SPECOM2016

Backchanneling via Twitter Data for Conversational Dialogue Systems

Michimasa Inaba and Kenichi Takahashi
Hiroshima City University, Japan

Dialogue Systems

- Task-oriented dialogue systems
  - Accomplish specific tasks
    - Reservation services
    - Directory-assistance services
- Non-task-oriented dialogue systems
  - Personal communication

Applications of Non-task-oriented Dialogue Systems

- Installation in humanoid robots
  - To build good relationships with humans
- Difficult dialogue tasks
  - To accomplish difficult tasks such as negotiation
- Entertainment

and so on…

Background

- Listener's active participation
  - Reaction and positive attitude (backchanneling) are essential for most speakers to talk and communicate effectively [Horiguchi 97].
- Backchanneling generation methods have been extensively studied

Related Works

- Previous methods for backchanneling
  - Using pitch patterns in the human user's utterances [Okato+ 96]
  - Using prosodic information [Ward+ 00]
  - Using estimated user's degree of interest about the topic [Kobayashi+ 13]

These studies employ a limited set of backchannels such as "hmm" or "sure."

Purpose

Generating a rich variety of backchanneling to realize smooth communication in non-task-oriented dialogue systems

- Approach
  - Employ Twitter data to train our model
    - Backchanneling is frequently used by Twitter users.
    - Easy to obtain a large amount of backchanneling data.
  - Use a recurrent neural network (RNN) to determine suitable backchannels

∗ backchanneling timing is ignored in this study.

Previous Works utilized RNN

Dialogue systems using RNN

- Response generation
  - Task-oriented [Tsung-Hsien+ 15]
  - Non-task-oriented [Cho+ 14] [Sordoni+ 15][Shang+ 15]
- These works utilized encoder-decoder model
  1. RNN encoder reads as input a variable-length word sequence and outputs a fixed-length vector
  2. Another RNN decodes a given fixed-length vector, producing an objective variable-length word sequence

In the proposed model, we use RNN as a feature extractor and classifier

Proposed Method

- Formulate the problem of what the backchanneling should return for given inputs as a multiclass classification problem
- Determine replies using this multiclass classifier
  - We use a Recurrent Neural Network (RNN) with long short-term memory (LSTM-RNN).
- Determine the reply (output) classes in advance to train the model

Example of Reply Classes

44 reply classes (a part is shown below)

The original Japanese replies are shown in parentheses.

That’s tough (すごいね)	I agree (同感です)	Sure (もちろん)	That's OK (大丈夫です)
So cute (かわいいよね)	So cool (かっこいいね)	I’m happy (嬉しいな)	That's good (よかった)
Thank you (ありがとう)	I’m sorry (ごめんね)	Awesome (さすがだね)	It's no go (だめだよ)
I see (そうなんだ)	You are right (確かにね)	Good luck (頑張って)	Is that true? (本当ですか)
Good for you (よかったね)	That's funny (笑えるね)	I'm jealous (羨ましい)	Sounds good (いいね)

Data Acquisition

- Tweet–reply pairs as training data
  - Ex. “So cool” tweet-reply data

so cool

Search

@ABC so cool

@DEF I wish my dog was human she is so cool

Does not begin with “so cool.”

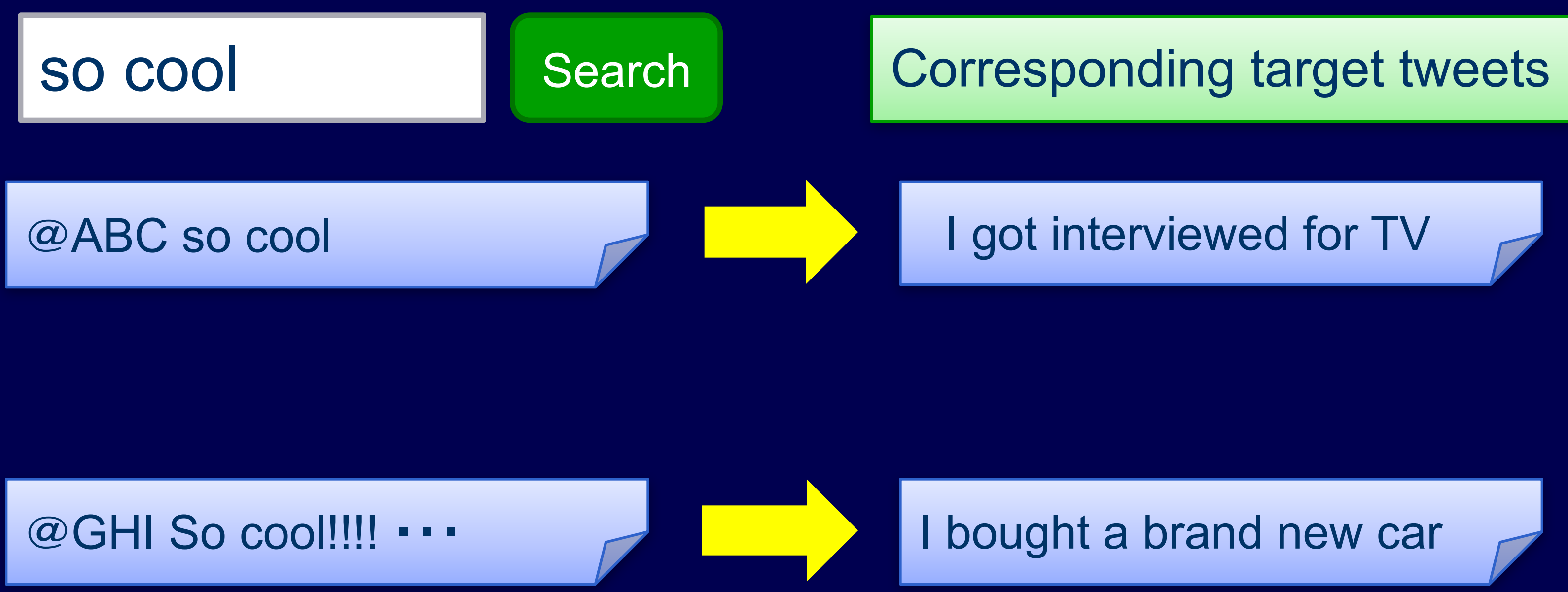
@GHI So cool!!!! ･･･

My grandma is so cool

This is not a reply.

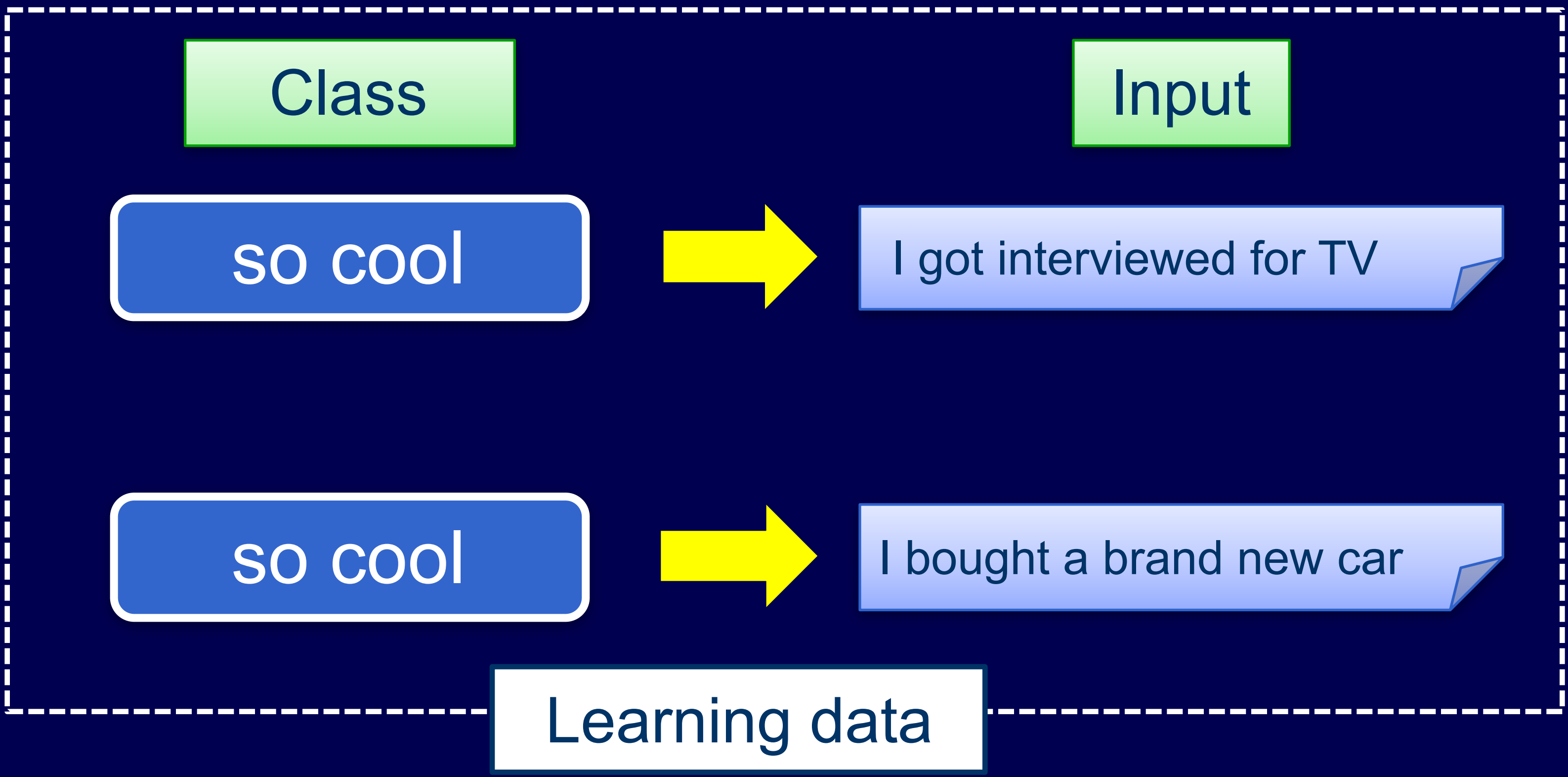
Data Acquisition

- Tweet–reply pairs as training data
  - Ex. “So cool” reply data



Data Acquisition

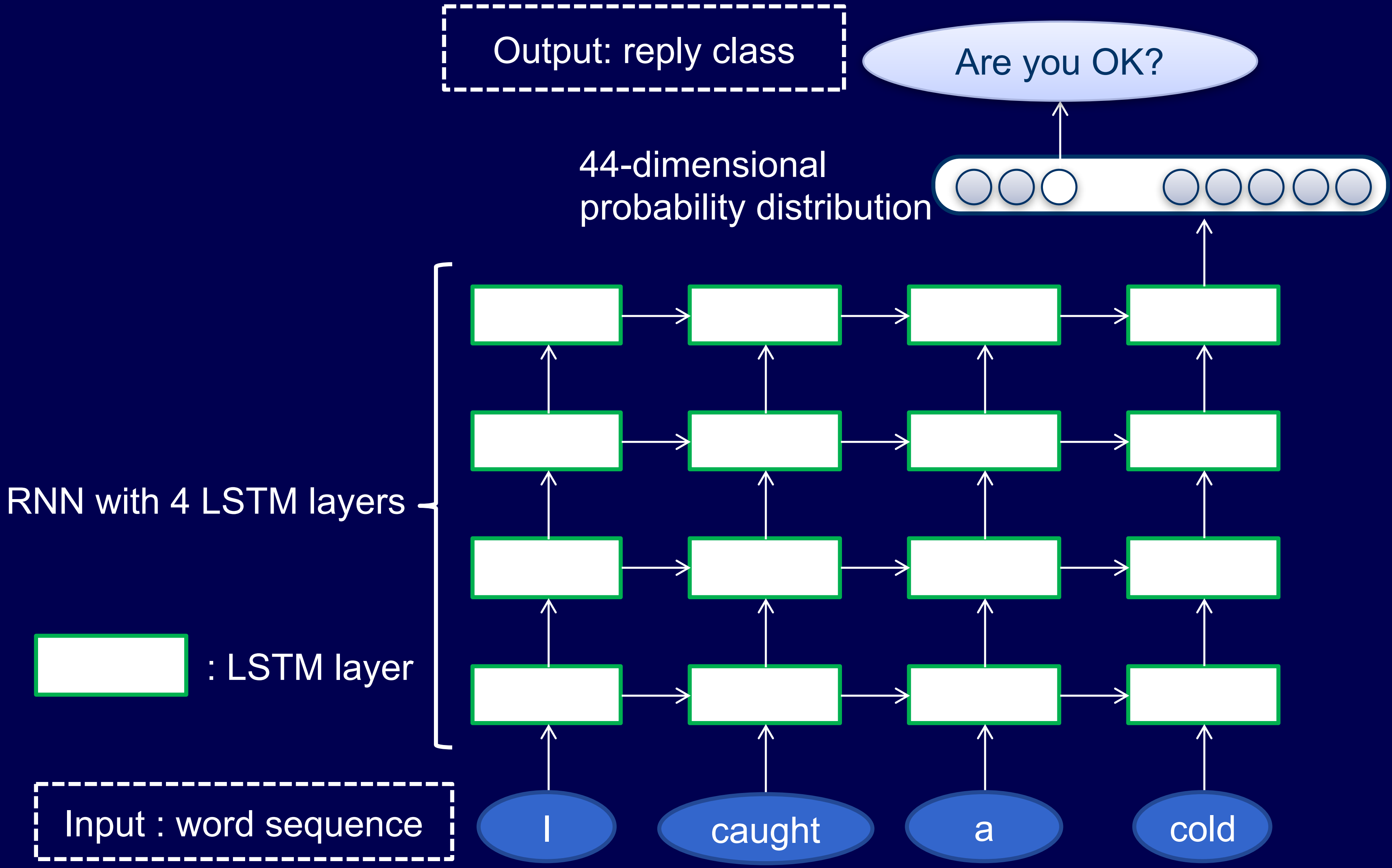
- Tweet–reply pairs as training data
  - Ex. “So cool” reply data



Long Short-Term Memory Recurrent Neural Network

- Recurrent Neural Network (RNN)
  - possesses an internal state
  - handles sequential data
- Long Short-Term Memory (LSTM)
  - Takes input and holds it selectively into a memory cell
- Use RNN with LSTM as a hidden layer (LSTM-RNN)

Proposed Model



Proposed Model

- Each LSTM layer has 1000 memory cells
- Input
  - 1000-dimensional distributed representation of words learned by Word2Vec [Mikolov+ 13]
- Output
  - 44-dimensional probability distribution corresponding to each reply class
- Trained using AdaGrad [Duchi+ 11]

Experiment

Experiment

- Automatic evaluation
  - Calculate the co-incidence ratio between our method’s outputs and the original replies in Twitter
- Manual evaluation
  - Human subjects evaluate our method’s outputs

Data

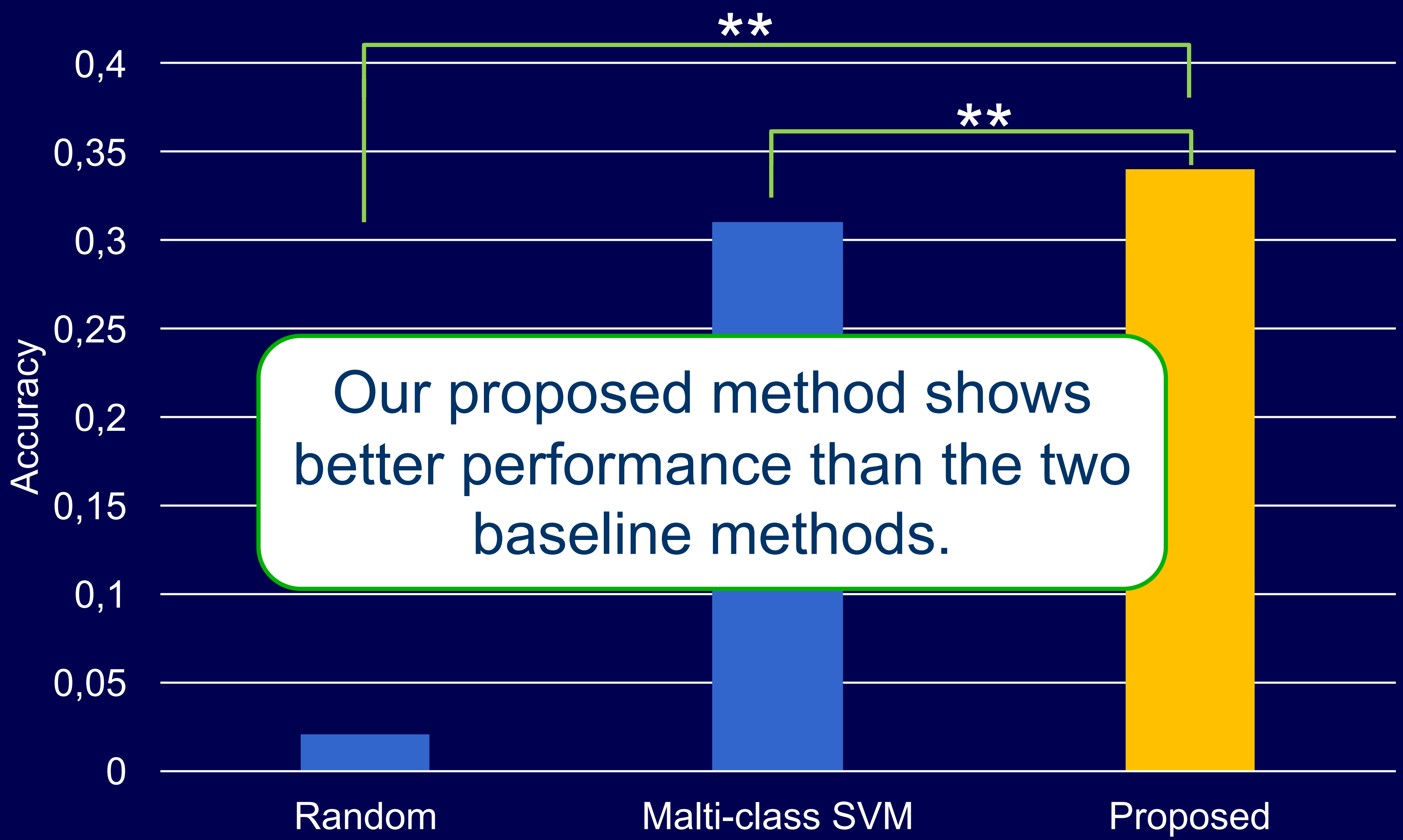
- 460,000 Japanese pairs of tweets and replies
  - 455,000 pairs for training the model
  - 5,000 pairs for evaluation
- Obtained 44 equally distributed reply classes

Baseline Methods

- Random
  - Randomly selects a reply class from among the 44 classes
- Multiclass Support Vector Machine
  - LIBSVM [Chung+ 11]
  - Unigram and trigram features
  - Linear kernel

Result

Result: Automatic Evaluation



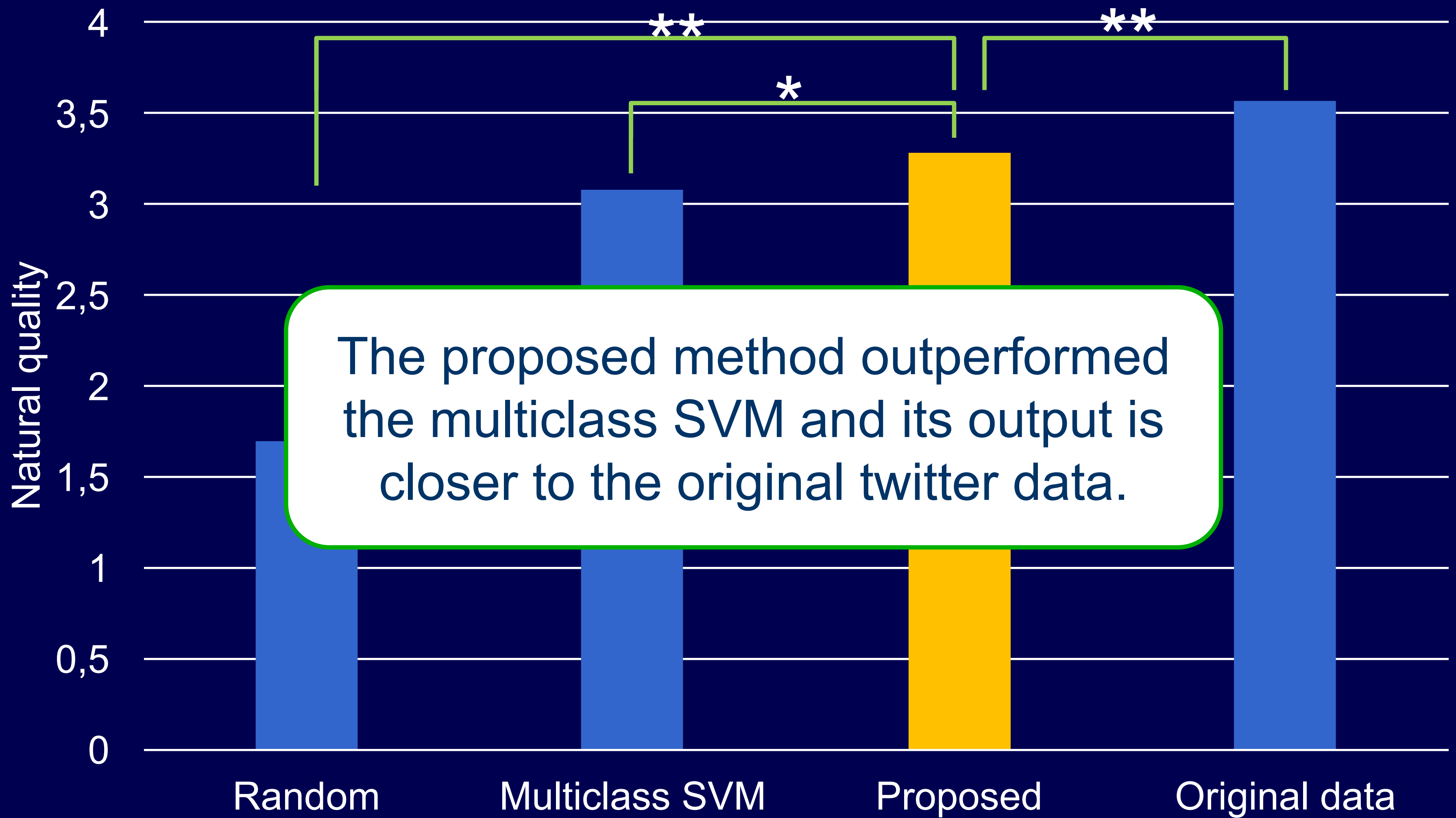
Our proposed method shows better performance than the two baseline methods.

** Significant difference at the 1% level by McNemar's test

Manual Evaluation

- Randomly selected 200 data pairs from 5,000 pairs
- Two human subjects evaluated outputs from each method for each given tweet
  - Judged the natural quality using a five-point Likert scale

Result: Manual Evaluation



The proposed method outperformed the multiclass SVM and its output is closer to the original twitter data.

* Significant differences at the 5% level by the t-test

** Significant differences at the 1% level by the t-test

Summary: Experimental Result

- Automatic evaluation
  - Our proposed method showed better performance than the two baseline methods.
  - Accuracy of our proposed method (0.34) is not very high.
- Manual evaluation
  - Natural quality of output of our proposed method is better than that of the multiclass SVM and is closer to the original twitter data.

Conclusions

- Proposed a method for generating a rich variety of backchanneling
- Formulated the problem of what backchannel to return for a given utterance as a multi-class classification problem
  - A suitable reply class is determined using an LSTM-RNN.
- Experimental results demonstrated that our method significantly outperformed baseline methods.

Future Work

- Reduce noise in the training data
  - Twitter data contain a substantial amount of noise
  - The proposed method could potentially be improved by decreasing noise from the training data by implementing a filtering technique.
- Backchanneling timing control to build a spoken conversational dialogue system