

事例検索と調査テーマ推薦に基づく ビジネス戦略支援 AI

A Decision Support AI for Business Operations via Information Retrieval and Query Suggestion

森下 皓文
Terufumi Morishita

尾崎 太亮
Hiroaki Ozaki

佐藤 美沙
Misa Sato

是枝 祐太
Yuta Koreeda

柳井 孝介
Kohsuke Yanai

日立製作所 研究開発グループ
Research & Development Group, Hitachi, Ltd.

We developed a system that supports decision making in business operations, which interactively assists iterative investigations. For example, for a decision “Should we buy Hitachi’s stock?”, the user may first query “What is the business domain of Hitachi?” to which the system responds “Electric power industry”. This encourages the user to query “How big is the power industry?”, and so on. Our system consists of two modules. The document retrieval module extracts pieces of texts given a “relation query” which consists of entities (e.g., “Hitachi”, “Electric power industry”) and their relationship (e.g., “be engaged in”). The query suggestion module proposes “relation queries” based on the previous query and the response. We developed a set of extraction rules for each of nine relationships in business domain to implement the two modules. We evaluated our rules in terms of precision of document retrieval and obtained significant gains over the baselines.

1. ビジネス上の意思決定支援

筆者らの研究グループでは、電子化されたテキストデータを意思決定支援に活用することを目的として、ディベートをする人工知能（以下ディベート型 AI）を開発している [Sato 15]. これを活用し、ビジネスにおける意思決定を支援するためのシステムを開発した。経営者の意思決定を考えれば分かるように、ビジネス上の意思決定は組織に対して大きな影響を与えるが、限られた時間の中で可能な限り正確に行わなければならない。このため、ディベート型 AI によって様々な観点での意見を自動で提示することが、有益である。本発表ではこのシステムのデモンストレーションを行う。

ディベート型 AI は、論題を入力として、経済・社会・科学などの“一般的な観点”に基づく事例をテキストデータから収集し、賛否意見として出力するシステムである。例えば「カジノを推進すべきか」という論題に対しては、経済の観点から「カジノは経済を活性化させる」といった事例や、社会の観点から「カジノによって犯罪が増加する」といった事例を、提示することができる。これらの観点は、事前に設定することができる。

ディベート型 AI をビジネス上の意思決定支援に活用するにあたり、前述の“一般的な観点”に加えて“ビジネス上の観点”で事例を収集する必要がある。更に、意思決定までの予備調査として、以下に例示するような一連のプロセスを支援する必要がある。

「日立に投資すべきか?」という意思決定の予備調査は以下のように行われるであろう：

1. 「まずは、日立が最近参入した事業を調べよう。」
⇒「日立は太陽光発電に参入したようだ。」
2. 「それでは、太陽光発電のニーズを調べよう。」
⇒「発展途上国のニーズが増えている。」
3. 「それでは、途上国市場の今後の成長度合いを調べよう。」
⇒「...」

この例から分かるように、「日立」「太陽光発電」「発展途上国」という調査対象を、「参入する」「ニーズがある」といったビジネス上の調査観点によって結びつけながら、新規の予備調査テーマ（「それでは、太陽光発電のニーズを調べよう。」）を次々と創案していく必要がある。よって、この調査テーマ創案を支援できることが望ましい。

そこで、“ビジネス上の観点”に基づいて関連事例を検索する機構と、予備調査プロセス支援のために新規調査テーマを推薦する機構を作成した。両機構の中核となるのは、調査対象間の関係性を特定するための、関係性抽出技術である。そこで、ビジネスドメインで重要な関係性を 9 つ定義し、関係性を特定するためのルールを作成した。また、これらの関係性を用いた事例検索の精度を計測することにより、効果を確認した。

2. 調査フローとシステム構成

2.1 調査フローとシステム構成

本システムが対象とする調査フローは、調査テーマの提案と、調査テーマに応じた事例の収集とを繰り返すことで、意思決定に有用な情報を探索的に収集するものである。

これを支援するための中核コンポーネントとして、調査テーマの推薦機構と、関連事例検索機構を、構築した。具体的な調査フローと共に図 1 に示す。

1. ユーザがテーマを考え、システムに入力する（図中①）。
（例：「日立に投資すべきか。」）
2. ユーザは、テーマに応じた予備調査を開始する。システムは予備調査テーマを推薦する（図中②）。（例：「日立が参入した事業を調べますか？」）
3. ユーザは予備調査テーマを選択する（図中③）。システムは、選択された予備調査テーマに関連する事例を検索しユーザに提示する（図中④）。（例：「日立は太陽光発電に進出する。」）
4. また、システムは、取得された事例を元に、次の予備調査テーマを推薦する（図中⑤）。（例：「太陽光発電のニーズを調べますか？」）

2～4 の調査サイクルを繰り返すことにより情報を蓄積し、意思決定に至る。

連絡先: 森下 皓文, 株式会社 日立製作所, 〒 185-8601 東京都国分寺市東恋ヶ窪一丁目 280 番地, 042-323-1111, terufumi.morishita.wp@hitachi.com

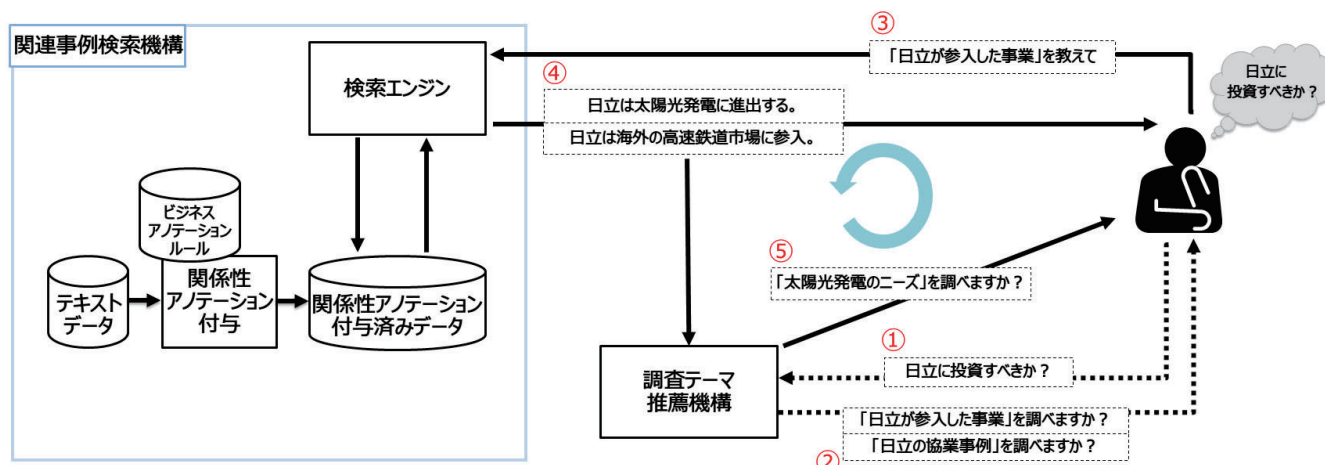


図 1: 本システムのコンポーネントと、ユーザーの調査フローとの関係

2.2 調査テーマに応じた関連事例の検索

2.2.1 ディベート型 AI における関係性抽出

調査テーマに応じた関連事例を検索するために、ディベート型 AI の中核技術である関係性抽出技術と、関係性に基づくテキスト検索技術を活用する。

関係性抽出技術では、構文パターンと述語辞書に基づくルールによって、テキスト中で述べられた関係性を特定する [Yanai 17]。例えば「カジノは経済を活性化する」というテキストでは、「カジノ」という主体が、「経済」という客体に対して、「Positive な影響を与える」という関係性が述べられている。この関係性を特定するために、「 x は y を活性化する」といった構文パターンを事前に用意しておき、入力テキストの構文と照合することで、主体 x と客体 y を特定する。また、Positive な影響を表す「活性化する」「促進する」「増やす」等の述語を含む辞書を構築する。辞書中の語句に全てに関して照合することにより、述語表現の揺れを吸収することができる。

関係性と主体・客体の情報（関係性情報）だけでなく、関係性情報を述べた元テキストにも、重要な情報が含まれる。そこで、関係性情報を元テキストに紐づけて検索エンジンへと登録する（この操作を、テキストへの“関係性アノテーションの付与”と呼ぶ）。これにより、関係性情報を指定して元テキストを検索する関係性検索が可能となる。例えば、主体を「カジノ」、関係性を「Positive な影響」と指定して検索することで、「カジノは経済を活性化する」といったテキストが取得される。元テキストの検索機能を持つことが、Virtuoso のような RDF ストアとの違いである。

ディベート型 AI では、主体が客体に対して「Positive/Negative な影響を与える」という影響関係 [佐藤 16] が重要となる。この関係性を利用して、「カジノは経済を活性化する」というように、議題語（カジノ）による影響を述べたテキストを検索し、賛否意見として提示する。

ビジネス上の意思決定においても、影響関係による検索は有用であると考えられる。影響関係で表された「日立は IoT 事業を行う会社を設立する。」といった事例を取得できるからである。しかし、次の問題が存在する。

まず「企業と企業が協業する」というように、影響関係ではない重要な関係性が存在する。

次に、ビジネスにおける意思決定支援では、単一観点での事例検索が重要である。複数の観点を含む事例が取得された場合、ユーザー側の追加作業として、不要な観点を述べた事例の

除去と、事例を観点毎に仕分け直すこと、が必要となる。結果として、調査の効率が著しく損なわれる。影響関係は、ビジネス上有用でない観点まで含めて、複数の観点を含んでしまっている。

そこで、ビジネス上重要で、粒度が小さく単一の観点を表せるような関係性を定義し、これらの構文パターン・述語辞書を作成した。

2.2.2 ビジネスドメイン関係性アノテーション

まず、ビジネスにおいて重要な存在である「事業主体」や「事業分野」等のエンティティ間の関係性を 5 つのグループに類型化した。次に、これら関係性グループにぶら下がる形で、実際の関係性を合計 9 つ定義した（表 1）。

例えば 3 行目 (enter(x,y)) に着目する。1 列目から関係性グループの意味が「(企業のような) 事業主体が事業を行う」であることが分かる。また、3・4 列目から、実際の関係性アノテーションは「事業主体が新規事業に参入する」と明示的に述べたテキストに付与される、と分かる。

2.3 調査テーマ推薦機構

システムが推薦調査テーマを構築する際に利用できる情報として、ユーザーが選んだ調査テーマの履歴や、取得された事例の履歴、また、検索対象となる文書集合の情報（例：文書間の類似度）等が挙げられる。2.1 節で述べた調査フローを支援するには「ユーザーが直近調べていること」に関係する情報が最も重要である。そこで、「直前の調査テーマ」と「ユーザーが注目中の事例」から推薦クエリを構築する。

本システムでは、調査テーマと、関係性検索のクエリとが一对一で紐付けられている。例えば「日立が手がける事業を調べる」という調査テーマは、enter(日立, ●) というクエリに対応する（●は、語句の指定が無いことを表す。）。

「直前の調査クエリ」と「ユーザーが注目中の事例」を入力とし、「推薦クエリ」を出力とする遷移テーブルを表 2 に作成した。各行が 1 つの遷移ルールを表す。ユーザーが現在調査中のクエリ（1 列目）と、ユーザーが興味を持ったエンティティ（2 列目）から、推薦クエリ（3 列目）を導出する。

遷移ルールは、2.2.2 項で定義した関係性グループの間に定義されている。粒度の大きな関係性グループ間の遷移という形で予備調査の大域的な方向を定めつつ、より粒度の小さい単一観点での関係性検索によって有用な事例情報を提供することができる。また、遷移ルールを変更することなしに、新たな観点

表 1: ビジネスドメイン 関係性アノテーション.

関係性グループ 略記	関係性グループ 意味	関係性アノテーション 略記	関係性アノテーション 付与条件	エンティティ
SERVE(x,y)	事業主体が、事業を行う。	serve(x,y)	x が y という事業を営んでいる。	x: 事業主体 y: 事業分野
		supply(x,y)	x が y を実際に売る・受注する。 実際に市場で取引がなされている。	x: 事業主体 y: 事業分野
		enter(x,y)	x が y という新しい事業を立ち上げる、 あるいは、y から撤退する。	x: 事業主体 y: 事業分野
COLLABORATE(x,y)	事業主体と、企業が 協力する。	collaborate(x,y)	x と y が協業関係にある。	x: 事業主体 y: 事業主体
		agree(x,y)	x と y がビジネスに関して合意する。	x: 事業主体 y: 事業主体
		M_and_A(x,y)	x が y を買収・売却する。	x: 事業主体 y: 事業主体
PROMOTE_DOMAIN(x,y)	何らかの原因が、事業分野への 追い風となる	demand(x,y)	x が y に需要を持つ。	x: 消費者 y: 事業分野
SUPPRESS_DOMAIN(x,y)	何らかの原因が、事業の 向かい風となる	regulate(x,y)	x が y を規制する。x が y を解禁する。	x: 規制の原因。 y: 事業分野
SUPPRESS_COMPANY(x,y)	何らかの原因が、事業主体への 向かい風となる	sanction(x,y)	x が y に対して、調査・警告・制裁等を行う。	x: 当局機関 y: 事業主体

を追加することも可能となる。

遷移ルール適用に係るフローを図 2 に示す。

1. ユーザが「日立が参入した事業を調べる (enter(日立,●))」というクエリで事例を取得する (図中①)。
2. ユーザは、最も興味のある事例 (「日立は太陽光発電に進出する」(enter(日立, 太陽光発電))) を選択する (図中②)。ユーザの興味は「太陽光発電」エンティティに向いている。
3. システムが次のクエリを推薦する。まず、enter(日立, 太陽光発電) を関係性グループで表すと SERVE(日立, 太陽光発電) である (図中③)。
4. 次に、適用可能な遷移ルールである表 2 中の 2 行目に従い、PROMOTE_DOMAIN(●, 太陽光発電) を導出する (図中④)。(SUPPRESS_DOMAIN も同様。)
5. 最後に、PROMOTE_DOMAIN グループの関係性である demand を用いて、「太陽光発電のニーズを調べますか？」demand(●, 太陽光発電) を推薦する (図中⑤)。(SUPPRESS_DOMAIN も同様。)

3. 評価

3.1 システム構築

今回のシステムを構築するにあたり、関係性検索の対象となるテキストデータとして、新聞記事 230 万件^{*1} と、RSS で取得した Web ニュース 60 万件を用いた。

これらのテキストデータに対して、2.2.2 項で述べたビジネスドメイン関係性アノテーション 9 種を付与した。計算機資源として CPU 30 コア (60 スレッド) を用いて、およそ 4 日間の実行時間であった。なお、この処理は事前にバッチ実行できるので、実運用上の問題は生じない。

検索エンジンは Apache Solr^{*2} を用いた。

3.2 事例検索の精度評価

3.2.1 評価手順

本システムのような検索システムを評価する上で重要な指標として、応答速度等のユーザビリティ、検索の再現率・精度などが挙げられる。

システムの応答速度に関して、まず、関連事例検索の速度は、Apache Solr の速度に準じる。次に、調査テーマ推薦では、

ユーザが選択した事例文へのアノテーション付与と遷移テーブルの参照のみを行うため、高速である。以上により、本システムの応答速度は十分に担保されていることが分かる。

次に、今回は網羅的な調査が難しいため、再現率の計測実験を行わなかった。再現率に関しては、[佐藤 16] が参考になる。佐藤らによれば、影響関係として同一な事例を述べた文のうち、構文パターン・述語辞書によって取得できる文は、全体の 8 % であった。

本システムを評価する上で重要な指標となるのは、図 1 中④での事例検索の精度である。これまでに述べてきたように、事例検索では、単一の観点に合致した事例を精度良く検索できる必要がある。更に、クエリ推薦では事例中のエンティティの情報を利用するので、事例が検索されるだけでなく、エンティティも精度良く特定される必要がある。

事例検索のクエリでは、エンティティのうちの片方が特定されており、もう片方が検索対象となる (例「日立が参入した事業を調べる」enter(日立,●))。

そこで、エンティティを表す語句を複数サンプリングし、精度評価を行った：

$$\frac{1}{|S|} \sum_{w \in S} \text{Precision}(\text{enter}(w, \bullet)) \quad (1)$$

S はサンプル語句の集合、 $|S|$ は集合の要素数を表す。

エンティティが事業主体の場合、企業名を日経一部上場企業から 20 社名を語句として選んだ。エンティティが事業分野の場合、日経一部のビジネスドメイン分類枠 (計 35 種類) を利用し、分類枠に対して評価者が想起した語句合計 20 種類を使用した (例：「通信」-「スマートフォン」)。

enter(s,●) の場合の評価手順例を示す：

1. 事業主体エンティティとして「日立」を採用する。
2. enter (日立, ●) というクエリで事例を検索する。「日立は太陽光発電に進出。」といった事例が取得される。
3. 各事例に対して、その事例が表 1 記載の関係性アノテーション付与条件と、エンティティ条件を満たすかを確認する。今回の条件は「事業主体が新規事業に参入する」なので、満たす。

ベースラインとして、影響関係アノテーションも評価した。

3.2.2 結果

表 3 に結果を示す。表中の各列が、1 つの事例検索タスクを表す。例えば enter(s,●) は、enter という関係性に関する事例検索タスクであり、事業主体 s をサンプリング語句で固定した上で、事業ドメインを検索するタスクである。

*1 毎日新聞記事データ集 91~2015 年版

http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html

*2 http://lucene.apache.org/solr/

表 2: 調査テーマ推薦のための、関係性クエリ間の遷移テーブル. ● は、検索対象の語句であることを表す.

現在調査中のクエリ	ユーザ興味エンティティ	推薦 関係性クエリ
SERVE(x,y)	x	COLLABORATE(x,●), SUPPRESS_COMPANY(●,x)
SERVE(x,y)	y	PROMOTE_DOMAIN(●,y), SUPPRESS_DOMAIN(●,y)
COLLABORATE(x,y)	x	SERVE(x,●)
COLLABORATE(x,y)	y	SERVE(y,●)
PROMOTE_DOMAIN(x,y)	y	SUPPRESS_DOMAIN(●,y), SERVE(●,y)
SUPPRESS_DOMAIN(x,y)	y	PROMOTE_DOMAIN(●,y), SERVE(●,y)

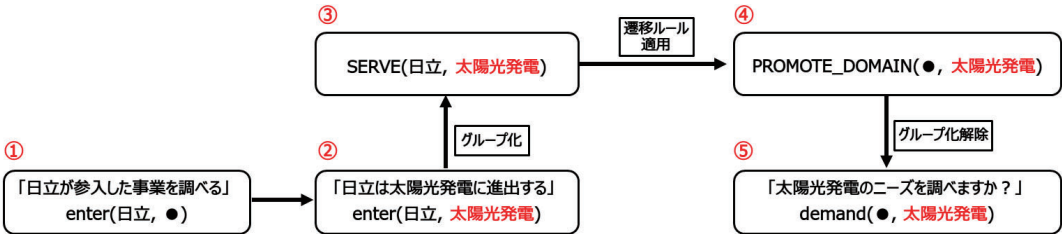


図 2: 遷移ルール適用に関するフロー.

表 3: 事例検索の精度評価. 各列が 1 つの検索タスクを表す. enter(s,●) は, s をサンプリング語句とするタスクである.

事例検索タスク	serve(s,●)	supply(s,●)	enter(s,●)	serve(●,s)	supply(●,s)	enter(●,s)
影響アノテーション (ベースライン)	43%	3%	51%	41%	14%	24%
ビジネスドメインアノテーション	47%	88%	56%	50%	61%	65%
collaborate(●,s)	agree(●,s)	M_and_A(●,s)	demand(●,s)	regulate(●,s)	sanction(●,s)	
15%	0%	0%	8%	3%	4%	
83%	50%	50%	31%	53%	48%	

全ての事例検索タスクにおいて、ビジネスドメイン関係性アノテーションはベースラインよりも、精度良く事例検索を行えることが分かった。

一方、精度は概ね 5 割程度であるので、改善の余地がある。そこで、偽陽性事例の解析を行い課題を精査した。特に精度が低い 2 つのタスクに関して以下に述べる。

demand(●,c) の事例検索精度は 31 % と最も低かった。例えば、demand(●, 半導体) クエリに関して「..鴻海はあらゆるモノがネットにつながる「IoT」の普及をにらみ、半導体メモリーの下需要が急増するとみて、..」*3 という事例が取得されていた。このように、需要者への言及が無い文が多かった。おそらく、事業の需要を述べる際には「特定のエンティティが事業分野を需要する」と直接述べることはせず、「ビジネス状況の変化（新規技術の登場、消費者趣向の変化、等）」という理由に応じて、商材の需要が変化する」と述べることが多いからであろう。上記の文は「IoT が半導体需要を押し上げる」と理解できるので、「IoT のニーズを調べますか？」(demand(●, IoT)) といったクエリを次に推薦すべきである。そのためには、「IoT」と「半導体」の関係性のように、明示的に述べられていない関係性を推論する必要がある。

serve(s,●) では、例えば「ヤフーは昨年 10 月、既卒者も含めた通年採用を始めた。」*4 というように、企業のアクションを述べた文が 5 割以上を占めた。これらの構文・述語は、「事業主体が事業をやっている」と述べた文と共通している。よって、エンティティに入る語句から文の意味を判定する必要がある。

4. おわりに

ビジネス上の意思決定支援システムを開発した。「会社が事業に参入する」というような“ビジネス上の観点”に基づいて関連事例を検索する機構を構築した。また、「日立が太陽光発電に参入する」事例に対して「太陽光発電のニーズを調べますか？」といった新規調査テーマを推薦する機構を作成した。両機構の中核となる、ビジネスドメインの関係性抽出ルールを作成した。これを用いた事例検索精度計測により、効果を確認した。

今後は、直接的な言及が無い語句間の関係を推論する手法や、構文構造だけでなく語句のタイプ（会社、事業領域、等）から文全体の意味の判定をする手法の開発を目指す。また、ユーザの現在までの調査状況を総合的に踏まえた上でクエリを推薦する技術の開発に、取り組む。

参考文献

[Sato 15] Sato, M., Yanai, K., Miyoshi, T., Yanase, T., Iwayama, M., Sun, Q., and Niwa, Y.: End-to-end Argument Generation System in Debating, in *Proceedings of ACL-IJCNLP 2015* (2015)

[Yanai 17] Yanai, K., Sato, M., Yanase, T., Kurotsuchi, K., Koreeda, Y., and Niwa, Y.: StruAP: A Tool for Bundling Linguistic Trees through Structure-based Abstract Pattern, in *Proceedings of EMNLP 2017* (2017)

[佐藤 16] 佐藤 美沙, 柳井 孝介, 柳瀬 利彦, 三好 利昇, 丹羽 芳樹: ディベート人工知能における影響関係認識のためのテキスト内の論理構造に関する考察, 人工知能学会全国大会 (2016)

*3 日本経済新聞 2017/04/20
*4 日本経済新聞 2017/05/12