音声対話アシスタントに関する最近の研究動向とYahoo! JAPAN研究所での取り組み

2017年11月9日

ヤフー株式会社 上席研究員 鍜治伸裕

YAHOO! JAPAN

自己紹介

名前

鍛治伸裕 博士(情報理工学)

略歴

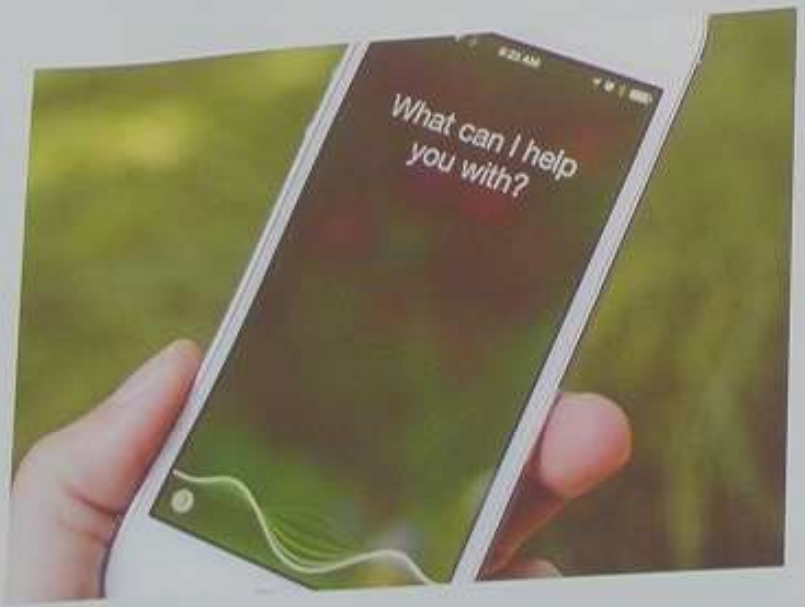
2005 東京大学 情報理工学系研究科 博士課程修了
2005～2015 東京大学 生産技術研究所 特任准教授等
2014～2015 情報通信研究機構 主任研究員
2015～現在 ヤフー株式会社 上席研究員

研究の興味

音声対話処理、言語の分散表現、ソーシャルテキスト解析

音声対話型インタフェースの時代

あらゆるデバイスに人工知能が搭載されるIoT時代の到来
ヒトとモノが音声対話でコミュニケーションできる世界が目前に

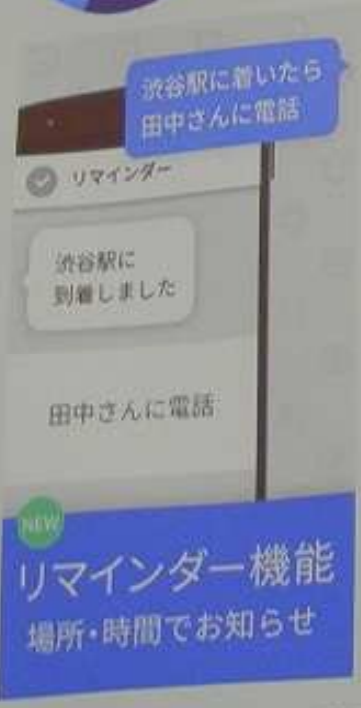


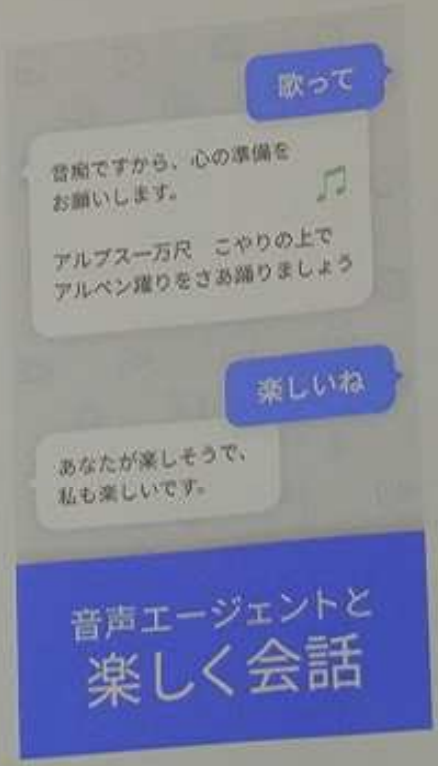
http://i.gzn.jp/img/2016/05/10/viv/s01.jpg
http://images.techhive.com/images/article/2017/01/echodothome-100704745-large.3x2.jpg

YAHOO! JAPAN

Yahoo! 音声アシスト

スマホ上で動作する音声対話システム
累計250万ダウンロード以上

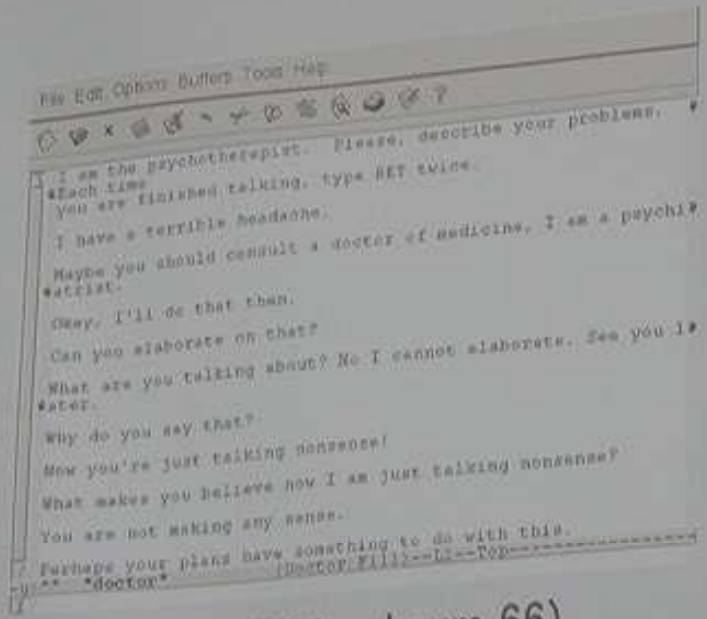




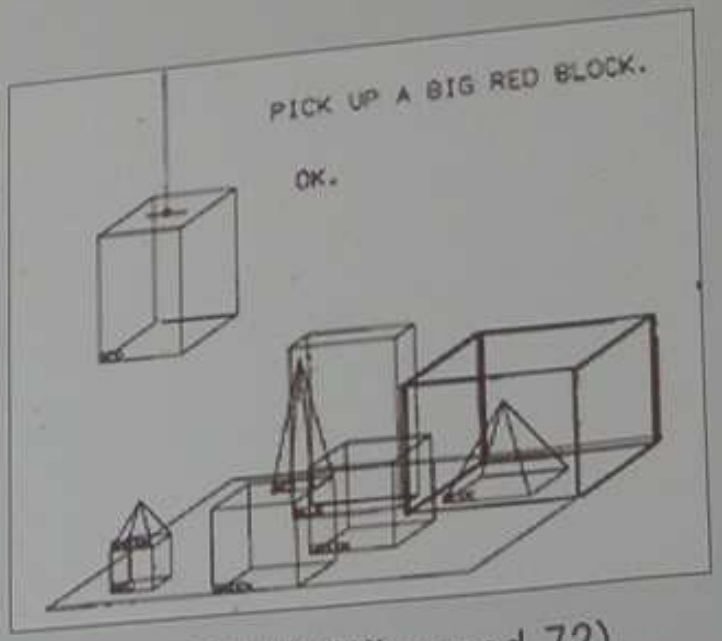
https://www.amazon.co.jp/gp/product/B000IWNED6

YAHOO! JAPAN

（音声）対話システムは半世紀以上の歴史を持つAIの古典的問題



Eliza (Weizenbaum 66)



SHRDLU (Winograd 72)

https://en.wikipedia.org/wiki/ELIZA
http://hci.stanford.edu/winograd/shrdlu

YAHOO! JAPAN

もう研究課題は残されていない？

強化学習に基づく理論的枠組みがすでに確立
対話システムに関する教科書が多数出版

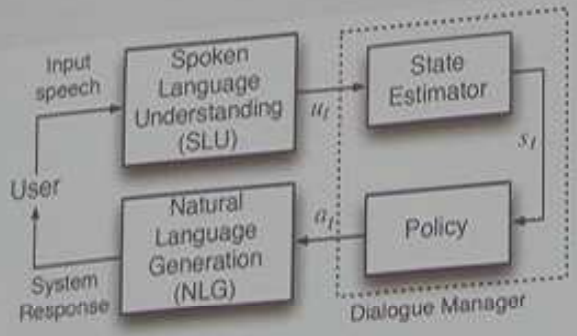


Fig. 1. Components of a finite state-based spoken dialogue system. At each turn the input speech is converted to an abstract representation of the user's intent u_t, the dialogue state s_t is updated and a deterministic decision rule called a *policy* maps the state into an action a_t in response.

有限状態音声対話モデル (Young et al. 2013)

対話システム(中野ら 2015)

http://www.coronasha.co.jp/np/isbn/9784339027570

ラボ環境から実環境へ

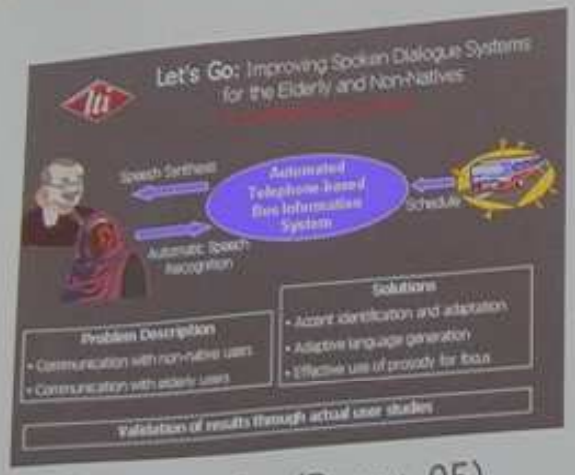
実際にサービスを運用して初めて顕在化する課題
- 新ドメインの迅速な追加 (Kim et al. ACL17)
- タスクと雑談の切り分け (Akasaki and Kaji ACL17)

本物の大規模ユーザを相手にする困難さ、面白さ
- ユーザ満足度の自動化 (Jiang et al. WWW15)
- システムエラー自動検出 (Sano et al. SIGDIAL17)

YAHOO! JAPAN

従来の対話システムはドメインが限定的

フライト情報案内 (Price 90)、バスの時刻表案内 (Raux+ 05)、
観光案内 (翠+ 11)など



Let's Go (Raux+ 05)

AssisTra (翠+ 11)

http://www.speech.cs.cmu.edu/letsgo
http://www.nict.go.jp/publication/NICT-News/1108/03.html

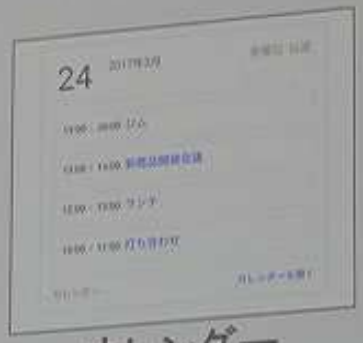
YAHOO! JAPAN

最近の対話システムは
多数のドメインをサポートする方向に

Yahoo! 音声アシストの場合：

天気・災害

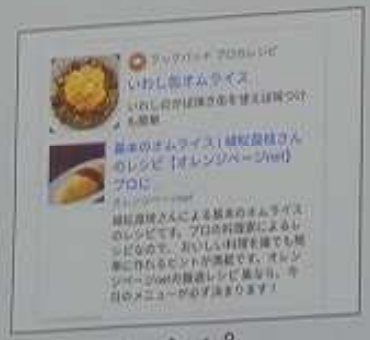
検索



カレンダー

アラーム

ニュース



レシピ

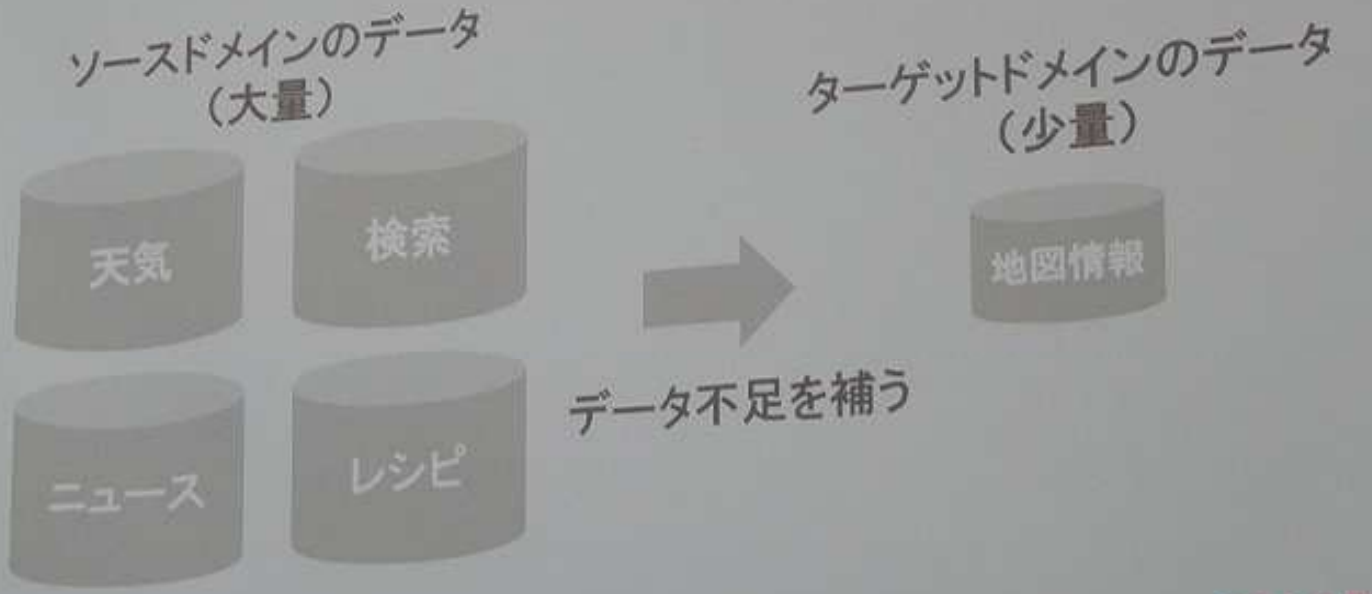
地図情報

他多数・・・

YAHOO! JAPAN

迅速な新ドメイン作成を実現するため ドメイン適応の研究が行われている

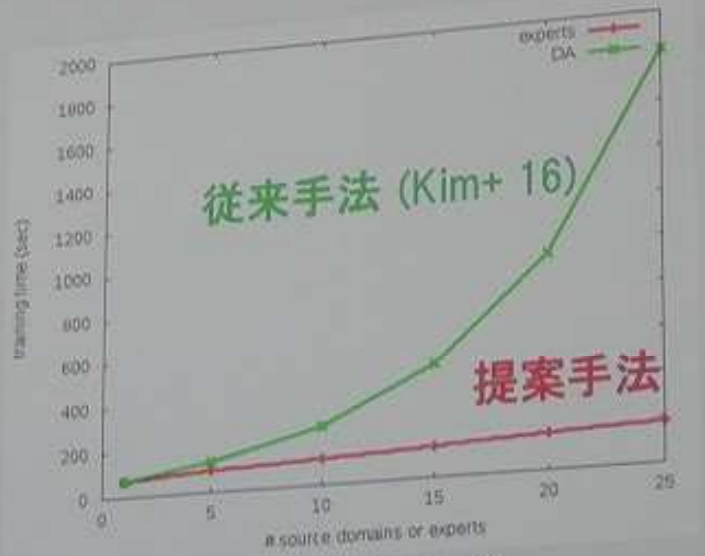
Feature augmentation (Daume+ 07; Kim+16)、Multi-task learning (Jaech+ 16)、Zero-shot learning (Chen+ 16) など



従来手法は再学習が必要なので大規模な対話サービスの運用には不向き

全てのデータ(ソース＋ターゲット)からの再学習が必要
ソースドメインの数が多い場合には非効率

学習時間



ソースドメイン数

YAHOO! JAPAN

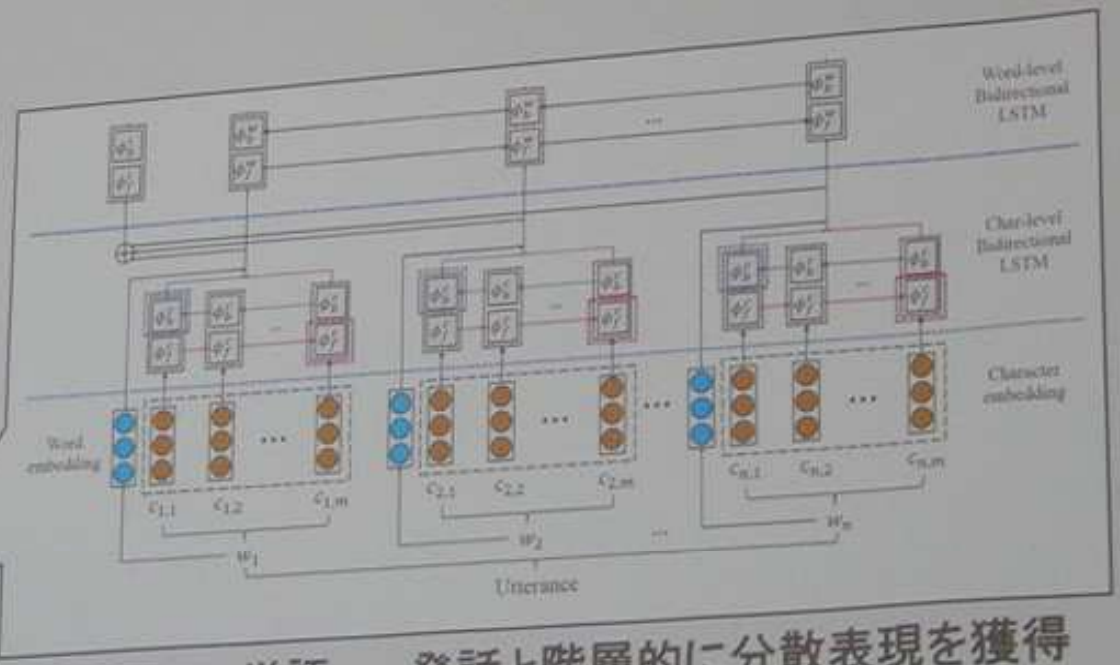
予備知識：提案手法の基となるモデル

出力ラベル
(発話意図など)

Feedforward

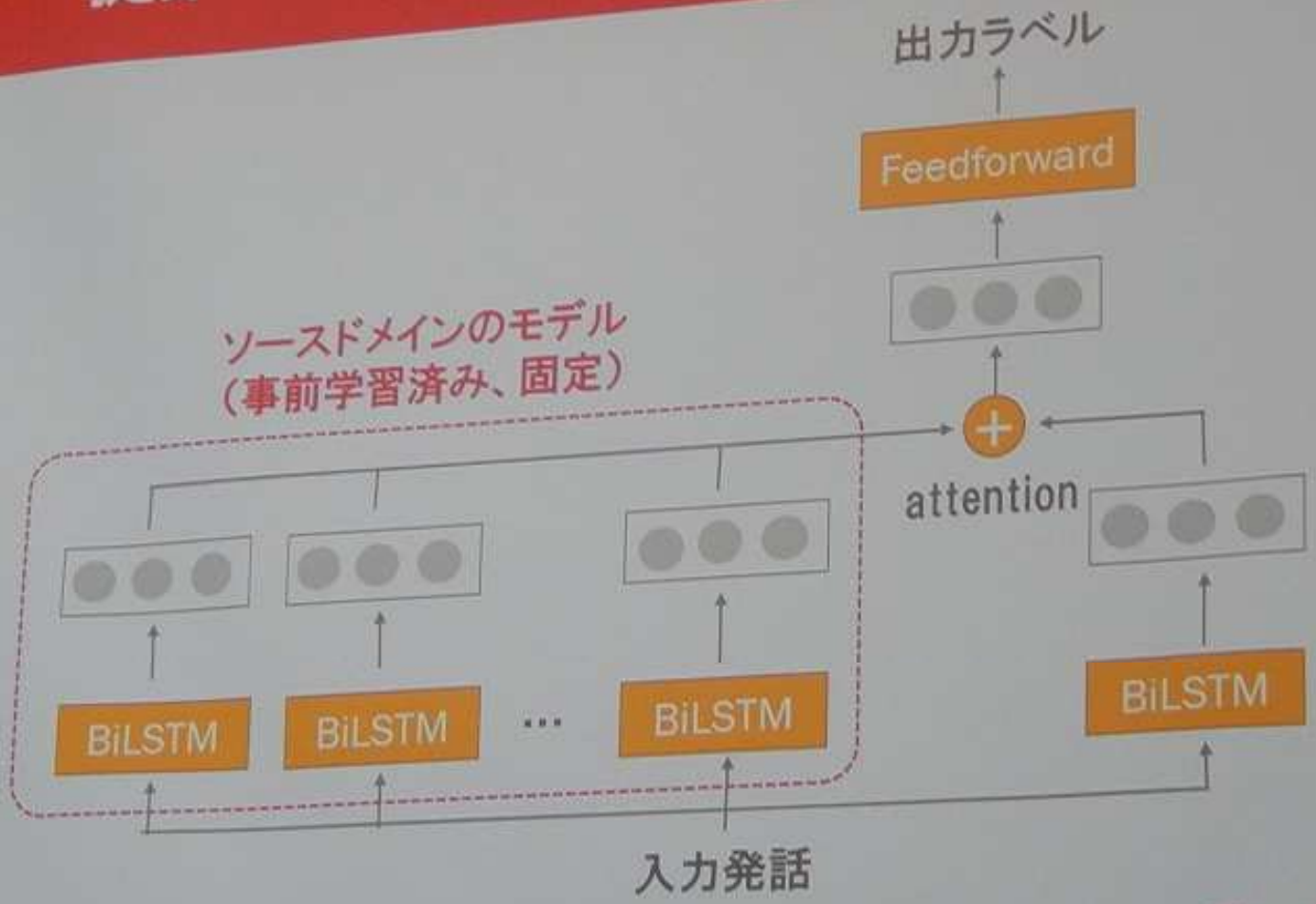
BiLSTM

入力発話



文字 → 単語 → 発話と階層的に分散表現を獲得

提案モデル：Domain Attention Model



再学習不要でなおかつ精度も良い

Microsoft Cortana の7ドメインのデータを使用
意図判定とスロット抽出の2つのタスク

		ベースライン			提案手法			
Task	Domain	TARGET	UNION	DA	DE^B	DE^1	DE^2	DE^{S2}
Intent	EVENTS	88.3	78.5	89.9	93.1	92.5	92.7	**94.5**
	FITNESS	88.0	77.7	92.0	92.0	91.2	91.8	**94.0**
	M-TICKET	88.2	79.2	91.9	94.4	91.5	92.7	**93.4**
	ORDERPIZZA	85.8	76.6	87.8	89.3	89.4	90.8	**92.8**
	REMINDER	87.2	76.3	91.2	90.0	90.5	90.2	**93.1**
	TAXI	87.3	76.8	89.3	89.9	89.6	89.2	**93.7**
	TV	88.9	76.4	90.3	81.5	91.5	92.0	**94.0**
	AVG	87.7	77.4	90.3	90.5	90.9	91.4	**93.6**
Slot	EVENTS	84.8	76.1	87.1	87.4	88.1	89.4	**90.2**
	FITNESS	84.0	75.6	86.4	86.3	87.0	88.1	**88.9**
	M-TICKET	84.2	75.6	86.4	86.1	86.8	88.4	**89.7**
	ORDERPIZZA	82.3	73.6	84.2	84.4	85.0	86.3	**87.1**
	REMINDER	83.5	75.0	85.9	86.3	87.0	88.3	**89.2**
	TAXI	83.0	74.6	85.6	85.5	86.3	87.5	**88.6**
	TV	85.4	76.7	87.7	87.6	88.3	89.3	**90.1**

YAHOO! JAPAN

Attention Weights の分析

ターゲットが TAXI の時 FLIGHTS が大きくなるなど直感に適合

従来の対話システムは タスク型と雑談型に分けて研究されてきた

- タスク型：システムに情報収集タスクを代行させる

この近くにある
イタリアン料理の店調べて

はい。本郷三丁目周辺の
イタリアンレストランは…

- 雑談型：システムとの会話そのものを楽しむ

機械学習って気になるけど
なんか難しそう。

思いきって勉強始めてみると
良いと思いますよ！

YAHOO! JAPAN

最近のサービスでは タスク型と雑談型という区別が曖昧に

タスク型

SHRDLU (Winograd 72)
ATIS (Price 90)
Let's GO (Raux+ 05)

雑談型

Eliza (Weizenbaum 66)
A.L.I.C.E. (Wallace 09)

アシスタント型
（タスクも雑談もこなす）

しゃべってコンシェル（吉村 12）
Yahoo! 音声アシスト（磯+ 13）
Siri (Bellegarda 14)
Cortana (Sarikaya 17)

YAHOO! JAPAN

雑談を意図したユーザ発話の検出が新しい課題になる

今日の大阪の天気は？ → 気象情報 API の呼び出し

8時にアラーム鳴らして → 携帯端末の操作　非雑談

明日の仕事休みたいなあ → 雑談生成モデルの駆動（seq2seqなど）　雑談

従来の雑談生成に関する研究では抜け落ちていたタスク

YAHOO! JAPAN

教師データを構築して分類器を学習

- 15160発話をクラウドソーシングを利用してラベル付与
- 各発話ごと7名の多数決(雑談/非雑談: 4833/10327)
- SVM と CNN の2つの分類器を学習、比較

ラベル	発話	得票数
雑談	お話ししよう	5
	趣味はなんですか？	7
	今月は休みがありません	5
非雑談	富士山の写真みせて	6
	近くのおいしいレストラン	7
	9時10分に起こして	7

YAHOO! JAPAN

ー工夫する：ツイートとウェブ検索ログを活用

- リプライのついたツイート≒雑談発話


USER1 @xxx
やっと仕事終わったよー！


USER2 @xxx
@USER1 お疲れ様！


USER3 @xxx
おはようございます～


USER4 @xxx
@USER3 おはよ！

- ウェブ検索ログ≒タスク要求発話

東京の天気

オムライスの作り方

君の名は。

ヤフー知恵袋

それぞれGRU言語モデルを学習し対数尤度を特徴量に

YAHOO! JAPAN

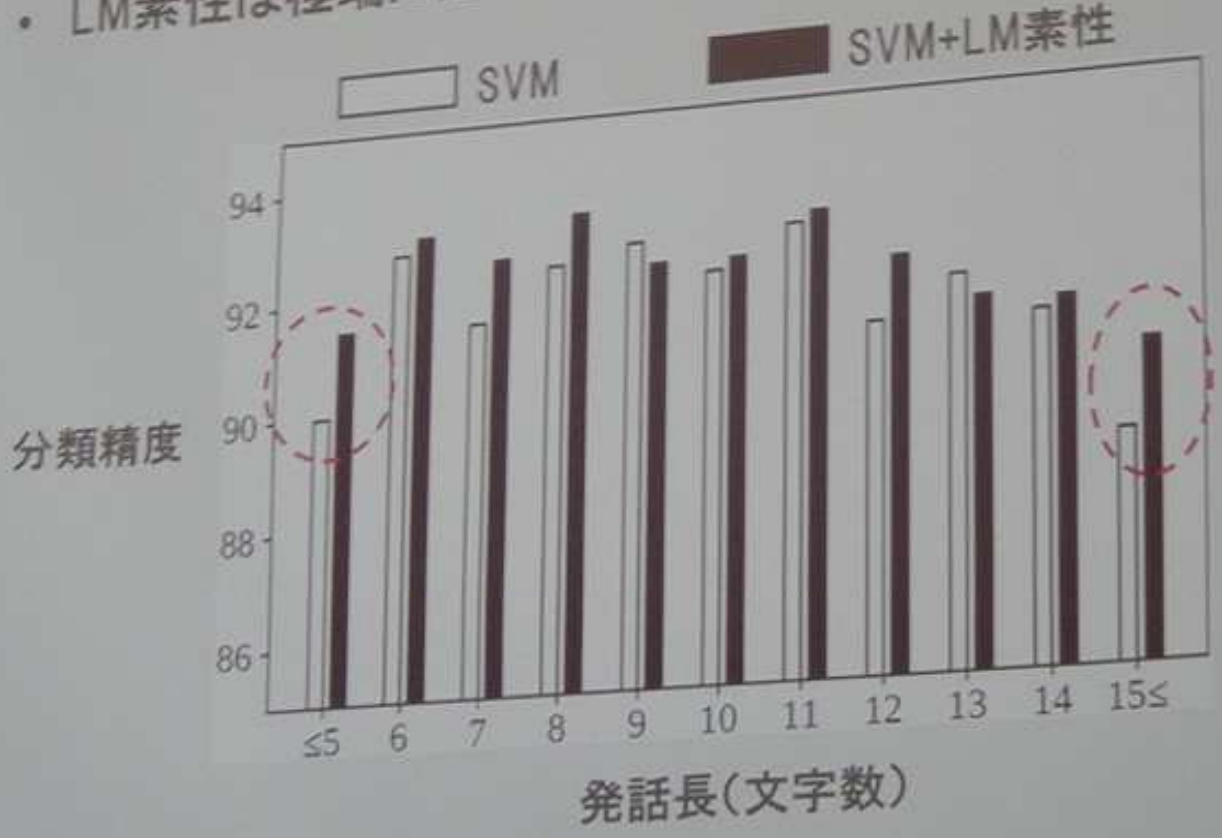
実験結果

- ベースライン(ツイートLM、内製)の精度を大きく改善
- 言語モデル(LM)素性の有効性を確認

手法	分類精度	適合率	再現率	F値
ツイート LM	72.07	54.54	74.48	62.94
内製の意図判定システム	78.31	62.57	79.51	70.03
SVM	91.35	87.62	84.88	86.21
SVM + ツイート/クエリ LM	92.15	88.61	86.50	87.53
CNN	90.84	87.03	83.80	85.36
CNN + ツイート/クエリ LM	91.48	87.78	85.18	86.56

発話長と分類精度の関係

- LM素性は極端に短いまたは長い発話に有効



YAHOO! JAPAN

本物のユーザを相手にした対話処理には正解がないので客観評価が難しい

多くの学術研究の進め方
- 客観評価が容易な部分問題を切り出して議論
- 訓練されたアノテータが適切な対話を天下り的に定義

素朴な疑問
- 部分問題が正しく解けてユーザ経験は向上するのか？
- 天下り的に決めた正解で本当にユーザは喜ぶのか？

YAHOO! JAPAN

セッションの満足度を推定による自動評価

ユーザの行動パターンから満足度を推定するモデルを学習

SAT(満足)

User	Cortana, call James.
Cortana	Sure, call James Smith mobile, is that right?
User	Yes
Cortana	Call James Smith mobile. [call the contact]

DSAT(不満)

User	Where is the nearest pharmacy?
Cortana	Here are 8 pharmacies near you. [show options on the screen]
User	Show me the direction to block sponsee (Clark's pharmacy)
Cortana	Sorry, I couldn't find (...) Doyou you wanna search the web for it?

YAHOO! JAPAN

Action系列を特徴量に利用

User	Where is the nearest pharmacy?	Command
Cortana	Here are 8 pharmacies near you. [show options on the screen]	Option
User	Show me the direction to block sponsee (Clark's pharmacy)	Select
Cortana	Sorry, I couldn't find (...) Do you you wanna search the web for it?	Confirm
User	No	NO
Cortana	Here are 8 pharmacies near you. [show options on the screen]	Option

満足度予測の実験結果

60名の被験者による300セッションの発話
満足度のratingをもとに SAT/DSAT ラベルを付与

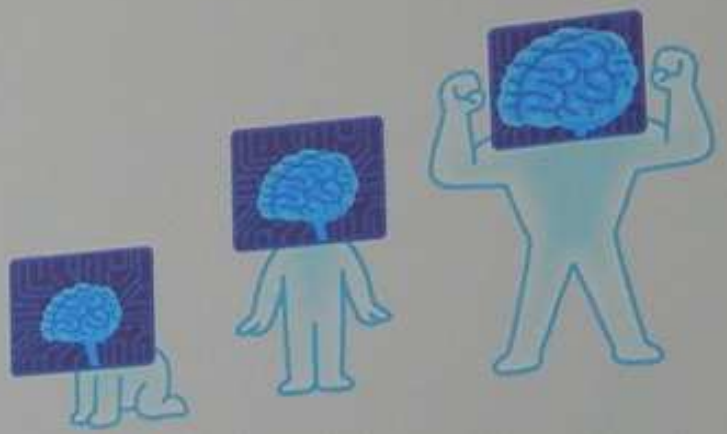
Features	User Satisfaction			
	Avg F_1	Pos F_1	Neg F_1	Accuracy
Click	0.524**	0.825**	0.222**	0.718**
Request	0.815	0.901	0.729	0.856
Response	0.758**	0.850**	0.665**	0.796**
Acoustic	0.743**	0.849**	0.638**	0.790**
Action Sequence	0.819	0.892	0.746	0.850
Best Feature Set	**All Features**			
	0.852**	0.920**	0.783*	0.886**

YAHOO! JAPAN

大勢のユーザとの対話を通じて自立的に学習するというシナリオが現実的に

ユーザがシステムを“普通に”利用
（明示的な教師信号や報酬は与えない）

ユーザとの対話の経験を通して
自立的に学習、成長



YAHOO! JAPAN

ユーザの修正発話に着目

アラーム

間違った応答

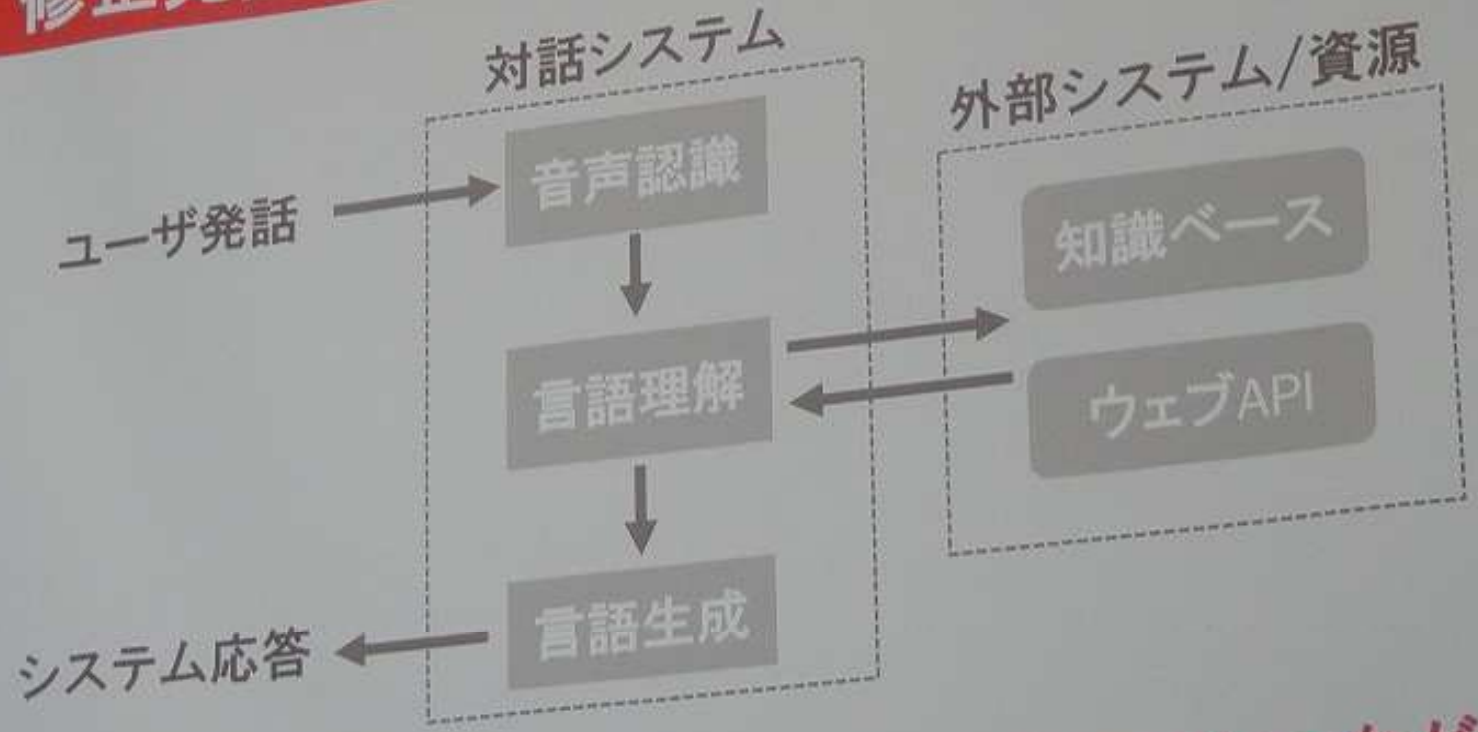
“アラームの”ウェブ検索結果
はこちらです…

アラームを起動して

修正発話

YAHOO! JAPAN

研究のゴール：修正発話の原因となるエラー源の自動検出



エラーの発生源が特定できれば（半）自動訂正につながる

YAHOO! JAPAN

様々な特徴量を設計

アラーム

"アラームの"ウェブ検索
結果はこちらです…

アラームを起動して

はい。アラームを起動します。

発話時間間隔
編集距離
修正タイプ(ADD, OMIT等)
…

ASRの確信度

ASR Conf.: 0.7
Domain: ウェブ検索
Intent: SEARCH(アラーム)

ASR Conf.: 0.9
Domain: アラーム
Intent: START

domain、intent(slot値など)などの一致

YAHOO! JAPAN

原因判定実験の結果

- SVM を用いて10分割交差検定
- 発話の表層情報だけでなくエラー原因ごとに作りこまれた素性を使うことでF_1値が向上

	エラー無し	音声認識エラー	言語理解エラー	言語生成エラー
ベースライン	0.58	0.59	0.36	0.03
＋音声認識素性	0.66††	0.67††	0.35	0.16
＋言語理解素性	0.71††	0.65	0.43	0.25†
＋言語生成素性	0.55	0.57	0.32	0.08
提案法（＋全素性）	0.75††	0.72††	0.49†	0.33††

YAHOO! JAPAN

まとめ

およそ半世紀の基礎研究の期間を経て、音声対話技術は、
音声対話アシスタントという形で実社会に巣立ちつつある

e.g., Google Home, Amazon Echo, Line Clova, Y! Voice Assist etc.

しかし、ラボ環境と実環境のギャップは依然として存在しており、
現在はその差を埋めている段階

これからは実環境での音声対話研究が面白くなるはず！

- 実サービスの運用から顕在化した新しい課題 (Kim et al. 2017; Akaskai and Kaji 2017)
- 大勢のリアルユーザの行動分析 (Jiang et al. 2015; Sano et al. 2017)