

G-4

投稿文からのシナリオ作成 Scenario made with messages

西室 洋介† 西田 豊明†
yousuke nishimuro toyoaki nishida

1. はじめに

人々は多くのコミュニケーションを電子空間上でやりとりする。そのため、コミュニケーションの軌跡を管理してまとめてくれるシステムは有用である。本稿では POC システムの中でやりとりされたコミュニケーションをニュース原稿形式に組み替えることで、ユーザに対しそれまでシステム内でやりとりされた内容を分かりやすく提示する方法を述べる。その際には話題ごとのまとまり、話しの一連の流れや分岐を、共起度やそれによる重要語の選別をもとに実現してゆく。

2. POC システム

本研究室では POC(Public Opinion Channel)というシステムがある[POC]。これはコミュニティウエアとしてユーザがそれぞれ自分の意見を配信できるよう開発されたものである(図1)。

自分の情報を発信したいユーザは、サーバーに自分のメッセージを投稿し、メッセージは POC カードとしてサーバーに蓄えられる。これが各ユーザに配信されることでメッセージの配信が行われる。今後、本稿では POC カードを扱い情報流通を高めるための実装を行ってゆくの、カードが実際どのような内容を保持しているかを示す。POC カードは Title、Body (50 単語程度)、url(参考 URL)、Image (画像へのパス、画像は別にファイルとして保持)の4つの内容を持つ。具体的には次のように保持される。

```
<opinion name="tanaka" date="2001/11/6"
img="http://imgSvr.ac.jp/u-tokyo_10052.bmp">
<title>湯島神社</title><comment>菅原道真をまつこの神
社は、学問の神様として有名。ここを訪れたときには、菊祭り
を行っていた。</comment>
<url>http://Reference.ac.jp/BluesB.html</url>
```

3. メッセージ群からのシナリオ作成

POC において、他のユーザが投稿したメッセージを一つ一つ順番に読んでゆくのは非常に能率が悪い。そこで、メッセージの配信はそれぞれのユーザが発信した情報をまとめた形にしてコミュニティに再送するという形を目指すべきである。これによりインタラクティブ放送が実現できると思われる。

インタラクティブ放送を実現するための一つとして投稿内容を話題ごとにまとめてニュースキャスターエージェントにしゃべらせるという方法が考えられる。そのための手順として本稿では投稿されたメッセージをある話題に沿ってまとめ、シナリオ原稿とする手順を提案する。

物語や脚本等のシナリオを書くときにはその作成手順がある。そのシナリオ作成は大きく分けて3つの構成から成る。その三つとはストーリー、シーケンス、シーンである

[scenario]。ストーリーとは話全体の大まかな流れであり、シーケンスとはストーリーよりは小さな分割だがひとまとまりの部分。シーンとは場面場面の個々動きである。例えば、POC において地域の話題があるとすれば、ストーリーとして「湯島」という大きな話題があり、湯島と言う話題の中にも「湯島天神」というシーケンスや、「湯島聖堂」というシーケンスが小さな話題としてあり、その個々のシーケンスの中に各 POC カードというシーンがある。このようにして、シナリオを作成する。

本稿では、POC に投稿された POC カードの集合に話題語が与えられた時にその話題に関するカードを選別する手法について述べる。さらに、選別されたカードを話が飛ばないように並び替えることによってユーザにとって聞きやすいテキスト並びを作成する。

3.1 単語同士の関連性の計算

本稿ではある話題語が与えられた時、それに関連の深い単語も考慮してその話題語に関連のあるテキストを集めることを目標とする。話題語に関連するテキストを選別するための前段階として、テキスト内での単語同士の共起度を計算する。語 w_i と語 w_j の共起度は以下のように計算する[Kitamura 97]

$$co(w_i, w_j) = \ln DF(w_i, w_j) \frac{2DF(w_i, w_j)}{DF(w_i) + DF(w_j)}$$

$DF(w_i)$ はテキスト集合全体 D での語 w_i の文章頻度。 $DF(w_i, w_j)$ は D において語 w_i と w_j が共に出現するテキスト数である。 $\ln DF(w_i, w_j)$ は出現頻度による重みであり、出現頻度の低い同士が偶然同時に出現し高い共起度を持つのを防ぐ。その語自身についての共起度は $co(w_i, w_i) = \ln DF(w_i)$ とする。

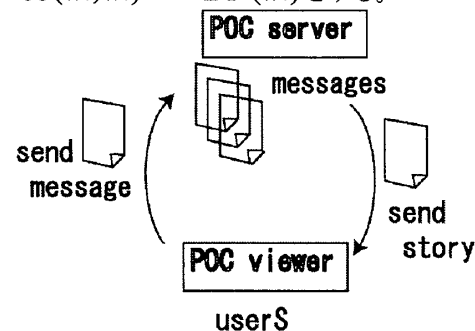


図1 POC システム

3.2 ストーリーの元となるテキストの抽出

ある話題語が与えられた時、それに関連の深い単語も考慮して、その話題語に関連のあるテキストを集めることでストーリーとなるテキスト抽出を行う。これは鎌田ら

† 東京大学大学院 情報理工学系研究科

[kamata]の方法を用いる。この話題語はユーザから与えられるか、または話題抽出手法で選出された単語のうちの1つとする。ここでは前述の共起度 co をもとに話題となる語 W_i に対してテキスト集合全体 D から関連するテキストを抽出する。まず、語 W_i と共起度 co が上位 10 位以内のものに対しては得点 co を与え、各テキストについて得点を持った語を含めばその得点を与える。この得点をもとに得点上位のテキストが語 W_i に関連するテキストとして抽出される。これにより、その話題語に関連する単語を含むテキストも抽出することが出来る。

3.3 シーケンスの作成

ある話題に沿って集められたテキストであっても、それを読み上げるだけではその話題中でいろいろな分野に話が飛んでしまう。そこで、短い時間内で話はまとまっている方がよい。ここでは読み上げるカードのつながりを強くし、読み上げる POC カードの話題がばらばらにならないようにする。これをシーケンスとする。

シーケンスは次のようにして作成する

- 1 話題語およびそれと共起度の高かった語、を重要語とする。
- 2 テキストのなかで、ストーリーの元となるテキスト抽出の際に付けられたスコアが一番大きなテキスト (T1) をとってくる。
- 3 2 で取ってきたテキスト (T1) に含まれる重要語を調べ、それらの重要語をより多く含むテキストをその次につながるテキスト (T2) にする。これを繰り返して一つのシーケンスを作る。
- 4 選ばれなかったテキストでスコアの高いものを先頭にして同じようにシーケンスを作っていく。

4. 会話型エージェント用のシーン作成

前章で選別、グループ分けされたメッセージをニュースキャスターエージェントが読み上げるための原稿に変換する手法を本章では述べる。

4.1 会話調への変換

投稿メッセージをニュースキャスターが読むためには会話調にしなくてはならないその為に、投稿されたメッセージの題名、投稿本文を順次読み上げてゆく方式をとる。また文末はそろえるよう変換する。

<ニュースキャスターが読み上げる本文の例>

<原文>

題:湯島天神

本文:菅原道真をまつるこの神社は、学問の神様として有名。ここを訪れたときには、菊祭りを行っていました。

<読み上げる本文>

次は湯島天神についての話題です。

菅原道真をまつるこの神社は、学問の神様として有名。ここを訪れたときには、菊祭りを行っていました。

4.2 話の分岐の作成

ニュースキャスターはニュースを選別された投稿メッセージを順次読んでゆくことを基本とするが、そのニュースキャスターエージェントが真面目な人型をしているか、キャラクター的な形をしているか等で流れる番組の雰囲気を変えることが出来る。番組を作成する場合は選別されたメッセージに対して、それと関連するメッセージを発する役者(コメンテーター)が必要と思われる。ここでは、その従

来選別されたメッセージに対してコメントするためのコメント用投稿メッセージを抽出してくる手順を述べる。

前提として、本システムは共起度をもとに重要語が存在する。重要語とは今、放送されている話題語に関連の深い語である。つまり、ここでは共起度をお互いに連想しあう単語としてみなすことで POC サーバー上のメッセージ群よりコメントをひっばってくる。

そこで、その重要語と関連の深い投稿テキストを役者が話すコメントとして用いる。

以下にコメント用の投稿メッセージ抽出の方法を述べる

- 1 今話している話題語と共起度の高い単語(重要語)を話題語として、共起度を用いた投稿メッセージの検索を行う。
- 2 抽出されたテキストをスコア順に並べる。
- 3 スコアの高いものからコメントの候補とし、その候補となった投稿メッセージがこの先に読まれる予定のメッセージでなければコメントとして採用する(図2)。

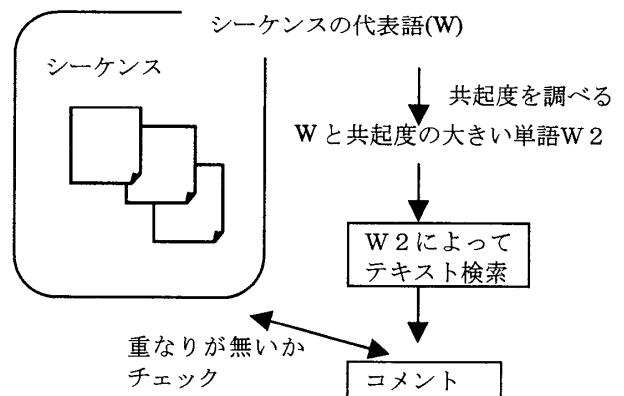


図2 コメントの作成

5 おわりに

本稿では、POC カードを特定の話題のもとで選別することによりニュースまたはラジオ形式でユーザに情報を届けることを提案した。ここではシーケンスという話の流れやコメントという話の分岐を共起度を用いて実現することを提案した。このシステムに対し「話題をどのように選別するか」ということを可能とするシステムをつなげると、自動放送が可能となる。もし、この自動放送が可能となれば、ユーザは何もすることなしに、ただ聞くだけでコミュニティの知識が得られることとなり、これまでの能動的な知識獲得欲を必要とするシステムより多くの人に使用してもらえるものと思われる。

6 参考文献

[POC] Synsophy Project

<http://www.synsophy.go.jp/>

[kitamura] 北村美穂子,松本祐治:対訳コーパスを利用した対訳表現の自動情報処理学会論文誌 Vol38 Num.4 pp. 727-736,1997

[kamata] 鎌田健一,黒橋禎夫,西田豊明:

雑多なテキスト集合からのストーリー生成 / 言語処理学会 第8回年次大会, pp.363-366 (2002.3).

[scenario] シナリオの技術 荒井一 ダヴィッド社 1986年初版