例.

タブレット端末では、分析結果の他にも、発話を音声認識してタイムラインとして並べて表示することが出来る。個々の発話は話者別に、かつ音声途切れたところで区切られて入力される。表示例を図 4-14 に示す。この例では、上から下に時系列で発話ログが表示されており、左、真ん中、右の 3 人の話者が交互に会話していることを示している。



図 4-14 発話のタイムライン表示.

4.7.2 展示会における来場者への聞き取り

本システムを株式会社イトーキ主催の展示会である「ITOKI PRESENTATION 2017」において公開し、一般の来場者に本システムについての意見の聞き取りを行った。

肯定的な意見としては、会議を効率化できそう、会議を評価する指標が得られるのはありがたい、客観的に会議状況を共有化できる、といった意見が多く見られた。

否定的な意見としては、発言の一つ一つまで監視されているのはプライバシー的に問題がある場合がある、システムに評価を下されること自体が受け入れがたいといった意見が散見された。プライバシーの問題は、もしこのシステムを導入するのであれば、利用者に必ず許可を取る仕組みを作ることや、データ保持の方針を組織として検討せねばならないことを示している。また、評価を受け入れがたいという点については、実際の使用場面でシステムと人間が共同するうえでは重大な問題であると思われ、人間にとって受け入れ易い見せ方を熟慮する必要があることを示している。また、今後実証実験を行っていく中で会議の良し悪しといった評価を行う際には、参加者の関係性(例えば上司、部下の関係など)や感覚(モチベーション)が強く影響することが考えられるため、評価データの取得方法についても考慮する必要がある。

4.8 まとめ

本章では、会議の言語情報に着目し、発話データをテキストデータに変換したものを対象として分析することで、会議の内容および時系列に基づく分析手法を提案した。本章で提案した方式は大きく二つあり、一つは会議を時系列構造解析し概要を可視化する手法、もう一つは、会議の状態遷移を可視化する手法およびその状態遷移から会議の進行を評価する手法である。

会議を時系列構造解析し会議の概要を可視化する手法としては、単一の会議を対象とした4つの可視化方式と、一連の複数の会議を対象とした一つの可視化方式を示した。これらの手法によって、「会議のどこの時間帯でどのような話題が話されているか」「どこの時間帯で重要な話をしているか」「全体として重要な単語はどのようなものであったか」を分析可能にした。実験によってこれらの可視化の例を示し、有用性を確認した。

また会議の状態遷移を可視化する手法およびその状態遷移から会議の進行を評価する手法を示した。会議における発散／収束の状態を表現する指標として、語の新規性に着目した「鮮度」と、全体から見た重要性を表す「代表度」の二つを定義した。さらに、これら二つの指標を軸として組み合わせることで、会議の状態として「探索」「深掘り」「合意形成」「停滞」の四状態を定義し、会議の進行を状態遷移図として可視化する方式を示した。提案手法を用いて実際の会議について分析を行い、状態遷移図を見ることで会議が順調に進行したのかを大まかに判別できることを示した。さらに会議の進行を評価する指標を定義し、定量的評価を行ったところ、その指標は合意形成を目標とする会議において合意形成がなされているかを評価することに用いることができることが分かった。さらにハードウェアを含めた実システムとしての実装のプロトタイプを示した。

今後の課題としては、会議のタイプの分類とそれ毎の評価指標の定義、参加者同士の関係性に影響しないアンケート等の設計とそれによる会議の評価を含めた実証実験、ユーザー毎の発話を分析対象とした会議支援システムの開発などが挙げられる。

第5章 結論

本論文では、知識を検証しその再創造を促すモデルである、知識創造サイクルモデルの定義と構成方法を示すとともに、その具体的な実現例を示した。

2章では、提案モデルである、知識創造サイクルモデルの定義と構成方法を示した。このモデルは、知識創造と知識利活用を組み合わせることで知識の再創造を行うためのモデルである。知識創造サイクルモデルによって、現在の知識が十分であるかを検証するための演算を構成できることを示した。

3章では、メディアコンテンツを対象とした統計的一般可逆作用素構成方式 iML を定義し、さらに、 ML と iML を組み合わせて検証を行う枠組みである $ML-iML$ 系の枠組みを提案した。これは知識創造サイクルモデルのメディアデータを対象とした実現例に相当する。また、 iML の具体的な実現例として、楽曲メディアコンテンツを対象として、印象メタデータから楽曲メディアコンテンツを生成する方式について示し、実装を行った。実験を行い、 $ML-iML$ 系が実現されていること、およびに、提案システムの楽曲生成システムとしての有効性を示した。さらに、異種メディア間の連結の事例として、音相理論と上記システムを組み合わせることで、任意の日本語の入力語の音相の印象に合致した楽曲を生成するシステムを構築した。このシステムについても実験を行い、印象に合った楽曲が生成されることを示した。

4章では、会議の言語情報に着目し、発話データをテキストデータに変換したものを対象として分析することで、会議の内容および時系列に基づく分析手法を提案した。4章で提案した方式は大きく二つあり、一つは会議の時系列構造解析し概要を可視化する手法、もう一つは、会議の状態遷移を可視化する手法およびその状態遷移から会議の進行を評価する手法である。これらは、知識創造サイクルモデルにおける知識創造に相当する。

今後の課題としては、4章で扱った会議の研究において、知識創造サイクルモデルにおける知識利活用系として会議のファシリテーションを支援するシステムを構築することで、参加者である人を含んだサイクルを実現すること、およびに、3章で示した異種メディア連結の事例にも検証のためのサイクルを実現し、複合分野に対する知識創造サイクルモデルを実現することが挙げられる。

謝辞

本研究の遂行にあたって、筑波大学大学院システム情報工学研究科 北川高嗣 教授には、研究の内容についてご指導いただいたのはもちろん、研究の在り方や、それをとりまく社会の情勢などについて、幅広く深い知見から様々なご助言をいただきました。さらに、公私を問わず、長期に渡り温かいご支援をいただきました。ここに深く感謝いたします。

筑波大学大学院システム情報工学研究科 櫻井鉄也 教授には、筑波大学情報数理研究室の在籍時に同室の先生としてお世話になったほか、数値解析の専門家としての深い知見からの多数のご助言をいただきました。ここに深く感謝いたします。

慶應義塾大学環境情報学部 清木先生 教授には、マルチメディアデータベースシステム、意味的連想検索システムに関して、多くのご助言、ご指導をいただきました。ここに深く感謝いたします。

筑波大学大学院システム情報工学研究科 北川博之 教授、加藤和彦 教授、今倉暁 准教授には、本論文をまとめるにあたり、多くのご助言をいただきました。ここに深く感謝いたします。

武蔵野大学データサイエンス学部および国際大学グローバル・コミュニケーション・センター(GLOCOM)の 中西崇文 准教授には、研究活動の多岐に渡って、多くのご助言・ご指導をいただき、また、日々多くの議論させていただきました。いつも研究を引っ張っていただき、筆者を励ましてくださいました。ここに深く感謝いたします。

株式会社イトーキの 田中勇一 氏、小笠原豊 氏、大橋一広 氏には、会議の分析・効率化についての共同研究の機会をいただき、多くの議論をさせていただきました。ここに深く感謝いたします。

東京医科歯科大学難治疾患研究所の 本間秀典 特任助教には、メディアデータからの印象抽出およびメディアデータ生成の研究について、筑波大学情報数理研究室の在籍時代に指導いただいたほか、その後も、筆者に迷いや悩みが生じた時に、公私に渡ってご支援をいただきました。ここに深く感謝いたします。

株式会社コルグの 芳村亮 氏には、メディアデータ生成の研究、特に自動作曲システムの構成にあたって多くのご助言をいただきました。氏が先鋒を付けた自動作曲の研究をこうして形に出来たことを嬉しく思います。ここに深く感謝いたします。

株式会社日本データ取引所の 上島邦彦 氏には、会議の分析における評価実験において必要な、会議への評定付けおよびコメント付けをしていただきました。ここに深く感謝いたします。

筑波大学情報数理研究室の 松本恭平 氏には、筑波大学および国際大学グローバル・コミュニケーション・センター(GLOCOM)での研究活動において、長期に渡って多くの議論をさせていただきました。ここに深く感謝いたします。

筑波大学情報数理研究室の皆様には、筆者の在籍中もその後も、研究やそれ以外の話題

について日々多くの議論をさせていただきました。大変楽しい時間でした。皆様に感謝の意を表します。

国際大学グローバル・コミュニケーション・センター(GLOCOM)の先生、リサーチアシスタント、その他スタッフの皆様には、研究活動のご支援をいただいたほか、情報分野やそれ以外の分野の議論にも各々の専門の見地から多くの示唆をいただきました。ここに深く感謝いたします。

最後に、筆者をここまで育てていただき、温かく見守っていただいた両親に深く感謝いたします。

参考文献

- [1] 掛谷 英紀, 学問とは何か: 専門家・メディア・科学技術の倫理, 大学教育出版, 2005.
- [2] ユヴァル ノア ハラリ, 柴田 裕之, サピエンス全史 (下) 文明の構造と人類の幸福, 河出書房新社, 2016.
- [3] 長沼 伸一郎, 物理数学の直感的方法 (普及版) 理工系で学ぶ数学の「難所突破」の特効薬, 東京: 講談社, 2011.
- [4] Y. Kiyoki, T. Kitagawa and T. Hayama, "A metadatabase system for semantic image search by a mathematical model of meaning," *ACM Sigmod Record*, vol. 23, no. 4, pp. 34-41, 1994.
- [5] K. Hevner, "Expression in music: a discussion of experimental studies and theories.," *Psychological review*, vol. 42, no. 2, p. 186, 1935.
- [6] K. Hevner, "Experimental studies of the elements of expression in music," *The American Journal of Psychology*, vol. 48, no. 2, pp. 246-268, 1936.
- [7] K. Hevner, "The affective value of pitch and tempo in music," *The American Journal of Psychology*, vol. 49, no. 4, pp. 621-630, 1937.
- [8] 岡田 龍太郎, 中西 崇文, 本間 秀典 and 北川 高嗣, "メディアコンテンツを対象とした統計的一般化逆作用素構成方式とその楽曲メディアコンテンツ生成への適用," *情報処理学会論文誌*, vol. 57, no. 5, pp. 1341-1354, 2016.
- [9] 小林 重順, カラーイメージスケール改訂版, 講談社, 東京, 2001.
- [10] 北川 高嗣, 中西 崇文 and 清木 康, "静止画像メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的画像検索への適用," *情報処理学会論文誌データベース (TOD)*, vol. 43, no. 12, pp. 38-51, 2002.
- [11] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan and D. Hassabis, "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," *Science*, vol. 362, no. 6419, pp. 1140-1144, 2018.
- [12] T. Kitagawa and Y. Kiyoki, "A mathematical model of meaning and its application to multidatabase systems," in *Research Issues in Data Engineering, 1993: Interoperability in Multidatabase Systems, 1993. Proceedings RIDE-IMS'93., Third International Workshop on*, 1993.
- [13] R. Okada, T. Nakanishi and T. Kitagawa, "A Method of Knowledge Creation and Knowledge Utilization by Generalized Inverse Operator," in *Advanced Applied*

Informatics (IIAIAI), 2014 IIAI 3rd International Conference on, 2014.

- [14] T. Nakanishi, R. Okada and T. Kitagawa, "Automatic media content creation system according to an impression by recognition-creation operators," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, 2016.
- [15] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American society for information science*, vol. 41, no. 6, pp. 391-407, 1990.
- [16] G. Salton, A. Wong and C.-S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613-620, 1975.
- [17] Pearson Education, *Longman dictionary of contemporary English*, 1987.
- [18] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross and K. J. Miller, "Introduction to WordNet: An on-line lexical database," *International journal of lexicography*, vol. 3, no. 4, pp. 235-244, 1990.
- [19] T. Mikolov, W.-t. Yih and G. Zweig, "Linguistic Regularities in Continuous Space Word Representations.," in *HLT-NAACL*, 2013.
- [20] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *CoRR*, vol. abs/1301.3781, 2013.
- [21] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, Eds., Curran Associates, Inc., 2013, pp. 3111-3119.
- [22] E. Negri, L. Fumagalli and M. Macchi, "A Review of the Roles of Digital Twin in CPS-based Production Systems," *Procedia Manufacturing*, vol. 11, pp. 939-948, 2017.
- [23] E. Glaessgen and D. Stargel, "The digital twin paradigm for future NASA and US Air Force vehicles," in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*, 2012.
- [24] 吉野 太智, 高木 秀幸, 清木 康 and 北川 高嗣, "楽曲データを対象としたメタデータ自動生成方式とその意味的連想検索への適用," *情報処理学会研究報告データベースシステム (DBS)*, vol. 1998, no. 57, pp. 109-116, 1998.
- [25] 北川 高嗣, 中西 崇文 and 清木 康, "楽曲メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的楽曲検索への適用," *電子情報通信学会論文誌.D-I, 情報・システム, I-情報処理 = The transactions of the Institute of Electronics, Information and Communication Engineers.D-I*, vol. 85, no. 6, pp. 512-526, 2002.

- [26] 本間 秀典, 中西 崇文 and 北川 高嗣, "任意の言葉を対象とした音韻印象変換作用素の構成とその感性検索への適用," *情報処理学会論文誌*, vol. 51, no. 5, pp. 1294-1309, 2010.
- [27] 岡田 龍太郎, 芳村 亮 and 本間 秀典, "任意の言葉の音相の印象に合致した楽曲の自動構成方式," *日本データベース学会 letters*, vol. 6, no. 2, pp. 25-28, 9 2007.
- [28] 木通 隆行, 日本語の音相: ことばのイメージを捉える技術, 表現する技術, 小学館スクウェア, 2004.
- [29] 北原 鉄朗 and 深山 覚, "自動作曲・自動編曲の現状と課題 (小特集 音楽情報処理技術: 分析から合成・作曲・利活用まで)," *電子情報通信学会誌 = The journal of the Institute of Electronics, Information and Communication Engineers*, vol. 98, no. 6, pp. 475-479, 6 2015.
- [30] L. A. Hiller Jr and L. M. Isaacson, "Musical composition with a high speed digital computer," in *Audio Engineering Society Convention 9*, 1957.
- [31] K. Ebcioglu, Computer counterpoint, Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 1980, pp. 534-543.
- [32] J. A. Biles, "GenJam: A genetic algorithm for generating jazz solos," in *ICMC*, 1994.
- [33] D. Ando, P. Dahlstedt, M. G. Nordahl and H. Iba, "Computer aided composition by means of interactive gp," in *International Computer Music Conference, Icmc 2006*, 2006.
- [34] S. M. Schwanauer and D. A. Levitt, Machine models of music, MIT Press, 1993.
- [35] 白井 亨 and 谷口 忠大, "階層 Pitman-Yor 言語モデルを用いたメロディー生成手法の提案," *研究報告音楽情報科学 (MUS)*, vol. 2011, no. 3, pp. 1-6, 2011.
- [36] 深山 覚, 中妻 啓, 酒向 慎司, 西本 卓也, 小野 順貴 and 嵯峨山 茂樹, "音楽要素の分解再構成に基づく日本語歌詞からの旋律自動作曲," *情報処理学会論文誌*, vol. 54, no. 5, pp. 1709-1720, 2013.
- [37] Y. Tuchiya and T. Kitahara, "Melodic outline extraction method for non-note-level melody editing," *Journal of Information Processing*, pp. 1302-1307, 2013.
- [38] 北原 鉄朗, 深山 覚, 片寄 晴弘, 嵯峨山 茂樹 and 長田 典子, "OrpheusBB: Human-in-the-loop 型の自動作曲システム," *情報処理学会インタラクシオン2011*, pp. 57-64, 2011.
- [39] 梅本 堯夫, 音楽心理学, 誠信書房, 1966.
- [40] 株式会社 NTT データ経営研究所, 「会議の革新とワークスタイル」に関する調査, 第巻 2017.

- [41] D. Cameron and 林 宅男, 話し言葉の談話分析, ひつじ書房, 2012.
- [42] R. Okada, T. Nakanishi, Y. Tanaka, Y. Ogasawara and K. Ohashi, "A Topic Structuration Method on Time Series for a Meeting from Text Data," in *International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, 2017.
- [43] T. Nakanishi, R. Okada, Y. Tanaka, Y. Ogasawara and K. Ohashi, "A Topic Extraction Method on the Flow of Conversation in Meetings," in *Advanced Applied Informatics (IIAI-AAI), 2017 6th IIAI International Congress on*, 2017.
- [44] R. Okada, T. Nakanishi, Y. Tanaka, Y. Ogasawara and K. Ohashi, "A Visualization Method of Relationships among Topics in a Series of Meetings," *IIAI IEEE:Information Engineering Express*, vol. 3, no. 4, pp. 115-124, 2017.
- [45] R. Okada, T. Nakanishi, Y. Tanaka, Y. Ogasawara and K. Ohashi, "A Time Series Structure Analysis Method of a Meeting Using Text Data and a Visualization Method of State Transitions," *New Generation Computing*, 2018, <https://doi.org/10.1007/s00354-018-0050-6>.
- [46] 市野 順子 and 田野 俊一, "会議状況判別のための発言の時系列的パターンの特徴分析," *電子情報通信学会技術研究報告.HCS, ヒューマンコミュニケーション基礎*, vol. 109, no. 224, pp. 23-28, 2009.
- [47] T. Bergstrom and K. Karahalios, "Conversation Clock: Visualizing audio patterns in co-located groups," in *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*, 2007.
- [48] D. O. Olguín, B. N. Waber, T. Kim, A. Mohan, K. Ara and A. Pentland, "Sensible organizations: Technology and methodology for automatically measuring organizational behavior," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 43-55, 2009.
- [49] J. P. Guilford, The nature of human intelligence., McGraw-Hill, 1967.
- [50] A. F. Osborn, Applied imagination; principles and procedures of creative problem-solving: principles and procedures of creative problem-solving, Scribner, 1963.
- [51] 川喜田 二郎, 発想法: 創造性開発のために, 中央公論社, 1967.
- [52] 堀 公俊, ファシリテーション入門, 日本経済新聞社, 2004.
- [53] 市野 順子 and 田野 俊一, "発言の時系列的パターンを用いた会議における発散/収束の判別の可能性," *人工知能学会論文誌*, vol. 25, no. 3, pp. 504-513, 2010.
- [54] 富山 健, 高瀬 裕 and 中野 有紀子, "議論中の言語・非言語情報に基づく発散的/収束的発話の識別," in *人工知能学会全国大会論文集 2017 年度人工知能学会全国大会*

(第 31 回) 論文集, 2017.

- [55] M. A. Hearst, "TextTiling: Segmenting text into multi-paragraph subtopic passages," *Computational linguistics*, vol. 23, no. 1, pp. 33-64, 1997.
- [56] 工藤 拓, "MeCab:Yet Another Part-of-Speech and Morphological Analyzer," 2006. [Online]. Available: <http://taku910.github.io/mecab/>. [Accessed 6 1 2019].

論文目録

本研究に関する著者の論文リストは以下の通りである。

学術雑誌論文

1. 岡田龍太郎, 中西崇文, 本間秀典, 北川高嗣, 「メディアコンテンツを対象とした統計的一般化逆作用素構成方式とその楽曲メディアコンテンツ生成への適用」, 情報処理学会論文誌 57(5), pp.1341-1354, 2016.
2. Ryotaro Okada, Takafumi Nakanishi, Yuichi Tanaka, Yutaka Ogasawara, Kazuhiro Ohashi, "A Visualization Method of Relationships among Topics in a Series of Meetings", IIAI IEEE:Information Engineering Express , Vol 3, No 4 pp.115-124, 2017.
3. 岡田龍太郎, 芳村亮, 本間秀典, 北川高嗣, 「任意の言葉の音相の印象に合致した楽曲の自動構成方式」, 日本データベース学会 letters6(2), pp.25-28, 2007.
4. Ryotaro Okada, Takafumi Nakanishi, Yuichi Tanaka, Yutaka Ogasawara, Kazuhiro Ohashi, "A Time Series Structure Analysis Method of a Meeting Using Text Data and a Visualization Method of State Transition", New Generation Computing, 2018.
DOI: 10.1007/s00354-018-0050-6

国際会議発表論文(査読あり)

1. Ryotaro Okada, Takafumi Nakanishi, Takashi Kitagawa, "A Method of Knowledge Creation and Knowledge Utilization by Generalized Inverse Operator", 2014 IIAI 3rd International Conference on Advanced Applied Informatics, IEEE, pp.253-258, 2014.
2. Ryotaro Okada, Takafumi Nakanishi, Yuichi Tanaka, Yutaka Ogasawara, Kazuhiro Ohashi, "A Topic Structuration Method on Time Series for a Meeting from Text Data", 発表: ACIS International SNPD 2017, IEEE, 収録: Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Studies in Computational Intelligence, Springer, pp.45-59, 2017.
3. Takafumi Nakanishi, Ryotaro Okada, Takashi Kitagawa, "Automatic media content creation system according to an impression by recognition-creation operators", IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), pp.175-180, 2016.
4. Takafumi Nakanishi, Ryotaro Okada, Yuichi Tanaka, Yutaka Ogasawara, Kazuhiro Ohashi, "A Topic Extraction Method on the Flow of Conversation in Meetings", 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), IEEE, pp.351-356, 2017.

国内学会発表論文(査読あり)

1. 岡田龍太郎,芳村亮,本間秀典,北川高嗣,「任意の言葉の音の印象に合致した楽曲の自動生成方式」,DEWS2007(日本データベース学会 データ工学ワークショップ), 情報処理学会研究報告データベースシステム (DBS) , 2007-DBS-143,p241-246.

国内学会発表論文(査読なし)

1. 岡田龍太郎,芳村亮,本間秀典,北川高嗣,「任意の言葉の音相の印象に合致した楽曲の自動生成方式」,電子情報通信学会技術研究報告.DE,データ工学 107(131),p241-246,2007.
2. 大高宏樹,岡田龍太郎,北川高嗣,「楽曲属性を反映する特徴量の動的選出に関する研究」,情報処理学会研究報告データベースシステム (DBS) 2011-DBS-153(17),p1-7,2011.
3. 井上尚,岡田龍太郎,北川高嗣,「色-印象データベースを利用した画像の印象を提示する画像作成支援アプリケーション」,情報処理学会研究報告データベースシステム (DBS) 2011-DBS-153(17),p1-7,2011.
4. 井上尚,岡田龍太郎,北川高嗣,「形状情報に着目した画像のメタデータ抽出」,情報処理学会研究報告データベースシステム (DBS) 2012-DBS-155(5),p1-5,2012.