隠れ変数モデルによる発話行為推定と人間によるラベルの比較と分析

大竹 清敬
情報通信研究機構 - ATR 音声言語コミュニケーション研究所
kiyonori.ohtake @ {nict.go.jp, atr.jp}

1 はじめに

対話システムを実現するにあたり，発話意図の推定は非常に重要である．そのため，発話意図の近似として発話行為の集合を作成し，タグ付きコーパスを整備してきた．それによって，発話行為の分析や，教師あり機械学習手法による発話行為推定器の実現など一定の成果をこれまで得てきた．しかしながら，ドメインやタスクによって要求される発話行為タグの種類や，粒度が異なる．そのため，これまでにも様々なドメインにおいて多くの発話行為集合が考案され，そのコーパスが整備されてきた (たとえば，[荒木 99, スコ 05] など)．

タグ付けされたデータを大量に用意できれば，それに対して教師あり学習手法を適用した高精度な発話行為推定器を構築できる．しかしながら，学習データの作成には，タグ付けは必須である．また，タグセットが異なる複数のコーパスをまとめたい場合など，タグセットの整合性をとることが困難な場合が多く，コストがかかる．本稿では，そういった教師あり学習手法を前提とした発話行為集合を考慮するのではなく，教師なし学習手法を前提として，計算機によるモデルがどのような分類を各発話に対して与えるのかを分析する．またそれが，人間が与えたある発話行為集合とどの程度異なるのかを明らかにすることを目的とする．

発話行為を推定することは，有限の発話行為ラベルを付けることとみなせる．したがって，ある発話が与えられたときに，そこから特徴ベクトルを作成し，その特徴ベクトルを分類することができれば，発話行為の分類が可能になる．発話行為分類ののち，適切なラベルを各分類に対して付けることで発話行為推定を行える．本稿では，特徴ベクトルを構成する要素として，発話に含まれる単語や形態素，ならびにその n-gram を考える．しかしながら，この特徴ベクトルをそのまま用いて各発話を分類するのは非効率である．なぜならば，広大な特徴ベクトル空間に対し，各発話に出現するその要素は非常に限られており，個々の発話に対応するベクトルは，非常に疎なベクトルとなるからである．また，特徴ベクトルの個々の要素は，独立であることが期待されるが，実際には，ある要素は別の要素と共起しやすい関係にあり，無駄な場合が多い．

そのような状況に対処するためのベクトル空間の次元を圧縮する技術として LSA (Latent Semantic Analysis) が提案された．実際に，Serafin らは，LSA を発話行為分類に適用し効果的であることを示している [Ser04]．本稿では，LSA のように直接的にベクトル空間の次元を圧縮するのではなく，隠れ変数モデルを用いる．隠れ変数モデルにおける隠れ変数が表現するトピックを用いて発話行為を分類する．

2 隠れ変数モデル

これまでにも隠れ変数を用いる統計的モデルがいくつか提案されているが，本稿では，plsi[Hof99] を用いる．ただし，plsi モデルにおいて，あるドキュメント (単語の集合) が与えられたときにそこから直接的にトピックを推定できないが，本稿では，文献 [Oht05] にて示した方法で発話 (単語の集合) から発話行為の近似であるトピックを推定する．

隠れ変数モデルでは，トピックを表現する隠れ変数の数はモデルを構築する際に与える．したがって，推定される発話行為は与えられた隠れ変数の数を次元とするベクトルで表現されることになる．これをトピックベクトルと呼ぶことにする．

3 対話データ

実験に用いる対話データとして ATR の対話データベース [Mor94] を用いた．このデータベースは旅行におけるさまざまな状況を想定した対話から構成される．使用したデータは，句読点が付けられた比較的書き言葉に近い程度まで形式化されたものを用いた．その理由は，教師なし学習を行う際に形態素解析器を用いて，学習素性を抽出するが，その精度を可能な限り，高くしたいからである．また，データベースに含まれる書き起こしには，あきらかなフィラー，言い淀みなどについてそれぞれタグが付けられており，そのタグに基づいてフィラーや言い淀みなどは削除して用いた．

今回使用したデータは 1,983 対話 (83,052 発話) からなる．このうち，1,970 対話 (82,563 発話) をパラメータ推定 (学習) に用い，13 対話 (489 発話) を評価用として用いる．学習データでは，1 発話平均約 12.1 の形態素から構成される．評価用データは，1 発話平均 11.7 形態素から構成される．さらに，評価用データには，文

献 [Tan99] にて言及されている発話行為タグを発話単位に人手で付与した．今回付与した発話行為タグは，かなり短いセグメントを単位として付与することを考慮して設計されている．今回使用したデータの発話単位では，想定しているセグメントより長い場合がしばしばあった．たとえば，YES か NO で答えるタイプの質問「…はありますか.」に対して「はい．…がございます.」と答えるような場合，応答には,「はい」(YES という発話行為と考える) と,「…がございます」(情報伝達の発話行為と考える) の両方の発話行為が含まれることになる．このような発話に対して，発話行為のタグ付けでは，両方のタグをつけることとした．また，実験においては，1 発話に対して 1 発話行為とすることで集計しやすくなるので，1 発話内の一番最後の発話行為をその発話を代表する発話行為として取り扱う．

発話行為ラベルは，全部で 26 種類あるが，評価用データにおいて使用されたのは，表 1 の左側にあげる 16 種である．

4 学習素性

隠れ変数モデルの plsi は，文書と単語の共起確率を与えるモデルである．したがって，学習 (パラメータ推定) に用いる素性は単語になるが，Serafin らが LSA を，任意の素性も含める拡張を行い効果を得ていた [Ser04] ことから，plsi を用いる本研究においても同様の拡張は有効であると考える．具体的には，各発話を形態素解析し，形態素単位として，その基本形と品詞をあわせて素性とする．ただし，固有名詞ならびに数詞は，汎化して個々の形態素を識別しない．つまり，その出現形を無視し，品詞のみの素性とする．さらに，その uni-gram 素性の他に，形態素の bi-gram も素性に含める．

形態素 bi-gram を素性に含めるのは，日本語において単語の認識が容易ではないことに配慮した結果である．形態素を単語の近似として用いた場合に，その単位が小さすぎる場合があると考えた．結果的に，素性数が増え，特徴空間の次元数が増えることになるが，plsi モデルを適用することで各素性と隠れ変数の関係が適切に設定され，次元圧縮がなされると期待する．

教師あり学習を発話行為推定に適用する場合では，推定しようとする発話の直前の発話行為ラベルを素性に加えて推定することが可能である．しかしながら，我々は，特定の発話行為集合を想定せずに，直接的に発話行為の推定を行おうとしているので，そのようなラベルを用いることはできない．そこで，直前の発話の素性をそれとわかるようにマークして推定しようとする発話の素性として追加する．

5 実験

学習用データを形態素解析し，plsi のパラメータ推定用素性を抽出した．形態素解析には mecab を使用した．この mecab は，京大コーパスをはじめ，対象と同一ドメインの対話データなども含めた形態素解析済みコーパスからパラメータ推定を行った独自のものである．使用した mecab の辞書は，活用を展開した状態で 69 万項目ほどの大きさである．素性は，13 万ほどが抽出された．これは，単純に単語を単位とした plsi に還元して考えると，語彙サイズが 13 万ということになるが，実際の学習データにおける形態素の語彙サイズは，約 7,800 である．さらに，固有名詞と数詞を汎化しているので実際に有効な語彙サイズ，つまり uni-gram 素性はこれよりも小さく約 6,800 である．

学習用データから抽出した素性を用いて plsi モデルのパラメータ推定を行った[1]．隠れ変数モデルを用いる場合の論点の一つとして隠れ変数の数をいくつにするかがある．最適な値を求める方法が確立されておらず，学習データの量をはじめさまざまな要因によって左右されるため試行錯誤が必要である．

隠れ変数の数は 10, 50, 100, 200, 300 とし，過学習問題を軽減するための温度パラメータは 0.9 とした．

まず，人間がラベル付けした結果が，トピックベクトルの空間内でどのように対応しているかの概要を調べた．評価用データに対して，トピックベクトルを作成し，各ラベル毎に平均ベクトルを作成した．その平均ベクトルがなす角の cos 値を求めた．これをそれぞれの発話行為ラベルについて行い，自分自身を除く，他のラベルとの cos 値の平均を表 1 にまとめる．

次に，隠れ変数の数を 100 としたモデルを用いて，評価用データからトピックベクトルを作成し，そのトピックベクトルを教師なしクラスタリング手法である K-means 法 (たとえば，[Dud00] 参照) を用いてクラスタリングした．

また，現状では，評価用の発話行為ラベルが付与されたデータが非常に少ないので機能することは期待できないが，実際に発話行為推定を試みた．発話行為ラベルが付与された評価データ 13 対話のうち，1 対話 (32 発話) を推定用データとし，残りの 12 対話を用いて，各発話行為の平均ベクトルを構成した．平均ベクトルならびに推定用データのトピックベクトルは，隠れ変数が 100，温度パラメータが 0.9 の plsi モデルを用いて作成した．この平均ベクトルと推定用の 1 対話内の各発話から得られるトピックベクトルのなす角の cos 値が最も大きい平均ベクトルのラベルを推定結果とすると，32 発話のうち正解したのは 12 発話であった．

6 考察

既に述べたようにいくつの隠れ変数を用いるかというのは，隠れ変数モデルを使用する場合の問題の一つである．表 1 から，今回行なった実験では，隠れ変数の数が 10 程度では，必要とする分解能が十分ではないよ

[1] 使用したツール:http://chasen.org/~taku/software/plsi/

表 1: ラベル間平均ベクトルの cos 値

	隠れ変数の数毎の平均ベクトル cos 値				
発話行為ラベル ＼ 隠れ変数の数	10	50	100	200	300
ACKNOWLEDGE (ACK)	0.602	0.446	0.330	0.333	0.345
ACTION-REQUEST (ACT-REQ)	0.657	0.420	0.387	0.396	0.347
ALERT	0.615	0.258	0.100	0.044	0.027
APOLOGY	0.679	0.253	0.232	0.200	0.176
CONFIRMATION-QUESTION (CONF-Q)	0.610	0.338	0.292	0.314	0.305
FAREWELL	0.631	0.326	0.291	0.299	0.287
GOOD-WISHES (G-WISHES)	0.610	0.267	0.258	0.280	0.251
GREET	0.666	0.311	0.274	0.266	0.267
INFORM	0.670	0.458	0.443	0.451	0.429
PERMISSION-REQUEST (PERM-REQ)	0.457	0.304	0.256	0.266	0.249
SUGGEST	0.532	0.327	0.312	0.288	0.269
THANK	0.656	0.242	0.241	0.257	0.245
THANKS-RESPONSE (THANK-RES)	0.592	0.250	0.170	0.185	0.158
WH-QUESTION (WH-Q)	0.582	0.390	0.312	0.368	0.367
YES	0.545	0.325	0.329	0.309	0.298
YN-QUESTION (YN-Q)	0.605	0.424	0.384	0.390	0.383
平均	0.607	0.334	0.288	0.290	0.275

うである．一方で，隠れ変数を増やしても，100 以降の数値に大きな変化はないことから，今回の実験条件の場合，100 程度あれば十分とみなせる．

次に，表 2 から，トピックベクトルが構成するベクトル空間では，人間によるラベル付けとトピックベクトルによるクラスター間に強い関連を見い出せない．これにはいくつかの要因がある．まず，今回用いた発話行為ラベルの設計の問題がある．今回用いた発話行為ラベルでは，INFORM として判断される発話が非常に多く，トピックベクトル空間において INFORM が占める割合は非常に大きなものとなる．次に WH-Q や CONF-Q などの質問に関するものが比較的多くのクラスタに分類された．これらの質問行為や，相手に動作を要求する ACT-REQ はかなり特徴的な表現を含むが，非常に多様である．たとえば，ACT-REQ には「…してください」「…していただく」「…してもらえませんか」など，それだけで ACT-REQ と判定できる表現が数多くある．しかしながら，今回の実験では，これらの表現を統一するようなことは行っておらず，すべて plsi のパラメータ推定によって特定のトピックと結びつけることを期待している．実験結果から，単純に形態素解析器し，そこから素性抽出した特徴ベクトルを用いるだけでは，plsi がその期待には思ったほど応えてくれないことがわかる．

規則に基づいて発話行為を推定する手法 [駒谷 99] がある一定の成果を納めていることから，事前知識をうまく取り込むことでさらに人間のラベル付けに近づけることが可能になると考える．一方で，こういった手法は既定の発話行為分類へ迎合するアプローチともとれ，安易に実装することは場当たり的な拡張となりやすい．

事前知識の反映のさせかたのひとつとして，ある種の制限言語を考え，その制限言語へ，表現を言い換える方法が考えられる．言い換えをほぼ同一の意味を伝達する表現への変換と考える限り，発話行為の判定に影響を与えるほどの大きな変換は考えにくい．したがって，このような言い換え処理は，隠れ変数を用いるモデルによって特徴空間の次元を圧縮する以前に，表現の多様性を圧縮する処理とみなせる．

発話のクラスタリングという観点から，各発話の特徴ベクトルをそのままクラスタリングすることも考えられる．しかしながら，実際に，学習データの特徴ベクトルに対してそのまま K-means 法 (クラスタ数 16) を適用した結果は，表 2 に示した結果とは異なるものだった．初期ベクトルの与え方の問題もあるかと思うが，特定のクラスタに集中する傾向が強まり，5 クラスタはまったく要素がない状態となった．そういう意味では，plsi による次元圧縮は十分に機能しているといえる．

わずかにラベル付けされたデータを用いて，発話行為推定を行った結果，37.5%(12/32) という正解率であった．データが少なく，不明確なところも多いが，結果を検証する中でいくつかの事に気がついた．ひとつは，音声から書き起こされた段階で失なわれた情報が有効

表 2: K-means 法によるクラスタリング (100 変数, 温度パラメータ 0.9, クラスタ数 16)

発話行為ラベル (頻度)	クラスタ ID:頻度
ACK (68)	1:7 6:24 7:2 8:1 12:3 13:7 14:8 15:16
ACT-REQ (44)	1:4 2:3 3:1 5:1 7:6 8:1 9:1 10:1 12:11 13:10 14:5
ALERT (1)	13:1
APOLOGY (2)	7:1 13:1
CONF-Q (29)	1:4 2:1 3:1 7:1 8:3 9:4 10:1 11:1 12:3 13:8 14:2
FAREWELL (16)	10:8 12:3 13:5
G-WISHES (1)	10:1
GREET (8)	1:3 12:1 13:4
INFORM (198)	1:27 2:11 3:10 4:6 5:2 6:2 7:35 8:7 9:12 10:7 11:14 12:26 13:27 14:12
PERM-REQ (1)	4:1
SUGGEST (6)	5:3 7:2 12:1
THANK (20)	0:16 8:2 10:1 13:1
THANK-RES (2)	10:2
WH-Q (40)	2:1 3:2 5:16 7:4 8:2 10:7 11:2 12:3 14:1 15:2
YES (18)	1:6 6:2 7:1 8:5 14:4
YN-Q (35)	1:2 2:4 5:4 7:6 8:1 12:8 13:7 14:3

な場合が多いこと．つまり,「…ですか」という文字列からは，発話者の意図が疑問にあるのか，確認にあるのかは判断できない．さらに，人間の判断のゆれから，ラベル付与に関して，複数ラベルも検討すべき場合がいくつかあった．たとえば，対話の終盤で「それではごゆっくりどうぞ」という発話があったとき，人間がつけたラベルでは，FAREWELL だったが，今回の方法では，ACT-REQ と推定された.「ゆっくりしてください」という意味での動作の要求ともとれ，まったく違うとは言い切れない．

7 関連研究

これまでにも様々な種類のアプローチによる発話行為推定手法が考案されてきた．たとえば，駒谷らはルールベースによる手法を用いた [駒谷 99]．大規模なタグ付けされた対話データベースが整備されるにしたがって，統計的機械学習手法が用いられるようになった (たとえば，[Sto00, Tan99] など)．既定の発話行為集合とそのラベル付けデータがあれば，これらのアプローチは非常に有効に機能する．

8 まとめと今後の課題

本稿では，既定の発話行為集合を想定するのではなく，隠れ変数モデルを用いた教師なし学習によって各発話をトピックベクトルで表現し，その性質について考察した．また，K-means 法による教師なしクラスタリング手法を適用し，その傾向についても概観した．今後は言い換え手法の適用，学習素性の工夫などを行い，詳細な分析を行う．

参考文献

[Dud00] Duda, R. O., Hart, P. E., and Stork, D. G.: *Pattern Classification*, A Wiley-Interscience Publication (2000).

[Hof99] Hofmann, T.: Probabilistic Latent Semantic Indexing, In *Proceedings of the 22nd Annual ACM Conference on Research and Development in Information Retrieval*, pp. 50–57, Berkeley, California (1999).

[Mor94] Morimoto, T., Uratani, N., Takezawa, T., Furuse, O., Sobashima, Y., Iida, H., Nakamura, A., Sagisaka, Y., Higuchi, N., and Yamazaki, Y.: A speech and language database for speech translation research, In *Proceedings of ICSLP '94*, pp. 1791–1794 (1994).

[Oht05] Ohtake, K.: Evaluating Contextual Dependency of Paraphrases using a Latent Variable Model, In *Proceedings of the Third International Workshop on Paraphrasing (IWP2005) conjunct with IJCNLP 2005*, pp. 65–72 (2005).

[Ser04] Serafin, R. and Eugenio, B. D.: FLSA: Extending Latent Semantic Analysis with features for dialogue act classification, In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04)*, pp. 692–699 (2004).

[Sto00] Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C. V., and Meteer, M.: Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech, *Computational Linguistics*, Vol. 26, No. 3, pp. 339–373 (2000).

[Tan99] Tanaka, H. and Yokoo, A.: An Efficient Statistical Speech Act Type Tagging System for Speech Translation Systems, In *Proceedings of the Thirty-Seventh Annual Meeting of the Association for Computational Linguistics (ACL'99)*, pp. 381–388 (1999).

[スコ 05] スコギンズリーバイ, 川嶋宏彰, 松山隆司：間の合った発話タイミング制御を目的とした漫才の動的構造の分析, インタラクション 2005, pp. D–404 (2005).

[駒谷 99] 駒谷和範, 荒木雅弘, 堂下修司：対話コーパスにおける発話単位タグの一推定手法, 人工知能学会誌, Vol. 14, No. 2, pp. 273–281 (1999).

[荒木 99] 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人：発話単位タグ標準化案の作成, 人工知能学会誌, Vol. 14, No. 2, pp. 251–260 (1999).