

DBpedia データセットを用いたステークホルダー表出とシナリオにおける関係推定

Externalization of Stakeholder Using DBpedia dataset and Estimation of Relationship in Scenarios

早矢仕晃章 大澤幸生

Teruaki Hayashi, Yukio Ohsawa

東京大学大学院 工学系研究科 システム創成学専攻

Department of Systems Innovation, School of Engineering, The University of Tokyo

Abstract: The potential expectation of companies has been increased for creating businesses by combining data from different domains, organizations, and sections. It is important to consider who stakeholders are and how they are involved in new businesses. However, it is difficult to create a business scenario by taking account of all the stakeholders in various domains, because the combination of stakeholders and their relationship with scenarios depend on the context of scenarios and have enormous patterns. In this paper, we propose a stakeholder recommender system for supporting scenario generation of data utilization. We implement a system for externalizing relevant stakeholders and estimating stakeholders' relationship with scenarios considering the context of scenarios, using DBpedia and scenarios of data utilization scenario generated in Action Planning as knowledge base.

1. はじめに

シナリオとは、データや知識から導かれた情報を元に、将来起こり得る事象を系列化したものである。意思決定者は創出されたシナリオを読み解き、理解することで実行におけるリスクを低減することができる。本論文におけるシナリオの創出は、すでに知識として確立した行動を策定するのではなく、まだ知識として確立していない行動、特にデータ利活用による新事業の創出におけるシナリオを対象としている。近年、異なる分野のデータ交換や保有データの利活用に対する期待は高まっている。しかし、データ利活用による新事業創出を目的としたビジネスでは、データの分析、欲しいデータの入手方法は知識として確立していないのが現状である。さらに、データ分析の結果などを実際のビジネスに活かすためには、各段階で様々な障壁が存在しており、特に、自社内を含むデータ利活用に関わる様々なステークホルダーの検討の重要性が指摘されている[1]。しかし、膨大な情報から問題解決に必要な知識を発見することが困難であるように[2]、複数の領域にまたがって存在する全てのステークホルダーを考慮することは難しい。

また、ビジネス環境の複雑化に伴い、マネジメン

トする対象のステークホルダーの複雑性は増している。そのため、意思決定者は異なる価値観や関心を持つ多様なステークホルダーとの関係、つまり事業シナリオへの関わり方を明確にすることが必要であると考えられる。しかし、ステークホルダーの関わり方は、ビジネスの文脈によって異なる。本論文における文脈とは、シナリオの内容及びシナリオが実現される状況を意味する。例えば、「街路灯情報を地図上にマッピングし、夜間に明るいルートを提案するアプリケーションの開発」という事業シナリオがあるとする。この文脈において、住民はシナリオ実現に協力的な立場を取るであろう。しかし、アプリケーションによって暗い地域が明らかとなり、そのエリアの夜間の犯罪発生件数が増加するという懸念から、夜間に暗い地域に生活する住民はシナリオ実現に反対する可能性が考えられる。一方、「街路灯情報から夜間に暗い地域を明らかにし、新たな街路灯設置エリアを発見する」という事業シナリオの場合、暗い地域の住民は協力的な立場になると考えられる。このように、同じステークホルダーでも事業シナリオによって、ステークホルダーのシナリオへの関わり方は異なる。つまり、ステークホルダーのシナリオへの関係はシナリオの文脈に依存する。

以上のように、新事業創出において、ステークホ

ルダーが誰であり、シナリオにどのように関わってくるのかということは、シナリオ実現において検討すべき重要な課題である。しかし、異なる分野を横断したデータ利活用に関わるステークホルダーとその関係についての組み合わせは膨大となり、全てのステークホルダーを考慮して実現し得るシナリオを策定することは困難である。そこで、人間のシナリオ創出の支援として、新事業に関わるステークホルダーを推薦するシステムが必要であると考えられる。

本論文では、データ利活用による新事業創出という既存の知識では解決が困難な問題に対して、ステークホルダー推薦システムを提案する。知識ベースとしてDBpedia[3]とデータ利活用シナリオのRDFストア[4]を用い、シナリオに関連すると考えられるステークホルダーについての情報を表出化し、シナリオ生成時に気付きえない潜在的に関係のあるステークホルダーを推薦し、シナリオにおける関係を推定するシステムを実装する。

2. ステークホルダー表出と推定

2.1 シナリオとステークホルダーの関係性

今までに人間の発言や行動から人間関係を推定する研究[5]や、論文の共著関係から人間の繋がりを可視化する研究[6]は行われてきた。しかし、文脈に応じて変わり得るステークホルダーの事業への関わり方についての研究はまだ十分行われていない。ここでいう関係とは、シナリオとステークホルダーの関係であり、ステークホルダー間関係を意味しているのではないことに注意してほしい。ステークホルダーという語は「利害関係者」という意味で用いられることが多いが、本論文ではFreemanの定義である「組織の目的達成に影響するまたは影響を受ける集団あるいは個人」[7]に基づき、シナリオ実現に関与する集団または個人と定義する。前章で述べたように、ステークホルダー及びそのシナリオへの関わり方は、シナリオ自体の文脈に依存する。つまり、同じステークホルダーであっても、シナリオによって協力者や反対者になり得るということである。

以上の議論を踏まえ、シナリオを*scenario*、ステークホルダーを*stakeholder*、関係を表す述語を*relationship*と表現すると、あるシナリオにおけるステークホルダーの関係は述語で*relationship(scenario, stakeholder)*と記述することができる。述語である*relationship*としてターゲット(*target*)や協力者(*supporter*)を定義すれば、シナリオにおけるステークホルダーの様々な関わり方を記述することができる。

続いて、データ利活用シナリオ生成手法アクション・プランニングにおいて創出されたシナリオを元にステークホルダーの関係推定のための訓練データを作成する。次節にて、アクション・プランニングの詳細について説明する。

2.2 アクション・プランニング

アクション・プランニング(以下、AP)は、データ市場の活性化とデータ利活用を促進するためのワークショップ *Innovators Marketplace on Data Jackets*[8]において創出されたデータ利活用案を元に、実行を促すシナリオを生成するワークショップ手法である[9, 10]。データ利活用案を実現する上で必要な要素の関係性やリスクを論理的に導き出すことで、意思決定を行う際に生じる盲点を低減させ、実行可能なシナリオの策定を行う手法である。

APには、シナリオ実現におけるステークホルダー(ターゲット、協力者、反対者など)や必要なリソース(分析技術、人的資源、時間配分、資金配分など)を論理的に検討していくシナリオ生成プロセスがあり、以下の3つのステップを設定している。

- 要求分析: データ利活用案から、消費者の要求について考察する。顕在的な要求から、論理的に要求の背景を検討し、潜在的なニーズを導く。
- 要素表出化: ステークホルダー、競合性、実現コスト、データ、実現までの時間や必要なリソースなどの関係性からシナリオ実現に関連する要素を導出する。要素とはシナリオを構成するステークホルダー及びリソースを意味する。
- 要素系列化: 要素表出化で表出した要素を系列化する。要素同士の関連性を時系列や因果関係で結合し、欠けていた要素の存在を明らかにする。

APは集団におけるコミュニケーションにより進行し、成果物としてシートに記述されたシナリオを得る。最近では、筆者らによってデータ利活用シナリオをRDF(*Resource Description Framework*)を用いて構造化し、RDFストアに格納することで知識として再利用する仕組みが提案されている[4]。本論文では、シナリオにおけるステークホルダーの関係推定のために、RDFによって構造化されたシナリオのうち、シナリオ及びステークホルダーとその関係を表す部分を抽出し、訓練データを作成する。

2.3 関係推定のための学習データ作成

シナリオの文脈把握には、各シナリオに含まれる単語の出現頻度や重要語の抽出などが必要である。これらを求めるために、文書のトピックを抽出する方法が考えられるが、APによって創出されるデータ

利活用シナリオは様々な領域の問題を扱っていることから、シナリオ内に含まれる単語は高次元かつスパースである。そのため、トピックを識別するのに十分な情報が含まれておらず、学習が収束しない恐れがある。また、異なる領域の問題を扱うシナリオの場合、一つのシナリオが複数のトピックを有することが考えられる。さらに、シナリオの種類によってトピックとなり得る主要部分の位置は様々である。

そこで本論文では、文脈をシナリオに含まれる単語の集合と仮定して計算を行うものとする。APのシナリオによる訓練データではシナリオに含まれる単語1つが独立して1つの文脈を表す仮定する。例えば、シナリオの中に「観光」と「研究」という単語が含まれているとすると、このシナリオは「観光」と「研究」という異なる2つの文脈を持つということである。以上の仮定により、ある文脈におけるステークホルダーとその関係を定めることができる。

例えば、「観光」の文脈でステークホルダーとして「外国人」がターゲット、「ホテル」が協力者という関係にあるとする。他のシナリオにおいても「観光」という文脈で同様のステークホルダーと関係が発見された場合、「観光」という文脈では「外国人」がターゲットらしい、また「ホテル」は協力者らしいということが学習可能となる。

以上の手続きにより、APによって創出されたデータ利活用シナリオ30件と、各シナリオに含まれる総ステークホルダー数322件とその関係から、文脈・ステークホルダー・関係のデータセットを31,538件作成し、訓練データとした。また、シナリオにおけるステークホルダーの関係は「ターゲット (*target*)」、「協力者 (*supporter*)」、「反対者 (*dissident*)」の3つを抽出した。なお、通常のAPのシナリオでは「協力者」と「反対者」について、それぞれ内部及び外部の視点を入れてステークホルダーを定めている。しかし、内部及び外部の視点はシナリオの文脈よりも、シナリオ作成者の社会的立場が反映されるため、本論文では「内部の協力者」と「外部の協力者」は「協力者」、「内部の反対者」と「外部の反対者」は「反対者」として統一した。

2.4 関係推定の計算

前節で述べたシナリオを訓練データとし、文脈を表す単語間の独立性を仮定することで、ある文脈が与えられたときのステークホルダーの関係をターゲット、協力者、反対者に分類する問題として扱うことができる。

今、あるシナリオ (*scenario*) が与えられたとき、*scenario*は文脈 (*context*) の集合と見なすことができるので、 $scenario = \bigcup_{i=1}^N context_i$ となる。各文脈

$context_i$ を用いて訓練データからステークホルダー (*stakeholder*) とその関係 (*relationship*) を取得する。あるステークホルダー ($stakeholder_j$) に注目したとき、各 $context_i$ における *relationship* (*target*, *supporter*, *dissident*の3種類) の出現頻度を計算することで $stakeholder_j$ の *scenario* における *relationship* を求めることができる。例えば、 $context_i$ において $stakeholder_j$ が *target* である頻度を求める場合は、訓練データである文脈・ステークホルダー・関係のデータセットから、 $context_i \cap stakeholder_j \cap target$ となるデータセットを取得し、取得数を数えることで求める。 $context_i$ において $stakeholder_j$ が *target* である頻度を $freq(target(context_i, stakeholder_j))$ と表すとすると、*scenario* において $stakeholder_j$ が *target* となる頻度は式(1)となる。*target* と同様の方法で *supporter* 及び *dissident* についても出現頻度を計算し、その中で最も出現頻度が高い関係が最終的に *scenario* における $stakeholder_j$ の *relationship* となる。

$$\begin{aligned} & freq(target(scenario, stakeholder_j)) \\ &= \sum_{i=0}^N freq(target(context_i, stakeholder_j)) \end{aligned} \quad (1)$$

3. DBpedia を用いたステークホルダー表出と関係推定方法

3.1 ステークホルダー情報の取得

あるシナリオが与えられたとき、その文脈においてどのステークホルダーが関連があるのか、ということを表出するため、DBpedia Japanese (以降、DBpedia と呼ぶ) の構造化された情報を用いる。DBpedia に含まれる構造化されたステークホルダーに関する情報を用いることで、過去に検討されたシナリオ内のステークホルダーだけでなく、シナリオに関係すると推定されるより多くの分野のステークホルダーを抽出できる。まず、DBpedia からステークホルダー群を抽出した。本論文では、ステークホルダー群として、DBpedia に含まれる職業一覧 (<<http://ja.dbpedia.org/resource/職業一覧>>) からリンク (<<http://dbpedia.org/ontology/wikiPageWikiLink>>) を有するリソースの職業名及び概要を取得した。以上に加え、Wikipedia において「職業」のカテゴリに属するリソースについても同様に取得し、合計 633

件のステークホルダー候補を DBpedia の SPARQL エンドポイントから取得した。

3.2 ステークホルダー表出と関係推定

本節では、新規シナリオの入力から、関連するステークホルダーとその関係を推定し、ユーザーに結果を提示するまでのプロセスについて説明する。事前処理として、DBpedia からステークホルダー情報として、職業リストを取得する。そして、過去のデータ利活用シナリオから、ある文脈におけるステークホルダーとその関係を格納した訓練データを用意する。以降、過去のシナリオから得られたステークホルダーを ST_{ap} 、DBpedia から取得したステークホルダーを ST_{db} と表記する。

- ① ユーザーは新規事業シナリオを提案し、本システムに入力する。システムはシナリオを分から書きし、文脈の集合を得る ($scenario = \bigcup_{i=1}^N context_i$)。
- ② 訓練データから、 $context_i$ における ST_{ap} とその関係 $relationship$ ($target$, $supporter$, $dissident$ の 3 種類) を取得する。
- ③ $context_i$ における ST_{ap} を元に、 ST_{ap} が有する語を含む ST_{db} を DBpedia から取得する。これにより、 $context_i$ における新しいステークホルダー ST_{db} が得られる。
- ④ 続いて、 ST_{db} の $context_i$ における $relationship$ を算出する。 ST_{db} は $context_i$ における ST_{ap} から取得したものであるため、 $context_i$ において ST_{ap} と同じ $relationship$ を ST_{db} に付与する。
- ⑤ 各文脈において ST_{db} のターゲット、協力者、反対者の出現頻度を数える。
- ⑥ 出現頻度が最大となる $relationship$ が $scenario$ における ST_{db} の $relationship$ として出力される。

4. 考察

4.1 ステークホルダー表出と関係推定の例

ユーザーが新しい事業シナリオを入力し、本提案システムによってステークホルダーとその関係を取得した例を二つ示す。一つは「外国人観光客に対して Twitter のつぶやき情報から隠れた観光スポットを抽出し、推薦するアプリケーションの開発」、もう一つは「電力料金の値上げに伴う料金最適化プラン」という事業シナリオである。表 1 及び表 2 はそれぞれのシナリオを提案システムに入力し、3.2 節の推定プロセスから得た出力である。表中の関係尤度は、式 (2) から求めた。ただし、 $rel(sc, st_j)$ は

$relationship(scenario, stakeholder_j)$ を表す。

$$\text{関係尤度} = \frac{\max \left(freq \left(rel(sc, st_j) \right) \right)}{\sum freq \left(rel(sc, st_j) \right)} \quad (2)$$

表 1 「外国人観光客に対して Twitter のつぶやき情報から隠れた観光スポットを抽出し、推薦するアプリケーションの開発」シナリオのステークホルダーとその関係

ステークホルダー	関係	関係尤度
音響監督	ターゲット	1.00
技術者	ターゲット	0.93
旅行作家	ターゲット	0.89
ネイリスト	ターゲット	0.62
入国審査官	ターゲット	0.54
海事代理士	ターゲット	0.47
通訳案内士	ターゲット	0.42
ランドスケープ コンサルタント	協力者	0.76
造園コンサルタント	協力者	0.70
ホステス	協力者	0.57
観光コンサルタント	協力者	0.64
バスガイド	協力者	0.55
舞妓	協力者	0.44

表 2 「電力料金の値上げに伴う料金最適化プラン」シナリオのステークホルダーとその関係

ステークホルダー	関係	関係尤度
通信士	ターゲット	0.50
造園コンサルタント	協力者	0.59
技術コンサルタント	反対者	0.80
アクチュアリー	反対者	0.80
PA エンジニア	反対者	0.71
電気工事士	反対者	0.67
公共政策 コンサルタント	反対者	0.56

4.2 提示結果についての考察

表 1 に表されるシナリオの主要な文脈は、「外国人」、「観光客」、「Twitter」、「観光スポット」、「アプリケーション」である。一方、表 2 のシナリオにおける主要な文脈は、「電気」、「料金」、「値上げ」となる。得られた結果は、それぞれシナリオの文脈において、概ね関連があると直観的に理解できるものとなった。しかし、音響監督、ネイリスト、ホステス、

造園コンサルタントが表出していることに違和感があるかもしれない。以下、これらの表出理由について考察を行う。

音響監督は、過去のシナリオの「外国人」という文脈において、 ST_{ap} に含まれる「アニメ」及び「ゲーム」という単語から表出したステークホルダーである。日本のアニメやゲームは海外の一部で人気があるコンテンツであるため、外国人をターゲットとしたアニメコンテンツを展開するという過去のシナリオから学習し、DBpedia内の職業リストから取得したものである。ネイリストは、「外国人」の文脈で、「日本」、「サロン」、「デザイン学校」という語からターゲットとして推定されたステークホルダーである。ホステスについては、DBpediaの概要情報に「飲食店」及び「旅館」という語が含まれており、これらの単語から事業シナリオに協力的なステークホルダーとして表出したものである。

また、両方のシナリオに表出した造園コンサルタントについては、それぞれ異なる文脈から表出し、協力者として推定されたステークホルダーである。表1のシナリオでは、文脈を「外国人」とし、「観光業者」及び「技術者」という ST_{ap} から造園コンサルタントが関連がある ST_{ab} だと判定された。一方、表2のシナリオでは、過去に検討された「電気」に関する文脈から、「エンジニア」及び「メーカー」を ST_{ab} として造園コンサルタントが協力者として表出した。以上のように、過去に検討されたシナリオを学習することで、新たな事業シナリオのステークホルダー及び関係が提示可能となる。

5. むすび

本稿では、データ利活用による新規事業創出のためのシナリオ生成支援手法として、ステークホルダー表出と関係推定システムを提案した。現在アクション・プランニングによるデータ利活用シナリオ生成の支援アプリケーション Resource Finder に本論文の提案システムを実装し、本提案システムの性能評価と推薦結果の有用性検証の実験を行っている。

本手法では、文脈抽出の複雑性を回避するため、シナリオ内の単語の独立性を仮定して文脈を定めた。そのため、「旅行」と「観光」のようにビジネスにおいて結びつきの強いと考えられる単語を異なる文脈として計算している。今後の課題として、共起性の高い単語を考慮した文脈把握方法を検討する必要があると考えられる。

また、本論文ではステークホルダーに着目し、職業のみを推薦対象としたが、データ利活用シナリオにはステークホルダーだけでなく、データ、分析手

法など他の様々な要素が含まれている。これらの要素についても、本システムを応用し過去のシナリオから学習させることで、ユーザーに推薦する仕組みを構築できると考えている。

謝辞

本研究は JST-CREST の一部です。また、本研究を支援してくださった構造計画研究所 (KKE) の皆様には感謝申し上げます。

参考文献

- [1] 経済産業省: 平成 26 年度経済産業省委託事業, 我が国経済社会の情報化・サービス化に係る基盤整備(データ駆動型イノベーション創出に関する調査事業) 調査報告書, (2015)
- [2] Simon, H.A.: A Behavioral Model of Rational Choice, *The Quarterly Journal of Economics*, Vol. 69, pp. 99-118, (1955)
- [3] Auer, S., Bizer, C., Kobilarov, G.: DBpedia: a Nucleus for a Web of Open Data, *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference*, (2007)
- [4] 早矢仕晃章, 大澤幸生: データ市場創造のための知識再利用と実行動を促すシナリオ生成手法の提案, *情報処理学会第 77 回全国大会*, pp. 39-40, (2015)
- [5] 神田崇行, 石黒浩: 対話型ヒューマノイドロボットからの日常生活の中の友達関係の推定, *情報処理, Vol.41, No.6*, pp.1000-1006, (2000)
- [6] 松尾豊, 友部博教, 橋田浩一, 中島秀之, 石塚満: Web の情報からの人間関係ネットワークの抽出, *人工知能, Vol.20, No.1*, pp.46-56, (2005)
- [7] Freeman, R.: *Strategic Management: A Stakeholder Approach*, Pitman, (1984)
- [8] Ohsawa, Y., Kido, H., Hayashi, T., Liu, C., and Komoda, K.: *Innovators Marketplace on Data Jackets, for Valuating, Sharing, and Synthesizing Data, Knowledge-based Information Systems in Practice, Springer-Verlag, Vol.30*, pp.83-97, (2015)
- [9] Hayashi, T., Ohsawa, Y.: *Processing Combinatorial Thinking: Innovators Marketplace as Role-based Game Plus Action Planning, International Journal of Knowledge and Systems Science, Vol.4, No.3*, pp.14-38, (2013)
- [10] Hayashi, T., Ohsawa, Y.: *Relationship between Externalized Knowledge and Evaluation in the Process of Creating Strategic Scenarios, Open Journal of Information Systems, Vol.2(1)*, pp.29-40, (2015)