IWSDS 11th Dec. 2009

Using Collaborative Filtering to Predict User Utterances in Dialogue

Ryuichiro Higashinaka, Noriaki Kawamae, Kohji Dohsaka, and Hideki Isozaki

NTT Communication Science Labs.

Motivation

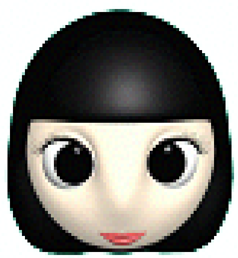
- Our goal is to improve closeness between a system and a user through dialogue
  - Useful for task-completion and continued use
- Inducing user agreement is important for improving closeness (Higashinaka et al., 2008)
- One possible way to induce user agreement is to say in advance what the user would say

Predicting what the user would say about cats and dogs.

Example

Closeness

I like cats because they are capricious, aren't they?



System

Yes I Agree!

User

And, I hate dogs because they bark at me.

Me too!

Suppose that (a) the user likes cats for their capriciousness.
(b) the user doesn't like dogs because they bark.

Previous Work

- Importance of closeness
  - Bickmore et al., (2001, 2005)
    - Closeness improves task-completion in possible face-threatening real-estate transactions
    - Closeness encourages the continued use of a health-care system
- Need for inducing user agreement
  - Higashinaka et al. (2008)
    - The number of dialogue acts corresponding to user agreement correlates with closeness

Goal of this work

- Predict user utterances to induce user agreement
- However, it is difficult to predict every user utterance

- Focus on predicting a user's evaluative expressions about entities
  - Entity: movies, books, animals, etc.
  - Evaluative expression:
    good, bad, interesting, terrible, cute, etc.

Approach

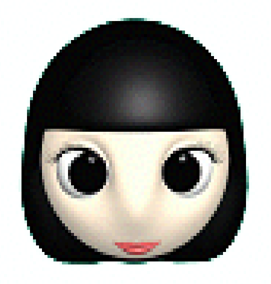
- Use collaborative filtering
  - A technique for using other users' information to model the behavior of a certain user (Breese et al., 1998)
  - Used in many recommendation systems
  - Has not been used to predict users' linguistic choices
- We predict a user's evaluative expressions about entities from other users' data

Approach (cont'd)

Current Topic
Entity Ej

How Ui used expressions for other entities

How other users used expressions for entities



System

Expression candidate	Score
Expr1	
Expr2	
ExprN	

User Ui

The score means how likely each expression can be used by Ui.

Approach (cont'd)

Current Topic

Entity Ej

How Ui used expressions for other entities

How other users used expressions for entities

System

Expression candidate	Score
Expr1	10
Expr2	30
ExprN	20

User Ui

The score means how likely each expression can be used by Ui.

Approach (cont'd)

- Two ways to use other users' data
  - Use similar users' expressions
    - When Ui is similar to Uj, assign large scores to the expressions used by Uj
  - Use similar entities' expressions
    - When Ej is similar to El, assign large scores to the expressions used for El by other users

Using similar users' expressions

- Assumption: similar users are likely to use similar expressions for the same entity
- Score of each expression:

Number of times Uh used ek

$$\mathrm{score}_{usim}(U_i, E_l, e_k) = \sum_{h=1}^{n} \boxed{\mathrm{sim}(U_i, U_h)} \cdot \boxed{\mathrm{freq}(U_h, E_l, e_k)}$$

Similarity between Ui and Uh

sim returns a variant of cosine similarity (Amatriain et al., 2009)

$$\mathrm{sim}(U_i, U_j) = \frac{\boldsymbol{u_i} \cdot \boldsymbol{u_j}}{|\boxed{\boldsymbol{u_i}}||\boldsymbol{u_j}|} \cdot \frac{2N_{i\cup j}}{N_i + N_j}$$

A vector indicating how Ui uses evaluative expressions

Example

User1

	good	bad	cute
cat	1	2	2
dog	2	0	7

User2

	good	bad	cute
cat	2	8	1
dog	0	1	2

User3

	good	bad	cute
cat	1	5	1
dog			

Suppose we want to assign scores for these

	good	bad	cute
User1	3	2	9
User2	2	9	3
User3	1	5	1

sim(User1,User3)
=cos([3,2,9],[1,5,1])
= 0.44
sim(User2,User3)
=cos([2,9,3],[1,5,1])
= 0.99

User3

	good	bad	cute
dog	0.88	0.99	5.06

2*0.44+0*0.99= 0.88

0*0.44+1*0.99 = 0.99

7*0.44+2*0.99= 5.06

System's predictions: cute>bad>good

Using similar entities' expressions

- Assumption: similar entities are likely to be expressed by similar expressions

- Score of each expression:

Similarity between Ej and El

$$\mathrm{score}_{esim}(U_i, E_j, e_k) = \sum_{l=1}^{m} \boxed{\mathrm{sim}(E_j, E_l)} \cdot \sum_{h=1}^{n} \boxed{\mathrm{freq}(U_h, E_l, e_k)}$$

Number of times users used ek for El

Example

User1

	good	bad	cute
cat	1	2	2
bear	0	1	7
dog	2	2	1

User2

	good	bad	cute
cat	2	3	1
bear	3	2	8
dog	3	2	1

User3

	good	bad	cute
cat	1	5	1
bear	2	1	2
dog			

Suppose we want to assign scores for these

	good	bad	cute
cat	4	10	4
bear	5	4	17
dog	5	4	2

sim(cat, dog)=
sim([4,10,4],[5,4,2])=
0.88
sim(bear, dog)=
sim([5,4,17],[5,4,2])=
0.62

User3

	good	bad	cute
dog	11.6	15.3	16.1

0.88*4+0.62*5+5
=3.52+3.1+5
=11.6

0.88*10+0.62*4+4
=8.8+2.48+4
=15.3

0.88*4+0.62*17+2
=3.52+10.54+2
=16.1

System's predictions: cute>bad>good

Using both similar users' and similar entities' expressions

- Assumption:
  - similar users use similar expressions
  - similar entities are likely to be expressed by similar expressions

- Score of each expression:

Number of times similar users used ek for El

$$\mathrm{score}_{\mathrm{usim+esim}}(U_i, E_j, e_k) = \sum_{l=1}^{m} \boxed{\mathrm{sim}(E_j, E_l)} \cdot \boxed{\mathrm{score}_{usim}(U_i, E_l, e_k)}$$

Similarity between entities Ej and El

Prediction Experiment

- Compare the prediction accuracy of four variations of our approach
  - (1) UserSim: use similar users’ expressions
  - (2) AnimalSim: use similar entities’ expressions
  - (3) UserSim+AnimalSim: use similar users’ and entities’ expressions
  - (4) Baseline: simple voting of other users’ expressions (UseSim without the user similarity weighting)
- Predict expressions of Ui for Ej by masking this information from data

Data

- Dialogue data
  - 1,000 human-computer dialogues (in text)
  - 50 users (25 males, 25 females)
  - Domain: Animal discussion
    - Participants talk about likes and dislikes about animals
  - Manually annotated with dialogue acts
- Extracted data for experiment
  - Sets of <Animal, Expressions> for 50 users
    - 47 evaluative expressions (mainly, adjectives) for 90 animals

Example Dialogue

	Utterance	Dialogue Act
System	Hello	Greeting
User	Hello	Greeting
System	Let's discuss likes and dislikes about animals.	Open-dialogue
User	Do you like insects?	Question
System	Do you like insects?	Question
User	I like fireflies.	Self-disclosure
System	You like fireflies.	Agreement
User	I like fireflies because they are beautiful.	Self-disclosure
System	I also like them very much.	Agreement
User	You too? They are evanescent, aren't they?	Question Self-disclosure

Entity:
firefly

Expressions:
beautiful
evanescent

Evaluation Criterion

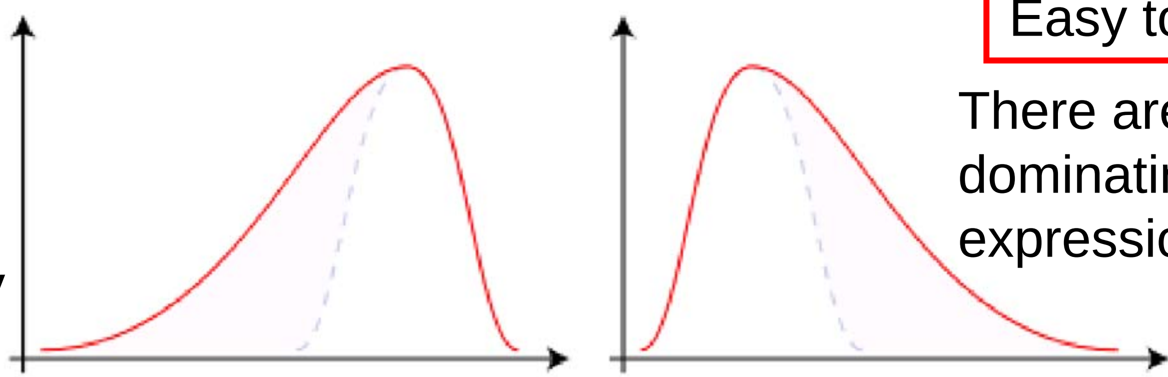
- Top-3 accuracy
  - the ratio of animals for which the top-3 predicted expressions contained those actually uttered by the user
- Limit the animals to make prediction
  - Some animals are expressed by many expressions → too difficult to predict
  - We set a skewness threshold to remove certain animals from evaluation

Skewness

- Capture the distortion of a distribution
- Calculate the skewness for the distribution of frequencies of evaluative expressions

Difficult to predict

There are many dominating expressions



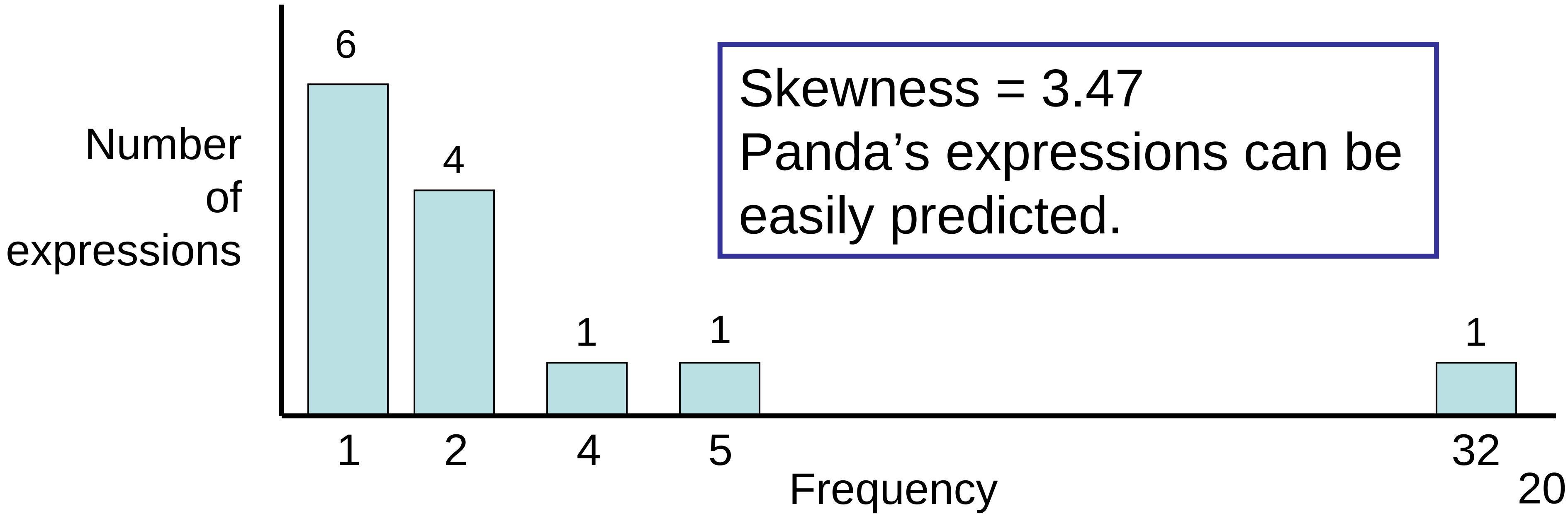
Easy to predict

There are only a few dominating expressions

(© Wikimedia)

Example

- Frequencies of expressions for Panda
  - cute(32), big(5), round(4), dangerous(2), white(2), good(2), eyes are cute(2), warm(1), sweet(1), black(1), small(1), strong(1), soft(1)



Skewness = 3.47
Panda’s expressions can be easily predicted.

Results

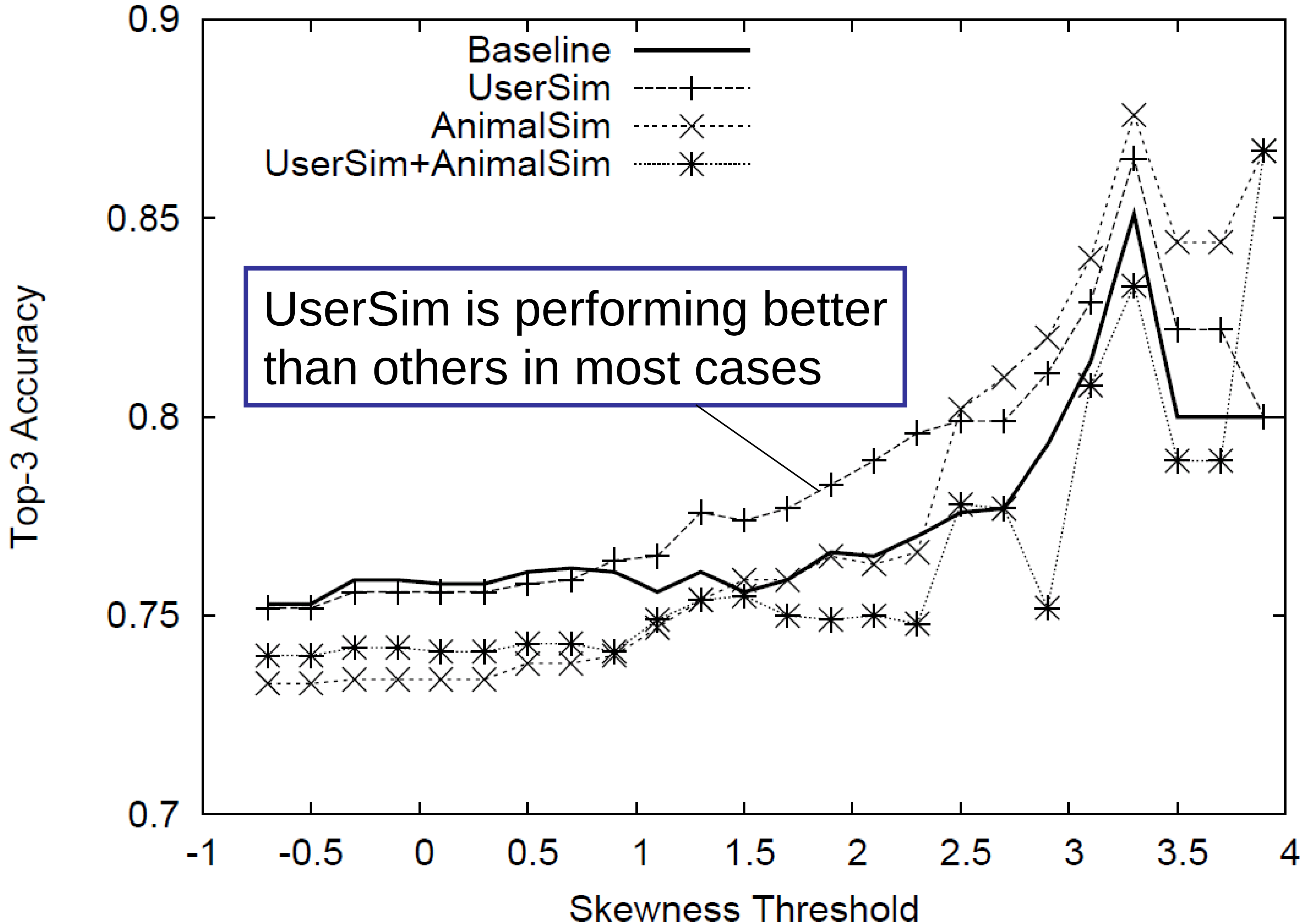
The number of animals over the threshold t

Top-3 accuracies depending on the skewness threshold t

	None(90)	t=1.0(74)	t=1.5(59)	t=2.0(35)	t=2.5(17)
Baseline	0.753	0.760	0.756	0.771	0.776
UserSim	0.752	0.765	0.774	0.794	0.799
AnimalSim	0.733	0.741	0.759	0.773	0.802
UserSim+AnimalSim	0.740	0.740	0.755	0.760	0.778

- Accuracy is generally high with 75~80%
- UserSim and AnimalSim significantly outperform Baseline for easy-to-predict animals
- AnimalSim performs poorly when no threshold is set
  ⇒ Some animals have specific expressions for them

Results (cont'd)



UserSim is performing better than others in most cases

Summary and Future Work

- Proposed using collaborative filtering to predict a user's evaluative expressions in order to increase closeness in dialogue
  - First to apply CF to the prediction of users' linguistic choices

- Similarity of users/entities is useful for improving the prediction accuracy

- Future work:
  - Verify our approach in an on-going dialogue
  - Expand to larger domains