

(1992. 7. 10)

同一事象を扱う複数の新聞記事を対象とする 質問応答システム

上田修一 神門典子 越塚美加

慶應義塾大学文学部図書館・情報学科

新聞というメディアに掲載される記事を対象として、繰り返して生起する社会事象について分析枠組みを作り、データベース化することにより、様々な質問に回答できるシステムを考案した。事例として「米価決定過程」を取り上げ、これに関して生起する事象を分析し、14種類のイベントに分け、各イベントを表現する要素ファイルに加え、用語、人物ファイルなどを用いて新聞記事をデータベース化した。これにより、人物、組織、用語、審議内容などの質問に答えることが可能となる。

A PROTOTYPE QUESTION-ANSWERING SYSTEM BASED ON ANALYSIS AND SYNTHESIS OF NEWSPAPER TEXT

Shuichi Ueda Noriko Kando Mika Koshizuka

Keio University School of Library and Information Science
2-15-45 Mita, Minato-ku, Tokyo 108

The newspaper is information media that supply news about social phenomenon. Readers of newspaper accept some fragmental facts from the newspaper from day to day and construct a story about an phenomenon. A periodic phenomenon consists of a series of events. We analyzed a typical phenomenon and developed a prototype question-answering system. This system designed to answer question about, and summarize events based on short newspaper text database (element file) and sorts of dictionary files.

1. はじめに

報道を主体とした新聞(newspaper)は、17世紀中葉に西欧で発明され、19世紀後半に欧米で発展し、20世紀には世界各国に普及した情報メディアである。日刊紙は、毎日の刊行、内容的にはニュースの伝達、形態としてはブランクでないштаブロイドで綴じられていないという特色を持っている。また新聞の利用される期間は、極めて短かいが、事件の記録としての価値を持っており、各新聞社は原紙の保存の他、縮刷版、マイクロフィルム、さらには電子媒体を使って蓄積している。

日本経済新聞社は1981年から同紙の新聞記事データベースの提供を始め、その後多数の新聞社が追随してデータベース作成、提供を行い、現在では「日経TELECOM」を通じて提供されている日刊新聞のデータベースは14種に達し国内で最もよく利用される全文データベースとなっている。

個々の新聞記事は単に断片的事実、意見を伝えるものであるが、新聞記事の一つの事象を連続的に一種の物語を伝えるものとしてとらえ、分析・統合操作を加えることによって、従来とは異なった形態の質問回答システムを考えることができる。

2. 社会事象と新聞記事の関係

2.1. 社会事象の繰り返し

多くの社会事象は繰り返し生起すると考えられる。社会事象の中には、あらかじめ定められた周期で定期的に繰り返されると不定期に繰り返されるものがある。例えば、参議院議員選挙や株主総会は前者であり、公定歩合の改定は後者である。衆議院選挙のように、両者の性格を持つものもある。

また、企業合併、火事、火山噴火などは、一見して1回ずつ独立して起きる事件のように思われるが、これらも、具体的な登場人物やそれ

ぞれの状況を変えて「繰り返して」起きる事象であると考えることができる。

たとえば、火事は、その場所・状況・具体的な人名などがその都度異なる一回ずつ独立した事象に思われる。しかし、どの火事も、発火・消火活動・救助・延焼・鎮火・原因究明・火事見舞い・保険審査・保険金支払いなどのイベントから構成され、火元・消防・警察・保険会社・被災者という属性を持った登場人物があり、原因・場所・発火日時・鎮火日時・延焼面積・被害額などということがらが関わっているというように、一連の流れを類型化することができる。

つまり、多くの社会事象は繰り返して生じており、事象を構成する個々のイベントの内容や具体的な登場人物名は変化するが、イベントの特性や生起順序、登場人物の属性は大きくは変わらずに繰り返されると考えられる。

2.2. 社会事象の単位と新聞記事の単位

社会事象は、地震のように数分間で終結する短期のものから、刑事事件とその裁判のように何年もの長期にわたるものまでである。さらに、地震のように一見非常に短時間で終結するように見えるものも、その予測から被害状況の把握、復旧作業などを含めた一連の大きな流れのなかに位置づけられるものが多い。

一方、新聞記事は、こうした一定の期間にわたって継続する社会事象について、刊行時点での最新の状況を報道するので、一つの事象は複数回に分けて新聞記事に掲載されることが多い。一つの新聞記事だけで事象の全容が伝達されるのはむしろ例外的である。(図1)

過去の事象について新聞記事から情報を得ることはできるが、個々の新聞記事は断片的事実を報道するのみであるので、事象の全体像を知るにはその事象の前後の記事を含めた探索が必要となる。

2.3. 社会事象の内容と新聞記事の内容

新聞記事は基本的に刊行時点での判断と他の記事との相対的關係によって掲載と掲載内容が決められている。したがって、ある事象を構成している全てのイベントを報道しているわけではない。また、その事象について読者が一定の知識を持っていることを前提として記事が書かれていることが多い。逆に、記者や編集者が、ある新聞記事の内容を報道するのに必要だと判断した場合には、その事象のそれまでの経緯や基礎的な知識にも言及されることもある。

つまり、新聞記事によって伝達される内容は、重複があるとともに欠落した部分がある。そこで、あるまとまりのある事象を単位とするだけでなく、重複を削除し、その事象に関わる基礎的な知識や事象を構成するイベントの生起順序の理解にもとづいて欠落している部分を予測することが望まれる。

すなわち、繰り返し生起する事象は、事象の種類に応じて、事象を構成するイベント群とその生起順序を規定し、しかもその事象に関わる知識をも組み込んだ分析枠組みを作成することができる。その枠組みに基づいて、複数の新聞記事の内容を分析・統合することによって、個々の新聞記事では断片的にしかわからない事象全体についての情報を、重複を削除して提供し、新聞記事だけではわからない背景となる事実についての質問に回答できたり、予測したりする

質問・回答システムを考えることができる。

3. SCISOR

同様な考え方で作られている前例として、米国のGE研究センターの Raoらが開発している SCISOR (System for Conceptual Information Summarization, Organization and Retrieval) をあげることができる。これは、企業合併・買収に関する新聞記事の全文を分析し、概念レベルでとらえることにより、質問に回答できるシステムを目指したものである⁽¹⁾⁻⁽⁵⁾。

Raoらは、従来の情報検索システムが自然言語処理に弱点を持っているとし、一方自然言語処理ではテキスト中の重要な語と句を全て含むような辞書を持っていないために限界があるとしている。そこで未知の語が出現しても柔軟に言語処理のできる処理戦略をたて、テキストから辞書的情報を自動的に獲得するという二つの手法を用いた人工知能システムの開発を進めている。

具体的には企業買収における短い新聞記事を分析し、新聞記事から抽出した概念情報を、自然言語インタフェースを通してアクセスするものである。人工知能技術をベースとし、自然語の構文解析技術、語単位のフルテキストサーチ技術を用いた検索技術などを組み合わせている。

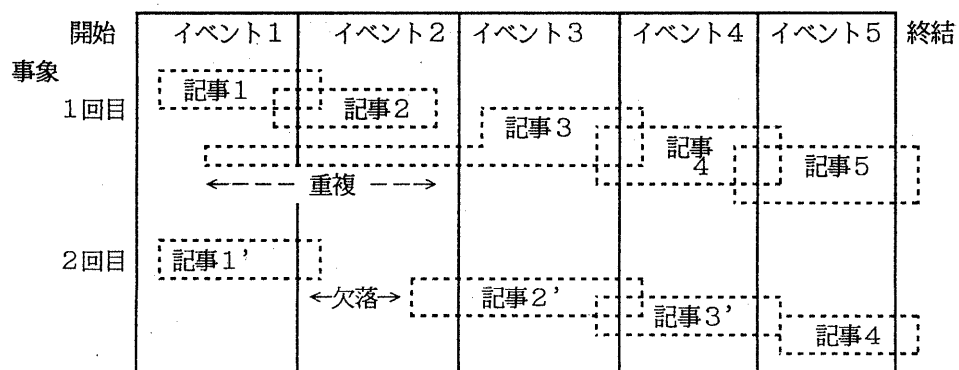


図1 事象・イベントと新聞記事の関係

4. 事象分析の枠組み

本研究は、自然言語処理や人工知能的システム構築よりも、ある事象について様々な質問に回答することができるような、事象の構造的理解とそのデータベース化に重点を置いている。

事象をデータベース化するために、事象と個々の知識に大きく分けた。

4.1 事象とイベント、要素

繰り返して生じる事象は、図2のように捉えることができる。事象はさらに詳しいレベルである「イベント」によって構成されている。事象が生じるたびに毎回同じイベントが生起する。イベントは省略されてもよく、また生起順序に異同がありうる。なおイベントは、その種類にかかわらず、主体、対象、時間、場所、行為、理由などを含んだ共通の構造を持っていると考えることができる。

イベントは、さらに個々の「要素」からなる。この要素は処理上の単位であり、データベース上のレコードに対応する。新聞記事は、この要素に対応するが、個々の新聞記事は一つの記事が一つの要素になることもあれば、一つの新聞記事が二つ以上の要素を含むことも多い。また、複数の記事が同一の内容を持つ要素を報じていることもある。

4.2 知識

各事象に登場する人物と用語などは知識部分として処理する必要がある。同じ事象について

も特定の役割を異なる人物が演じる。また、新しい用語が出現する場合がある。質問・回答システムに最小限必要とされるものとしては、人物と組織、用語がある。

4.3 ファイルの構造

以上の検討から、以下のような種類のファイルを用意した。

- (1) イベントファイル
- (2) 用語ファイル
- (3) 人物ファイル

組織と人名からなる。

- (4) デフォルトファイル

各イベントごとに使用する項目を指示

5. 事例：米価審議・決定過程

5.1 事例の選択

事例として新聞上で報道されることが多く、定期的に生起し、過程が複雑であり、またその過程で関与する組織、個人が多い事象として、「米価審議・決定過程」を取り上げた。

米価審議は毎年一定の時期に行われ、政府、政党、農業団体等が関与する。最初に農水省からシナリオが提示されるが、必ずしもその通りには進行しない。しかし過去10年間をみるとイベントの種類には変動がみられない。

5.2 イベントの種類

この米価審議・決定過程は表1のようなイベントから構成されている。

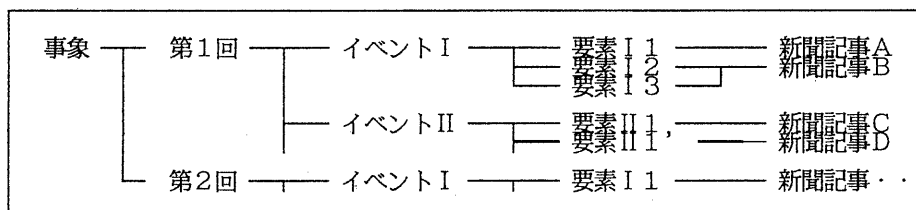


図2 イベントと新聞記事の関係

表1 イベントの種類

1.	要求米価
2.	米価審議会日程
3.	農産物価格安定基金
4.	自民党米価委員会
5.	事前米価
6.	政府米価(生産者米価)
7.	米価(生産者米価)
8.	政府米価(生産者米価)
9.	政府米価(消費者米価)
10.	米価(消費者米価)
11.	政府米価(消費者)
12.	影響
13.	状況・意見
14.	予測・意見

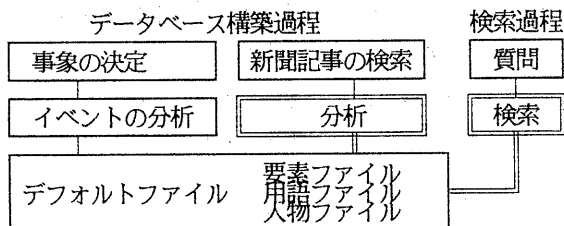


図3 データ構築過程と検索過程

要素ファイル

	例
記事コード	19910613M02
イベント種類	要求米価
主体(誰が)	全国農業協同組合中央会
時間(いつ)	910612
場所(どこで)	自民党本部
対象(誰に)	農林三役
内容(何を)	「現行価格維持」とする据え置き要求を提出
理由(なぜ)	生産コストは4.8%上昇しているが、コメ開放問題で消費者の理解を得るため
予測(見通し)	政府は引き下げをねらう

デフォルトファイル

	例
デフォルトコード	08
イベント種類	政府決定生産者米価
主体	政府
時間	*
場所	*
対象	*
内容	*
理由	*
予測	*

用語ファイル

	例1
用語コード	020
用語	新算定方式
意味	平均生産費以下の農家の生産費
同義語への参照	(90年度方式)
関連語への参照	(米価算定方式)
上位語への参照	
正式名への参照	
組織への参照	
役割への参照	
有効年月日開始	19900601
有効年月日終了	19911231

人物ファイル(組織)

	例
組織コード	102
組織名	農林水産省
役割1	大臣
役割2	事務次官
役割3	
役割4	
上位組織	政府
有効年月日開始	
有効年月日終了	

人物ファイル(人名)

	例
人物コード	1021
役割コード	1
氏名	近藤元次
組織コード	102
有効年月日開始	19900101
有効年月日終了	19901231

図4 各ファイルの項目と例

6. 新聞記事の分析

米価審議会を4日に控え、自民党は3日朝、同党本部で総合農政調査会（山本富雄会長）の	予測	主体	時間	場所	（組織・人名）
米価委員会（太河原太一郎委員長）を開き、2年続きの引き下げを目指す政府側との深夜の	対象	（人名）			（要求米価の別の「要素」）
最終調整に向け、党内論議に入った。					

図5 新聞記事分析例 （1991年6月3日朝日新聞夕刊）

7. 想定される質問の種類と例

7.1 用語

質問：「コメ議員」とは何か

使用ファイル：用語ファイル、人物ファイル（組織、人物）

回答：

コメ議員とは、農村地帯に選挙区を持ち農業政策に積極的な議員である ←(意味)
 ベトコン議員、自民党農林族、農林議員ともいう。 ←(同義語・省略形)
 また、コメ議員には以下のものが含まれる。 ←(下位概念・構成員⁵)
 自民党内のコメ議員の集まり¹である「農村振興議員協議会」²（略称
 ：農振協³。85年度会長：桜内義雄⁴）
 良質米産地の国会議員が中心である「新農政研究会」（略称：新農政
 研。85年度会長：長谷川峻）
 自民党内コメ議員の会で、生産者米価引き上げ強硬派である「日本農
 政刷新同志会」（85年度会長：桜井新）

1:下位概念をあらわす用語の意味 2:下位概念 3:省略形 4:組織の構成員

質問：「生・所方式」とは何か

使用ファイル：用語ファイル

回答：

生・所方式とは、生産費・所得補償方式⁵のことである。 ←(完全形⁵)
 生産費・所得補償方式とは、生産者米価算定方式⁶の一つで、稲作にか ←(上位概念⁶)
 かる肥料等の物材費、家族労働費、地代等を積み、過去3年間の米の10 ←(意味⁷)
 アールあたりの収量でわった数字⁷である。
 生産・所得補償方式、生・所方式ともいう。 ←(同義語)

7.2 人物と組織

質問：「全中の1990年度の会長」は誰か

使用ファイル：用語ファイル、人物ファイル（組織、人物）

回答：

全中とは、全国農業協同組合連合会⁸のことである。
全国農業協同組合連合会の1990年度の会長は堀内巳次⁹。

←(組織の正式名称⁸)

←(当該年の会長名⁹)

質問：「米価委員会」は何に属しているか

使用ファイル：人物ファイル（組織）

回答：

米価委員会は、自由民主党¹⁰総合農政調査会¹¹に属している。

←(直上の組織¹¹、
その上位組織¹⁰)

7.3 イベントの生じた年月

質問：「1985年度の米価審議会」の年月日

使用ファイル：デフォルトファイル、要素ファイル

回答：

1985年度の米価審議会は、1985年7月9日から10日まで。

7.4 イベントの内容

質問中のがかりとなる語をまず用語ファイル、人物ファイルで検索し、その同義語および下位概念が出現する全てのイベントを要素ファイルから検索し、新聞記事の掲載日や記事の物理的単位をこえて、イベントの種類、時間、主体ごとに整理し、重複を整理して回答する。

質問：昭和60年度生産者米価決定におけるコメ議員の動静

使用ファイル：用語ファイル、人物ファイル、デフォルトファイル、要素ファイル

回答：

米審日程	6月11日	日本農政刷新同志会は米価審議会の開催を7月7日の東京都議選のあとにするよう求める方針を決定。
状況・意見	6月18日	農村振興議員協議会は、①基本米価は引き上げ、良質米奨金は現行制度を維持する②市場開放の行動計画作成に関しては農作物を例外とすべきだと決議した。
米価委員会	7月5日	米価委員会の中で、新農政研究会を中心に奨励金の削減に反対し、現行行水準の維持を求める強い意見が相次いだ。

質問：１９９１年度生産者米価決定における米価審議会の審議内容

使用ファイル：人物ファイル、デフォルトファイル、要素ファイル

回答：

事前米審	4月	食糧庁は米審委員懇談会に対して９０年の算定方式を再度説明したが「強い反対意見は出なかった」ことから、変更しないことにした。９０年の生産者米価答申時に、この新算定方式に対し「現状追認的で、適地適産による生産性向上を図るうえの効果か疑問」などの否定的な意見を付け、方式の再検討を促した
	6月27日	近藤農林水産相は米価審議会に生産者米価の算定方式を諮問した。
	6月28日	米価審議会は２日間の日程を終えた。
政府諮問	7月 4日	近藤農林水産相は生産者米価を農水省分庁舎での米価審議会に諮問した。
米審答申	7月 4日	米価審議会は農林水産相に基本米価で０．６５％引き下げ、６０キロあたり１万６３９２円（現行１万６５００円）とすることを、事実上、諮問どおりに答申した。
状況・意見	7月 4日	米価審議会は生産者委員を中心に反対意見があった。米価の算定方式は、稲作の担い手を重視するという国の政策からみて、不明確だとする委員が多く、付帯意見として速やかに再検討するよう求めた。
	7月 4日	米価審議会はその他に農村の実態を踏まえて生産・構造・地域政策を充実、強化し、稲作の担い手の確保・育成のため具体策を総合的に検討、実施するよう求めた。

引用文献

- (1) Rau, L.F. Spontaneous retrieval in a conceptual information system. Proceedings of International Joint Conference on Artificial Intelligence. 10th. Vol.1, p.155-162(1987)
- (2) Rau, L.F. Information retrieval from neverending stories. Proceedings of National Conference on Artificial Intelligence. 1987. Vol.1, p.317-321(1987)
- (3) Rau, L.F., Jacobs, P.S., Zernik, U. Information extration and text summerization using linguistic knowledge acquisition. Information Processing and Management. Vol. 25, No.4, p.419-428(1989)
- (4) Rau, L.F., Jacobs, P.S. NL ∩ IR: Natural language for information retrieval. International Journal of Intelligence System. Vol.4, No.3, p.319-343(1989)
- (5) Jacobs, P.S., Rau, L.F., Sibley, E.H. SCISOR: Extracting information from on-line news. Communication of ACM. Vol.33, No.11, p.88-97(1990)
- (6) 上田修一, 越塚美加, 神門典子. 新聞記事の分析・統合・質問回答システム. 三田図書館・情報学会研究大会予稿集. 1991-11-16. p.41-44.