

教師なし手法による対話行為分類への換言器の適用

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、利用者の対話行為ならび発話意図の推定は非常に重要である。そのため、対話行為タグ付きコーパスがこれまでも整備されてきた。それによって、対話行為の分析や、教師あり機械学習手法による対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、対話のドメインやタスクによって要求される対話行為タグの種類や、その粒度が異なる。これまでも様々なドメインにおいて多くの対話行為集合が考案され、タグ付きコーパスが整備されてきた (たとえば, [Ara99, Jur97] 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対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タグ付けは必須であり、コーパスとしての一貫性を持った高品質なタグ付けは、非常に高価である。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、この場合にもコストがかかる。本稿では、そういった教師あり学習手法を前提とせず、教師なし手法を前提とし、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。

対話行為を推定することは、有限の対話行為ラベルを付けることとみなせる。したがって、ある発話を与えられたときに、そこから特徴を抽出し、それに基づいて発話を分類することができれば、対話行為の分類が可能になる。対話行為分類ののち、適切なラベルを各分類に対して付けることで対話行為推定を行える。本稿では、発話の特徴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。これらから、特徴ベクトルを作成するが、これ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, たとえば文献 [Lan98] 参照) が提案された。Serafinらは、LSA を対話行為分類に適用し効果的であることを示している [Ser04]。

本研究でも、同様のアプローチをとる。発話に対して換言器を適用し、表現の多様性を抑える。そして、そこから特徴ベクトルを作成し、LSA の確率的拡張である $plsi$ [Hof99] を用いて次元圧縮を行う。次元圧縮されたベクトルを教師なしクラスタリング手法によって分類する。

2 換言器

人間は、同一の意味を表すために、多様な表現を用いるが、計算機上では、これらの多様な表現はすべて別のものとして扱われる。そこで、本研究では、特徴ベクトルを次元圧縮する前に、換言器を適用し、表現の多様性を圧縮することを考える。特に日本語においては、敬語の運用によって文末表現の多様性が顕著であり (たとえば、ください-いただけないでしょうか-もらえませんか、など)、これらの表現を換言しても、対話行為推定には、ほとんど影響がない。そのため対話行為分類においては、このような換言が有効に機能すると期待できる。

我々はすでに、敬語を主な対象として単一の表現へ換言する規則に基づく換言器を発表している [Oht01] ので、本研究では、これに若干の修正を加えたものを用いる。より具体的には、換言器は、形態素解析ならびに文節まとめあげまでの処理がされた入力を前提としている。換言規則は、形態素解析結果と文節境界をコンパクトにまとめた形式に対して正規表現によるマッチングを行い目的とする表現へと置換する形式で記述する。この換言器は、Perl によって記述され、現在 600 ほどの規則がある。

3 教師なし手法による対話行為の分類

特徴ベクトルを次元圧縮するために、教師なし学習手法を用いる $plsi$ を適用する。隠れ変数モデルである $plsi$ は、文書と単語の共起確率を与えるモデルである。 $plsi$ モデルにおける隠れ変数は、トピックを表すと言われる。本研究では、このトピックによって対話行為を近似できると考える。 $plsi$ モデルでは、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 (隠れ変数の値) 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) から対話行為の近似であるトピックを推定する。

$plsi$ では、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。この圧縮された特徴ベクトルをトピックベクトルと呼ぶことにする。

3.1 素性

plsi は文書と単語の共起確率を与えるモデルであるため、学習 (パラメータ推定) に用いる素性は単語になる。しかし、Serafin らが LSA に対して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。

一方で、対話行為を分類する上で、どのような素性が有効なのか現段階では、はっきりしていない。本研究では、各発話を形態素解析し、形態素を単位として、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せず、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。

また、対話行為を分類しようとする発話の直前の発話に含まれる情報も有効であると考え、直前の発話から得られる素性に、直前の発話のものであるというシンボルをつけて、素性に含めた。これは、教師あり学習を対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対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が、特定の対話行為集合を想定せずに、対話行為を推定しようとしているので、そのようなラベルを用いることができないためである。

また、韻律情報の利用が対話行為推定に貢献することもわかっているが (たとえば, [Shr98]), 本研究の現状では、韻律情報を考慮せず、音声の書き起こしから得られる情報のみを扱う。

4 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) を評価に用いる。学習データは、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対話行為タグを発話単位に人手で付与した。今回付与した対話行為タグは、比較的短いセグメントに対して付与すること前提に設計されている。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 という対話行為と考える) と、「…がございます」(情報伝達の対話行為と考える) の両方の対話行為が含まれることになる。このような発話に対して、対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 対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対話行為をその発話を代表する対話行為として取り扱う。

対話行為ラベルは、全部で 26 種類あるが、評価用データにおいてタグ付与されたのは、表 1 にあげる 16 種である。

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 mecab¹ 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ため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²。plsi や LDA [Ble03] のような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。

そこで、まず用いるべき隠れ変数の数 (トピックベクトルの次元) をいくつにするかを決定するための実験を行った。評価用データに対して、トピックベクトルを作

¹<http://mecab.sourceforge.net/>

²使用したツール:<http://chasen.org/~taku/software/plsi/>

表 1: 換言器を用いた場合とそうでない場合のクラスタリング結果

対話行為 ラベル (頻度)	換言なし (クラスタ ID: 頻度)	換言あり (クラスタ ID: 頻度)
ACK (68)	B:7, G:24, H:2, I:1, M:3, N:7, O:8, P:16	a:12, c:14, f:2, g:4, k:3, n:8, p:25
ACT-REQ (44)	B:4, C:3, D:1, F:1, H:6, I:1, J:1, K:1, M:11, N:10, O:5	b:1, c:1, d:1, f:2, g:2, i:7, j:25, k:4, l:1
ALERT (1)	N:1	o:1
APOLOGY (2)	H:1, N:1	n:1, o:1
CONF-Q (29)	B:4, C:1, D:1, H:1, I:3, J:4, K:1, L:1 M:3, N:8, O:2	c:2, d:1, e:6, f:1, g:2, i:2, n:5, o:10
FAREWELL (16)	K:8, M:3, N:5	g:1, i:2, k:2, n:1, o:10
G-WISHES (1)	K:1	o:1
GREET (8)	B:3, M:1, N:4	g:3, o:5
INFORM (198)	B:27, C:11, D:10, E:6, F:2, G:2, H:35, I:7, J:12, K:7, L:14, M:26, N:27, O:12	a:11, b:2, c:1, d:4, e:6, f:16, g:28, h:13, i:10, j:3, k:16, m:18, n:5, o:63, p:2
PERM-REQ (1)	E:1	g:1
SUGGEST (6)	F:3, H:2, M:1	c:2, f:2, m:1, o:1
THANK (20)	A:16, I:2, K:1, N:1	a:1, g:2, i:2, l:10, n:4, o:1
THANK-RES (2)	K:2	o:2
WH-Q (40)	C:1, D:2, F:16, H:4, I:2, K:7, L:2, M:3, O:1, P:2	a:1, c:19, f:6, g:1, i:10, k:2, n:1
YES (18)	B:6, G:2, H:1, I:5, O:4	a:6, g:9, o:1, p:2
YN-Q (35)	B:2, C:4, F:4, H:6, I:1, M:8, N:7, O:3	a:2, b:4, c:12, f:4, g:2, h:1, i:4, j:1, n:1, o:4

成し、各対話行為ラベル毎に平均ベクトルを作成した。この平均ベクトルが自分自身を除く他の対話行為ラベルの平均ベクトルとなす角の \cos 値を求め、それらの値の平均値を求めた。すべての対話行為ラベルに対する平均値を平均し、それで評価した。言い換えると、この値が低ければ低いほど、あるラベルと他の対話行為ラベルをはっきりと区別できることになる。また、plsi において過学習問題を軽減するための温度パラメータは予備検討をした結果から 0.9 とした。実験結果は、隠れ変数の数を 10, 50, 100, 200, 300 と変化させた場合に、それぞれ 0.607, 0.334, 0.288, 0.290, 0.275 となったため、クラスタリング時の計算量も考慮し、隠れ変数は 100 で充分と判断した。

次に、換言器を適用した場合と、しない場合を比較した。コーパス全体へ換言器を適用した結果、83,052 発話のうち 56,027 発話が換言された。換言器を適用したコーパスと、元のコーパスからそれぞれ隠れ変数の数を 100 とした plsi モデルを作成した。そして、同じく換言器を適用した評価用データと、そうでない評価用データからそれぞれ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。クラスタ数は、人間による対話行為ラベル数と同一の 16 である。この両者の結果を、表 1 にまとめる。なお、クラスタ ID に、換言器を適

用した場合とそうでない場合ともにアルファベットを用いているが、両者に直接的な関係はないため、換言器を適用しなかった結果には、大文字を、換言器を適用した場合の結果には小文字をクラスタ ID として使用した。紙面の都合から、頻度が 0 となったクラスタは表示していない。

6 考察

隠れ変数モデルを利用する際の問題として、隠れ変数の数をいくつにすべきかがある。今回の実験においては、16 程度の対話行為ラベルを識別するために必要となるトピックベクトルの次元を 100 と見積もった。より高次元の方がより高い性能を出せる可能性があるが、クラスタリングにおける計算量も大きくなる。

換言器を適用した場合と、そうでない場合の、K-means 法によるクラスタリング結果を、人間による対話行為ラベルと比較した。残念ながら、今回の実験結果からは、換言器が期待するように機能したと結論することができない。たとえば、ACT-REQ のクラスタリングにおいては、換言器を用いない場合、大きく 2 つのクラスタ (M と N) にまたがっているが、換言器を用いることで、一つのクラスタ (j) に集約される傾向が強まり、換言器が期待するように動作していると見ることができる。一方で、THANK のクラスタリングでは、換言器を用いない場合、特定のクラスタ (A) と

強い関係があるのに対し、換言器を適用した場合、より多くのクラスタに分類される結果となった。

また、今回使用した対話行為タグの設計の問題もある。特に INFORM に関しては、特定のクラスタと関連づけることができなかった。これは、INFORM タグそのものが、他のタグに分類できないような発話に対して付与されており、INFORM に対応するようなクラスタは、寄せ集めとなることから必然的に巨大になる。

さらに、現状では、対話行為を推定するための特徴として使用している素性が、対話行為を識別するために充分であるかどうかを十分に検討できていない。対話行為を識別するために、必要十分な特徴空間を設計できないかぎり、いかなる次元圧縮手法、クラスタリング手法を適用しても高精度な対話行為分類は困難と考える。たとえば、換言器を適用しなかった結果におけるクラスター ID が N のクラスタでは、「お願いします」という表現が多用されている。ACT-REQ や FAREWELL などの対話行為がラベル付けされる発話で、「～してください、お願いします」(ACT-REQ) や、「よろしくお願いします」(FAREWELL) とこの表現を多用することは容易に想像が付く。このように、いずれの対話行為においても用いる表現を特徴とすべきではないが、plsi による次元圧縮の結果どのように取り扱われているかまで、分析できていない。クラスタリングの結果から、おそらく特徴として扱われているのではないかと想像する。一方で、次元圧縮をせずに、高次元かつ疎なベクトルに対する効率のよいクラスタリングを適用するアプローチもあり、その検討は今後の課題である。

我々は、現在、対話システム構築を前提としたコーパス整備をすすめている [Oht08]。そのコーパスで付与しているタグセットは、対話システム構築を考慮して設計している。今後は、このコーパスに対して、対話行為の教師なし手法による分類を検討する予定である。本稿で使用したコーパスならびに、現在整備をすすめているコーパスともに、2 者による対話であり、それぞれの役割が限定された対話である。そのことから、これらの役割を前提として、対話における発話の特徴づけることができれば、さらに対話行為推定の精度が向上する可能性がある。

7 まとめと今後の課題

本稿では、教師なし手法による対話行為分類において換言器を適用し、表現を圧縮することを試みた。一部の結果では、期待するように換言器が機能し、よりよい結果を得られた。一方で、評価に用いた対話行為タグセットの本質の問題や、特徴空間の設計に付随する問題もあることがわかっている。今後は、現在整備しているコーパスに基づいて分析をすすめ、特徴空間の設計、次元圧縮によらずに、高次元かつ疎なベクトルをに対して効率的なクラスタリングを適用し、教師なし手法による対話行為分類が可能かどうか検討する。

参考文献

- [Ara99] 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).
- [Ble03] BLEI, D. M., NG, A. Y., and JORDAN, M. I.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022 (2003).
- [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 (1999).
- [Jur97] JURAFSKY, D., SHRIBERG, E., and BIASCA, D.: Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Draft 13, Technical report, University of Colorado at Boulder & SRI International (1997).
- [Lan98] LANDAUER, T. K., FOLTZ, P. W., and LAHAM, D.: An Introduction to Latent Semantic Analysis, *Discourse Processes*, Vol. 25, pp. 259–284 (1998).
- [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 IC-SLP '94*, pp. 1791–1794 (1994).
- [Oht01] OHTAKE, K. and YAMAMOTO, K.: Paraphrasing Honorifics, In *Workshop Proceedings of Automatic Paraphrasing: Theories and Applications (NLPRS2001 Post-Conference Workshop)*, pp. 13–20 (2001).
- [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).
- [Oht08] 大竹清敬, 堀智織, 柏岡秀紀, 中村哲: 京都観光案内対話コーパスにおける対話行為の分析, 言語処理学会第 14 回年次大会 発表論文集 (2008).
- [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).
- [Shr98] SHRIBERG, E., BATES, R., STOLCKE, A., TAYLOR, P., JURAFSKY, D., RIES, K., COCCARO, N., MARTIN, R., METEER, M., and ESS-DYKEMA, C. V.: Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?, *LANGUAGE AND SPEECH Special Issue on Prosody and Conversation*, Vol. 41, No. 3–4, pp. 439–487 (1998).
- [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).