Captioning Events in Tourist Spots by Neural Language Generation

Mai Nguyen[1], Koichiro Yoshino[1,2], Yu Suzuki[1,3], Satoshi Nakamura[1]
[1]Nara Institute of Science and Technology
[2]Japan Science and Technology Agency
[3]Gifu University

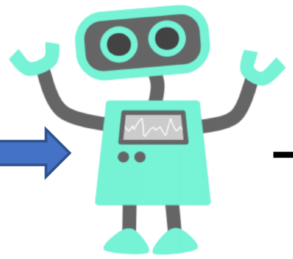
Outline

- Motivations
- Related works
- System Architecture
- Experiments
- Conclusion

Motivations

Data sources

Language Generation



Fushimi Inari is very crowded, you should go there in early morning

Natural Language Generation (NLG) will be a good interface to output info. for user
- User can understand information immediately
- Giving textual description is one of the easiest ways to present information

Objectives

The task of generating textual descriptions about tourist spots to support user decision making in tourist domain:

Important factors:

1. **Informativeness**
2. **Naturalness**

Related works

The task of generating text from data has been investigated in many domains

1. Weather forecast (Belz et al., 2008)
2. Navigation assistance (Dale et al., 2003)
3. Sports (Liang et al., 2009)
4. Market Comments (Murakami et al., 2017)

NLG approaches:
Rule-based: Define a set of rules to map frames to NL

Pros: simple, error-free, easy to control

Cons: time-consuming, poor scalability

=> Neural-based NLG

Pros: Access millions of traveler reviews
Cons: Out-of-date information

This paper introduces …

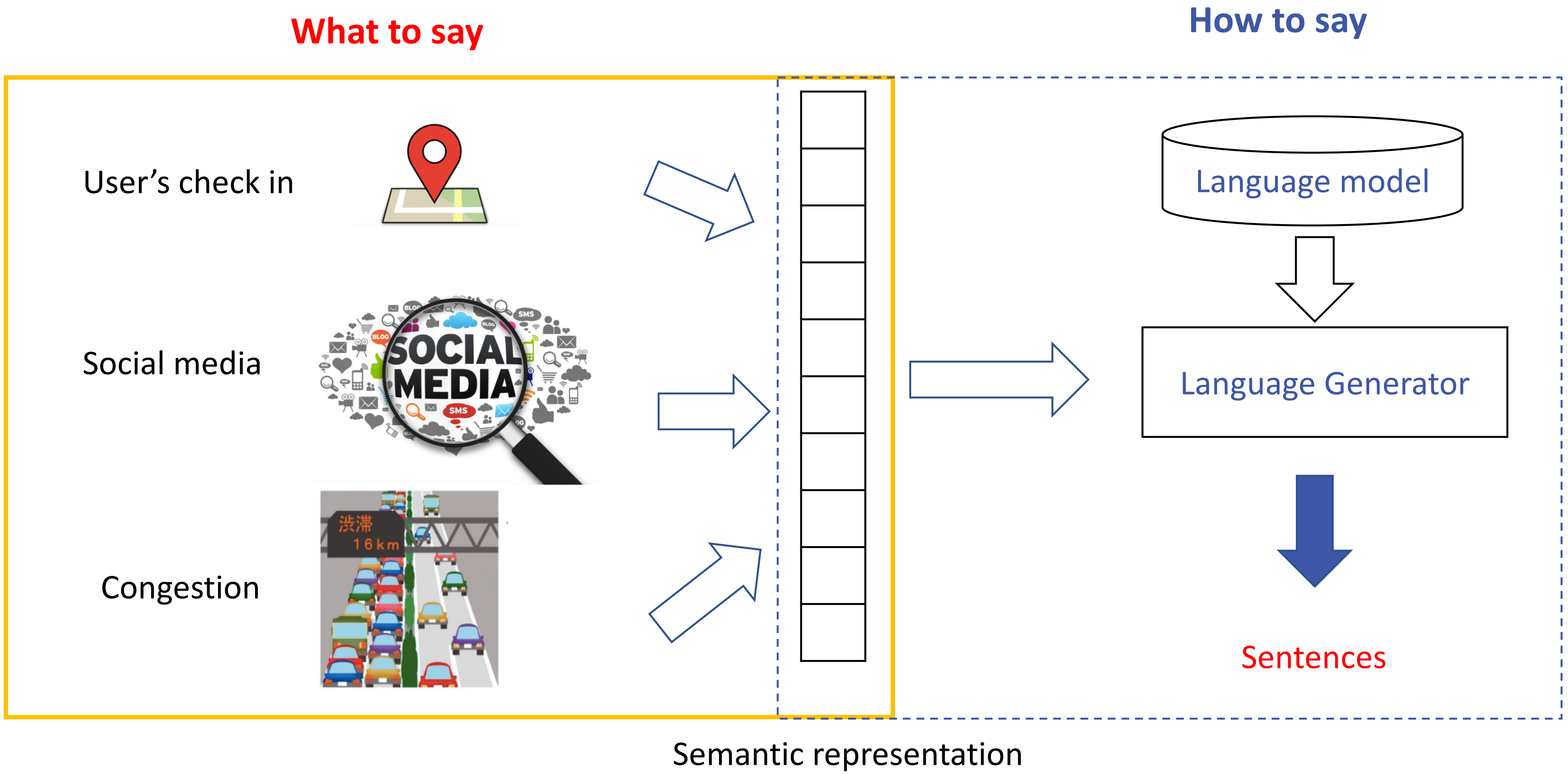
- **A system for producing textual description about tourist attractions**

✓ Integrate a neural language generation

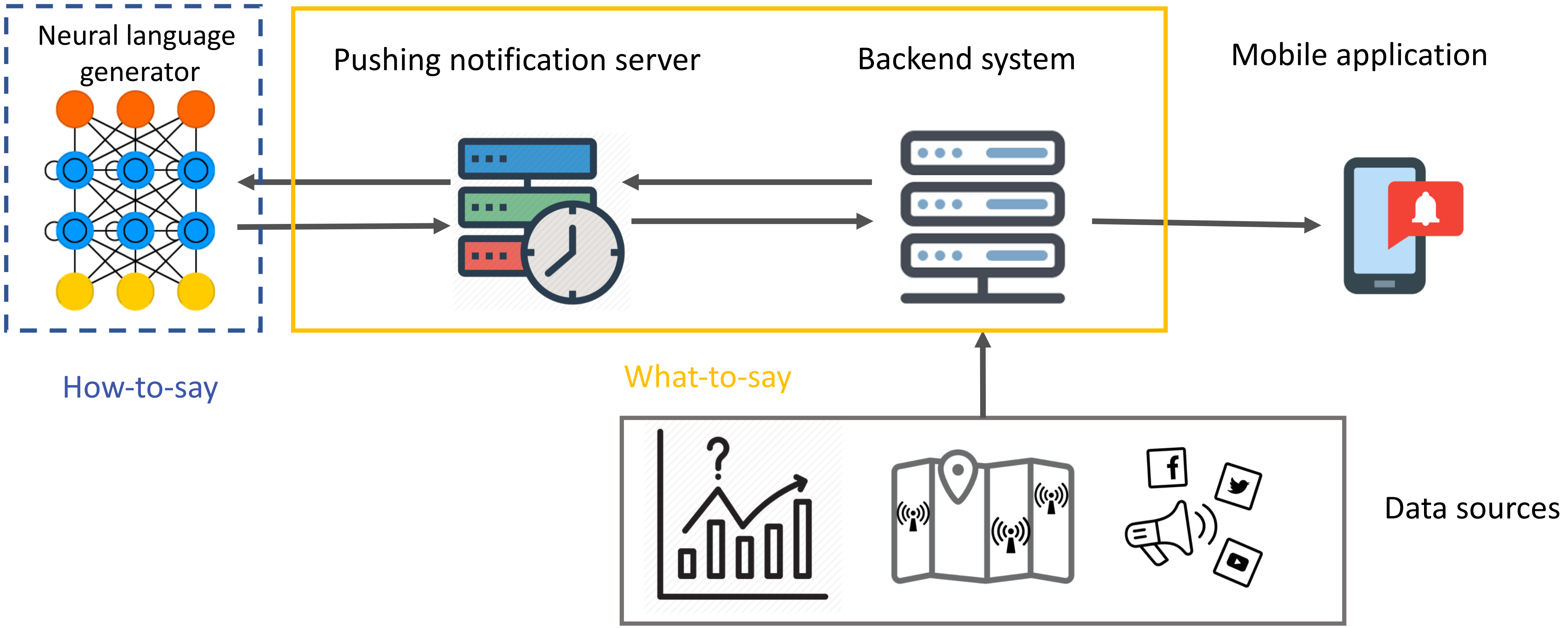
✓ Summarize a multiple data resources such as infrared sensor, social media, user’s check-in

✓ coupling with backend-server, we developed a real-time application with up-to-date information.

NLG Pipeline



System Architecture

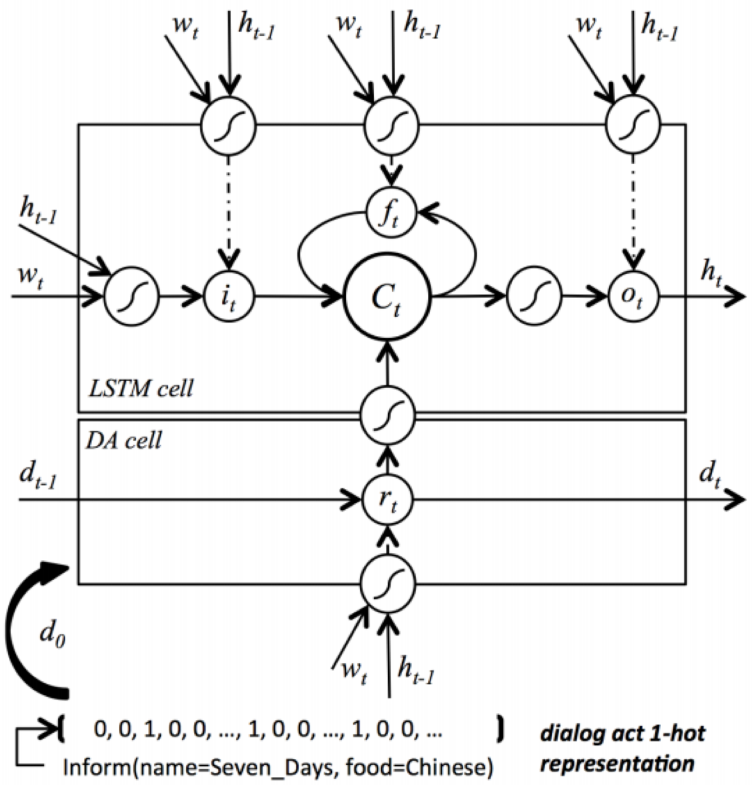


What-to-say

- Backend server: extract and summarize data from several information sources
  - Infrared sensor, Twitter, user's check-in
- Pushing server:
  - Check new-update information from backend with a specific time
  - Create a semantic expression, N-hot vector of frames that have slot values expression.
  - Send the pushing request to the back-end server

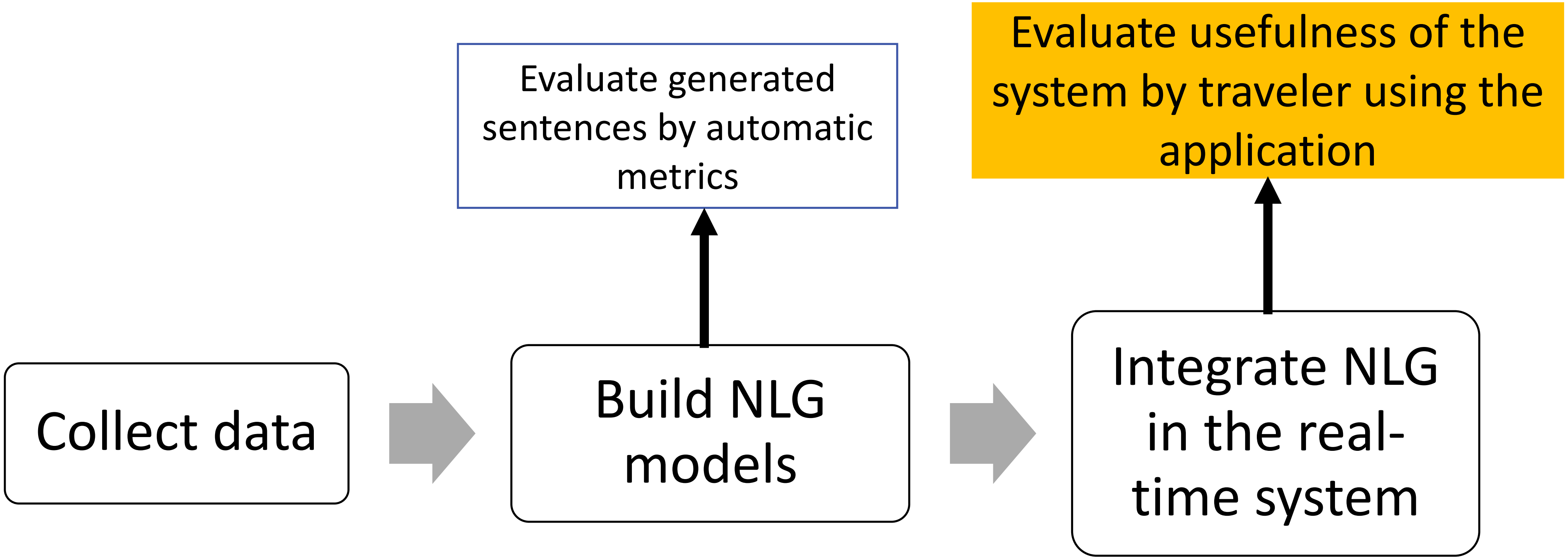
How-to-say, Neural Language Generation

- **Utilize semantically controlled LSTM, proposed by** [Wen et al, 2015]
- Core Idea: using a gate mechanism to control the generated semantics (dialog act/slots)
- Works well in limited domain, if the "what to say" is decided



Experiments

The target application is a system that describes comprehensive, up-to-date information about tourist attraction by NLG



Dataset

Attribute	Data Type	Example Value	Data Sources
name	verbatim string	kyoto tower	POIS
event	verbatim string	cherry blossom, gion festival	official websites
state_event	dictionary	happening, finished	official websites
crowded	dictionary	high, low, average	infrared sensors
recommended	boolean	yes/no	recommendation system
time	enumerable	now, holidays	
popular	boolean	yes/no	users check-in

List of possible attributes in the our dataset, data sources and examples

Data collection

1. Corpus: A pair of meaning representation (MR) and corresponding references pairs in the tourist domain.
   - Meaning representation (MR): a set of pairs key-value pairs
   - Reference: natural language utterance describing MR

MR	name=[Fushimi Inari], crowded=[high], time=[festival days], recommended=[no]
Reference	It is a good idea to avoid Fushimi Inari during festival days because it is extraordinary crowded.

2. Human annotation: workers were recruited via crowd sourcing service.
3. Results: ~ 3300 data points collected in English

Pre-process dataset

- Remove useless instances: which provide not enough information (**Informativeness**)
- Grammar correction: detected by Grammarly and fixed manually, ~ 15%. (**Naturalness**)
- Delexicalization: some properties such as name and event to avoid data sparsity.

"name[maizuru park], event[cherry blossoms], event_state[happening], crowdedness[2]"
"maizuru park is not very crowded, the cherry blossoms festival is currently held there."

↓

X-name is not very crowded , the X-event festival is currently held there .

Evaluation of NLG

- Corpus is split into training, validation, test set in the ratio 85%:6%:9%

	Training	Validation	Test
REF	2800	200	296
MR	283	80	83

- Training:
  - character-based version of SC-LSTM (hidden size 1024, batch size 256)
  - Adam optimizer,
  - Dropout,
  - Beam search: beam size = 10
- Baseline: rule-based
- Metrics: BLEU, NIST, METEOR, ROUGE-L, CIDEr

Result

- Automatic metrics

System	BLEU	NIST	METEOR	ROUGE-L	CIDEr
Rule-based	0.41	5.67	0.36	**0.67**	2.23
Our generator	**0.43**	**6.03**	**0.37**	0.64	**2.70**

Improvement:
- ✓ BLEU: 0.02
- ✓ NIST: 0.36
- ✓ METEOR: 0.01
- ✓ CIDEr: 0.47

Examples of generated text

N. of slots	Example of description generated corresponding to MR
3	*name[nanzenji temple], event[autumn leaves], popular[yes]* Autumn leaves in nanzenji temple is popular.
4	*name[Yoyogi Park], crowded[no], time[now], recommended[no]* Yoyogi Park is not crowded right now, but it is not recommended to visit.
5	*name[yanagidani kannon], event[autumn leaves], state[happening], crowded[high], recommended[no]* Autumn leaves is happening in yanagidani kannon, it is extremely crowded, you should not go there.
5	*name[kyoto imperial palace], event[aoi festival], state[happening], crowded[low], recommended[yes]* Aoi festival is happening in kyoto imperial palace, it is medium crowded, you should go there.

The **NLG model can produce fluent text** and would often repeat the same sentence structure multiple times

Results analysis

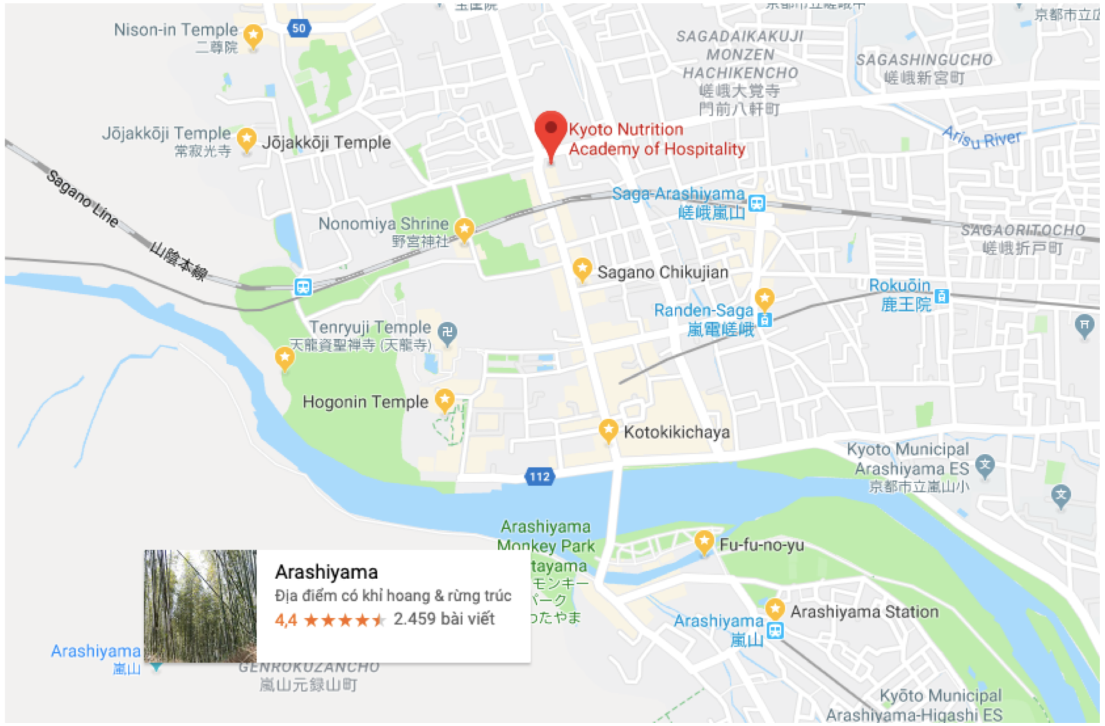
Human Evaluation - in term of informativeness

No. slots	Total of MR	Number of correct generated text	Percentage
3	18	17	94.4%
4	54	49	90.7%
5	11	5	45.4%
Total	83	70	85.54%

Neural LG have difficulty capturing long-term structure MRs

Experiment in Real Field

- Place: Arashiyama area, Kyoto
- Participants
  - 12 students
  - 19 ordinary people
- Task descriptions
  - Use the application during their trip
  - Visit at least 3 POIs from 10 AM to 3 PM
  - Answer the questionnaire



Application

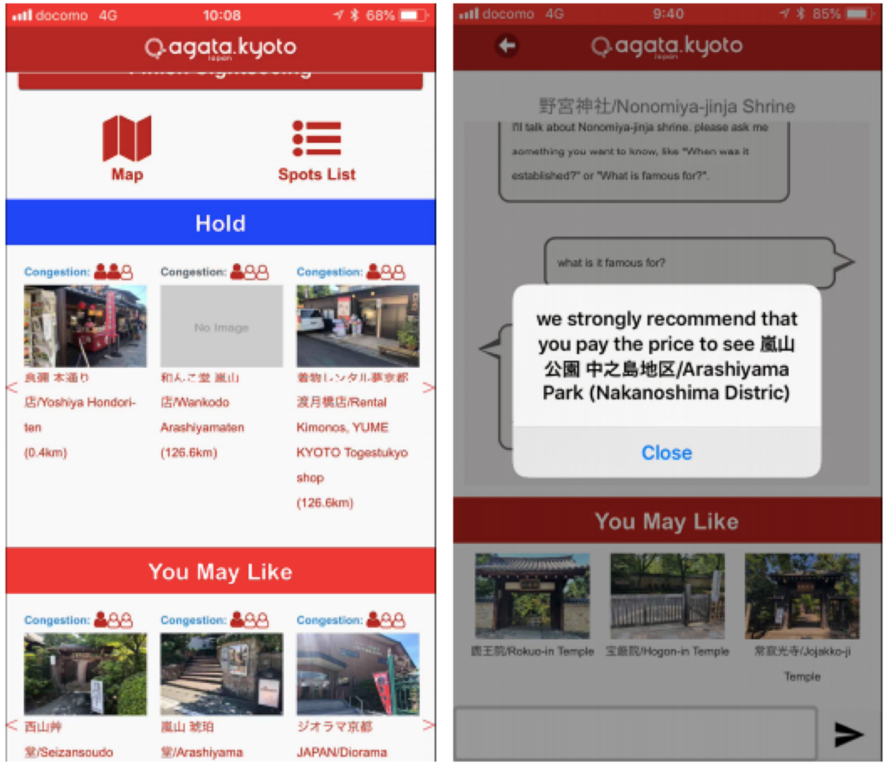


図 3　提案システムの画面.

Fig. 3 Screen shots of the proposed application system.

Experimental Results

Evaluate usefulness of the pushing function in range 1-5:

Experiment	Answers					Ratio of satisfied
	5	4	3	2	1	
Students	1	1	5	1	4	58%
Ordinary people	2	1	3	3	1	60 %

- 5: Very satisfied
- 4: Rather Satisfied
- 3: Satisfied
- 2: Not Satisfied
- 1: Not Satisfied at all

Conclusions

- Our generation model produces fluent text overall, work smoothly on real-time application.
- In term of usefulness, our system could support decision-marking for tourists.
- Limitations:
  - Neural LG have difficulty capturing long-term structure because the training data is quite small
  - The application's notification was too much for travelers.

Reference

1. Tsung-Hsien Wen, Milica Gasic, N. M. P. S. D. V. and Young, S.: Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems, Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing(EMNLP), pp. 1711-1721 (2015).
2. Sripada, S. G., Reiter, E. and Davy, I.: SUMTIMEMOUSAM: Congurable Marine Weather Forecast Generator (2003)
3. Belz, A.: Automatic Generation of Weather Forecast Texts Using Comprehensive Probabilistic Generation-space Models, Nat. Lang. Eng., Vol. 14, No. 4, pp. 431-455 (online), DOI: 10.1017/S1351324907004664 (2008).
4. Dale, R., Geldof, S. and Prost, J.-P.: CORAL: Using Natural Language Generation for Navigational Assistance, Proceedings of the 26th Australasian Computer Science Conference - Volume 16, ACSC '03, Darlinghurst, Australia, Australia, Australian Computer Society, Inc., pp. 35{44 (online), available from (2003)
5. Liang, P., Jordan, M. and Klein, D.: Learning Se-mantic Correspondences with Less Supervision, Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Suntec, Singapore, Association for Computational Linguistics, pp. 91-99 (2009).