

# 分析結果とその解釈の統合を支援する 知識創発インタフェース

## Knowledge Creation Support by Integrating Analysis Results and Interpretations

西山 知志  
Satoshi Nishiyama

砂山 渡  
Wataru Sunayama

広島市立大学 情報科学部  
Faculty of Information Sciences, Hiroshima City University

**Abstract:** In the process of convergence of knowledge creation in text mining, by repeating the interpretation and integration of the collected analysis result, we are aiming to acquire a straightforward knowledge represented as cause-and-effect relationship. In this study, by the “multi-stage of interpretation” that make giving a repeat interpretation in one of the results, provide assistance leading to the acquisition of general knowledge by extending the width of the interpretation. In addition, by the “cause and result of labeling” to the interpretation, provide assistance to integrate knowledge as a cause-and-effect relationship.

### 1 はじめに

近年、情報技術の発展により、私たちはブログや SNS などを通して、テキスト形式で情報発信を手軽に行えるようになった。そうした中で、インターネット上には膨大なテキストデータが氾濫している状況にある。テキストデータには様々な情報が埋もれており、その中から活用できる有用な知識をいかにして獲得するのが、ビジネスの現場において課題となっている。例えば、企業の視点からテキストデータを利用する際に、顧客の声や風評を抽出した分析結果を、マーケティングや商品開発に活用することが求められている。ただ単に、テキストデータの分析結果を得ることにとどまるだけでは、知識の獲得につなげていくのは難しい。収集した分析結果に対して、それぞれが何を意味しているのかについて、効果的な解釈を行うことが必要になる。

そこで本研究では、テキストデータの分析結果に対する解釈の与え方として 2 つの方法を考えた。1 つ目の方法は解釈の多段化で、解釈を与える際に論理的思考の飛躍を促すことで解釈の幅を広げる支援を行う。2 つ目の方法は解釈への原因、結果のラベル付けで、解釈同士に潜む因果関係を捉える支援を行う。これらの支援をもとに、知識創発を促すことを目的とする。

本研究における知識創発とは、テキストマイニングツールを用いて収集したテキストデータの分析結果に対して、それらを総合的に解釈して 1 つの新たな知識に統合することを指す。多くの分析結果とその解釈が

得られても、まとまりがなければそれらを実際に活用できないので、根本的に信頼性が高い知識を得るために、解釈の統合を行う。本研究の目的が達成された場合、より幅広い解釈の中で解釈同士の因果関係を意識した知識創発を実現できる。

### 2 関連研究

テキストから知識を獲得するための研究として、テキストを対象とした評価情報の分析に関する研究 [1]、出来事の望ましさ判定を目的とした語彙知識獲得 [2]、軽卑表現の情報を活用した知識発見 [3] がある。以上に挙げた 3 つの研究に対して、テキストデータの分析結果から知識を獲得する点で類似しているが、本研究は、得られた複数の結果から最終的に 1 つの新たな知識へ結びつける点で異なる。また、テキストに含まれる文章やキーワード同士の因果関係に関する研究として、接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得 [4]、共起パターンの学習による事態間関係知識の獲得 [5]、構文パターンを用いた因果関係の抽出 [6] がある。以上に挙げた 3 つの研究に対して、分析結果からテキスト内に含まれる因果関係を捉える点で類似しているが、本研究は、分析結果の解釈同士の因果関係を捉える点で異なる。

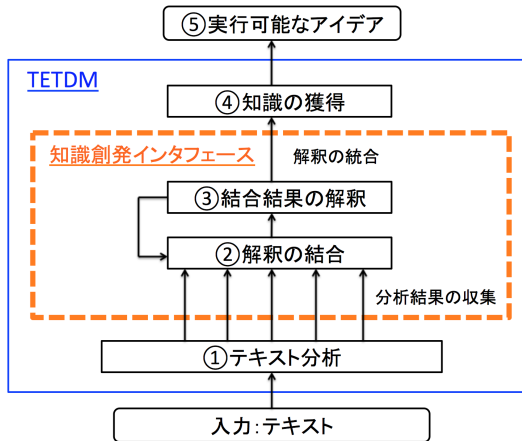


図 1: 従来の知識創発インタフェースの枠組み



図 2: 従来の知識創発インタフェースの画面

### 3 従来の知識創発インタフェース

#### 3.1 TETDM(テキストデータマイニングのための統合環境) への実装

本研究では、知識創発インタフェースを TETDM[7] 中の 1 つの機能として実装する。TETDM とは、テキストデータマイニングのための統合環境のことで、複数のテキストマイニングツールを組み合わせて統一的に扱うことができる。

#### 3.2 知識創発の 5 つのプロセス

従来の知識創発インタフェースの枠組みを図 1 に示す。知識創発を行う際には、TETDM にテキストを入力して最終的に得た知識を実行可能なアイデアにつなげていくまでに、図 1 の丸数字が示す 5 つのプロセスを踏むことになる。また、従来の知識創発インタフェースのシステム画面を図 2 に示す。

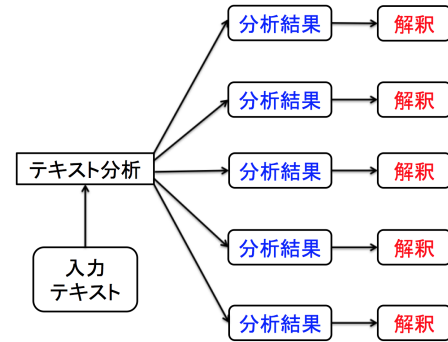


図 3: 知識創発における発散的思考



図 4: 共通点および解釈の記入欄 (赤枠)

#### 3.2.1 テキストの分析結果の収集

TETDM に分析したいテキストを入力する。テキストマイニングツールを用いてテキスト分析を行い、複数の分析結果を得る。このとき、それぞれの分析結果に対して、20 文字程度の文章で解釈を与える。

知識創発において、最終的に 1 つの知識を得るために分析結果を得る行為を繰り返すことを、発散的思考と呼ぶ。図 3 に発散的思考のイメージを示す。

#### 3.2.2 分析結果の解釈の結合

収集した分析結果のそれぞれの解釈の中から、共通点を含む解釈同士を探す。共通点を見つけた解釈同士は後に結合する。このとき、共通点の内容を図 4 赤枠上段のテキストフォームに記入する。

#### 3.2.3 共通点の解釈

解釈同士で見つけた共通点から考えられる解釈を、図 4 赤枠下段のテキストフォームに記入する。図 4 最下の結合ボタンを押すと、記入した共通点の解釈が新たに生成される。

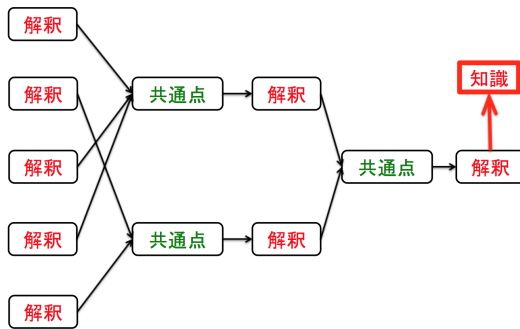


図 5: 知識創発における収束的思考

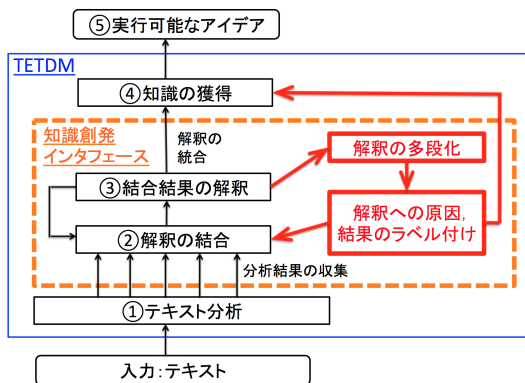


図 6: 提案する知識創発インタフェースの枠組み

### 3.2.4 全ての解釈の統合による知識の獲得

解釈の結合と共通点の解釈を繰り返し、最終的に 1 つに統合された解釈を新たな知識として獲得する。知識創発において、最終的に 1 つの知識を得るために全ての解釈を統合することを、収束的思考と呼ぶ。図 5 に収束的思考のイメージを示す。

### 3.2.5 得られた知識をもとにした実行可能なアイデアの考案

全ての解釈を統合して知識を獲得するだけにとどまってしまうと、その知識のままでは実際に役に立たない。そこで最後に、得られた知識をもとに実際に現場で活用できる、実行可能なアイデアを考案する必要がある。

## 4 提案する知識創発インタフェースのシステム構成

提案する知識創発インタフェースの枠組みを図 6 に示す。図 1 に示した従来の知識創発インタフェースに対して新たに、解釈の多段化、および、解釈への原因、結果のラベル付けの 2 つの機能を実装した。また、提



図 7: 提案する知識創発インタフェースの画面



図 8: 解釈の多段化のイメージ

案する知識創発インタフェースのシステム画面を図 7 に示す。

### 4.1 解釈の多段化

創発とは、部分の性質の単純な総和にとどまらない性質が、全体として現れることを意味する。収集した分析結果の解釈を、ただ単に共通点で整理しながら統合するだけでは、知識の創発にはつながりにくい。知識の創発を促すには、共通点の解釈の幅を広げながら統合していく必要がある。そこで、共通点に対する幅広い解釈を目指すために、解釈の多段化を考えた。図 8 に解釈の多段化のイメージを示す。

共通点に対して、連続して解釈を与えることを可能にした。まず、共通点に対して解釈 1 を与える。解釈 1 に対してさらに発展させた解釈 2 を与える。解釈 2 に対してさらに発展させた解釈 3 を与える。発展させた解釈のパターンは、例えば次の 3 つが考えられる。1 つ目は、「AKB48 が好き」を「アイドル好き」と解釈することで、「AKB48」が「アイドル」に抽象化されるパターンがある。2 つ目は、「スポーツを観戦する」を「サッカーを観戦する」と解釈することで、「スポーツ」が「サッカー」に具体化されるパターンがある。3 つ目は「友達が多い」を「フレンドリーな性格」と解釈することで、「友達が多い」が「フレンドリー」に言い換えが行われるパターンがある。

知識創発において、解釈を行う思考が堅実なままでは一般的な知識の獲得にとどまることになる。一方で、解釈を行う思考に論理的な飛躍が大きいほど、1 度の解釈だけではたどり着きにくかった解釈を得ることが

解釈1 :

解釈2 :

解釈3 :

図 9: 解釈の多段化を行うテキストフォーム

でき、解釈の幅が広がりやすくなる。解釈の多段化により、知識創発を行う上で、解釈の幅を広げることが期待できる。

#### 4.1.1 解釈の多段化のインターフェース

解釈の多段化のインターフェース部分を図9に示す。共通点から考えられる新たな結合結果の解釈を1段目のテキストフォームに記入する。2段目のテキストフォームには、共通点から考えられる2つ目の新たな結合結果の解釈、または、1段目に記入した解釈をさらに発展させた解釈を記入する。3段目のテキストフォームには、共通点から考えられる3つ目の新たな結合結果の解釈、または、2段目に記入した解釈をさらに発展させた解釈を記入する。

## 4.2 解釈への原因，結果のラベル付け

知識創発において価値ある創発を行うためには、収集した分析結果の本質に迫りながら、解釈を統合していくことが必要になる。そのためには、収集した分析結果の解釈から主要な知識を探る力が求められる(図10)。主要な知識とは、根本的な原因と結果の全体像を意味する。結果の全体像が分かると、分析結果から何が起きているのかを理解しやすくなる。また、全ての結果を引き起こした根本的な原因が分かると、知識を活用する明確な方針を立てやすくなる。最終的に、根本的な原因と結果の全体像の関係性を読み取ることで、分析結果の解釈同士に潜む因果関係を捉えることができる。このとき、未知の因果関係の創発結果が得られると、未知知識の発見につながる。一方で、既知の因果関係の創発結果が得られると、既知知識の理解や確認につながる。

そこで、解釈同士に潜む因果関係を意識した知識創発を目指すために、解釈への原因，結果のラベル付けを考えた。知識創発を行う際には、分析結果から何の知識を得たいのかについて、創発テーマを設定する。創発テーマに関して原因を表す解釈には原因ラベルを付ける。一方で、創発テーマに関して結果を表す解釈には結果ラベルを付ける。ラベル付けにより、原因，結果を表す解釈を整理しながら統合することが可能になる。原因ラベルがついた解釈同士を統合することで根本的な原因が分かりやすくなり、結果ラベルの付いた

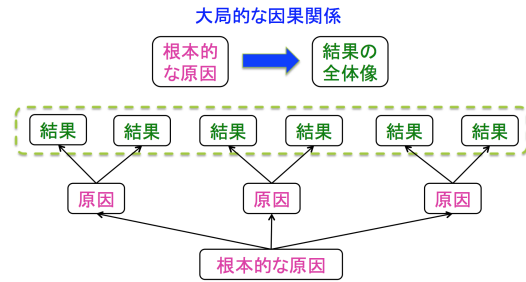


図 10: 主要な知識の探索



図 11: 解釈への原因，結果のラベル付けを行うチェックボックス

解釈同士を統合することで結果の全体像が分かりやすくなる。解釈同士に潜む因果関係を意識した知識創発が期待できる。

#### 4.2.1 解釈への原因，結果のラベル付けのインターフェース

解釈を記入するテキストフォーム右端にある「原因」と「結果」のチェックボックス(図11)をチェックすると、チェックした段の解釈に原因と結果のラベルが結合後に付与される。原因，結果のそれぞれラベル付けがされた解釈は図12の通りにラベルが表示される。3段記入した解釈のうち、創発テーマに関して原因を表す解釈の段には、原因のチェックボックスにチェックを入れる。原因のラベル付けがされた解釈は図13左の通り、システム画面の左側に表示される。一方で、創発テーマに関して結果を表す解釈の段には、結果のチェックボックスにチェックを入れる。結果のラベル付けがされた解釈は図13右の通り、システム画面の右側に表示される。

## 5 知識創発インターフェースの有効性検証実験

### 5.1 実験目的

本実験では、提案した知識創発インターフェースによって、次の2つの有効性を検証することを目的とする。



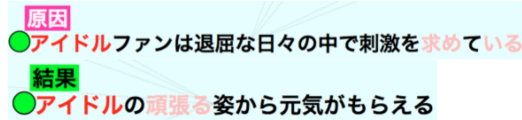


図 12: 原因, 結果のラベル付けがされた解釈



図 13: 解釈結合後の知識創発インタフェースの画面

1. 解釈の多段化により, 解釈の幅を広げることができたのか.
2. 解釈への原因, 結果のラベル付けにより, 解釈同士に潜む因果関係を意識した知識創発が行われたのか.

## 5.2 実験方法

テキストの分析結果の解釈 10 個に対して, 次の 2 つの創発テーマに従い各 2 回ずつ, 合計 4 回の知識創発を被験者に行ってもらった.

1. アイドルのファン心理について述べたテキストの分析結果の解釈 10 個を「ファンがアイドルに求めていること」をテーマに知識創発を行う.
2. ソーシャルゲームの魅力について述べたテキストの分析結果の解釈 10 個を「ユーザがソーシャルゲームに求めていること」をテーマに知識創発を行う.

比較システムは, 解釈の多段化および解釈への原因, 結果のラベル付けの, 2 つの機能を除いたインタフェース (図 2) を用意した. 被験者は大学生 10 人を A と B の 2 グループに分けた. 各被験者が 4 回の知識創発を行った際の創発テーマと使用したシステムについて表 1 に示す.

表 1: 各被験者が知識創発を行った順番

	A グループ 5 人	B グループ 5 人
1 回目	テーマ 1 提案システム	テーマ 1 比較システム
2 回目	テーマ 1 提案システム	テーマ 1 提案システム
3 回目	テーマ 2 比較システム	テーマ 2 提案システム
4 回目	テーマ 2 提案システム	テーマ 2 提案システム

表 2: 解釈の回数 (平均)

	比較システム	提案システム
解釈 1	7.2	6.5
解釈 2	-	5.3
解釈 3	-	3.4
計	7.2	15.2

## 5.3 実験結果と考察

### 5.3.1 解釈の多段化による有効性の実験結果と考察

表 2 に, 各段階ごとの解釈の回数の平均を調べた結果を示す. 表 2 より, 比較システムよりも提案システムの方が解釈を行った回数が増えた. これは, 連続した解釈を可能にしたことで, 1 つの解釈に対して解釈を与える回数が増えたためと考えられる.

次に, 最初に用意された分析結果の文章中に含まれていない名詞が, 解釈を行った文章中にどれだけの種類が新たに登場したかを調べた. 表 3 に, その結果を示す. 表 3 より, 比較システムよりも提案システムの方が, 解釈で新たに登場した分析結果に含まれない名詞数が多くなった. これは, 解釈を重ねることで, 論理的思考の飛躍を促せたためと考えられる.

テーマ 1 において, 解釈の多段化が行われた具体例を挙げる. 1 段階目で「下手でも一生懸命に歌やダンスをする姿がいい」と解釈された後に, 2 段階目で「努力によってパフォーマンスに成長があれば嬉しい」と解釈され, さらに 3 段階目で「ステップアップの過程が見られて更に応援したくなる」と解釈された. このとき, 1 段階目から 2 段階目の解釈で, 「一生懸命」が「努力」に言い換え, 「歌」や「ダンス」が「パフォーマンス」に抽象化, 「いい」が「嬉しい」に具体化が行われた. また, 2 段階目から 3 段階目の解釈で, 「成長」が「ステップアップの過程」に言い換え, 「嬉しい」が「応援したくなる」に具体化が行われた.

以上の実験結果より, 解釈の回数が増えると, 解釈で新たに登場した分析結果に含まれない名詞数が多くなることが分かった. これは, 解釈の多段化により, 連続した解釈を行うことで, 1 度の解釈だけではたどり着きにくかった新たな解釈が生まれやすくなったためと考えられる. このことから, 提案システムに実装した解釈の多段化により, 解釈の幅を広げられる有効性が示された.

表 3: 解釈で新たに登場した分析結果に含まれない名詞数 (平均)

	比較システム	提案システム
解釈 1	6.4	5.8
解釈 2	-	3.0
解釈 3	-	2.1
計	6.4	10.9

表 4: ラベルが付与された回数と付与率 (平均)

	原因	結果	ラベル付与率
1 回目	2.8	2.6	92.4
2 回目	3.1	3.0	95.5
3 回目	3.4	3.5	97.3

### 5.3.2 解釈への原因, 結果のラベル付けによる有効性の実験結果と考察

表 4 に, 提案システムにおいて, 原因と結果のそれぞれのラベルが付与された回数と, 共通点で結合した回数に対するラベルの付与率を調べた結果を示す. 表 4 より, 提案システムを 3 回使用したいずれの場合においても, 9 割を超える共通点に対する解釈にラベルが付与されていた. これは, 解釈同士の因果関係を探るべく積極的に原因と結果のラベルが使用されたためと考えられる.

次に, 提案システムにおいて, 分析結果を最終的に原因ラベルが付いた解釈 1 個および結果ラベルが付いた解釈 1 個として統合した被験者数を調べた. 表 5 に, その結果を示す. 表 5 より, 提案システムを 3 回使用したいずれの場合においても, 7 割を超える被験者が原因ラベルが付いた解釈 1 個および結果ラベルが付いた解釈 1 個として統合していた. これは, 解釈に原因ラベルを付けたことで, 原因を表す解釈が明示され, 根本的な原因を発見しやすくなったためと考えられる. また, 解釈に結果ラベルを付けたことで, 結果を表す解釈が明示され, 結果の全体像が分かりやすくなったためと考えられる.

テーマ 2 において, 原因ラベルが付いた解釈 1 個と結果ラベルが付いた解釈 1 個に統合された因果関係の具体例を挙げる. 収集した分析結果が最終的に, 原因ラベルが付いた解釈「有名なアニメやゲームとのコラボ企画が楽しい」と, 結果ラベルが付いた解釈「中毒性がありハマると毎日起動する」に統合された. この因果関係から「面白いイベント企画を毎日発生させて飽かさせない工夫をする」実行可能なアイデアが考案された.

以上の実験結果より, 原因と結果のラベルが積極的に使用され, 因果関係を捉えられる形でまとめられていたことが分かった. これは, ラベル付けにより, 解釈を原因と結果に整理することで, 因果関係に基づい

表 5: 最終的に原因 1 個, 結果 1 個として統合した被験者数 (/10 人)

	被験者数
1 回目	7
2 回目	9
3 回目	9

た統合を行いやすくなったためと考えられる. このことから, 提案システムに実装した解釈への原因と結果のラベル付けにより, 解釈同士に潜む因果関係を意識した知識創発が行われる有効性が示された.

## 6 結論

本研究では, 「解釈の多段化」「解釈への原因, 結果のラベル付け」を用いた知識創発インタフェースを提案した. 検証実験により, 解釈の幅を広げ, 因果関係を意識した知識創発を促す有効性を確認した. 今後は, 創発された因果関係の価値を自動評価する方法について検討していきたい.

## 参考文献

- [1] 乾孝司, 奥村学: テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol.13, No.3, pp.201-241 (2006)
- [2] 乾孝司, 乾健太郎, 松本裕治: 出来事の望ましさ判定を目的とした語彙知識獲得, 言語処理学会第 10 回年次大会発表論文集, pp.91-94 (2004)
- [3] 荻野紫穂, 那須川哲哉, 金山博, 榎美紀: 軽卑表現の情報を活用した知識発見, 言語処理学会第 18 回年次大会発表論文集, pp.58-61 (2012)
- [4] 乾孝司, 乾健太郎, 松本裕治: 接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得, 情報処理学会論文誌, Vol.45, No.3, pp.919-933 (2004)
- [5] 阿部修也, 乾健太郎, 松本裕治: 共起パターンの学習による事態間関係知識の獲得, 自然言語処理, Vol.16, No.5, pp.79-100 (2009)
- [6] 坂地泰紀, 竹内康介, 増山繁, 関根聡: 構文パターンを用いた因果関係の抽出, 言語処理学会第 14 回年次大会論文集, pp.1144-1147 (2008)
- [7] 砂山渡, 高間康史, 西原陽子, 徳永秀和, 串間宗夫, 阿部秀尚, 梶並知記: テキストデータマイニングのための統合環境 TETDM の開発, 人工知能学会論文誌, Vol.28, No.1, pp.1-22 (2013)