

会話連動ナビゲーションシステムのための辞書構築

Building a dictionary for Information navigation system based on conversation

境美樹 横関大子郎 高田慎也

SAKAI Miki YOKOZEKI Daigoro TAKADA Shinya

日本電信電話株式会社 NTT 情報流通プラットフォーム研究所
NTT Information Sharing Platform Laboratories, NTT Corporation

1. はじめに

電話によるコミュニケーションを行う際に、ユーザが発話した内容について、関連のある情報を参考にする場合がある。例えば、発話内容に応じて、不明な点があれば検索を行い、関連ある新たな情報を取得する、といったプロセスを踏む。しかし、このような行為を会話中に行うことは、ユーザにとって負担であり、PCの操作は会話を中断させるなどといった問題が発生する。

そこで、著者らは、発話内容に応じて、発話語・発話語に対する連想情報や話題となっている語を提示する、会話連動ナビゲーションシステムを提案した[1]。

会話連動ナビゲーションシステムでは、知識NW(図1)と呼ぶ、固有名詞を中心とした語彙の関連性が記述されたネットワークを利用して、連想語や話題語を提供している。一方で、流行の話題や個人の好みに応じた情報提供を行うため、この知識NWを自動構築することが課題のひとつである。

本稿では、知識NWの要件を定義し、用語に対する説明文に対して、2つの構築手法を試し、特性の違いを明らかにする。更に説明文以外のリソースへの適用可能性を検討する。

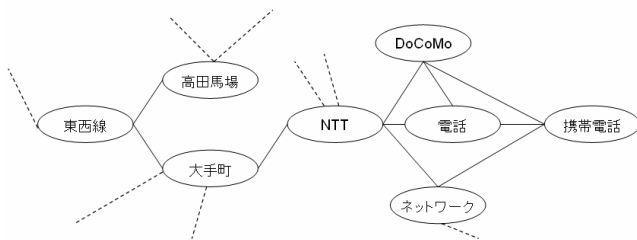


図1：知識NWの例

2. 知識NWの要件

会話連動ナビゲーションシステムでは、ユーザの発話から認識された、複数の発話語間の距離と最短経路上に出現する語の頻度を元に連想語抽出を行っている。具体的には、以下の条件を満たす単語に重み付け、閾値以上の値を持つ単語を連想語として抽出している。

(ア) 2つの発話語が知識NW上で1ホップ離れている場合に、双方に隣接する単語

(イ) 2つの発話語が知識NW上で2~3ホップ離れている場合、双方の中間に位置する単語

ここで特に(ア)の連想語抽出に着目した場合、知識NW上に、図2の様に3ノードが共に繋がるようなリンクが多数存在している事が望ましい。このような特性を持つNWの要件として、スモールワールド[2]性が挙げられる。スモールワールドとは、以下の条件を満たすものを指す。

- I. 総ノード数に対して総リンク数が数倍程度しかない
- II. 任意の2ノード間の距離がノード数に比べて著しく小さい
- III. 同一のノード数、リンク数を持つランダムネットワークのクラスタリング係数よりも非常に大きな値を持つ

ここで言う、クラスタリング係数とは、任意のノードvがk個のノードと隣接している時、k個のノード間に存在するリンク数(kC_2)と実際に存在するノード数の割合を全てのノードについて平均したものである。

また、任意の2点間の平均距離が3~4程度であれば、(イ)の方式で連想語が抽出される可能性は高まる。

つまり、会話連動ナビゲーションシステムで求められる知識NWの要件として、以下の2つがあげられる。

- (a) スモールワールド性を有する
- (b) ノード間の平均距離が3~4である

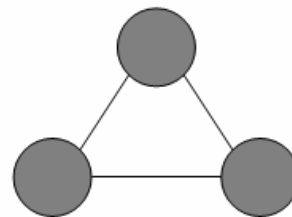


図2：(ア)で求められるノード間の関係

3. 知識NWの構築

本稿では、簡単な2つの手法を用いて、知識NWの構築を試みた。その結果、スモールワールド性が満たされているかの評価を4章で紹介する。

NTTグループ
日本におけるグローバルな情報流通サービス企業グループとして、市内、長距離、国際、移動体通信、データ通信を提供しています。

図3：用語解説の例

<構築手法>

(1) 定義参照

一般的な辞典の様に、見出し語とその説明文を元に知識NWを構築する。

例えば、「NTTグループ」、「移動通信」、「データ通信」が見出し語として存在する時、図3の「NTTグループ」という用語の説明文から、「NTTグループ」と「移動通信」、「NTTグループ」と「データ通信」をリンクで結んで、知識NWを構築する。

(2) 語彙の共起

今回は、同じ文中に登場する語彙(主に固有名詞)同士をリンクで結ぶ事でネットワークを構築する。

例えば、「岡田屋のケーキはクリームがおいしい。」という文章があれば、「岡田屋」と「ケーキ」、「岡田屋」と「クリーム」、「ケーキ」と「クリーム」をリンクで結ぶ。

但し、名詞を抽出する際に誤抽出する可能性や、たまたま同じ文中に現れたが、関係の無い語彙がリンクされる場合がある。これらを回避するため、ある語彙のペアが複数文章に表れたときのみリンクを結ぶ事にする。

4. 評価

3章にて提示した、(1)定義参照の手法で構築された知識NWは、見出し語とそれを説明した文中に出現した単語とをリンクで結んでおり、確実に見出し語と関連のある単語を関連付けている。そのため、連想語の質も高いだろうと推測される。この知識NWを実際に会話連動ナビゲーションシステムへ適用してみても、違和感の無い連想語提示が行えている。

そこで、今回、用語解説のWebページを利用して、知識NW構築を行った結果を示し、手法による結果の違いを検証する。但し、(2)の手法は、抽出した単語数が10万を超えた時点で処理を打ち切り、他の単語とのリンクの無いノードを削除している。その結果、76,788語が残った。

結果を表1に示す。表1から、各方式共に、総ノード数に対する総リンク数が少なく、ノード間の平均距離もノード数に対して小さく、クラスタリング係数がランダムネットワークに比べて非常に大きくなっているため、スモールワールド性を保っている事が分かる。

つまり、知識NW上で隣接した2単語が発話されると、双方に共通した隣接ノードが存在する可能性が高く、(ア)の方式を用いて連想語が提示される可能性が高くなるだろう。

一方、平均ノード間距離を見ると、(1)(2)共に、3前後の値であり、各ノードの平均距離の分布を見ても、3~4に値が集中している。このことから、(イ)の連想語抽出方式に対しても有効な知識NWが構築できていると言える。

これらの結果から、(1)と同様、(2)の手法を用いてもスモールワールド性は保たれ、平均ノード間距離も想定していた範囲にあるので、連想語抽出が可能であり、見出しの無い一般的なWebページ(ニュースやBlogなど)を用いても構造上の要件を満たす知識NWが構築可能であると考える。

5. おわりに

今回、会話連動ナビゲーションシステムにおける知識NW構築手法に関する検討を行い、各手法に対するスモールワールド性に着目した評価を行った。その結果、それぞれの手法でスモールワールドとなることが分かり、隣接した単語の連想語抽出が上手く機能する知識NWは容易に構築できることが分かった。

今後の課題としては、距離2~3ホップの単語に対する連想語抽出に適した知識NWの考察、知識NW自体の質(連想語が発話語に対して適切であるかなど)の評価などが挙げられる。

参考文献

- [1] 向垣内 他, “会話連動ナビゲーションシステム”, FIT2005 (第4回情報科学技術フォーラム), pp.393-394, 2005.
- [2] Watts, “Six Degrees: The Science of a Connected Age”, 2003.

表1：スモールワールド性の評価

手法	ノード数	リンク数	平均リンク数	平均ノード間距離	クラスタリング係数	DR(*1)	CR(*2)
(1)	39344	594697	15.12	3.15	0.2648	3.47	0.000379
(2)	76788	745454	9.71	2.95	0.7361	4.07	0.000129

(*1) ノード数・リンク数が同数のランダムネットワークの平均ノード間距離

(*2) ノード数・リンク数が同数のランダムネットワークのクラスタリング係数