Leveraging Behavioral Patterns of Mobile Applications for Personalized Spoken Language Understanding

YUN-NUNG (VIVIAN) CHEN
MING SUN
ALEXANDER I RUDNICKY
ANATOLE GERSHMAN

Introduction

- Global Digital Statistics (2015 Jan)

Global Population	Active Internet Users	Active Social Accounts	Active Mobile Users
7.21B	3.01B	2.08B	3.65B

Spoken language interfaces become important are incorporated in smart devices as intelligent assistants (IAs).

Spoken language understanding (SLU) is a key component of IA, which predicts users' intended apps by understanding input utterances.

Introduction

- Task: user intent prediction
- Challenge: language ambiguity

send to vivian

Communication

v.s.

Email?

Message?

① User preference
  - ✓ Some people prefer “Message” to “Email”
  - ✓ Some people prefer “Outlook” to “Gmail”

② App-level contexts
  - ✓ “Message” is more likely to follow “Camera”
  - ✓ “Email” is more likely to follow “Excel”

Considering behavioral patterns in history to model SLU for intent prediction.

Data Collection

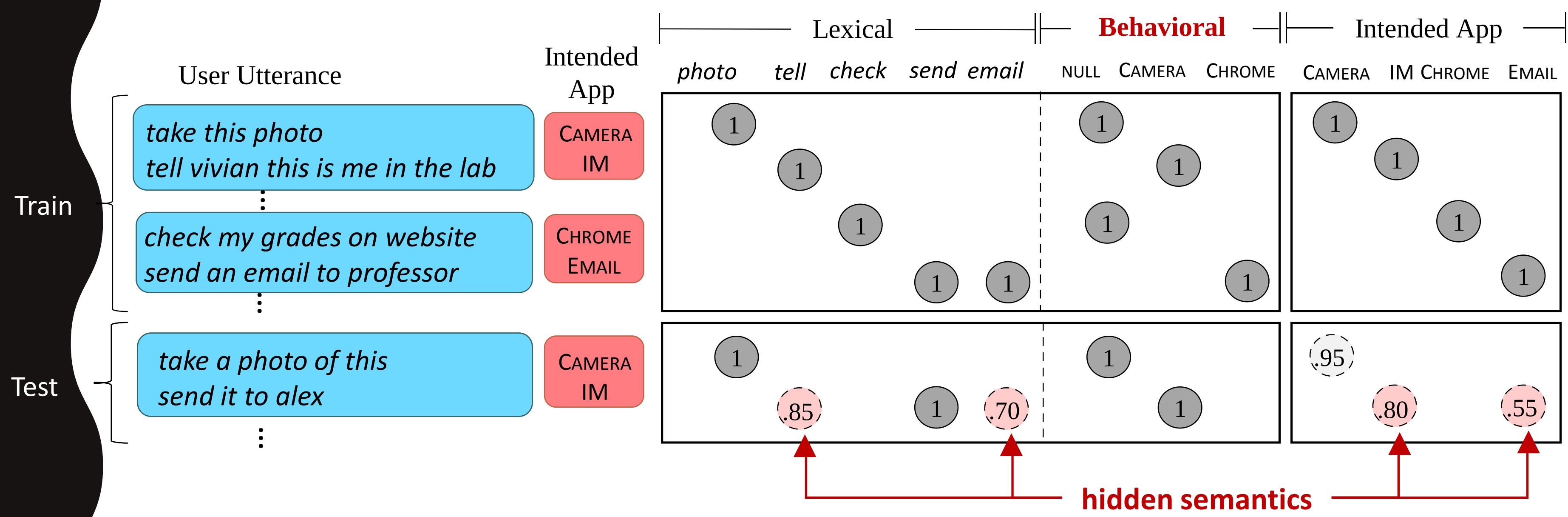
- Subjects' app invocation is logged on a daily basis
- Subjects annotate their app activities with
  - **Task Structure**: link applications that serve a common goal
  - **Task Description**: briefly describe the goal or intention of the task

Meta	TASK59; 20150203; 1; Tuesday; 10:48
App	com.android.settings → com.lge.music
Desc	play music via bluetooth speaker

- Subjects use a wizard system to perform the annotated task by speech

Dialogue			
	W_1:	Ready.	
	U_1:	Connect my phone to bluetooth speaker.	SETTINGS
	W_2:	Connected to bluetooth speaker.	
	U_2:	And play music.	MUSIC
	W_3:	What music would you like to play?	
	U_3:	Shuffle playlist.	MUSIC
	W_4:	I will play the music for you.	

SLU for Intent Prediction

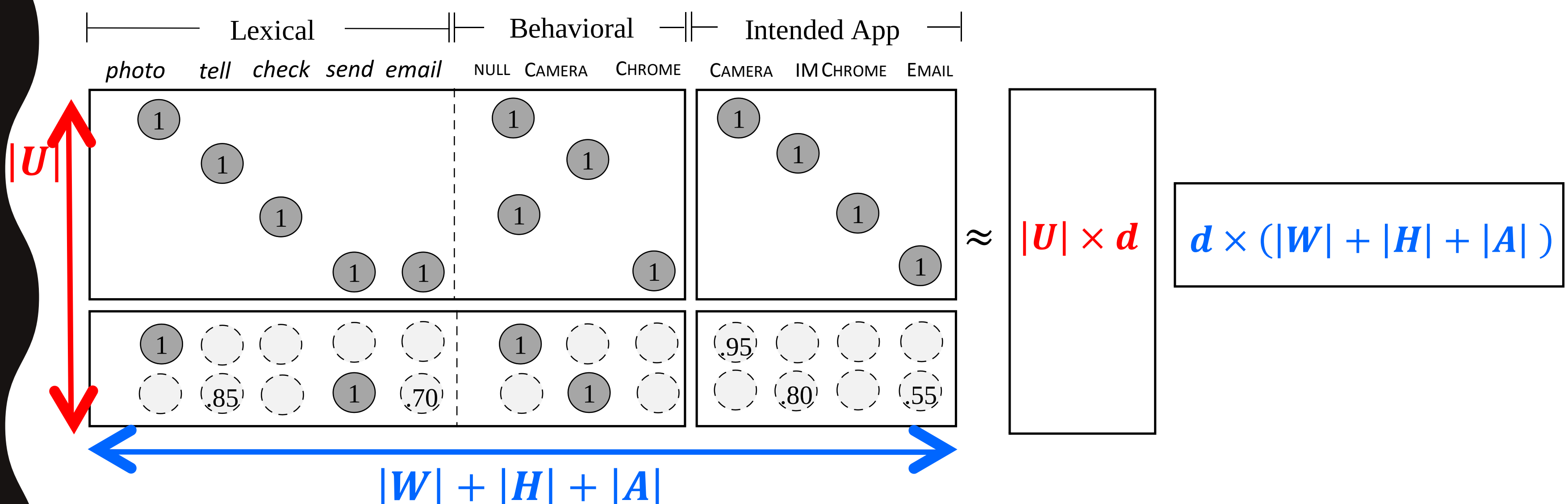


Issue: unobserved hidden semantics may benefit understanding

Solution: use matrix factorization to complete a partially-missing matrix based on a low-rank latent semantics assumption.

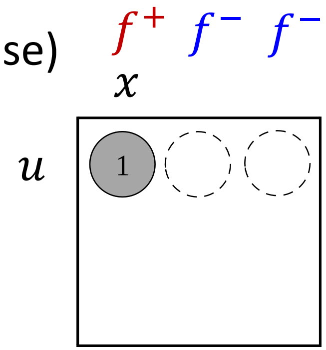
Matrix Factorization (MF)

- The decomposed matrices represent low-rank latent semantics for utterances and words/histories/apps respectively
- The product of two matrices fills the probability of hidden semantics



Parameter Estimation

- Model implicit feedback by completing the matrix
  - not treat unobserved facts as negative samples (true or false)
  - give observed facts higher scores than unobserved facts



$$f^+ = \langle u, x^+ \rangle \quad f^- = \langle u, x^- \rangle \Rightarrow p(f^+) > p(f^-)$$

$$p(M_{u,x} = 1 \mid \theta_{u,x}) = \sigma(\theta_{u,x}) = \frac{1}{1 + \exp(-\theta_{u,x})}$$

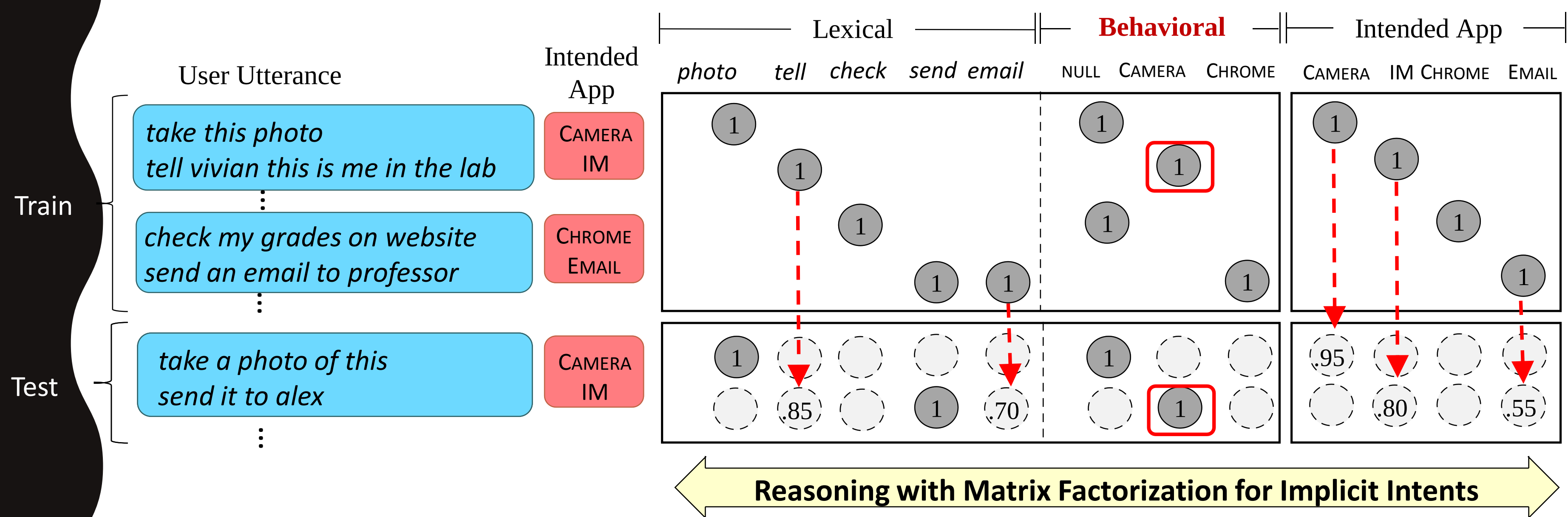
- Objective:

$$\sum_{f^+ \in \mathcal{O}} \sum_{f^- \notin \mathcal{O}} \ln \sigma(\theta_{f^+} - \theta_{f^-})$$

  - the model can be achieved by SGD updates with fact pairs

The objective is to learn a set of well-ranked apps per utterance.

SLU Modeling by MF



Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC)
  mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (**MLE**)

Approach			Lexical	Behavioral	All
(a)	MLE	User-Indep		13.5 / 19.6	
(b)		User-Dep		20.2 / 27.9	

The user-dependent model is better than the user-independent model.

Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC)
  mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (**MLE**)
  Multinomial Logistic Regression (**MLR**)

	Approach		Lexical	Behavioral	All
(a)	MLE	User-Indep		13.5 / 19.6	
(b)		User-Dep		20.2 / 27.9	
(c)	MLR	User-Indep	42.8 / 46.4	14.9 / 18.7	
(d)		User-Dep	48.2 / 52.1	19.3 / 25.2	

Lexical features are useful to predict intended apps for both user-independent and user-dependent models.

Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC)
  mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (**MLE**)
  Multinomial Logistic Regression (**MLR**)

Approach			Lexical	Behavioral	All
(a)	MLE	User-Indep		13.5 / 19.6	
(b)		User-Dep		20.2 / 27.9	
(c)	MLR	User-Indep	42.8 / 46.4	14.9 / 18.7	**46.2^{+} / 50.1^{+}**
(d)		User-Dep	48.2 / 52.1	19.3 / 25.2	**50.1^{+} / 53.9^{+}**

Combining lexical and behavioral features improves performance of the MLR model, which models explicit information from observations.

Experiments

- Dataset: 533 dialogues (1,607 utterances); 455 multi-turn dialogues
- Google recognized transcripts (word error rate = 25%)
- Evaluation metric: accuracy of user intent prediction (ACC)
  mean average precision of ranked intents (MAP)
- Baseline: Maximum Likelihood Estimation (**MLE**)
  Multinomial Logistic Regression (**MLR**)

	Approach		Lexical	Behavioral	All
(a)	MLE	User-Indep		13.5 / 19.6	
(b)		User-Dep		20.2 / 27.9	
(c)	MLR	User-Indep	42.8 / 46.4	14.9 / 18.7	46.2^{+} / 50.1^{+}
(d)		User-Dep	48.2 / 52.1	19.3 / 25.2	50.1^{+} / 53.9^{+}
(e)	(c) + Personalized MF		**47.6 / 51.1**	**16.4 / 20.3**	**50.3^{+*} / 54.2^{+*}**
(f)	(d) + Personalized MF		**48.3 / 52.7**	**20.6 / 26.7**	**51.9^{+*} / 55.7^{+*}**

Personalized MF significantly improves MLR results by considering hidden semantics.

Conclusion

- An MF model exploits both lexical and behavioral features for SLU, which considers **implicit semantics** to **enhance intent inference** given the noisy ASR inputs.
- We are able to model users' **contextual behaviors** and their **app preference** for better intent prediction.
- The proposed multi-model personalized SLU effectively improves intent prediction performance, achieving about **52% on turn accuracy** and **56% on mean average precision** for ASR transcripts with 25% word error rate.

Carnegie Mellon

THANKS FOR ATTENTIONS!

Q&A

Data Available at

http://AppDialogue.com