

# Word-Based Dialog State Tracking with Recurrent Neural Networks

Matthew Henderson, Blaise Thomson, Steve Young  
University of Cambridge, U.K.

吉野 幸一郎 (京大 河原研)



@caesar\_wanya

Kyoto University  
Media Archiving Research Laboratory

# SIGDIALの報告

- Venue: Philadelphia, PA, USA
- INLG (言語生成の会議)と共催
  - Pre-event に YRRSDS
- 67 件の投稿
  - Long: (13+9)/43
  - Short: 8/20
  - Demo: 4/4
- Dialog State Tracking Challenge (DSTC)
  - 7 long and 2 short
  - 今回読む論文はDSTCのもの



# Best paper

- **Best Paper Award Winners**

*Web-style Ranking and SLU Combination for Dialog State Tracking*

Jason D. Williams

*Identifying Narrative Clause Types in Personal Stories*

Reid Swanson, Elahe Rahimtoroghi, Thomas Corcoran and Marilyn A. Walker

## **Best Paper Award Nominees**

*Word-Based Dialog State Tracking with Recurrent Neural Networks*

Matthew Henderson, Blaise Thomson and Steve Young

*Situated Language Understanding at 25 Miles per Hour*

Teruhisa Misu, Antoine Raux, Rakesh Gupta and Ian Lane

# 流行りなど

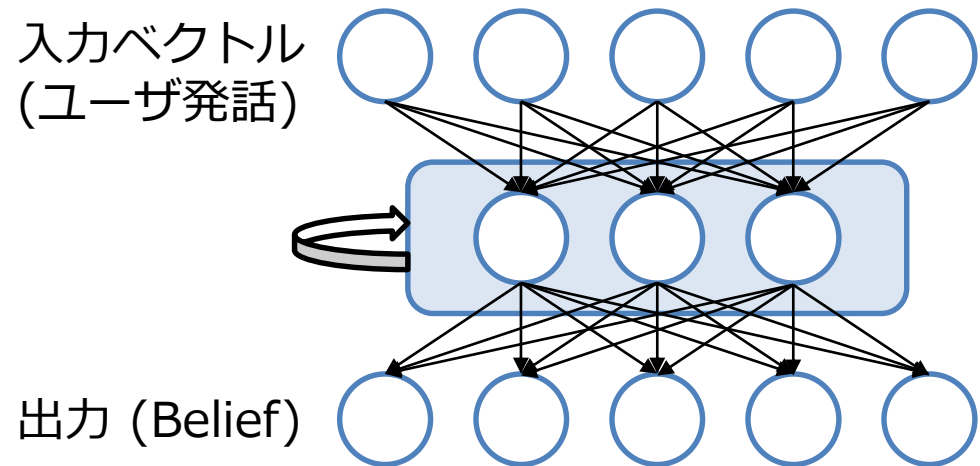
- **DSTC で Task-oriented な対話制御は Red Ocean化**
- **Situated dialogue**
  - 地図情報をReferできる対話 など
- **Non-task-oriented:**
  - Goal指向ではないが緩やかな目的を持つ
    - 情報案内システム
    - 人との信頼関係の構築
- **システムのドメイン適応**

# DSTC (Dialog State Tracking Challenge)

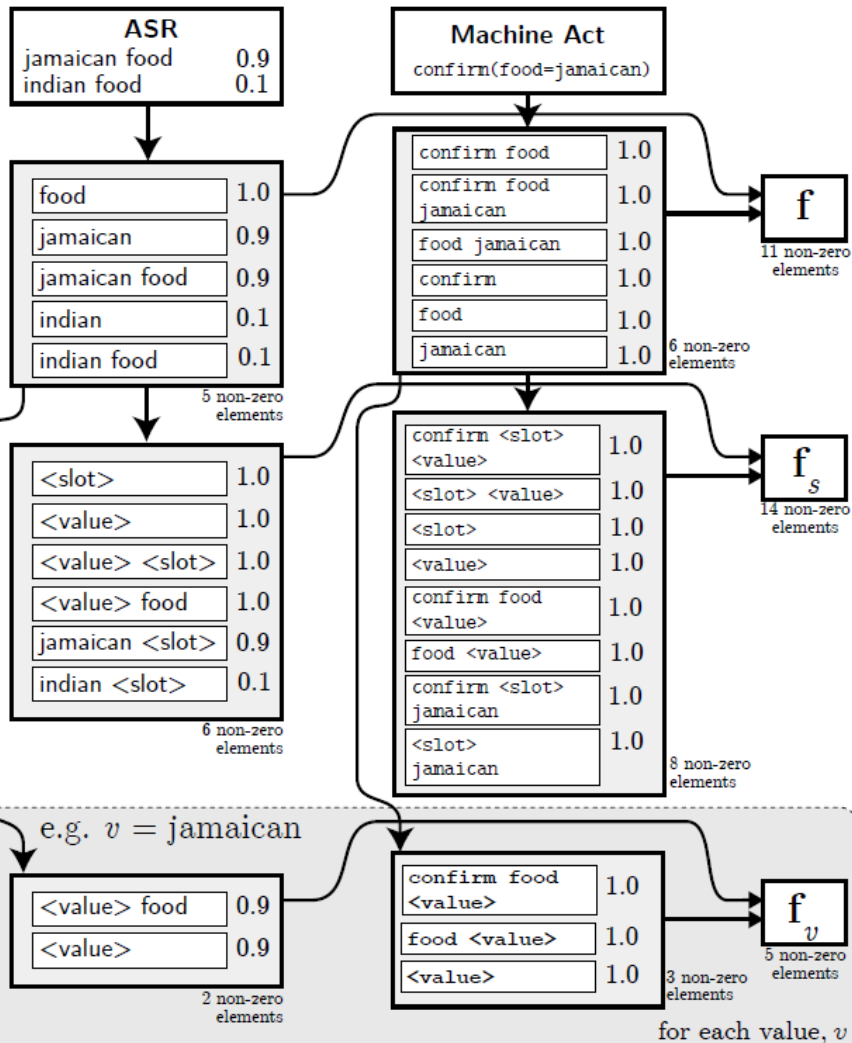
- **タスク達成対話のユーザ意図の認識精度を競う**
  - A cheap place in the north.  
→ `inform(area=north, pricerange=cheap)`
  - What is their phone number and address?  
→ `request(phone), request(address)`
- **今回は2回目、タスクはレストラン案内**
  - Amazon Mechanical Turk で収集
- **評価尺度**
  - Accuracy: 各ターンのTrackerの精度 (1-best)
  - L2: 出力するBeliefの二乗誤差
  - ROC: Receiver-operating characteristics  
N-best 中の correct-accept の割合 (MRRっぽい)

# DST with RNN

- Joint Goal Constraints, Requested Slots で最高精度
- アイディアはとてもシンプル
  - Belief Update の更新をRNNに置き換えるイメージ
  - スパースにならないよう工夫をしている



# ベクトルの作り方



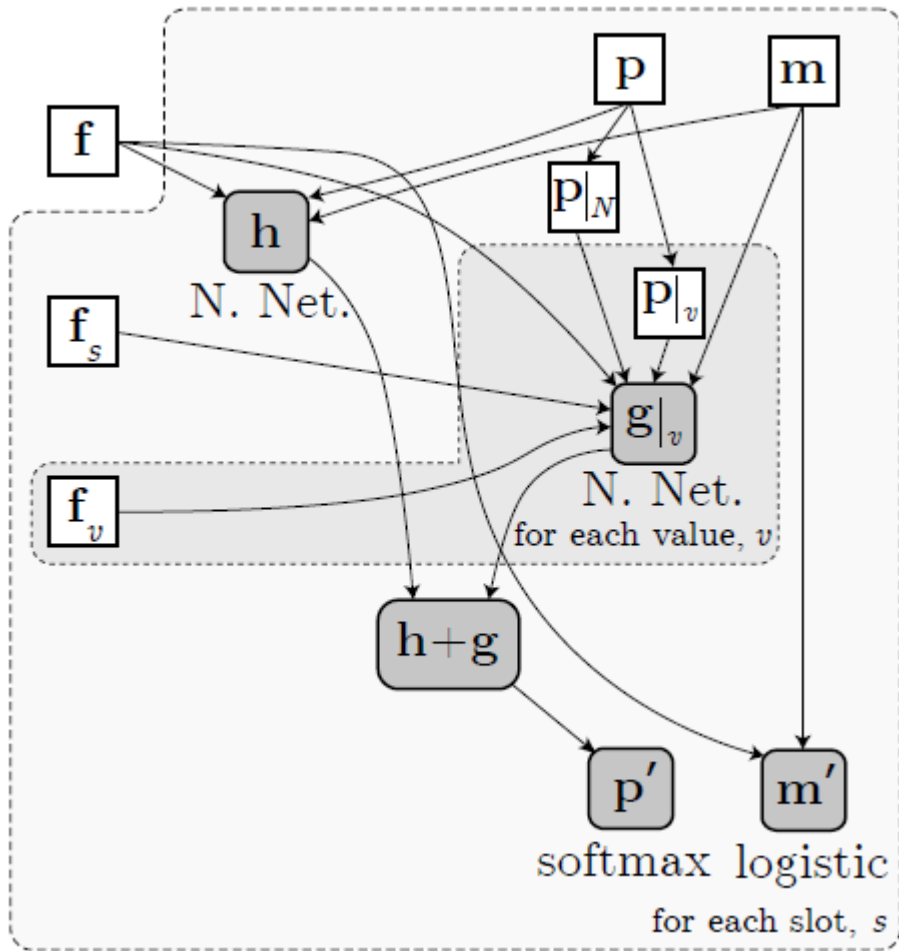
- ASRの結果

- N-best中のn-gram (重みつき) → f
- 取りうるスロット → f<sub>s</sub>
- 取りうる値 → f<sub>v</sub>

- Machine act

- 取りうる行動 (+value) → f
- 取りうる行動 (+スロット) → f<sub>s</sub>
- 取りうる行動 (+値) → f<sub>v</sub>

# NNの構築



- NN f
  - 素性ベクトル  $f$
  - 全ての仮説の確率  $p$
  - RNNの一時メモリ  $m$
- NN  $g|_v$  (汎化用)
  - 仮説にある全ての値  $v$  について作った汎化ベクトル  $f_{v1}, f_{v2}, \dots, f_{vn}$
  - $f_s$
  - 汎化クラス的确率  $g|_v$
  - RNNの一時メモリ  $m$
- $p$ は $h$ と $g|_v$ 全てのsoftmax
- $m'$ は前の $m$ と $f$ のみから更新



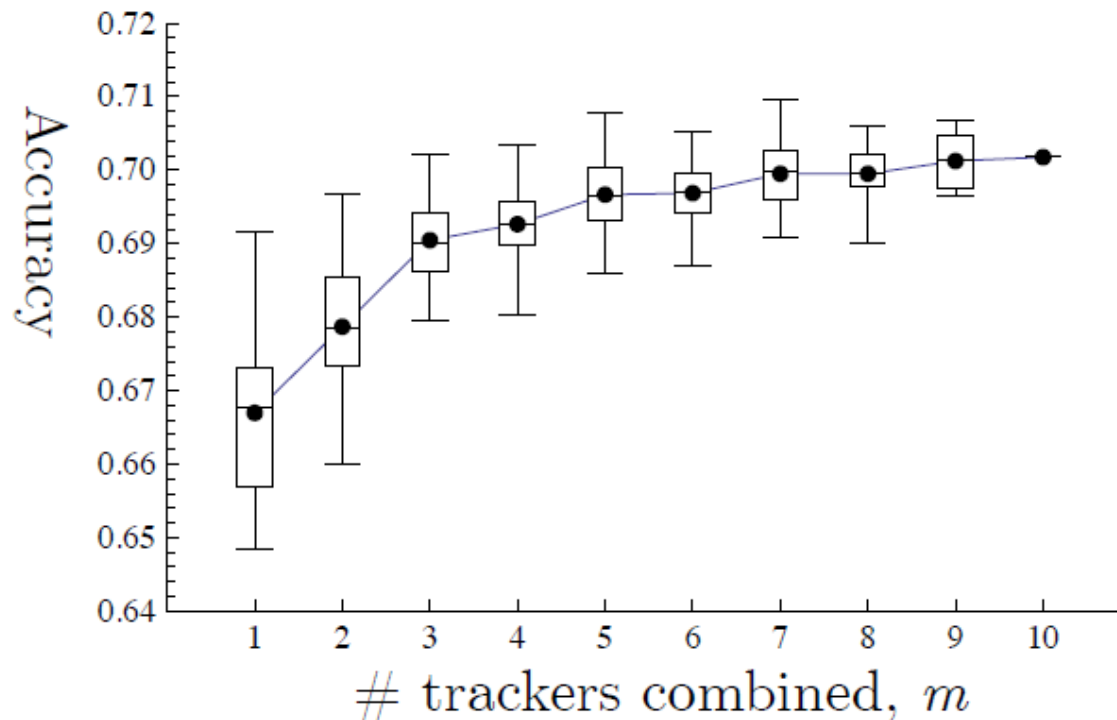
# NNの学習

- $m'$  の更新の際の重みパラメータをAutoencoder による Denoising (dA)で求める 中間層は圧縮
  - 入力と出力が同じのNNで得られる中間層
$$\mathbf{h} = \sigma (W_{dA} \mathbf{f}_{noisy} + b_0)$$
$$\mathbf{f}_{rec} = \sigma (W_{dA}^T \mathbf{h} + b_1)$$
  - 目的関数は $f$ と $f_{rec}$ のクロスエントロピー

		Joint Goals		Method		Requested	
Shared init.	dA init.	Acc	L2	Acc	L2	Acc	L2
		0.686	0.477	0.913	0.147	0.963	0.059
	✓	0.688	0.466	0.915	0.144	0.962	0.059
✓		0.680	0.479	0.910	0.152	0.962	0.059
✓	✓	<b>0.696</b>	<b>0.463</b>	<b>0.915</b>	<b>0.144</b>	<b>0.965</b>	<b>0.057</b>
Baseline:		0.612	0.632	0.830	0.266	0.894	0.174

# 学習のパラメータ

- 正則化パラメータは色々なものを作って、それらが出力するスコアからいいものを選ぶ方が精度が上がる



# 結果

		Tracker Inputs		Joint Goals			Method			Requested		
entry	Include h	Live ASR	Live SLU	Acc	L2	ROC	Acc	L2	ROC	Acc	L2	ROC
0	✓	✓		<b>0.768</b>	<b>0.346</b>	0.365	<b>0.940</b>	<b>0.095</b>	0.452	<b>0.978</b>	<b>0.035</b>	0.525
1		✓		0.746	0.381	0.383	0.939	0.097	0.423	0.977	0.038	0.490
2	✓		✓	0.742	0.387	0.345	0.922	0.124	0.447	0.957	0.069	0.340
3			✓	0.737	0.406	0.321	0.922	0.125	0.406	0.957	0.073	0.385

- Live SLU: 既存のSLUベースの出力を用いる
- Live ASR: ASRの認識結果を突っ込む
- Include h: サブネットワーク h を挟んで使うかどうか