

データ活用戦略創造支援とデータ価値評価手法の提案

A Method for Supporting Creation of Data Use-scenarios and Valuation of Data

劉 暢
Chang Liu

大澤幸生
Yukio Ohsawa

東京大学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

We introduced Innovators' Marketplace on Data Jackets (IMDJ) to create scenarios of data use. Innovators' Marketplace (IM) is a creative support process on the metaphor of a marketplace and Data Jacket (DJ) is a meta-data format. IMDJ is a place where goals of data analysis or scenarios of data use are created through intense communication of stakeholders and also proper data and analysis methods for expected outcomes are specified without exposing the full contents of data sets. In this paper we propose a method to enhance the scenarios made in IMDJ with emphasis on specifying proper valuables and analysis methods to use

1. はじめに

情報技術の発展により、日々取り扱われる情報量は膨大になり、ビッグデータに一部で注目が集まる一方、扱いやすい小さなデータの巨大なネットワークとしてこれを捉え、必要なデータを必要な人間が活用できる仕組み作りが重要不可欠であると考えられる。小さなデータとデータ、あるいはデータ内の変数と変数に着目し、膨大な情報の中から適切な情報を入手して組合せ、的確な問題解決並びに意思決定に生かすことが望ましい。

近年、Linked Open Data[1]のように、公開データを整理してデータのウェブを作る動きがある一方で、一般に公開はされないものの、有用なデータ(医療検査結果や経済状況等)が交渉によって利用可能になれば、データの活用が大いに広がると期待できる。そこで、市場の原理によって利用者が必要なデータを選び、所有者と交渉の末に入手できるような「データの市場」の発展が期待される。

データの活用シナリオの創出支援技法である Innovators' Marketplace on Data Jackets[2,3]によって創出された活用シナリオは参加者や専門家から高い評価を得るものがある一方で、提案されたシナリオでは、データの分析やシミュレーション、その他活用方法に関して具体的に示されない場合があることが課題となっている。具体的な分析手法や用いる変数に関する十分な議論なしには、データの価値評価を行うことも不可能である。本論文では、IMDJにて、要求に応えるためのシナリオとして提案され、評価されたものを、変数に着目して再構築し、より具体的な活用シナリオへと仕上げるための手法を提案する。

2. 関連研究

2.1 データジャケット

Data Jacket(s)[2,3](以下、DJ)は、データの所有者がデータの題名、概要、共有姿勢に加えてデータ中の変数のみを書き出した、データのダイジェストである。コンセプトとしては、CDやDJのジャケットのように、ユーザが概要を手にとって吟味し、中身に価値があるか、購入すべきかどうかを検討していく。データの所有者は、自らのデータに関してDJを書いて公開することで、

自らのデータがどのように活用できるのかを市場に吟味してもらう一方で、データの中身は実際に取引が成立するまでは秘密にしておくことができる。

2.2 データ上のイノベーションゲーム

Innovators' Marketplace on Data Jackets(以下、IMDJ)は、DJを用いた、多様なデータの価値発見と市場創成のための技法である。テーマに基づいて集められた情報の繋がりを可視化するアルゴリズム KeyGraph[4]を用いてマップを作成し、市場を模したワークショップにて、抽出されたキーワードの関係に人間が解釈を与え、議論を行うことにより、「意思決定において重要な判断材料となる事象、状況またはそれらについての情報」を発見[5]していきながら、要求を満たす提案を創出していく。IMDJは、主に以下の4段階からなる。

- ① データ所有者が概要説明と変数群を記入した DJ を作成し公開する。
- ② DJ 同士を、共通する変数や単語を介して結びつけたマップを KeyGraph で作成する。
- ③ ワorkshop形式で、参加者が要求を出し、他の参加者が DJ と DJ を組み合わせて要求に応えるようなデータ活用シナリオの提案を提案する。市場を模したゲーム内のシナリオの値付けによってデータの利用法と評価価値を決めていく。また、シナリオがよりよく要求を満たすものになるように、要求者と提案者の間で議論が行われる。
- ④ ユーザは自分にとって価値のあるデータの提供を求め、データの所有者は条件を吟味して、取引が行われる。

3. 提案手法

要求に応えるために活用シナリオを創出する際、①データの題名と概要を用いてシナリオの概要を構想する段階と、②データに含まれる変数にまで留意し、どういった変数同士を、何を結びけとして、どういった分析手法によって分析をするかを特定する段階の、二つの段階が考えられる。その橋渡しとして、図1のようなマップを提案する。図の中心に提案シナリオが位置し、その周りを使用データが囲む。また、データに含まれる変数が階層構造を為しながら、周囲を円形に囲んでいる。変数とその階層構造はDJに記されているものを用いている。

このマップを用いて、IMDJで提案されたデータの組合せについて、変数にまで着目した、より具体的で論理的なデータ活

連絡先: 劉暢, 東京大学工学系研究科システム創成学専攻大澤研究室, puerta.japan@gmail.com

This research was supported by JST, CREST.

用シナリオを創出する。ユーザには、データに含まれる変数と変数同士の紐づけに改めて着目することによって、提案されたシナリオの再構築を行ってもらうことで、実行可能なシナリオへと仕上げてもらおう。

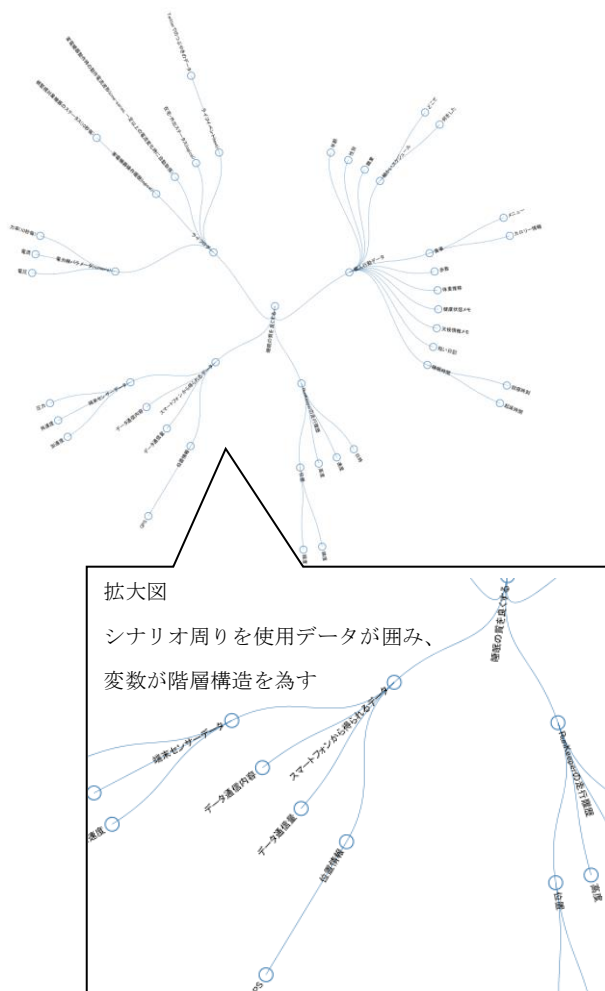


図1 可視化マップ

4. 実験と考察

4.1 実験概要

被験者は7名(うち大学生2名, 大学院生5名)であった。大学関連就職関連, 交通関連, エネルギー関連, 気象関連, 医療関連, 災害関連, その他, 計44個のDJを用いた。これらのDJは全てデータの所有者が実際に記入したものである。

- ① IMDJの手順に従い、「大学生活をより良くするために」というテーマの下、被験者に単独で要求とそれを満たすデータ活用シナリオを提案してもらった。(10分程度)
- ② 被験者の提案に沿って、シナリオを構成するデータ(DJ)とそのデータに含まれる変数(DJに記述されているもの)を階層的に可視化した図を用意した。
- ③ その図を見て、データに含まれる変数と変数の結びつき、紐づけ、分析方法、分析によって得られる知見を吟味しながら、シナリオを再構築してもらった。(30分程度)

4.2 実験結果

被験者には、図2のように、可視化マップへと直接書き込みをしてもらった。シナリオを再構築する作業の中で、被験者全員

が、提案シナリオで用いられたデータに含まれている変数を、シナリオに必要な変数と不必要な変数に分類した。また、変数に着目することで、具体的な分析方法を提案することや、変数の対応(紐づけ)に言及することが多く見られた。また、もとのシナリオの不足点に気づき、新たなデータあるいは変数を追加したり、データそのものや、分析事例、分析手法に関して追加情報への需要を挙げたりしていた。

(1) シナリオの整理・再構築

全ての被験者は可視化マップを用いて、データに含まれる変数に着目し、自らが提案したシナリオの再構築することで、どのデータの変数からどのような知見を得るのかをより明確にした。ある被験者の例を示すと、この被験者のもとの提案シナリオは「学生が低学年のうちから就職を意識できるようにするために、『入学時の学生全数アンケート』、『学生の就職データ』、『テレビ番組視聴履歴』、『進路(進学学科)に関する意識調査』のデータを用いて、身近な授業やテレビ番組を見ていた人の就職先を表示し、モチベーションを高める。」というものであった。これに対し、再構築後のシナリオは、「『進路(進学学科)に関する意識調査』のデータに含まれる進学学科の名称から、希望者数、内申点から学科の人気度を割出し、インターネットにある企業別就職難易度偏差値と掛け合わせて、学科の人気度と就職先の難易度の相関を知る。また、『入学時の学生全数アンケート』に含まれる進路希望に関する事項と組み合わせれば、入学当時の進路希望と実際の進学先、就職先の対応関係を見ることが出来る。『学生の就職データ』に含まれる就職先名称と合わせて、学生の履修講義や単位取得講義のデータを追加することで、講義と就職先の関連を見る。また、『テレビ番組視聴履歴』のデータが、個人を識別可能なデータであれば、学生が見ている番組名称と、『学生の就職データ』就職先に関連が分かるかもしれない。」となり、用いる変数と、分析から得ようとする知見がより明らかになった。

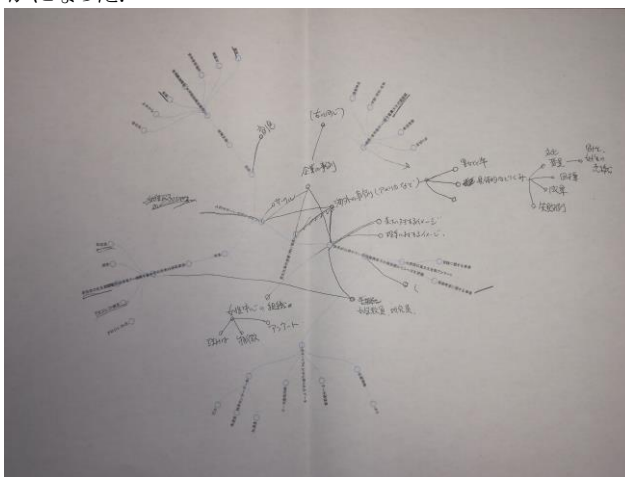


図2 書き込み作業後の可視化マップ(図の中央に書き込まれた線は、異なるデータに含まれる変数と、分析から得たい、あるいは得られると考える知見を結んでいる。図の右上の部分では、追加データとそのデータに含まれる変数が樹形図で書かれているのが見受けられる。)

(2) データの対応・紐づけ

4名の被験者が、データ同士の紐づけに関して新たな気づきがあった。例えば、固体が識別可能なデータであるか、あるいは集計データであるかについてもとのシナリオの提案段階では考慮していなかったと認めた事例が見られた。ある被験者は、「入学時の学生全数アンケート」、「学生の就職データ」、「進路(進

学学科」に関する意識調査」, 「個人の行動に関するデータ」を用いて, 「大学の歳入を増やすための, データバンクとしての大学」というシナリオを提案していたが, 「個人の行動に関するデータ」のみが個人を識別できるデータであり, 他のデータは集計データであることに, 再構築の作業を行って初めて気が付いた。そこで, 当初提案したシナリオを見直し, 集計データと個人データを合わせて活用するように工夫し, シナリオを修正した。修正後のシナリオは, 「学生の将来に関する意識を知るためのもの」として, 『入学時の学生全数アンケート』に含まれる進路希望に関する事項から, 目標とする学部や職種, 『進路(進学学科)に関する意識調査』に含まれる学部(学科)名称と希望者数から人気の学部学科, また, 『学生の就職データ』に含まれる所属学部, 就職先名称, 人数から学部ごとの特色を抽出し, 卒業年度とも照らし合わせて年ごとの特色を見ることができる。これらの集計データを用いて, 企業はアプローチすべき学部・学科, または学年を把握することができる。一方, 『個人の行動に関するデータ』から, 個別の学生の所属学科や研究室, 就活スケジュールや応募先, 内定先が分かれば, 企業は, 興味を持った学生を惹き付けておく工夫ができる。」というものになった。

(3) 具体的な分析方法

ある被験者は, 当初のシナリオには含まれていなかった, 具体的な分析方法とそれに用いる変数を明確に示した。当初, 「主要地点の交通量」, 「電力基幹システムを考慮した電源構築系統のコスト」, 「ライフログ」のデータを用いて, 節電に役立つ旨のシナリオを提案したが, 具体的な分析方法への言及はなかった。可視化マップを用いた作業の後, 「道路の計画に役立つ基礎資料を参考にすることで, おそらく様々な制約の下での混雑が少なくなるような道路計画をするであろうと予想される。電気製品の電力使用量を道路の混雑度やピーク比率に見立てた分析を行うことができると期待される。また, 『電力基幹システムを考慮した電源構築系統のコスト』に含まれる発電設備の効率やコスト, 負荷追従率等を交通量のデータと組み合わせると, コストを最小化する線形計画法が行えるのではないか。」と提案した。

(4) 具体的な追加データ

ある被験者は, シナリオを提案する当初は, 使えるデータが少なく, シナリオを提案することが難しく感じたという感想を述べていた。この被験者は, 自らの提案したシナリオに, 「海外の事例」, 「企業の事例」, 「育児」などのデータが必要だと認知し, 追加を行った。他の被験者も同様に, 自らが提案したシナリオを再構築していく途中で, 不足点を発見し, 追加データの必要性に気が付く事例が多く見られた。同時に, これらの被験者は具体的にどのようなデータを補うべきかを記していた。

(5) 追加情報の必要性

自らの当初の提案が十分に実行可能であると判断した 1 名の被験者を除き, 他の被験者の全員が, データそのものに関するより詳しい変数のリストや背景知識, 類似データの有無といった追加情報, および分析方法に関する知識, 分析の事例などの情報の必要性に気が付いた。

4.3 考察

被験者が変数に着目して, 提案を再構築することで, 当初の発想では気付かなかった点(変数の適不適, データ同士の紐づけ, 具体的な分析方法, シナリオの補足, 追加データの必要

性など)を発見する効果が見られた。また, データに関する背景知識の不足やデータの更なる詳細, 分析手法や分析事例といった参照情報の不足に被験者が気付いた。データや分析に関する知識を補い合えるような, 複数の参加者によるコミュニケーションの場がこれらの問題点を解決するのに有効であると思われる。

また, 被験者が終始単独で作業をしているため, 他者の評価を受けることや意見を反映することができない。本来の IMDJ では, 複数の参加者が議論する中で, 好意的な評価や建設的な批判が行われる。複数の参加者が議論を通じて提案に潜む様々な制約を発見する技法も研究されている[7]。提案手法においても, 複数の参加者が互いのシナリオを見直していく方法が有効であると思われる。

今後, データの分析シナリオを見直すのに有効な着眼点(変数, 紐づけ, 分析手法など)を予め提示しておくようなフレームワークを整理し, ガイドラインや作業シート等を作成して活用するとういと思われる。

5. 今後の展望

本研究は, IMDJ で提案されたデータ活用シナリオを再構築することを可視化マップとフレームワークによって支援し, 結果としてよりシナリオを論理的にすることを旨とする。

知識の共有・創造・獲得プロセスである野中らの SECI モデル[8]によれば, 組織の人々が共同者とともに言葉にしにくい経験知を共有し(S:共同化), 言語の形に表し(E:表出化), 知識を組み合わせて新たな知識を作り(C:連結化), 新知識を個人尾内面に定着させ(I:内面化), また共同化に向かう。シナリオを論理的に再構築するプロセスにおいて, 複数の参加者で共同作業をする場では, SECI のプロセスを以下のように読み替えることができる。

- ① 表出化:IMDJ での要求獲得ならびにシナリオ提案, 提案手法でのシナリオ内の変数可視化, 変数同士の紐づけ, 分析方法, 追加データの書き出し, および変数間の繋がり背景を表出
- ② 連結化:コミュニケーションによって互いが提案するシナリオを補強し, 変数の扱いや分析方法に関する互いの経験や背景知識を結び付けて, より論理的なシナリオを創出
- ③ 内面化:シナリオに沿って分析を実施, データを実際に利用していく中での, 新たな発見や知識の個人への定着
- ④ 共同化:実際のデータ利用を通じて個人が得た経験知の, 現場あるいは IMDJ の場での共有

このプロセスにおいて, 本稿は, ①表出化の段階を実施と, ②連結化への準備と位置付けることができる。以降の段階について今後の研究で進めていく。

また, 提案手法の開発を進めるとともに, 再構築されたシナリオの評価を行っていく。IMDJ では, ゲーム中の仮想紙幣を用いてシナリオの評価を行うが, 過去 7 回の IMDJ で出された 125 個のシナリオのうち, 論理的シナリオ(具体的な分析に言及していると認められるシナリオ)は 32 個(全体の 25.6%)であり, 一方で, 論理的なシナリオが全体のシナリオに占める割合が高い IMDJ では, ゲーム中のシナリオの売上が活発である傾向があるという知見が得られている。(図 3 参照)

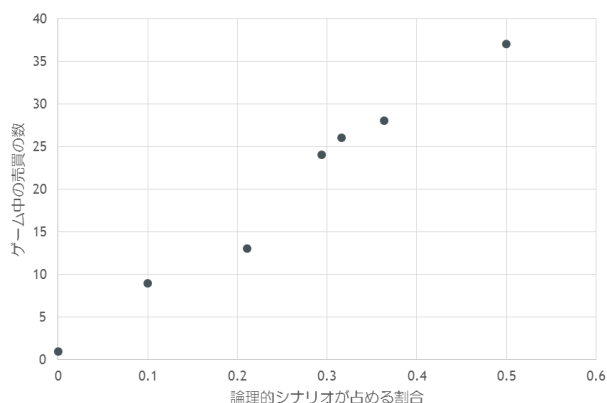


図3 論理的シナリオが占める割合と、ゲーム中の売買の数

また、直近に実施した IMDJ について、論理的シナリオ(6 個)の購入総額は 35(平均 5.83), それ以外のシナリオ(13 個)の購入総額は 22(平均 1.69) という差が見られている。少数ながらも、変数に着目し、具体的な分析手法にまで言及したシナリオは、IMDJ のゲーム中に高い評価を得ている傾向があるという仮説が立てられ、今後この検証も進めていく。

本研究によって得られた知見を、IMDJ ならびに DJ の改良に生かすことも視野に入れる。

参考文献

- [1] Yu, L. 2011. A Linked Open Data, Developer's Guide to the Semantic Web Chapter 11, Springer: 409-466
- [2] Ohsawa, Y., Kido, H., Hayashi, T., and Liu, C., Data Jackets for Synthesizing Values in the Market of Data, International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies, vol.22, pp.709-716 (2013)
- [3] Liu, C., Ohsawa, Y., and Suda, Y., Valuation of Data through Use-Scenarios in Innovators' Marketplace on Data Jackets, 1st Workshop on Market of Data, ICDM 2013
- [4] Ohsawa, Y., Benson, N.E., and Yachida, M., KeyGraph: Automatic Indexing by Co-occurrence Graph based on Building Construction Metaphor, Proc. Advanced Digital Library Conference (IEEE ADL'98), pp.12-18 (1998)
- [5] Ohsawa, Y. and McBurney, P., (2003) 'Chance Discovery', Springer.
- [6] Ohsawa, Y., and Nishihara, Y., (2012) 'Innovators' Marketplace: Using Games to Activate and Train Innovators', Springer-Verlag.
- [7] Ohsawa, Y., Nishihara, Y., Nakamura, J., Kushiro, N., and Nitta, K., Tsugology for revealing intentions and constraints, IEEE Conference on Systems, Man and Cybernetics (SMC2010) pp.1332-1337 (2010)
- [8] Nonaka, I., and Takeuchi, H., The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation, Oxford University Press (1995)