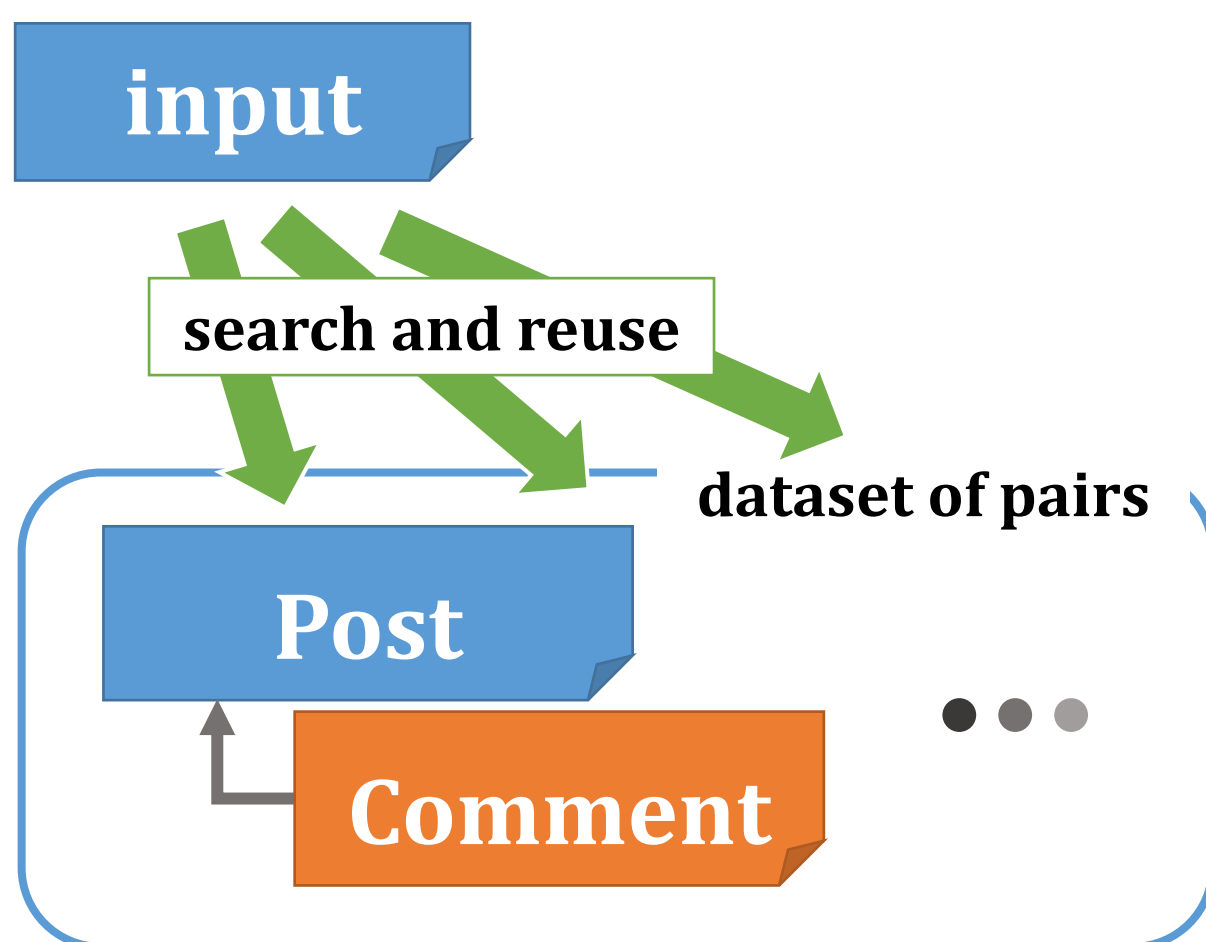


# Scoring of response based on suitability of dialogue-act and content similarity

Kyoto Institute of Technology

Sota Matsumoto, Masahiro Araki

## Short Text Conversation



When an utterance is given, we search for the appropriate utterances as the response from the repository of pairs consisting of post and comment on it.

## Scoring of Reply Utterance

To evaluate an utterance, we use the score obtained by multiplying following two evaluation values.

### ● Interactive Functional Suitability

The suitability in function of conversation with the utterance as a response to the input utterance.

### ● Content Similarity

Such as the story of soccer and lecture of university, similarity of topics with utterances.

$$Score(p, t_a) = ifs(p, t_a) * csim(p, t_a)$$

$ifs(p, t_a)$  and  $csim(p, t_a)$  are evaluation values of the above mentioned of any one of the utterance in the repository  $t_a$  to new given utterance  $p$ .

## Interactive Functional Suitability

The estimated values of interactive functional suitability are learned by the relationship of dialogue-act.

### □ Dialogue-act

Dialogue-act indicates the function of utterance in the conversation.

e.g. greeting, question, desire, etc.

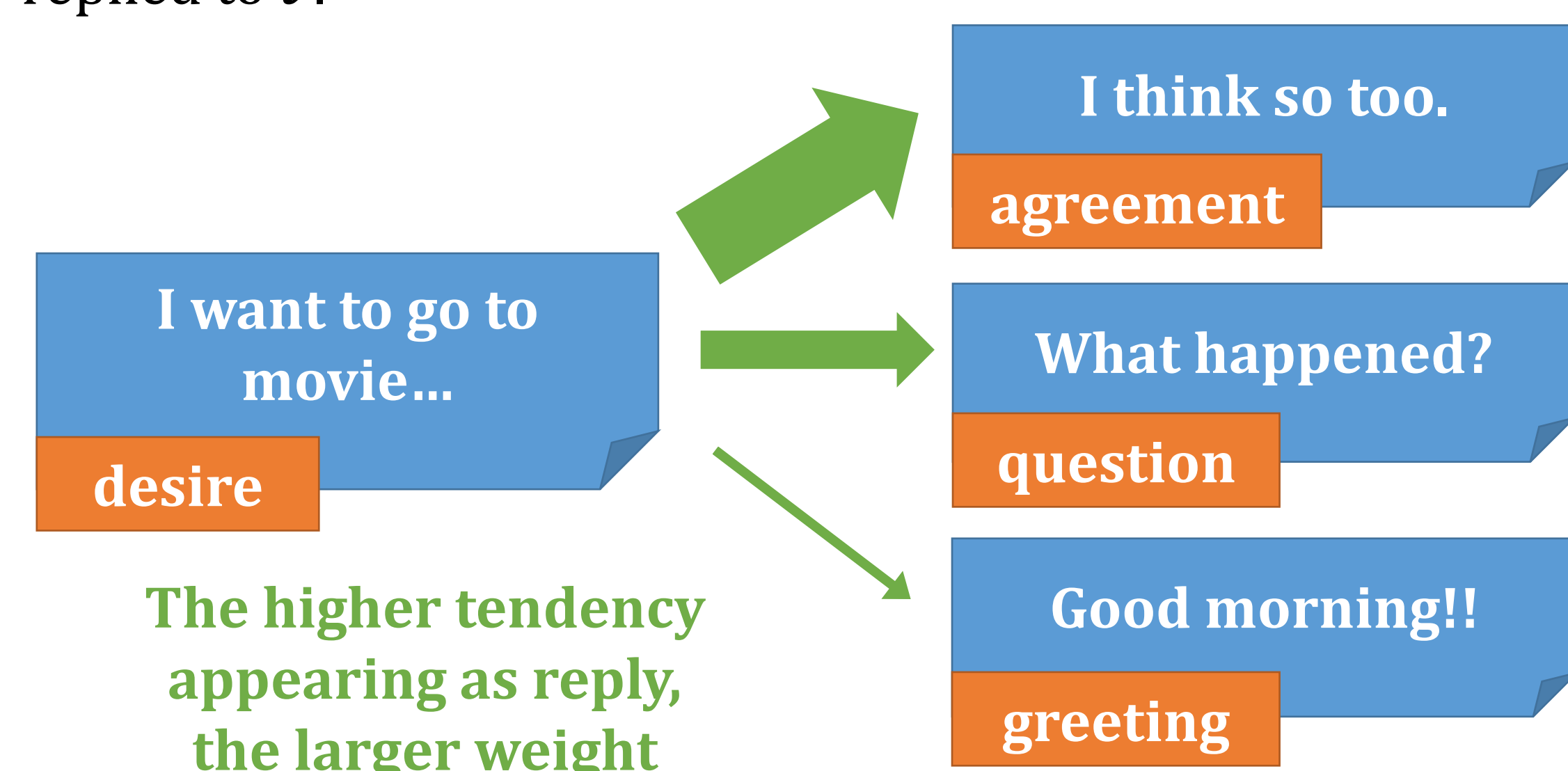
In order to design dialogue-acts adapted to the domain, we classify it by using the **Chinese Restaurant Process (CRP)**.

### □ Relationship weights

learn the relationships between dialogue-acts as weight table.

$$W[i][j] = \frac{count(i, j)}{N}$$

$count(i, j)$  is the number of times that dialogue-act  $j$  is replied to  $i$ .



The higher tendency appearing as reply, the larger weight

Using this table, it is possible to determine the suitability of utterances in the repository to a new one.

$$ifs(p, t_a) = W[dae(p)][dae(t_a)]$$

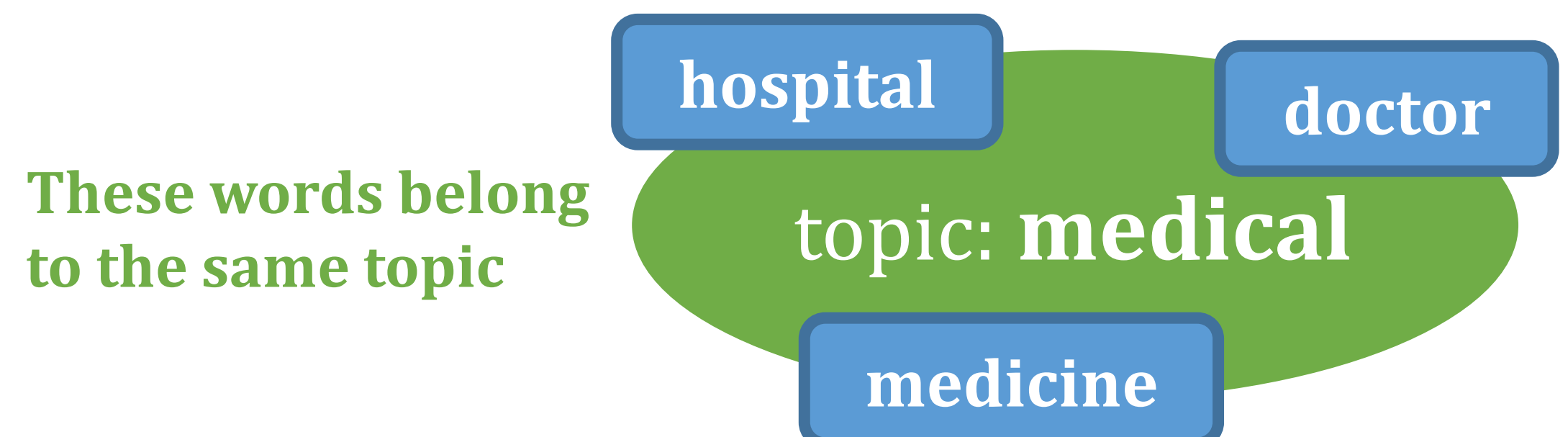
$dae(*)$  is a dialogue-act estimator for utterance.

## Content Similarity

Content similarity is obtained from cosine similarities of documents for the following 2 type of vectors.

### □ LDA (Latent Dirichlet Allocation)

estimate the potential topics to which the document belongs



### □ IDF (Inverse Document Frequency)

search documents with common words having high informativeness

$$csim(p, t_a) = \alpha * lsim(p, t_a) + (1 - \alpha) * isim(p, t_a)$$

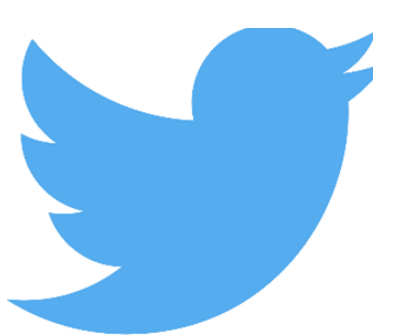
$\alpha$  is the parameter for adjusting the ratio of LDA and IDF

## Experimental Results

### □ Data

We used the posts to Twitter in training and testing.

- training data : 822,254 posts (411,127 pairs)
- testing data : 202 posts



### □ Dialogue-act Classification

As a result, training data were classified into 41 dialogue-act.

About half of the data belonged to one cluster, and factors that make up the cluster mainly were case particles that are not directly related to the function of utterance.

### □ Evaluation Results

10 annotators annotated in the three-degree evaluation of 0-2 according to the appropriateness of the response.

| $\alpha$ | case 2-1 | case 2-5 | case 12-1 | case 12-5 |
|----------|----------|----------|-----------|-----------|
| 0.0      | 0.2297   | 0.2050   | 0.5589    | 0.5380    |
| 0.4      | 0.1817   | 0.1743   | 0.4748    | 0.4535    |
| 0.5      | 0.1812   | 0.1660   | 0.4614    | 0.4317    |
| 1.0      | 0.0787   | 0.0787   | 0.2114    | 0.2130    |

In "case X-Y", X is a set of evaluation values to determine that the response is appropriate, and Y is the lowest rank number of the candidates evaluated in each utterance.

## Conclusion

- Select appropriate replies to a given utterance from the conversation data repository.
- Score an utterance by interactive functional suitability and content similarity.
- As a result, dialogue-act didn't basically work well, also content similarity was obtained a higher evaluation in the case of not using the LDA.
- We will filter the words used in determining factor in the classification of dialogue-act in the part-of-speech, also try increasing the number of dimensions of the vector of LDA.