2F1-03

データから物語の生成

Story Generation for Aiding Data Analysis

中田 豊久　　國藤 進
Toyohisa Nakada　Susumu Kunifuji

北陸先端科学技術大学院大学 知識科学研究科
School of Knowledge Science, Japan Advanced Institute of Science and Technology

The purpose of our study is to aid exploratory data analysis by using story. Assuming that one of the problems in data analysis is to cling to a viewpoint, we have developed a system that creates stories from data in order to become aware of other viewpoint. We describe the architecture of the prototype system and discuss results from preliminary experiment.

1. はじめに

研究の目的は，データ分析を支援する仕組みを作ることである．データ分析とは，例えば，顧客情報を分析して売り上げを向上させるマーケティングや，家計簿を見渡して家庭の支出を改善するなどの行為のことである．一般的には統計的手法がこれを支える技術として挙げられる．最大値，最小値，平均値などの算出や，相関関係を見つけ出す事は，この統計の範囲内で可能であるのだが，我々の研究の目的は，この先の支援を目指すものである．

データマイニングで有名な話である，あるスーパーにおいて週末にビールと紙おむつが良く売れる，というルールをデータから導き出した例を使って話をしたいと思う．結果から先に示すと，週末に子供を持つ夫が妻に「紙おむつを買ってきて」とお願いされ，スーパーに行ったついでにビールを購入するという現象がよくあったからである．これに気づいたスーパーの店主は，紙おむつの横にビールを陳列し，紙おむつのみを買って帰るこれまでの顧客にビールを買わせ売り上げを増進したという話である．

統計手法による解析では，紙おむつとビールの売り上げに相関がある，という結果を導き出す事は出来るが，そこから先に示したストーリーを導き出すには，例えば，その対象の顧客を捕まえてアンケートを取るなどの新しいデータが必要になる．このときに重要なのが，仮説を持ってアンケートを行うか，仮説を持たずにアンケートを行うか，の違いである．意識的に現象を見ようとする前者の方が，良い結果をもたらす可能性が高いだろう．この場合では，紙おむつとビールの相関から，紙おむつからビールへの因果へ着目したことによる成功であろう．この視点の想起は，例えば一度紙おむつとビールの相関はただの偶然である，と思ってしまった分析者には簡単に出来るものではない．そこで我々は，システムが自動的に因果の関係を示唆する情報を分析者に提供できないかと考えた．この因果の関係を示すためには，売り上げデータだけではもちろん不可能であり，予め用意された知識ベースが必要になる．この知識ベースと観測データとから仮説生成を支援するストーリーを生成する事が本研究の目的である．この生成されたストーリーは，先に示した仮説そのものになれば尚良いのだが，そうでなくても仮説を想起させる事が出来れば良いと考えている．

連絡先: 中田 豊久, 北陸先端科学技術大学院大学, 知識科学研究科, 923-1292 石川県能美郡辰口町旭台 1-1, TEL 0761-51-1699, FAX 0761-51-1775, Mail t-nakada@jaist.ac.jp

図 1: Rabbit and duck

2. データ分析について

データ分析を以下の 2 つに分類して考える．

- 検証的データ分析
- 探索的データ分析

前者の検証的データ分析は，統計における仮説検定に相当する分析作業のことである．予め分析者自身が持っている仮説を検証するという目的でデータを分析する．後者の探索的データ分析は，明示的な仮説を持たずにデータを分析する作業のことである．データの傾向把握や，予期せぬ知識発見を期待してデータを見渡す．これら 2 つの分析は，明確にその差が分類できるものではない．検証的データ分析の最中であっても予期せぬ発見を生むこともあるし，探索的データ分析の中から仮説が生成され，検証的データ分析に移っていくということは，データ分析において一般的なことであると考えれる．よってデータ分析を支援する事を目的とした場合，両者を支援する事が理想であるが，まず我々は後者の探索的データ分析を支援するという目的に集中して研究を行うこととした．

2.1 探索的データ分析における問題点

図 1 は，だまし絵と呼ばれる絵である．この絵は，右側の 2 本に分かれる物体を耳と見ることでウサギに見え，くちばしと見ることによりアヒルに見える絵である．この絵を最初に見た人は，ウサギかアヒルか先に気づいた方に気をとられ，他方を気づかないということも少なくはない．これが，データ分析における問題点である．仮説を立てるのは，データ分析において重要であるのだが，その仮説が他の視点を阻害してしまっているという問題である．図 1 のようなだまし絵の場合，他者からそのヒントを提示されると，比較的簡単にもう片方の視点に気づく事が多い．このことから，我々のストーリー生成には，分析者へ新たなデータ分析の視点を喚起する効果を期待している．しかしこれは，分析者の現在の頭の中の状態を知るこ

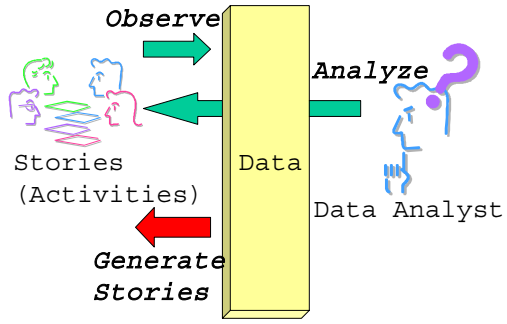

図 2: Data analysis model and story generation

とが困難であることから，完全にこの問題をクリアできるという保障はない．しかしながら，システムが複数のストーリーを生成して複数の視点を分析者に与えることにより，新たな視点への可能性を上げられると考えている．

3. ストーリーによる気づき

ここでは，なぜデータ分析者への提供情報にストーリーという形を取るかを議論する．これは，図 2 に示すデータ分析のモデルに起因する．図 2 は，右にデータ分析者，真中に観測されたデータ，左に観測される事象が示されている．通常，データ分析者は，真中のデータを分析するのであるが，その目的は，そのデータを通して見える，奥に潜むストーリーを知ることである．よって，データ分析者からの分析という矢印は，データで止まらずその先まで突き抜けている．このようにデータ分析を捉えた時に，ストーリー生成は，図 2 の Data から Stories をつなぐ矢印である．複数考えられる Data から Stories への矢印をシステムが提供する事が出来れば，分析者の支援となり得ると考えた．

また，[福田 01] は，想起における提供される情報の抽象度について報告している．抽象度低の物語，抽象度中の諺，抽象度高の 2,3 文節からなる命題の中で，最も正確に想起を可能とするのは，抽象度中の諺であるとしている．ここで言う正確な想起とは，高次の関係構造が一致している物語を想起できるかどうかという意味である（詳細は [福田 01]）．一般的に ”ストーリー ”という言葉は，”物語 ”とは違う．物語の英語は narrative であり，その中の意味の部分がストーリーである．言い換えると物語を，その表現方法を捨て意味の部分のみに抽象化したものがストーリーであると言う事が出来る．このような理由から，我々は，諺ほどではないが，より正確な想起をもたらす可能性の高いストーリーという形をデータ分析の新しい視点形成のために使用出来ると考えた．

4. Web アクセスログからのストーリー生成

初期のプロトタイプシステムとして Web アクセスログの分析を支援するストーリー生成システムを構築した．

4.1 ストーリー作成

ストーリーの記述方法には，[小方 03] にて提案されている図 3 のように枝に動詞的概念（事象概念），葉に物語事象を持つ物語木を使用している．

ストーリー生成のためのアーキテクチャを，図 4 に示す．

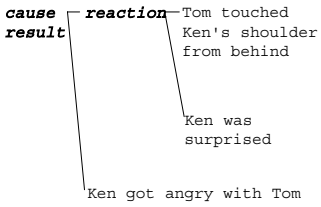

図 3: A sample of story tree

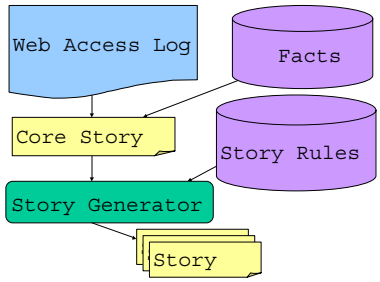

図 4: Architecture of the system that create stories

システムは知識データベースとして 2 つのデータベースを使用する．1 つは Facts と呼ばれ，1 つの閲覧 Web ページと 1 つの物語事象とを対応付けるものであり，もう 1 つは，Story Rules と呼ばれ，物語事象を拡張するルール集となっている．まずシステムは，Web アクセスログから IP アドレス毎の閲覧ページシーケンスを抽出し，Facts データベースを使って，最も原始的な Core Story を生成する．次に，その Core Story を Story Rules の中の適用可能ルールを適用して拡張していく．この際，どのルールを，どこの物語事象に適応するか，という競合解消が問題となるが，本研究では基本的に競合解消を行わない事としている．これにより複数のストーリーが作成される事になる．この方法は，人工知能分野でのプロダクションシステム [大原 88] と等価である．

4.2 実験

著者らの所属する北陸先端科学技術大学院大学の公式ホームページ (http://www.jaist.ac.jp/) のアクセスログからストーリー生成の実験を試みた．紙面の都合上，結果の 1 例のみを表 1 に示す．第一列は，ある Web 閲覧者のアクセスログである．この実験では一連の Web アクセスログを前の Web ページを見てから 30 分以内に次のページを見たときに同一セッションであると定義している．また，ストーリールールの競合解消には，ストーリのどこにルールを適用するかはランダムに 1 つを選択するとし，そこに適用するルールの選択は行わず，適用可能な全てのルールを適用することとした．しかし，現実には多すぎるストーリーが生成されてしまうため，8 個を上限としてストーリー生成を制御することとした．表 1 の場合は，最大の 8 個のストーリーが生成され，その中から最も異なる意味を持っていると本稿の第一著者が判断をした 2 つのストーリーを示している．

4.3 考察

生成された 2 つのストーリーは，Web 閲覧者が知識科学研究科に好印象を持った Story1 と，そうでない Story2 とを想起させる事が出来ると考えられる．Story1 では，最初に知

表 1: Result from preliminary experiment at School of Knowledge Science in JAIST

Web Access Log	Story1	Story2
Tue Feb 24 12:46:01 JST 2004 GET /index-jp.html	継起	継起
Tue Feb 24 12:46:03 JST 2004 GET /index-j.html	原因-結果	原因-結果
Tue Feb 24 12:46:08 JST 2004 GET /ks/index.html	原因-結果	原因-結果
Tue Feb 24 12:46:29 JST 2004 GET /ks/aboutKS/ks1120.files/frame.htm	(探す トップページ)	(探す トップページ)
Tue Feb 24 12:46:30 JST 2004 GET /ks/aboutKS/ks1120.files/outline.htm	(見つける トップページ)	(見つける トップページ)
Tue Feb 24 12:46:30 JST 2004 GET /ks/aboutKS/ks1120.files/slide0020.htm	(見る トップページ)	(見る トップページ)
Tue Feb 24 12:46:30 JST 2004 GET /ks/aboutKS/ks1120.files/slide0020.htm	反応	反応
Tue Feb 24 12:46:31 JST 2004 GET /ks/aboutKS/ks1120.files/outline.htm	(見る 知識科トップページ)	(見る 知識科トップページ)
Tue Feb 24 12:46:31 JST 2004 GET /ks/aboutKS/ks1120.files/outline.htm	(興味を持つ 知識科)	(興味を持つ 知識科)
Tue Feb 24 12:46:50 JST 2004 GET /ks/aboutKS/ks1120.files/slide0270.htm	理由	原因-結果
Tue Feb 24 12:46:54 JST 2004 GET /ks/aboutKS/ks1120.files/slide0270.htm	(知りたい 知識科)	反応
Tue Feb 24 12:47:10 JST 2004 GET /ks/aboutKS/ks1120.files/slide0262.htm	(見る 知識科の紹介 (スライド))	(見る 知識科の紹介 (スライド))
Tue Feb 24 12:47:15 JST 2004 GET /ks/aboutKS/ks1120.files/slide0262.htm	(見る 育成する人材像 (KS))	(分からない 知識科の紹介 (スライド))
Tue Feb 24 12:47:19 JST 2004 GET /ks/aboutKS/ks1120.files/slide0293.htm	反応	(飽きる 知識科の紹介 (スライド))
Tue Feb 24 12:47:24 JST 2004 GET /ks/aboutKS/ks1120.files/slide0293.htm	(見る 知識科 進路状況)	(見る 育成する人材像 (KS))
Tue Feb 24 12:47:34 JST 2004 GET /ks/aboutKS/ks1120.files/slide0294.htm	(思う 就職状況は良い)	反応
Tue Feb 24 12:47:40 JST 2004 GET /ks/aboutKS/ks1120.files/slide0294.htm	反応	(見る 知識科 進路状況)
Tue Feb 24 12:47:45 JST 2004 GET /ks/aboutKS/ks1120.files/slide0263.htm	(見る 知識科トップページ)	(思う 就職状況は悪い)
Tue Feb 24 12:47:49 JST 2004 GET /ks/aboutKS/ks1120.files/slide0263.htm	(興味を持つ 知識科)	原因-結果
Tue Feb 24 12:47:56 JST 2004 GET /ks/aboutKS/ks1120.files/slide0264.htm		反応
Tue Feb 24 12:48:01 JST 2004 GET /ks/aboutKS/ks1120.files/slide0264.htm		(見る 知識科トップページ)
Tue Feb 24 12:48:07 JST 2004 GET /ks/aboutKS/ks1120.files/slide0265.htm		(分からない 知識科)
Tue Feb 24 12:48:10 JST 2004 GET /ks/aboutKS/ks1120.files/slide0266.htm		(飽きる 知識科)
Tue Feb 24 12:48:15 JST 2004 GET /ks/aboutKS/ks1120.files/slide0266.htm		
Tue Feb 24 12:48:31 JST 2004 GET /ks/bg.html		
Tue Feb 24 12:48:31 JST 2004 GET /ks/menu.html		
Tue Feb 24 12:48:31 JST 2004 GET /ks/jinzai.html		
Tue Feb 24 12:49:23 JST 2004 GET /gakusei/guidance/sinro.html		
Tue Feb 24 12:51:19 JST 2004 GET /ks/index.html		

識科学研究科に興味を持ち，その興味を最後まで維持している．一方，Story2 では，同じように最初に知識科学研究科に興味を持つのであるが，最後にはよく分からずに飽きてしまっている．この 2 つのストーリーの内，どちらが現実をよく説明しているかは，このシステムの範囲外である．重要なことは，意味の異なるストーリーを同じログから作成出来たことである．これにより，例えば，片方しか想起しなかった分析者に新たな視点を喚起する事が可能となると考えられる．

5. 関連研究

ストーリー生成には，[小方 03] のように文学の観点からの研究がある．本研究と同じくボトムアップ型のストーリー生成であるが，その目的は，物語理解のために物語を作る機構を作る，というものである．ストーリーを展開するルールには，文学研究で得られた知識が入っているため，全体として意味のある文章になる方向にストーリー生成が行われている．我々のストーリ生成は，想起を目的としているため，全体としてのまとまりよりも生成されるストーリーの多様性に重点を置いている事が相違点である．

また，[坂本 02] では，経験の伝達のために展示場などでの自分の行動履歴からストーリーを生成し，漫画表現を用いて表示してくれる．このストーリー生成法は，テンプレートを用いたトップダウン的方法で，我々のボトムアップの方法とは異なる．これもその目的の違いからであり，我々の提案するストーリ生成法ではボトムアップによる創発効果を期待している．

さらにチャンス発見 [大澤 03] においてもシナリオ創発ワークショップ [シナリオ創発ワークショップ] によってシナリオの重要性が指摘されている．”ストーリー ”ではなく ”シナリオ ”という言葉が示すように，当初は発見したチャンスを生かすシナリオ作り，という未来に対する志向が強かったと思われるが，最近では，チャンス発見のためのシナリオ作り，というチャンス発見前のシナリオの重要性も指摘されている．後者の意味の ”シナリオ ”は，我々が使う ”ストーリー ”と同じである．

6. まとめ

本稿では，探索的データ分析における最も重要な仮説生成を支援するために，データからストーリーを作るということについて議論し，作成したプロトタイプシステムを示し，その機能を紹介した．プロトタイプシステムの出力結果がデータ分析者に有効となる可能性を見る事が出来たが，その評価が現段階では行えていない．今後は，実際のデータ分析作業におけるプロトタイプシステムの有効性を調べていきたいと考えている．

参考文献

[福田 01] 福田 健: 第 5 章 アナロジーと想起, 類似性から見た心, 大西 仁 and 鈴木 宏昭 編著, 共立出版株式会社 (2001)

[小方 03] 小方 孝: 第 5 章 物語の多重性と拡張文学理論の概念, 複雑系社会理論の新地平, 吉田 雅明 編, 専修大学出版局 (2003)

[大原 88] 大原 育夫: 人工知能の基礎知識, 近代科学社 (1988)

[大澤 03] 大澤 幸生: チャンス発見の情報技術, 大澤 幸生 監修・著, 東京電機大学出版局 (2003)

[坂本 02] 坂本 竜基, 角 康之, 中尾 恵子, 間瀬 健二, 國藤 進: コミックダイヤリ：漫画表現を利用した経験や興味の伝達支援, 情報処理学会論文誌, Vol. 43, No. 12 (2002)

[シナリオ創発ワークショップ] http://www.chancediscovery.com/index.html