

対話システムの研究課題

第8回対話システムシンポジウム チュートリアル
2017年10月12日

(株)ホンダ・リサーチ・インスティテュート・ジャパン

中野 幹生

nakano@jp.honda-ri.com

- (株)ホンダ・リサーチ・インスティテュート・ジャパン プリンシパル・リサーチャ
- 仕事：対話システムの基礎研究 (> 25 years)
- 職歴
 - 1990-2004：日本電信電話(株) 基礎研究所, コミュニケーション科学基礎研究所
 - 2004-現在：(株)ホンダ・リサーチ・インスティテュート・ジャパン
- 書籍執筆（共著）
 - 対話システム（コロナ社, 2015）
 - 話し言葉対話の計算モデル（電子情報通信学会, 2014）
- 対話システム関係の主な学会活動
 - SIGDial Board Member (2009-2013)
 - SIGDIAL 2010 General Chair
 - Areal Chair (Dialogue and Discourse Area)
 - ACL 2012, IJCNLP 2017
 - 人工知能学会論文誌 論文特集「知的対話システム」
 - 編集委員長（2013） 編集委員（2015, 2017）
 - 国際会議査読・プログラム委員
 - SIGDIAL, ACL, NAACL, EMNLP, IJCNLP, IWSDS, IUI, HRIなど

本チュートリアルの目的

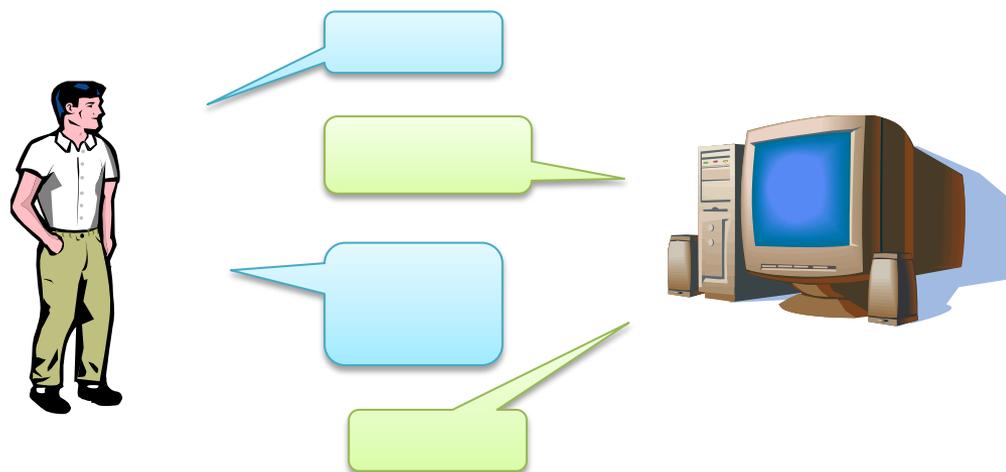
- 対話システムの研究をしようとしている方に、対話システム研究の目的と研究テーマの見つけ方を紹介したい
- 詳細よりも概観
- これまでよりもこれからに重点をおく
- サービス・製品開発には直接役立たないかもしれないが、今後どんな技術が出てくるかを考えるのに役立つと嬉しい

- 対話システム研究の目的
- 対話システムの研究テーマの分類
- 今後の展望

対話システム研究の目標

対話システムとは

- 人間と言語で情報を授受する機械
- 複数のやりとりを行う（1問1答ではない，履歴を使う）



- システムとユーザの知識に齟齬があってもやりとりの中で解消
 - 対話することの楽しみを提供
- ⇒ 社会の様々な問題に適用可能

対話システム研究の目標

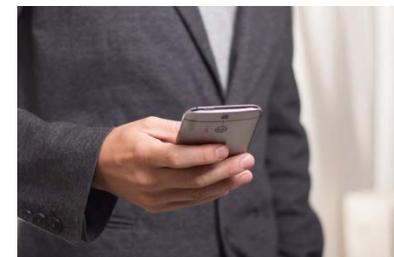
- 対話システム構築の**ベストプラクティス**
 - どうやったら良い対話システムが作れるか？
 - アーキテクチャ
 - 要素技術
 - 知識の記述方法
 - 統計モデルの訓練データの集め方

- 以下の2つが異なればベストプラクティスは違ってくる！
 - 作ろうとしている対話システムの**タイプ**
 - 開発時・運用時の**制約条件**

- 一人で全部できないので、コミュニティで**共有**
→ 新しいシステムを作るときの直観を持つ

対話システムのタイプ

- 入出力の**モダリティ**
 - テキスト
 - 音声
 - マルチモーダル（音声，画像，その他のセンサ，CGエージェント，ロボット）
- 達成すべき**目標**の有無・種類
 - タスク指向型（情報検索，説明，インタビュー，説得，交渉，...）
 - 非タスク指向型（雑談，傾聴，クイズ，ゲーム，...）
 - 混合型
- 対話の**ドメイン**
 - 単一ドメイン（フライト，ホテル，野球，レストラン，...），マルチドメイン，オープンドメイン
- 対話**参加者の数**
 - 1対1，マルチパーティ

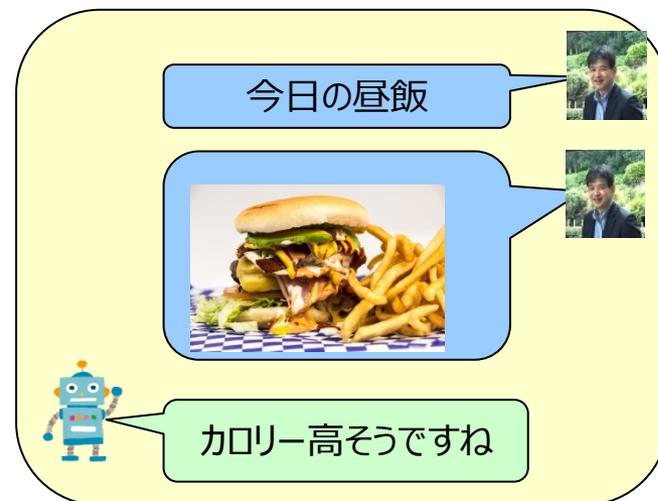


開発時・運用時の制約条件

- 開発コスト・保守コスト
 - 対話知識記述
 - コーディング
 - 統計モデル用訓練データの収集・アノテーションコスト
- 運用コスト
- ハードウェアの制約
- 想定ユーザ・設置場所

対話システムのタイプと制約条件の例

モダリティ：テキスト＋画像認識・表示
 タスク：雑談＋情報検索＋インタビュー＋推薦
 ドメイン：単一（料理）
 参加者数：1対1
 制約条件：スマートフォン＋サーバで動作



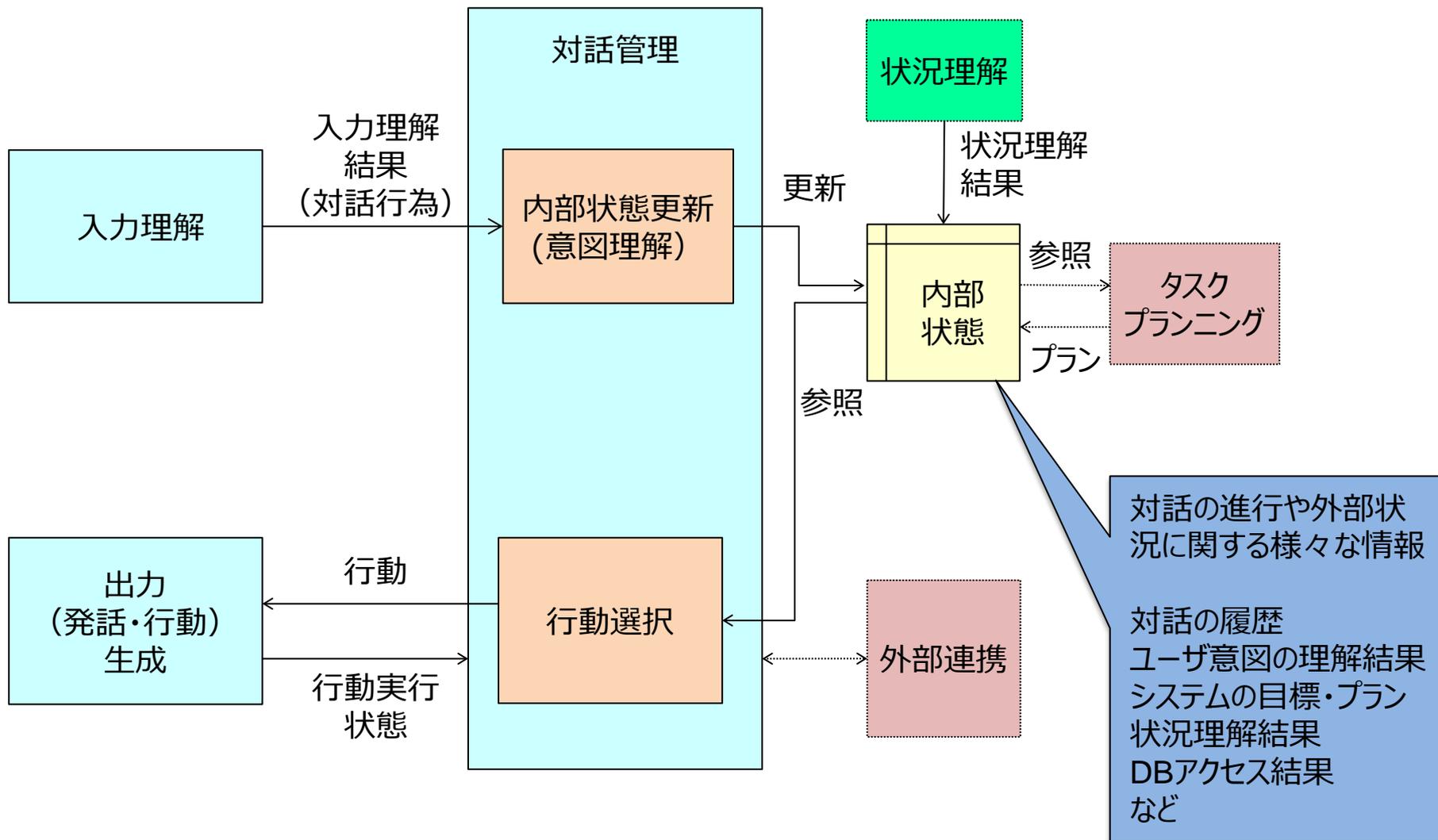
モダリティ：音声入出力
 ＋大画面上の画像表示
 ＋バーチャルエージェント
 タスク：案内＋説明
 ドメイン：観光
 参加者数：1対多
 制約条件：対話シナリオの作成
 コストがかけられない



[中野 15]

対話システムの研究テーマの分類

対話システムの構成



研究テーマのタイプ

1. 新規特長型
 - 対話システムに新たな特長を持たせる → システムが良くなる
2. 開発・運用コスト削減型
 - 少ないコストで開発・運用ができる
3. 性能向上型
 - 新しい特長はないが，モジュールの改良でシステムが良くなる
4. 評価コスト削減型
 - 対話システムの良さを少ない労力で評価する方法を提案

評価方法 (1) システム全体の良さ

- 主観評価
 - 対話後にアンケート
 - 第三者による評価
- 客観評価
 - タスク達成率
 - 例：情報提供できたか，説得できたか，情報取得できたか，適切に広告出せたか？
 - 対話の長さ
 - 短い方が良い場合も，長い方が良い場合もある
 - 対話中のユーザの行動を分析
 - 例：発話長，発話回数
 - 人間同士の対話との類似度(?)
- システムの良さの定義は作りたいシステム次第

評価方法 (2) モジュール毎の評価

- そのモジュールが良くなればシステム全体が良くなりそうなのが
明らかな場合
- 評価指標
 - 一般的に使われているもの
 - 自分で定義
- あらかじめ収録しておいた人システム間対話データで
オフライン評価も可能

1. 新規特長型テーマ

- 新しい特長：人間がやっていることで、システムがまだできていないこと



- 特長の見つけ方
 - 人間同士の対話の分析
 - 人文系の対話研究

- まったく新しい特長を見つけるのは難しい
 - 他のタイプのシステムで研究されている特長を横展開

- タイプ・条件×特長は大きいマトリックス
 - Open Problemが多い
 - 例：スムーズな話者交替: マルチモーダル多人数対話
 - テキストchatbot
 - 例：ヘルプ生成：情報検索対話→ インタビュー対話

特長の例

スムーズな話者交替

漸次的な理解・生成

対話中の知識獲得

状況に依存した理解・生成

感情・態度・対話参加度

同調

ユーザ適応

パーソナリティ表出

ラーハンドリング

ヘルプ生成

インタラクションデザイン

特長の例（詳細1）

- スムーズな話者交替・発話開始タイミング
 - 誰に向かって話しているかの推定 [Bohus 10, Nakano 13]
 - 話し始めるべきときに話し, 黙るべきときに黙る
 - 話者交替 (turn-taking) [DeVault 09]
 - 対話参加度 (engagement) [Ishii 13, Bohus 14]
 - 相槌(backchannel)をうつ [Lala 17]
- 漸次的理解・生成
 - ユーザが話しているそばからどんどん理解 [Manuvinakurike 16]
 - 発話の途中でもユーザの反応や状況の変化に対応 [Skantze 10]
- 対話中の知識獲得
 - 対話しながら知らない単語や知識を獲得 [Ono 17]

特長の例（詳細2）

- 状況依存理解・生成
 - 周りの状況を考慮にいれてユーザの意図を理解 [Kennington 14]
 - 周りの状況を考慮にいれて発話
- 感情や態度の利用
 - ユーザの感情や態度に応じて行動を変える [Hasegawa 13]
 - システムが感情を表出 [Niewiadomski 09]
- 同調
 - 対話のリズムや語彙をユーザに合わせる [Lopes 15]
- インタラクションデザイン
 - 認知負荷を減らす, いらいらさせない, 信頼関係を構築するためのシステム行動 [Gašić 12 Huang 11, Zhao 16]

特長の例（詳細3）

- パーソナライゼーション・ユーザ適応
 - ユーザの個人の興味や知識レベルを獲得して，それに応じた行動をする
[Komatani 09b]
- パーソナリティ表出
 - システムの話し方をパーソナリティに応じて変更 [Liu 16]
 - 話す内容の一貫性を保つ
- 破綻の検出・エラーハンドリング
 - うまくいかない対話を検出してリカバリ [Higashinaka 15, Bohus 09]
- ヘルプ生成
 - ユーザがシステムの使い方がわからない時，システムがヘルプを出す
[Komatani 09a]

2. 開発・運用コスト削減型テーマ

- 知識記述のコスト削減
 - (辞書, オントロジ, 言語理解・生成用知識, 発話候補文, 対話シナリオ, システムパーソナリティなど)
 - 知識記述を容易にするための記述形式やツール
 - Webなどからなるべく自動的に知識獲得
 - 対話ログから知識獲得
 - End-to-end learning (人間同士の対話から全部学習) [Vinyals 15, Li 16]
- インテグレーションのコスト削減
 - ツールキット・フレームワーク (MMDAgent, PyDialなど)
 - モジュールのマイクロサービス化 (Microsoft LUISなど)

3. 性能向上型テーマ

要素技術の例

入力理解

意図理解

行動選択（対話管理）

言語・行動生成

ドメイン選択

話者交替タイミング

発話参加度推定

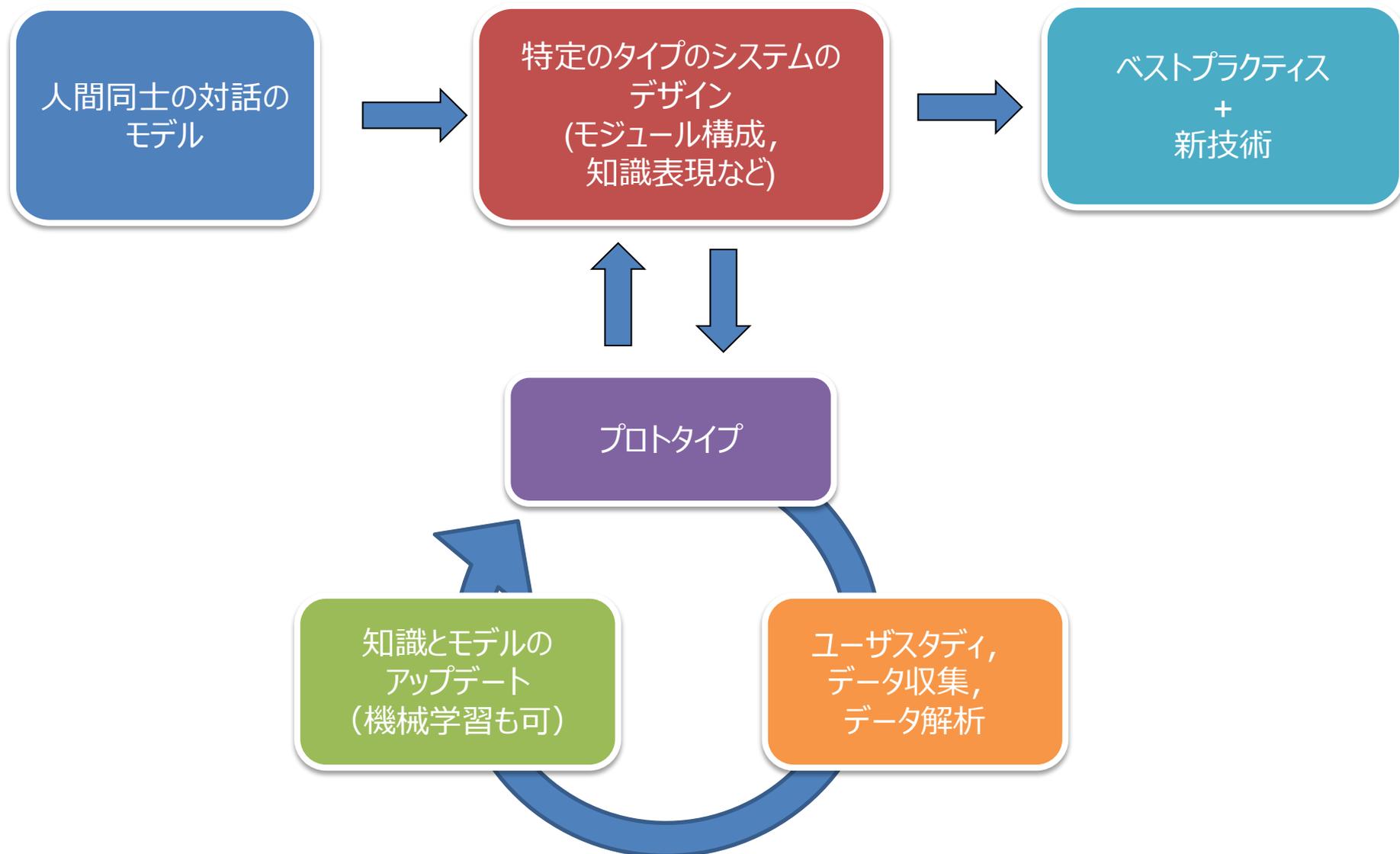
感情・態度・興味の推定

- 要素技術の性能の向上が目標
- タイプ1：従来使われていなかった情報を使う
 - 例：ユーザの位置，視線，DBの内容を意図理解に利用 [Misu 14]
- タイプ2：使う情報は同じだが新しい手法
 - 新しい機械学習の手法など
 - 音声言語理解
 - 雑談対話システムの応答選択 [Lowe 15, Inaba 16]
 - 強化学習による対話制御の学習
 - データセットが共有されているものもある
 - 対話破綻検出チャレンジ [東中 16]
 - 意図理解：Dialogue State Tracking Challenge [Kim 16]
 - 言語生成：E2E NLG Challenge [Novikova 17]
 - Facebook bAbI

4. 評価コスト削減型テーマ

- アンケートをとらなくても評価したい
 - ユーザ行動の客観指標とアンケート結果の相関を使う [Walker 98]
- 被験者実験せずに評価したい
 - ユーザシミュレータ [Schatzmann 06]
 - 統計モデルの訓練やバグ発見にも使える
- あるモジュールのオフライン評価指標を決めたい
 - システム全体のパフォーマンスとの相関を使う [Higashinaka 04]

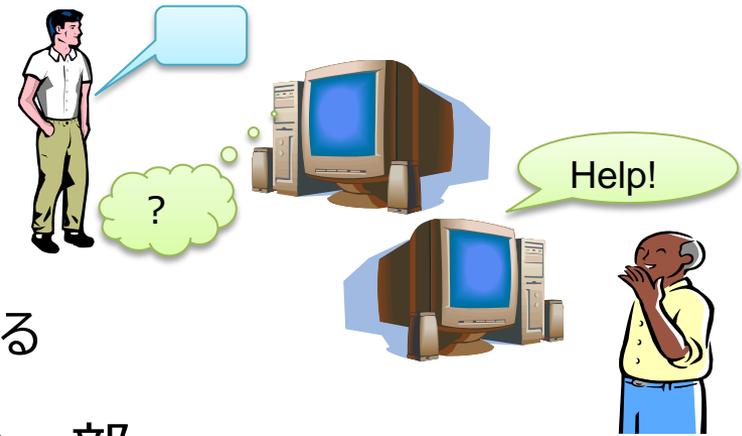
対話システム研究のサイクル



今後の展望

今後発展しそうなシステム

- マネタイズできそうなシステム
 - 広告・宣伝
 - 人手不足解消・サービスの均一化
 - 問い合わせ対応, 傾聴 [Meguro 13]
 - インタビュー [Kobori 16]
 - ゲーム [Kobayashi 15]
- 人間と協調するシステム
 - 機械が頑張ってもできないことだけ人間が助ける
- 対話システム単独ではなく, 他のシステムの一部
 - インタラクティブ情報検索 [Kelly 09]
 - インタラクティブ機械学習 [Amershi 14]
 - 移動ロボット [Nakano 11]



今後考慮すべき制約条件

- 提供する情報の正確さ, 倫理性
 - Web上の不正確な記事に基づく発話を避ける
 - 公序良俗に反する内容を避ける
- 個人情報保護
 - 発話内容に含まれている個人情報や入力画像などの個人情報を保護しつつ, パーソナライゼーション
- バリアフリー
 - e.g., 対話型デジタルサイネージの視覚・聴覚に障害がある人への対応

今後取り入れられそうな技術

- 社会的信号処理
 - 視線, ジェスチャー, 言語, パラ言語などに現れる意図
 - 人間同士のマルチモーダルインタラクションのモデル化

- 人以外の画像の理解
 - アップロードされた写真の理解
 - 対話環境の認識
 - 実世界の参照表現理解・生成

- 新しい入出力モダリティ
 - IoT, 高精度GPS,
 - VR, AR

おわりに

- まとめ
 - 対話システム研究の目的はベストプラクティスの共有
 - 対話システムのタイプと制約条件に応じて多くの未解決問題
 - お手本は人間
- 是非新しい対話システム, 新しい課題にチャレンジしてください!
- ご意見・ご質問・間違いの指摘は nakano@jp.honda-ri.com まで
- 謝辞: 日頃より情報共有して下さる対話システム研究コミュニティの皆様へ感謝します.

- 質疑応答の際に言及した本は、川添愛著「働きたくないイタチと言葉がわかるロボット」（朝日出版社，2017）です。現状の統計的自然言語処理技術による言語理解の可能性と限界が寓話で分かりやすく説明されています。

論文以外に参考になる資料

- 中野 幹生「対話システム研究の進め方」SLUD第60回研究会
 (2010年10月) チュートリアル (2015年8月改訂版)
<https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxkaWFsb2dzeXN0ZW1zanB8Z3g6MThmNGZjNWZiMjQ5OTVzMg>
- 河原達也：音声対話システムの進化と淘汰：歴史と最近の技術動向,
 人工知能学会誌 28(1), 45-51, (2013)
- 東中 竜一郎, 船越 孝太郎：対話システムの理論と実践, 言語処理
 学会 第22回年次大会 チュートリアル資料 (2016)

なるべく新しく、かつ、研究のイメージが湧くものを少数あげました。サーベイではないので網羅的ではありません。

- [Amershi 14] Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T.: Power to the People: The Role of Humans in Interactive Machine Learning, *AI Magazine*, Vol. 35, No. 4, pp. 105-120 (2014)
- [Bohus 09] Bohus, D. and Rudnicky, A. I.: The RavenClaw dialog management framework: Architecture and systems, *Computer Speech and Language*, Vol. 23, No. 3, pp. 332-361 (2009)
- [Bohus 10] Bohus, D. and Horvitz, E.: Facilitating Multiparty Dialog with Gaze, Gesture, and Speech, in *Proc. 12th ICMI* (2010)
- [Bohus 14] Bohus, D. and Horvitz, E.: Managing Human-Robot Engagement with Forecasts and... um... Hesitations, in *Proc. 16th ICMI*, pp. 2-9 (2014)
- [DeVault 09] DeVault, D., Sagae, K., and Traum, D.: Can I Finish?: Learning When to Respond to Incremental Interpretation Results in Interactive Dialogue, in *Proc. 10th SIGDIAL Conference*, pp. 11-20 (2009)
- [Gašić 12] Gašić, M., Tsiakoulis, P., Henderson, M., Thomson, B., Yu, K., Tzirkel, E., and Young, S. J.: The Effect of Cognitive Load on a Statistical Dialogue System, in *Proc. 13th SIGDIAL Conference*, pp. 74-78 (2012)
- [Hasegawa 13] Hasegawa, T., Kaji, N., Yoshinaga, N., and Toyoda, M.: Predicting and Eliciting Addressee's Emotion in Online Dialogue, in *Proc. 51st ACL*, pp. 964-972 (2013)
- [Higashinaka 04] Higashinaka, R., Miyazaki, N., Nakano, M., and Aikawa, K.: Evaluating Discourse Understanding in Spoken Dialogue Systems, *ACM Transactions on Speech and Language Processing*, Vol. 1, No. 1, pp. 1-20 (2004)

- [Higashinaka 15] Higashinaka, R., Mizukami, M., Funakoshi, K., Araki, M., Tsukahara, H., and Kobayashi, Y.: Fatal or not? Finding errors that lead to dialogue breakdowns in chat-oriented dialogue systems, in Proc. EMNLP 2015, pp. 2243-2248 (2015)
- [東中 14] 東中 竜一郎 : 雑談対話システムに向けた取り組み, 人工知能学会研究会資料SIG-SLUD-70 (2014)
- [Huang 11] Huang, L., Morency, L.-P., and Gratch, J.: Virtual Rapport 2.0, in Proc. 11th IVA, pp. 68-79 (2011)
- [Inaba 16] Inaba, M. and Takahashi, K.: Neural Utterance Ranking Model for Conversational Dialogue Systems, in Proc. 17th SIGDIAL Conference, pp. 393-403 (2016)
- [Ishii 13] Ishii, R., Nakano, Y. I., and Nishida, T.: Gaze awareness in conversational agents: Estimating a user's conversational engagement from eye gaze, ACM Trans. on Interactive Intelligent Systems, Vol. 3, No. 2 (2013), Article No. 11, pp. 1-25
- [Kelly 09] Kelly, D.: Methods for Evaluating Interactive Information Retrieval Systems with Users, Foundations and Trends in Information Retrieval, Vol. 3, No. 1-2, pp. 1-224 (2009)
- [Kennington 14] Kennington, C., Spyros, K., and Schlangen, D.: Situated Incremental Natural Language Understanding using a Multimodal, Linguistically-driven Update Model, in Proc. 25th COLING (2014)
- [Kim 16] Kim, S., D'Haro, L. F., Banchs, R. E., Williams, J. D., Henderson, M., and Yoshino, K.: The Fifth Dialog State Tracking Challenge, in Proc. SLT 2016 (2016)
- [Kobayashi 15] Kobayashi, H., Tanio, K., and Sassano, M.: Effects of Game on User Engagement with Spoken Dialogue System, in Proc. 16th SIGDIAL Conference, pp. 422-426 (2015)
- [Kobori 17] Kobori, T., Nakano, M., and Nakamura, T.: Small Talk Improves User Impressions of Interview Dialogue Systems, in Proc. 18th SIGDIAL Conference, pp. 370-380 (2017)
- [Komatani 09a] Komatani, K., Ikeda, S., Fukubayashi, Y., Ogata, T., and Okuno, H.G.: Ranking help message candidates based on robust grammar verification results and utterance history in spoken dialogue systems, in Proc. 10th SIGDIAL Conference, pp. 314-321 (2009)

- [Komatani 09b] Komatani, K. and Rudnicky, A. I.: Predicting Barge-in Utterance Errors by using Implicitly Supervised ASR Accuracy and Barge-in Rate per User, in Proc. 47th ACL/4th IJCNLP, pp. 89-92 (2009)
- [Lala 17] Lala, D., Pierrick, M., Inoue, K., Ishida, M., Zhao, T., and Kawahara, T.: Attentive listening system with backchanneling, response generation and flexible turn-taking, in Proc. 18th SIGDIAL Conference, pp. 127-136 (2017)
- [Li 16] Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., , and Dolan, B.: A Persona-based Neural Conversation Model, in Proc. 54th ACL, pp. 994-1003 (2016)
- [Liu 16] Liu, K., Tolins, J., Tree, J. E. F., Neff, M., and Walker, M. A.: Two Techniques for Assessing Virtual Agent Personality, IEEE Transactions on Affective Computing, Vol. 7, No. 1, pp. 94-105 (2016)
- [Lopes 15] Lopes, J., Eskenazi, M., and Trancoso, I.: From rule-based to data-driven lexical entrainment models in spoken dialog systems, Computer Speech and Language, Vol. 31, No. 1, pp. 87-112 (2015)
- [Lowe 15] Lowe, R., Pow, N., Serban, I. V., and Pineau, J.: The {Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems, in Proc. 16th SIGDIAL Conference, pp. 285-294 (2015)
- [Manuvinakurike 16] Manuvinakurike, R. R., Kennington, C., DeVault, D., and Schlangen, D.: Real-Time Understanding of Complex Discriminative Scene Descriptions, in Proc. 17th SIGDIAL Conference, pp. 232-241 (2016)
- [Meguro 13] Meguro, T., Minami, Y., Higashinaka, R., and Dohsaka, K.: Learning to control listening-oriented dialogue using partially observable markov decision processes, ACM Trans. on Speech and Language Processing, Vol. 10, No. 4, p. 15 (2013)
- [Misu 14] Misu, T., Raux, A., Gupta, R., and Lane, I.: Situated Language Understanding at 25 Miles per Hour, in Proc. 15th SIGDIAL Conference, pp. 22-31 (2014)
- [Nakano 11] Nakano, M., Hasegawa, Y., Funakoshi, K., Takeuchi, J., Torii, T., Nakadai, K., Kanda, N., Komatani, K., Okuno, H. G., and Tsujino, H.: A multi-expert model for dialogue and behavior control of conversational robots and agents, Knowledge-Based Systems, Vol. 24, No. 2, pp. 248-256 (2011)

- [Nakano 13] Nakano, Y. I., Baba, N., Huang, H.-H., and Hayashi, Y.: Implementation and evaluation of a multimodal addressee identification mechanism for multiparty conversation systems, in Proc. 15th ICMI, pp. 35-42 (2013)
- [中野 15] 中野 幹生, 駒谷 和範, 船越 孝太郎, 中野 有紀子 : 対話システム, コロナ社 (2015)
- [Niewiadomski 09] Niewiadomski, R., Hyniewska, S., and Pelachaud, C.: Modeling Emotional Expressions as Sequences of Behaviors, in Proc. 9th IVA, pp. 316-322 (2009)
- [Novikova 16] Novikova, J., Duvšek, O., and Rieser, V.: The E2E Dataset: New Challenges For End-to-End Generation, in Proc. 18th SIGDIAL Conference, pp. 201-106 (2016)
- [Ono 17] Ono, K., Takeda, R., Nichols, E., Nakano, M., and Komatani, K.: Lexical Acquisition through Implicit Confirmations over Multiple Dialogues, in Proc. 18th SIGDIAL Conference, pp. 50-59 (2017)
- [Schatzmann 06] Schatzmann, J., Weilhammer, K., Stuttle, M., and Young, S.: A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies, The Knowledge Engineering Review, Vol. 21, No. 2, pp. 97-126 (2006)
- [Skantze 10] Skantze, G. and Hjalmarsson, A.: Towards Incremental Speech Generation in Dialogue Systems, in Proc. 11th SIGDIAL Conference, pp. 1-8 (2010)
- [Vinyals 15] Vinyals, O. and Le, Q.: A Neural Conversational Model, in Proc. ICML 2015 Deep Learning Workshop (2015)
- [Walker 98] Walker, M. A., Litman, D. J., Kamm, C. A., and Abella, A.: Evaluating Spoken Dialogue Agents with {PARADISE: Two Case Studies, Computer Speech and Language, Vol. 12, No. 3 (1998)
- [Zhao 16] Zhao, R., Sinha, T., Black, A. W., and Cassell, J.: Socially-Aware Virtual Agents: Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior, in Proc. 16th IVA, pp. 218-233 (2016)