

意味ネットワークによる非タスク指向型対話システムの評価

稲葉 通将[†] 磯村 直樹[†] 鳥海不二夫[†] 石井健一郎[†]

[†] 名古屋大学大学院情報科学研究科
愛知県名古屋市千種区不老町

E-mail: †{inaba,isomura}@kishii.ss.is.nagoya-u.ac.jp, ††{tori,kishii}@is.nagoya-u.ac.jp

あらまし 近年、人間と会話のできる対話システムが社会に広がりつつある。しかし、人間と雑談をし、話を盛り上げるための非タスク指向型対話システムの研究は少なく、知的な応答が可能な非タスク指向型対話システムは皆無である。本論文では知的な応答が可能な非タスク指向型対話システムに必要な知識データ構築を目指し、意味ネットワークを対話から構築する。さらにその応用として非タスク指向型対話システムの客観的・定量的な評価法を提案する。実験の結果、複数の非タスク指向型対話システムの性能を比較評価できることが確認された。また、人間同士の対話についても比較評価できる可能性が確認された。

キーワード 対話システム, 意味ネットワーク, 評価法, 非タスク指向

Evaluation method of Non-task-oriented dialogue system using Semantic Network

Michimasa INABA[†], Naoki ISOMURA[†], Fujio TORIUMI[†], and Kenichiro ISHII[†]

[†] Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya City

E-mail: †{inaba,isomura}@kishii.ss.is.nagoya-u.ac.jp, ††{tori,kishii}@is.nagoya-u.ac.jp

Abstract Recently, computerized dialogue systems have been actively investigated and used in various fields. However, There are few researches of the systems to talk with a human, and there are no systems that can make an intellectual reply. In this paper, we build a semantic network from a dialogue aiming at the knowledge data construction that is necessary for non-task-oriented dialogue systems. Furthermore, we propose a new criterion which can evaluate non-task-oriented dialogue systems objectively and quantitatively using semantic network. As a result of an experiment, it was clarified that our method can evaluate dialogue systems. In addition, we also carried out a similar experiment with human-human dialogues. The result suggested the possibility that our method can evaluate the dialogues using semantic network.

Key words Dialogue System, Semantic Network, Evaluation Method, non-task-oriented

1. はじめに

近年、人間と会話のできる対話システムが社会に広がりつつある。例えば、チケット予約受付など、何らかのタスク達成を目的としたタスク指向型対話システムが様々な場面で使用されている。また、人間と雑談のできる非タスク指向型対話システムも少しずつではあるが、使われ始めている。しかし、非タスク指向型対話に関する研究は少なく、その多くはタスク指向型対話システムに関するものである。そこで、我々はインタビューを模し、対話を盛り上げる対話システム作成を目標にする[4]。プロのインタビューは、対話を盛り上げることにより、相手を

楽しませ、相手から様々な話を引き出す。また、インタビューは事前に対話相手についての情報を大量に収集し、十全の準備を行っている。我々は、対話相手の情報を対話相手のホームページやブログなどから獲得し、大量の発話集(スクリプト集)を用意して対話を進める対話システム作成を目指している。そのような高度な対話システムのためには、知識データが不可欠である。本研究では、知識データとして意味ネットワークを対話から構築する。この意味ネットワークは、

- (1) 対話システムの作成
 - (2) 作成した対話システムの評価
- の2点に应用可能であると考えられる。

本論文では、意味ネットワークを上記の(2)に相当する非タスク指向型対話システムの客観的・定量的な評価に適用する。なお、対象とする対話はテキスト対話に限定する。

既存の意味ネットワークとしては、Wordnet [5] や free association norms [6]、日本語のものでは日本語語彙大系 [7] などがあるが、どれも語彙数が少なく、膨大な数の語彙が使用される非タスク指向型対話への応用には不向きである。本研究では、Web 日本語 N グラム [8] を用いて意味ネットワークを構築する。Web 日本語 N グラムの語彙数は約 256 万語であり、語彙数は十分である。

タスク指向型対話システムの評価には、タスク達成率やタスク達成までに要した発話回数などを用いた客観的・定量的な評価法が提案されている [9] [10] [11]。一方、非タスク指向型対話システムには対話後のアンケートなど、主観的な評価法が主に用いられている。以前我々は、テキスト対話に対してタグ付与を行い、付与されたタグ系列を学習した HMM (Hidden Markov Model) の出力確率を計算することにより対話の自然さを評価する方法を提案した [4]。この方法は、発話の繋がりの自然さを評価しようというものである。本論文では、これまでの研究を進展させ、発話の繋がりに留まらず、発話の焦点や主題など、対話の意味的な評価ができるような方法を提案する。

2. 意味ネットワークの構築

2.1 概要

Grice の協調の原則 [12] にもあるように、対話において人間は、お互いに関連した事柄を話している。特に、1対1の対話においては、連続する発話と発話の意味的な関連性は非常に強いものと思われる。ここでいう発話とは、一方が話し始めてから、終わりまでの連続した1回の発言であり、対話とは、会話の始まりから終わりまでの発話の系列である。本研究では、発話の意味を強く反映する名詞をノード、連続した発話中に出現する名詞の意味的な関連をリンクとすることにより、対話を意味ネットワークとして表現する。この意味ネットワークは、話題の広がりや話の焦点を反映したものとなり、対話の意味による評価のために有用であると考えられる。なお、ネットワークが複雑になるのを避けるためリンクに重みや関係などの情報は与えない。意味ネットワーク構築の大まかな流れは以下の通りである。

- (1) 対話から名詞のペアを獲得する
 - (2) 語の共起情報によって名詞ペアを絞り込む
 - (3) 残った名詞ペアからネットワークを構築する
- 以降では、意味ネットワーク構築手法の詳細について述べる。

2.2 ネットワーク構築に使用した対話

本研究で使用する対話には、条件を統一するため以下の制約を課した。

- 30分のテキスト対話とする
- 1対1で交互に発話する
- 話題は制限しない
- 顔文字や方言は使用しない

対話は、PCを介したチャット形式で行った。そのため話者に

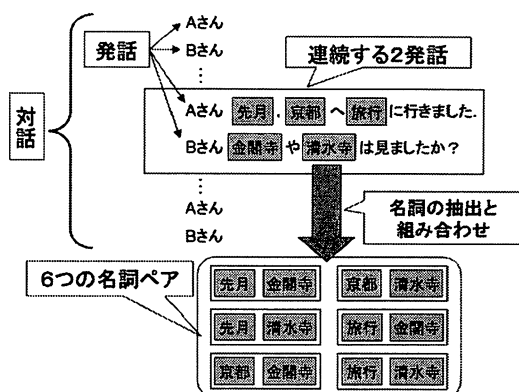


図1 名詞ペアの獲得

よっては、1つの発話を句読点で区切って連続して画面上に出力するような場合もあったが、分析の際には1発話として分析できるよう前処理を行った。

2.3 名詞ペアの獲得

名詞ペアは、対話中の全ての連続する2発話から獲得する。つまり、3つの発話からなる対話からは、1番目の発話と2番目の発話、及び2番目の発話と3番目の発話の計2つの連続する発話から名詞ペアを獲得することになる。連続する発話は、前述した対話の制約により、それぞれが別々の話者による発話となる。

手法としては、まず、2つの発話をそれぞれ MeCab [13] により形態素解析し、発話中の名詞を抽出する。次に、同じ発話内で抽出された名詞のペアを除いた、名詞の全通りのペアを、名詞ペアとして獲得する。

図1に名詞ペアの獲得例を示す。図1の連続する2発話を形態素解析すると、Aさんの発話からは「先月」、「京都」、「旅行」の3つの名詞が、Bさんの発話からは「金閣寺」、「清水寺」の2つの名詞が抽出される。次に、それらの組み合わせを求めることにより、図1に示されている6つの名詞ペアが獲得できる。

2.4 共起情報の抽出

名詞ペアの意味的な関連をリンクとして意味ネットワークを構築していくのであるが、獲得した名詞ペアの中には、意味的な関連が弱い名詞ペアも多く含まれている。したがって、そのような名詞ペアを除外する必要がある。発話間の名詞の意味的な関連の強弱を判断するためには、発話間の名詞の共起の強さをいれればよい。しかし、共起情報を得るために必要な大規模雑談対話コーパスは整備されていないのが現状である。本研究では、名詞と名詞の意味的な関連の強さを記述するものとして、Web 日本語 N グラム [8] から抽出した名詞の共起情報を用いた。

2.4.1 共起頻度の獲得

共起頻度の獲得のため用いた Web 日本語 N グラムは、Google によって約 200 億文の日本語データから作成された n-gram データであり、長さ 1~7 の単語 n-gram が収録されている。この Web 日本語 N グラムを用いる理由は、まず語彙数が約 256 万

語と非常に大きいことが挙げられる。タスク指向型対話では、対話中に出現する語彙はタスクに応じてほぼ固定であるが、非タスク指向型対話では、出現する語彙数は非常に多い。当然、対話を正確に分析するためには、幅広い語彙をカバーする必要がある。一般的な辞書の収録語彙数が数万～数十万語であることと比較しても、約 256 万語というサイズは非常に大きいといえる。もう一つの理由は、n-gram を作成する際に使われた日本語文に、ブログや掲示板などが含まれていることが挙げられる。ブログや掲示板は非タスク指向型対話に近い表現が使われており、それらを基に作られた n-gram は本研究に適していると考えられる。つまり、仮に新聞のコーパスから共起頻度を獲得した場合、「車」に対して共起しやすい語は「生産」や「事故」などになると考えられるが、そのような共起関係は非タスク指向型対話ではあまり見られない。非タスク指向型対話においてはそのような語よりも、「旅行」や「免許」のような語との共起関係の方が重要である。

共起頻度の獲得には、長さ 1～7 の単語 n-gram の中で、最も長い 7-gram を用いた。ただし、n-gram データから共起頻度を獲得する場合、単純に n-gram 内の名詞と名詞の共起数を数えるという方法では正しい値は得られない。仮に、n-gram を獲得するために用いた文として「昨日太郎は東京に行きました」があったとすると、7-gram には「昨日 / 太郎 / は / 東京 / に / 行き / まし」と「太郎 / は / 東京 / に / 行き / まし / た」の 2 つが収録されることになる。よって、ここから単純に「太郎」と「東京」の共起頻度を求めると、元は 1 文である 2 つの 7-gram を重複して数えてしまうことになるからである。そこで、このようなことを防ぐために、共起頻度を求めたい語の 1 つが 7-gram の先頭か末尾に含まれているデータのみを用いて、頻度を計算する。すると、先ほどの例でも「昨日 / 太郎 / は / 東京 / に / 行き / まし」は「太郎」と「東京」の共起頻度の計算からは除外され、正しい頻度が計算できることがわかる。

2.4.2 対数尤度比による共起強度

共起関係の強さを測る指標として、対数尤度比 LLS (Log-Likelihood Score) [14] を用いる。対数尤度比はさまざまな指標と比べ、単語の出現頻度の多寡に関わらずバランスよく共起関係が検出可能な指標であるといわれている。本手法では、出現頻度に関わらず、より多くの語彙に関する共起情報を獲得するため、対数尤度比を用いる。

対数尤度比の計算式は次の通りである。

$$LLS = 2(a \log \frac{aN}{(a+b)(a+c)} + b \log \frac{bN}{(a+b)(b+d)} + c \log \frac{cN}{(a+c)(c+d)} + d \log \frac{dN}{(b+d)(c+d)})$$

ここで N は収録されている文書集合の総文数、 a は 2 語の共起する文数、 b は共起関係を求めたい一方の語の単独での出現文数、 c はもう一方の語の単独での出現文数、 d は共起関係を求めたい 2 語が共に出現しない文数である。つまり、 a, b, c, d と N には以下の関係が成り立っている。

$$N = a + b + c + d$$

LLS の値は共起関係が強ければ大きな値をとる。本研究では LLS の値が n 以下であった名詞ペアを意味的な関連が弱いペアであるとして除外する。今回の実験では、 $n=10000$ として名詞ペアを絞り込んだ。

2.5 ネットワークの構築

最後に、残った名詞ペアからネットワークを構築する。ノードは 2.4 節で除外されなかった全名詞ペアに含まれる名詞からなり (重複無し)、名詞ペアが存在するノード間にはリンクを張る。ただし、同一の名詞ペアが複数存在しても、張られるリンクは 1 つのみとする。リンクは無向とする。

3. 意味ネットワークの評価

3.1 対話システム評価法

対話から意味ネットワークを構築し、そのネットワークから対話を評価する。評価は、構築されたネットワークの規模を示すノード数、及びリンク数を比較することで行う。構築されたネットワークの規模が大きいほど話が盛り上がり、自然な対話となされたと考えられる。なお、第 2 節で述べた意味ネットワーク構築手法では、複数のノードからなる大きなネットワークとは別に、ノードが 2 つとその間にリンクが 1 つあるのみのような、小さなネットワークが独立して構築される場合がある。通常、ネットワーク分析には最大連結成分のみを用いるが、評価にはノード数とリンク数のみを利用するので、今回は特別な処理などは行わない。

3.2 対話システム評価法の頑健性

本評価法は単純な対話戦略を持つ対話システムに対する頑健性を備えている。本評価法で対話システムが高い評価を得るためには

- 直前の発話と関連したことを発話として出力する
- 話の流れや主旨に即した発話を出力する

の 2 点をともに満たすことが必要である。

ここで、「直前の発話と関連したことを発話として出力する」は満たすが、「話の流れや主旨に即した発話を出力する」は満たさない対話システムを考えてみる。例えば、人間の発話に含まれる名詞に着目し、その名詞と意味的な関連のある名詞を含む発話を続けるだけの対話システムが考えられる。実際にそのような発話をする対話システムとの対話を考えてみると

人間：今日は雨が降っていますね。

対話システム：雨といえば、天気、晴れ、傘。

このような発話が考えられる。ここだけを見ると、意味ネットワークが広がりそうに思える。しかし、これに続く人間の発話としては、「何を言っているんですか？」や「それがどうかしたのですか？」等が想定される。つまり、単純な対話戦略しか持たない対話システムでは意味的に自然な対話は続かず、結局、意味ネットワークは広がることなく対話は終了する。

4. 対話システム評価法の評価実験

4.1 実験に使用した対話

表 1 の対話ログから意味ネットワークを構築し、そのネット

対話の種類	対話数	対話の自然さ
(1) 人間同士の対話	24	3.54
(2) ELIZA 型 KELDIC と人間との対話	24	1.67
(3) インタビュー型 KELDIC と人間との対話	24	2.29

表 1 対話システム評価法の評価実験に用いた対話

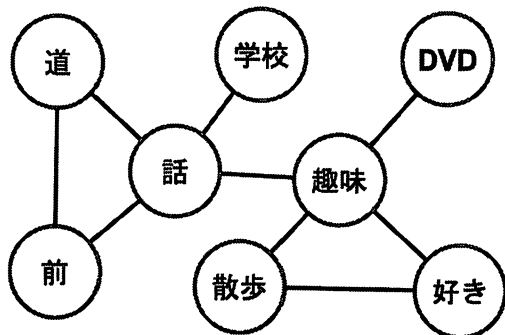


図 2 ELIZA 型 KELDIC と人間の対話の意味ネットワーク

ワークの規模によって対話システムの評価を行う。実験で用いた対話は表 1 の合計 72 対話である。

対話は人間同士の対話 24 対話 (表 1 (1)), 対話システムと人間の対話 48 対話 (表 1 (2)(3)) より構成されている。

人間同士の対話は、互いに面識のない被験者により行われた。また、対話システムと人間の対話についても、同じ対話システムとは 1 度しか対話は行っていない。対話システムには、我々が考案した対話システムの KELDIC (Ken's Laboratory Dialogue Computer) のうち、ELIZA [15] を模して作られた ELIZA 型 KELDIC と、プロのインタビューを模したインタビュー型 KELDIC [16] を用いた。ELIZA 型 KELDIC はあらかじめ用意された発話生成ルールの中から、対話相手の発話と一致するルールを探し、そのルールに応じた発話を返す対話戦略をとっている。インタビュー型 KELDIC は、ELIZA 型 KELDIC を改良し、大量の発話集 (スクリプト集) を用意して対話を行う対話システムである。

全ての対話は、対話を行った人間によって対話の自然さを 1 (極めて不自然) から 5 (極めて自然) の 5 段階で主観評価されており、表 1 に示されている。この結果から、人間同士の対話が最も自然であり、続いてインタビュー型 KELDIC、ELIZA 型 KELDIC の順に自然な対話が可能であるとわかる。

4.2 実験結果

以下に実験結果を示す。

表 2 は表 1 にある計 72 対話全てから意味ネットワークを構築し、そのネットワークのリンク数の平均と標準偏差、及びノード数の平均と標準偏差を示したものである。この表より、ELIZA 型 KELDIC とインタビュー型 KELDIC、及びインタビュー型 KELDIC と人間との間に有意な差があることが確認

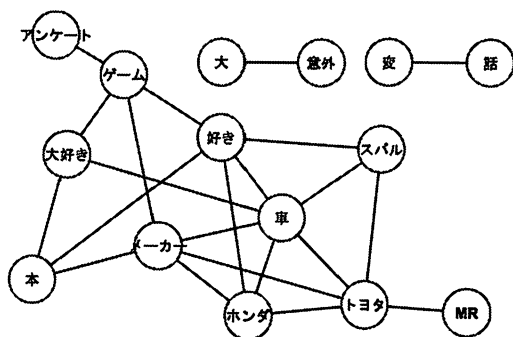


図 3 インタビュー型 KELDIC と人間の対話の意味ネットワーク

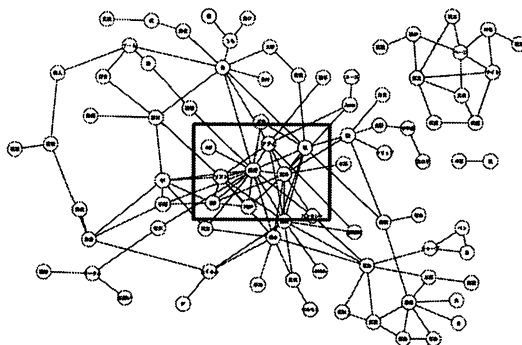


図 4 人間同士の対話の意味ネットワーク

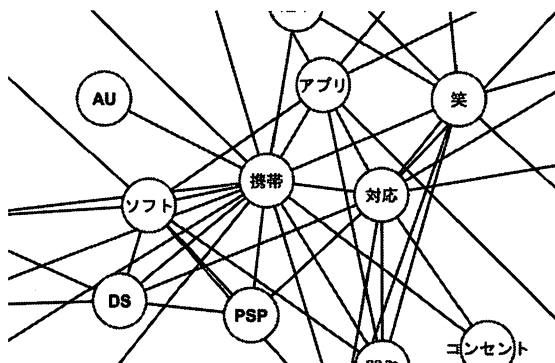


図 5 「図 4」の枠内の拡大図

できる。

図 2 は ELIZA 型 KELDIC と人間との対話 (表 1 (2)), 図 3 はインタビュー型 KELDIC と人間との対話 (表 1 (3)), 図 4 は人間同士の対話 (表 1(1)) からそれぞれ構築したネットワークのうち、1 つを抜粋したものである。図 5 は図 4 の中央枠内の拡大図である。

ELIZA 型 KELDIC は、人間の発話に対して自然な発話を返すことがほとんどできず、「なるほど」や「そうですか」のような単純な頷きが多く、また自分から話題を提供することもしないため、図 2 のような小規模な意味ネットワークが構築された。インタビュー型 KELDIC は、人間の発話に対してある程

対話の種類	リンク数の平均	リンク数の標準偏差	ノード数の平均	ノード数の標準偏差
(1) 人間同士の対話	114.38	53.67	62.75	15.59
(2) ELIZA 型 KELDIC と人間との対話	15.58	8.08	15.17	5.89
(3) インタビュー型 KELDIC と人間との対話	7.38	4.01	8.38	3.17

表 2 対話システム評価法の評価実験結果

度は自然な発話を返すことができ、また、積極的に話題を提供することから、図 3 のような ELIZA 型 KELDIC と比べて大きな意味ネットワークが構築された。しかし、インタビュー型 KELDIC も不自然な発話を返すことが少なからずあり、そこで話題が途切れてしまうことがあったため、大きなネットワークとはならなかった。図 4 の人間同士の発話から構築された意味ネットワークはノード数・リンク数の多い大きなネットワークとなった。人間同士の対話では、単純な顔きのような発話はほとんど無く、話題が無意味に途中で途切れることも無いため、このような結果となったと考えられる。図 5 は図 4 の中央部の拡大図である。この意味ネットワークを構築するために用いた対話は、携帯ゲームに関する話題を中心に進んでいた。図 5 を見ると、携帯、アプリ、PSP などから多数のリンクが張られており、意味ネットワークだけで携帯ゲームが対話の中心であったことが読み取れる。

規模を比較すると、人間同士の対話から構築した意味ネットワークが最も大きく、インタビュー型 KELDIC、ELIZA 型 KELDIC と続いており、表 1 の主観評価と同様の傾向が表れている。また、ELIZA 型 KELDIC とインタビュー型 KELDIC の意味ネットワークの規模の差に比べ、インタビュー型 KELDIC と人間の差が大きいという同じ傾向を示している。

以上のことから、意味ネットワークの規模による対話システムの評価は妥当であるといえる。

5. 人間同士の対話の評価実験

5.1 概要

意味ネットワークによる対話の評価法が、人間同士の対話にも適用できるかどうかについても実験を行った。

実験で用いた対話は表 3 の合計 30 対話である。対話はプロのインタビューによるインタビュー型対話 15 対話 (表 3 (1))、素人によるインタビュー型対話 15 対話 (表 3 (2)) より構成されている。

対話は全てインタビュー形式の対話である。一方のインタビュー (インタビューをする側の発話者) には、もう一方のインタビューイ (インタビューを受ける側の発話者) に関するプロフィールデータが渡されている。プロフィールデータはインタビューイが事前に自分のことを自由作文形式で記述したものである。インタビュー役は、そのプロフィールデータを元に質問をし、対話を進めていく、という形式をとった。また、対話は互いに面識のない被験者 30 人により行われ、全て異なるペアによる対話とした。

対話の種類	対話数	対話の自然さ
(1) プロによるインタビュー型対話	15	3.67
(2) 素人によるインタビュー型対話	15	3.45

表 3 人間同士の対話の評価に使用した対話

プロのインタビューは

- 対話を盛り上げるよう努める
- 対話相手を楽しませる話をする
- 話の腰を折らない
- 相手の意見に理解を示す
- YES/NO ではなく 5W1H 型の疑問文で話を膨らませるといった特徴的な対話法を身につけている。一方、インタビュー役の素人は、そのような対話法で対話を進めるよう直前に指示されているのみとした。

また、対話システムの評価実験と同様に、対話の自然さについても、5 段階で主観評価され、表 3 に示されている。プロの方が自然な対話が可能であるという結果であったが、誤差とも考えられる程度の差であり、素人と大きな差は見られなかった。

5.2 実験結果

結果を表 4 に示す。

実験の結果、プロによるインタビュー型対話の方が、ノード数の平均、リンク数の平均ともに大きいという結果になった。ノード数の平均にはプロと素人の間に大きな差はなかったが、リンク数の平均では差があった。そこでリンク数について、有意水準 0.05 とした t 検定 (片側検定) を行ったところ、プロと素人の間で有意差が認められた。しかし、有意水準 0.01 とした t 検定では有意差は認められなかった。実際の対話を見ると、プロによる対話も、素人による対話も、人間が目で見てもすぐに見分けが付くような大きな違いは無いため、この実験結果は妥当であるという見方もできる。しかし、データ数も少ないことから、意味ネットワークで人間同士の対話の自然さについての微妙な違いが評価できる可能性がある、というのが本実験の妥当な結論であると思われる。

6. 終わりに

6.1 まとめ

意味ネットワークにより、非タスク指向型対話システムを、話の繋がりや形式など浅いレベルではなく、意味を考慮した深いレベルで評価できることを示した。また、人間同士の対話に

対話の種類	リンク数の 平均	リンク数の 標準偏差	ノード数の 平均	ノード数の 標準偏差
(1) プロによるインタビュー型対話	124.50	62.19	64.47	17.51
(2) 素人によるインタビュー型対話	84.73	31.78	57.73	3.71

表 4 人間同士の対話の評価実験結果

対しても、意味ネットワークによる評価が有効である可能性を示した。

本手法では、対話中の連続した発話の意味的な関連性に着目し、発話の意味を強く反映する名詞をノード、連続した発話中に出現する名詞の意味的な関連をリンクとすることにより、対話を意味ネットワークとして表現した。次に、対話別に構築した意味ネットワークの規模を比較することにより、対話の自然さを評価した。

6.2 今後の課題

今後の課題として、大きく分けて以下の3つの課題がある。

- (1) 意味ネットワーク構築手法の課題
- (2) 対話評価法の課題
- (3) 対話システム構築に向けての課題

まず、(1)の意味ネットワーク構築手法の課題であるが、名詞間の意味的な関連性の獲得法を改良することが挙げられる。今回は共起情報を用いたが、図3の右上部の「大」と「意外」のように、意味的な関連が弱い名詞ペアでも共起関係が強いとされるものも散見された。改良案としては、共起関係に加えて、シソーラスなどの他の意味ネットワークを補助的に用いる方法が考えられる。

次に、(2)の対話評価法の課題について述べる。規模による評価だけではなく、ネットワークの複雑性や広がりなどの特徴を用いた評価ができれば、さらに詳細な評価が可能であると考えられる。本論文の手法では、1対話から獲得されるリンク数及びノード数が、対話システムごと、さらには人間同士でも非常にばらつきがあった。したがって、獲得された意味ネットワークの密度や平均経路長などを単純に分析するだけでは比較評価できない。分析のためには、ノード数やリンク数を揃えるなど、何らかの形でネットワークの正規化が必要である。

最後に、(3)の対話システム構築に向けての課題について述べる。我々は大量のスクリプト集を用意して対話を行うインタビュー形式の対話システムの作成を目指している。本研究で構築した意味ネットワークは、スクリプト集からの適切な発話選択に有用であると思われる。スクリプト集からの適切な発話選択は、意味ネットワークが成長していく過程を時系列で分析することにより、次の瞬間にリンクで繋がりがやすい名詞を発見し、その名詞をキーとしてスクリプトから発話を選択する手法が考えられる。また、話を重ねるほど対話システムの性能が向上する学習機能の実店にも意味ネットワークは有用であると思われる。例えば、実際の対話で出現した名詞と名詞のリンク関係に、重み付けをすることにより、意味的関連性の強弱が表せる。この強弱を使えば発話選択性能のさらなる向上が見込まれる。

文 献

- [1] J. Gustafson, N. Lindberg, and M. Lundeberg. The August

- Spoken Dialogue System. In *Sixth European Conference on Speech Communication and Technology*. ISCA, 1999.
- [2] B. Pellom, W. Ward, J. Hansen, R. Cole, K. Hacioglu, J. Zhang, X. Yu, and S. Pradhan. University of Colorado dialog systems for travel and navigation. In *Proceedings of the first international conference on Human language technology research*, pp. 1–6. Association for Computational Linguistics Morristown, NJ, USA, 2001.
- [3] J. Allen, G. Ferguson, and A. Stent. An architecture for more realistic conversational systems. In *Proceedings of the 6th international conference on Intelligent user interfaces*, pp. 1–8. ACM New York, NY, USA, 2001.
- [4] 磯村直樹, 鳥海不二夫, 石井健一郎. HMMによる非タスク指向型対話システムの評価. 電子情報通信学会論文誌, Vol. J92-D, No. 4 掲載予定.
- [5] G.A. Miller. WordNet: a lexical database for English. *Communications of the ACM*, Vol. 38, No. 11, pp. 39–41, 1995.
- [6] D.L. Nelson, C.L. McEvoy, and T.A. Schreiber. The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, Computers*, Vol. 36, No. 3, pp. 402–407, 2004.
- [7] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩己, 小倉健太郎, 大山芳史, 林良彦. 日本語語彙大系-全5巻-. 岩波書店, 1997.
- [8] 工藤拓, 賀沢秀人. Web 日本語 N グラム第1版. 言語資源協会発行.
- [9] D. Jurafsky, J.H. Martin, A. Kehler, K. Vander Linden, and N. Ward. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. MIT Press, 2000.
- [10] M.A. Walker, D.J. Litman, C.A. Kamm, and A. Abella. PARADISE: a framework for evaluating spoken dialogue agents. In *Proceedings of the 35th annual meeting on Association for Computational Linguistics*, pp. 271–280. Association for Computational Linguistics Morristown, NJ, USA, 1997.
- [11] R. Soricut and E. Brill. A Unified Framework for Automatic Evaluation using N-gram Co-Occurrence Statistics. In *ACL 2004*, pp. 613–620, 2004.
- [12] P. Grice. *Studies in the way of words*. Harvard University Press.
- [13] 工藤拓. MeCab: Yet Another Part-of-Speech and Morphological Analyzer. *Version 0.7 b*, 2002.
- [14] T. Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, Vol. 19, No. 1, pp. 61–74, 1993.
- [15] J. Weizenbaum. ELIZA - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, Vol. 9, No. 1, pp. 36–45, 1966.
- [16] 岡田隴二, 鳥海不二夫, 石井健一郎. インタビューを模した対話エージェントのための質問文自動生成. *Proceedings of JAWS 2007*, 2007.