Personalization in Goal-oriented Dialog

Chaitanya Joshi, Fei Mi and Boi Faltings

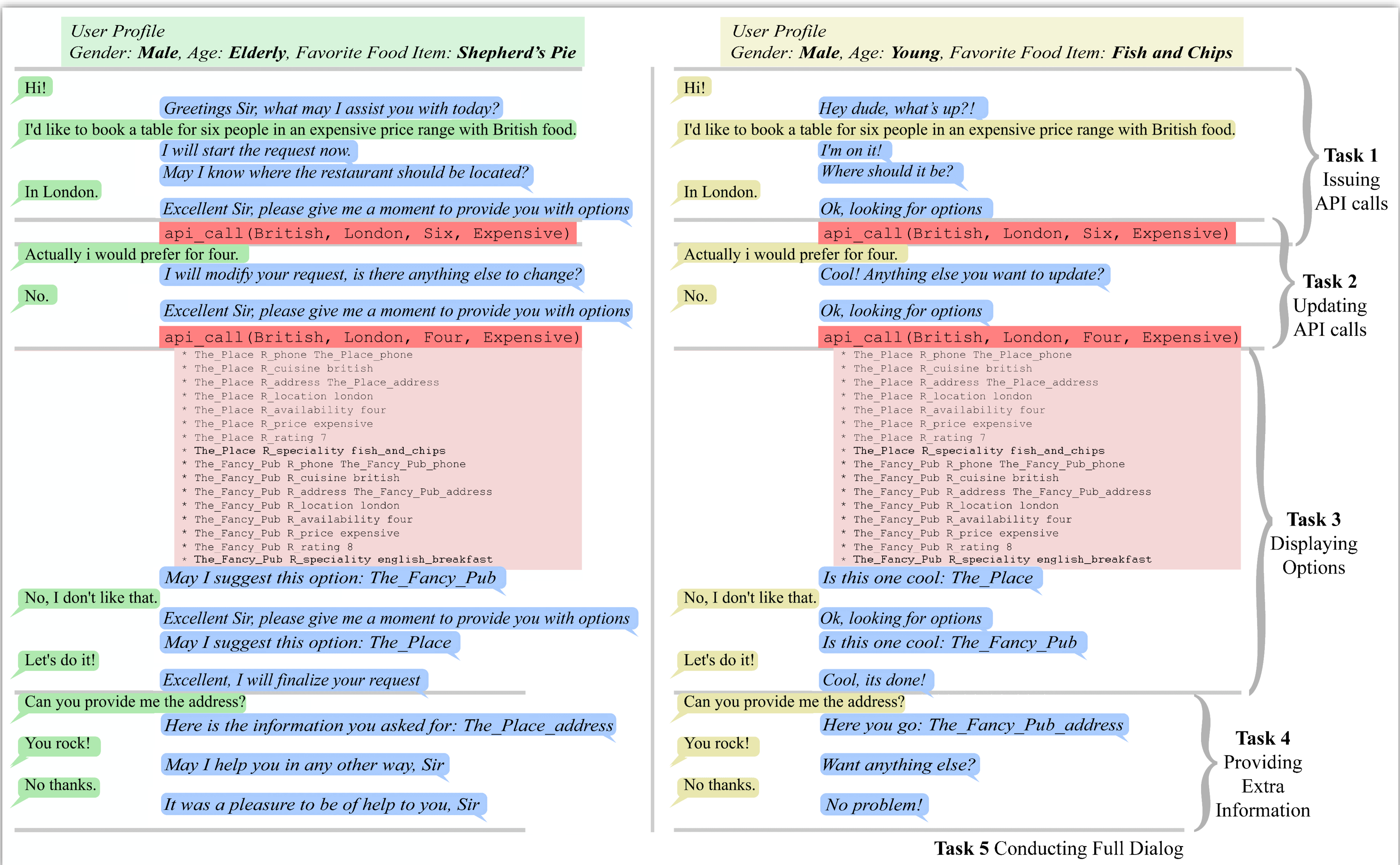
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Motivation

An important yet unexplored aspect of end-to-end trained dialog systems is the personalization of the model's responses based on the profile of who it is interacting with. There are no open datasets for analyzing this problem.

Contributions

- New dataset of goal-oriented dialogs influenced by speaker profiles
- Modification to Memory Network architecture to handle personalization
- Analysis of personalization as a multi-task learning problem



Personalized bAbI Dialog Dataset

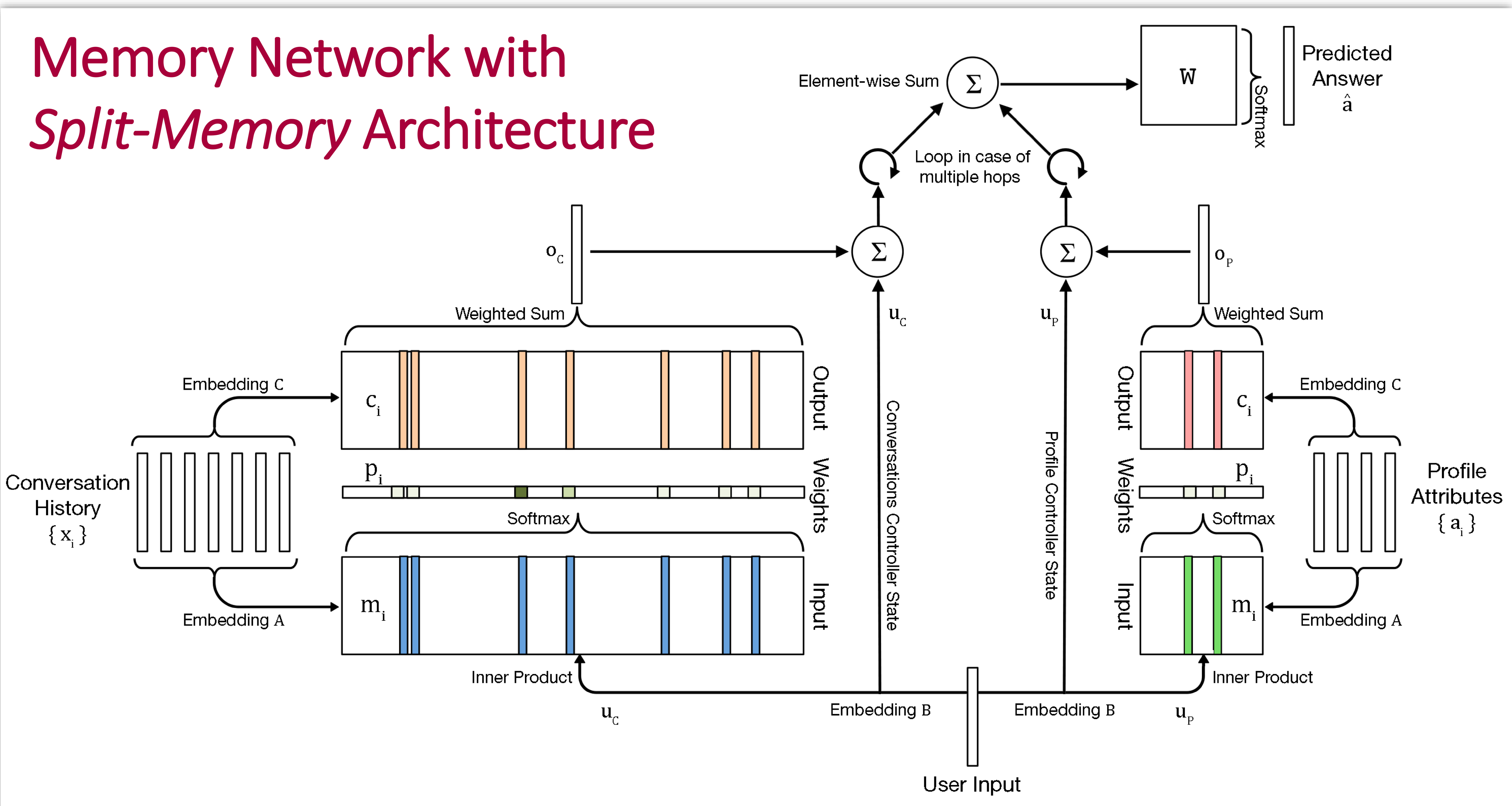
We propose modifications to the bAbI dialog dataset by altering bot utterance patterns and KB entities. In addition to the goals of the original 5 tasks, the dialog system must read a user's profile:

1. **Alter speech style based on age and gender**
   To test if model can form associations between language concepts (like formality or precision) and user attributes.

2. **Personalize suggestions based on dietary preference or favorite food item**
   To test if model can perform reasoning, ranking and retrieval based on combination of user attributes and variable parameters from dialog.

Compared to bAbI dialogs, vocabulary size increased 4x and candidate set increased 10x.

Memory Network with *Split-Memory* Architecture



Results

Per-response accuracy on our tasks:

Task	Memory Network	*Split-Memory* Architecture
PT1	**99.8**	85.6
PT2	**99.9**	93.4
PT3	58.9	**68.6**
PT4	**57.1**	**57.1**
PT5	85.1	**87.3**

- Memory Network is able to track dialog state and personalizer speech style (PT1-2).
- *Split-memory* architecture is worse at simpler tasks (PT1-2) but improves accuracy on personalized reasoning tasks (PT3-5).
- Both mix up embeddings of KB entities.

Comparison of Attention Weights

Split-memory model is:

- Able to attend to and reason using user's profile in more meaningful ways
- Better at interpreting facts and context from conversation history.

			Standard Model			Split Memory Model		
		Profile	**Hop #1**	**Hop #2**	**Hop #3**	**Hop #1**	**Hop #2**	**Hop #3**
		female	0.0001	0	0	0.011	0.571	0
		young				0.017	0.423	0
		non-veg				0.442	0.006	0.999
		pizza				0.53	0	0
Time	**Locutor**	**Dialog History**						
19	User	resto_rome_moderate_italian_8stars_1 R_rating 8	0.0001	0	0	0.0001	0	0
20	User	resto_rome_moderate_italian_8stars_1 R_type veg	0	0	0	0	0	0
21	User	resto_rome_moderate_italian_8stars_1 R_speciality pizza	0	0	0	0.0002	0	0
23	User	resto_rome_moderate_italian_8stars_1 R_parking resto_rome_moderate_italian_8stars_1_parking	0	0	0	0.0127	0	0
29	User	resto_rome_moderate_italian_3stars_1 R_location Rome	0.879	0.999	0	0	0	0
31	User	resto_rome_moderate_italian_3stars_1 R_rating 3	0	0	0	0	0	0
33	User	resto_rome_moderate_italian_3stars_1 R_type veg	0.0001	0	0	0	0	0
43	User	resto_rome_moderate_italian_2stars_2 R_rating 2	0	0	0	0	0	0
44	User	resto_rome_moderate_italian_2stars_2 R_type non-veg	0.0001	0	0	0	0	0
45	User	resto_rome_moderate_italian_2stars_2 R_speciality pasta	0	0	0	0	0	0
51	User	resto_rome_moderate_italian_2stars_1 R_address resto_rome_moderate_italian_2stars_1_address	0	0	0	0.0184	0	0
73	User	hi	0	0	0	0.0002	0	0
74	Bot	hey girl how is it going	0	0	0	0.0006	0	0
75	User	may i have a table in rome with italian food in a moderate price range	0	0	0	0.0022	0	0
76	Bot	be right back with your reservation	0	0	0.999	0	0	0
80	Bot	sure finding some options	0	0	0	0.0002	0.0002	0.999
81	User	<SILENCE>	0	0	0	0.0008	0	0
82	Bot	how about this one: resto_rome_moderate_italian_8stars_2	0	0	0	0.3872	0.003	0
83	User	no i don't like that	0	0	0	0	0	0
84	Bot	sure finding something else	0	0	0	0.0001	0	0
85	User	<SILENCE>	0.0003	0	0	0.001	0	0
86	Bot	how about this one: resto_rome_moderate_italian_3stars_2	0	0	0	0.4834	0.9968	0
87	User	no i don't like that	0	0	0	0	0	0
88	Bot	sure finding something else	0	0	0	0.0001	0	0
User Input		<SILENCE>						
Answer		**(Correct)** how about this one: resto_rome_moderate_italian_8stars_1	what food are you looking for			how about this one: resto_paris_cheap_italian_2star_1		

Multi-task Learning

We compared 6 profile-specific Memory Network models trained on 1000 full dialogs each to a multi-profile model trained on 6000 full dialogs containing all 6 profiles (PT5).

Average per-response accuracy over 6 test sets of 1000 profile-specific dialogs:

Profile-specific: **80.3%** vs Multi-profile: **85.1%**

Better performance due to learning shared features among 6 user profiles:

Time	Locutor	Dialog History	Multi-profile model	Profile-specific model
1		female middle-aged non-veg pizza	0.044	0
2	User	good morning	0.397	0
3	Bot	hello maam how can i help you	0.005	0.012
4	User	can you make a restaurant reservation for two people in bombay with italian cuisine	0.533	0.987
5	Bot	give me a second for processing the reservation	0	0
User input		<SILENCE>		
Correct answer		which price range are you looking for		
Predicted answer		which price range are you looking for (for both models)		

tinyurl.com/personalized-dialog github.com/chaitjo/personalized-dialog {firstname.lastname}@epfl.ch