# CUTKB at NTCIR-14 QALab-PoliInfo Task

Toshiki Tomihira and Yohei Seki
University of Tsukuba, Japan

June 12th, 2019@NTCIR-14

# INDEX

1. **Motivation**
2. Classification task
3. Our approach
4. Evaluation results
5. Summary

# Motivation

The rise of social media -> democratized content creation and has made it easy for everybody to share and spread information online.

## ON POSITIVE SIDE

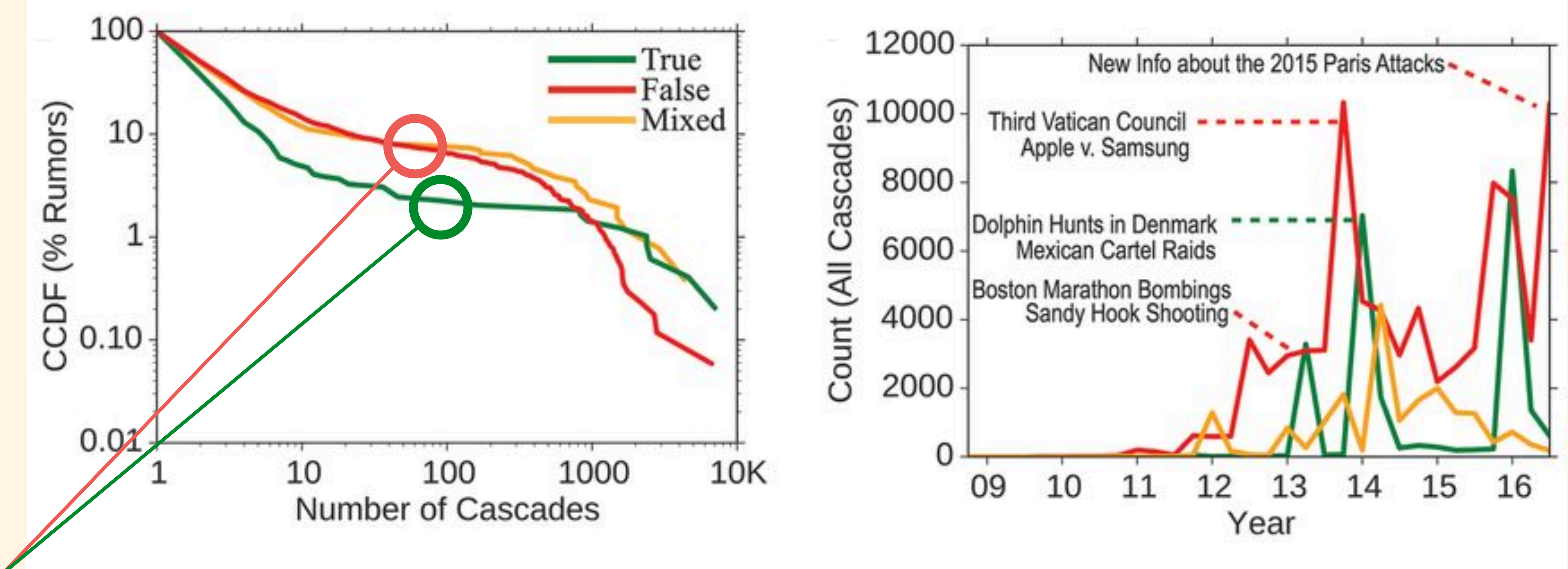
We enable much faster dissemination of information compared to what was possible with newspapers, radio, and TV.

## ON NEGATIVE SIDE

Stripping traditional media from their gate-keeping role has left the public unprotected against **the spread of misinformation**, which could now travel at breaking-news speed over the same democratic channel.

# Background (1)

The graph shows the results for the spread of true, false, mixed rumors using *Twitter* dataset [Vosoughi et al., 2018].

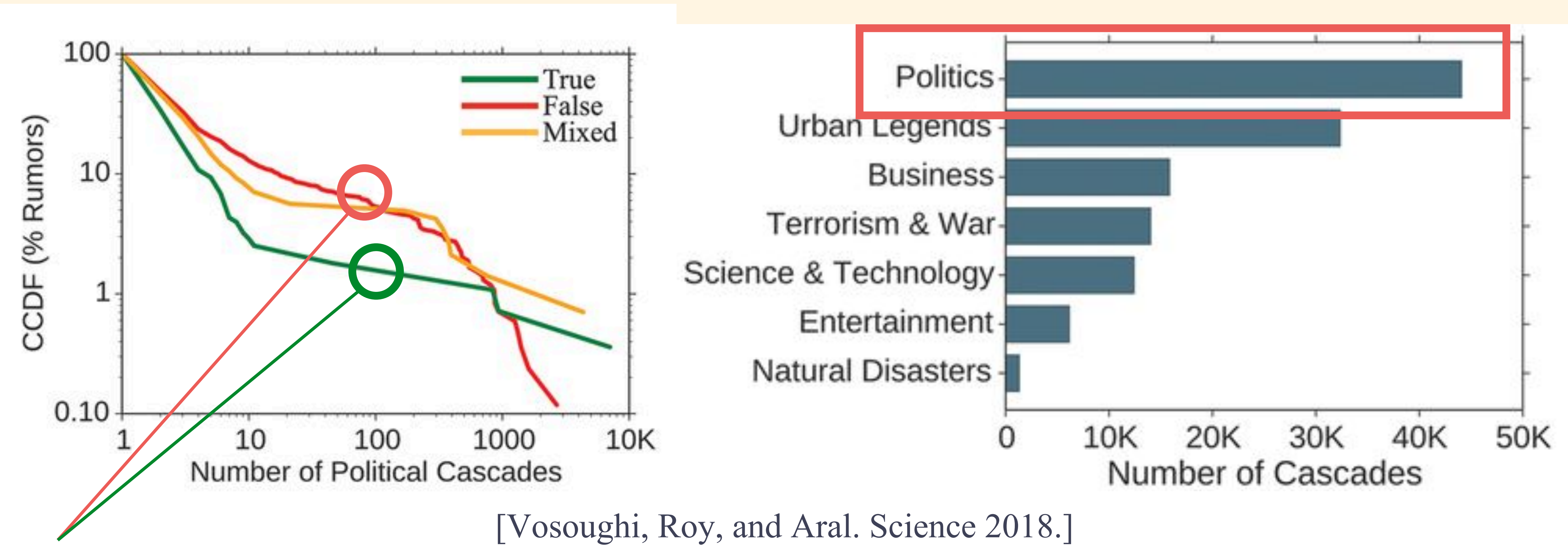



[Vosoughi, Roy, and Aral. Science 2018.]

False news reached at more people and diffused faster than the truth.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. Science, 359(6380):1146–1151.

# Background (2)




[Vosoughi, Roy, and Aral. Science 2018.]

Much politics rumors are in circulation, but less true.

**→Fake news has become a social problem.**

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. Science, 359(6380):1146–1151.

# INDEX

1. Motivation
2. **Classification task**
3. Our approach
4. Evaluation results
5. Summary

# Task Definition

## Goal

To find “opinion with a factual verifiable basis” from politician’s utterance.

## Inputs and outputs

Inputs: “Topics” and “Politicians’ utterance”

Output: labels for three attributes

## Labels

1. Relevance: 0 or 1
2. Fact-checkability: 0 or 1
3. Stance: support, against or other

# Label Examples

| ID | utterance | Relevance | Fact-checkability | Stance |
|---|---|---|---|---|
| 1 | I do not agree with the transfer of the new bank Tokyo or the Tsukiji market. | TRUE | FALSE | against |
| 2 | The Tokyo Metropolitan Government conducted construction work on soil contamination of Toyosu on August 30th. | TRUE | TRUE | other |
| 3 | Toyosu is an area where visitors can expect customers by new market relocation. | TRUE | TRUE | support |

# INDEX

1. Motivation
2. Classification task
3. **Our approach**
4. Evaluation results
5. Summary

# Our approach

**Fact-checkability**

→LSTM+CNN

**Relevance**

→ LSTM Model of two input

**Stance**

→ Simple LSTM Model

# Approach: Fact-checkability

| |
|---|
| 空き家の活用につきましては、これまで福島県空き家・古民家相談センターを設立し、県内への定住希望者に対する情報提供や改修相談等を実施してまいりました。 |
| 空き家の活用事例としましては、綾町と諸塚村において、空き家を改修した再生利用が行われております。 |
| 厚生環境委員会で奄美大島を視察させていただいたとき、空き家対策として、人口減対策として市が所有者から空き家を借り上げ、空き家を改修し、Ｕターン、Ｉターン、奄美への移住者に低家賃で貸し出して人口減対策に取り組んでいるという話を聞いてまいりました。 |

Blue underline : important verbs to confirm factuality.

Green underline : fact checkable parts.

Red underline : clauses shared between documents.

Common clauses or words between documents are important clues
→ LSTM + CNN

# Approach: Fact-checkability

Improve judgment by performing convolution and time series prediction:

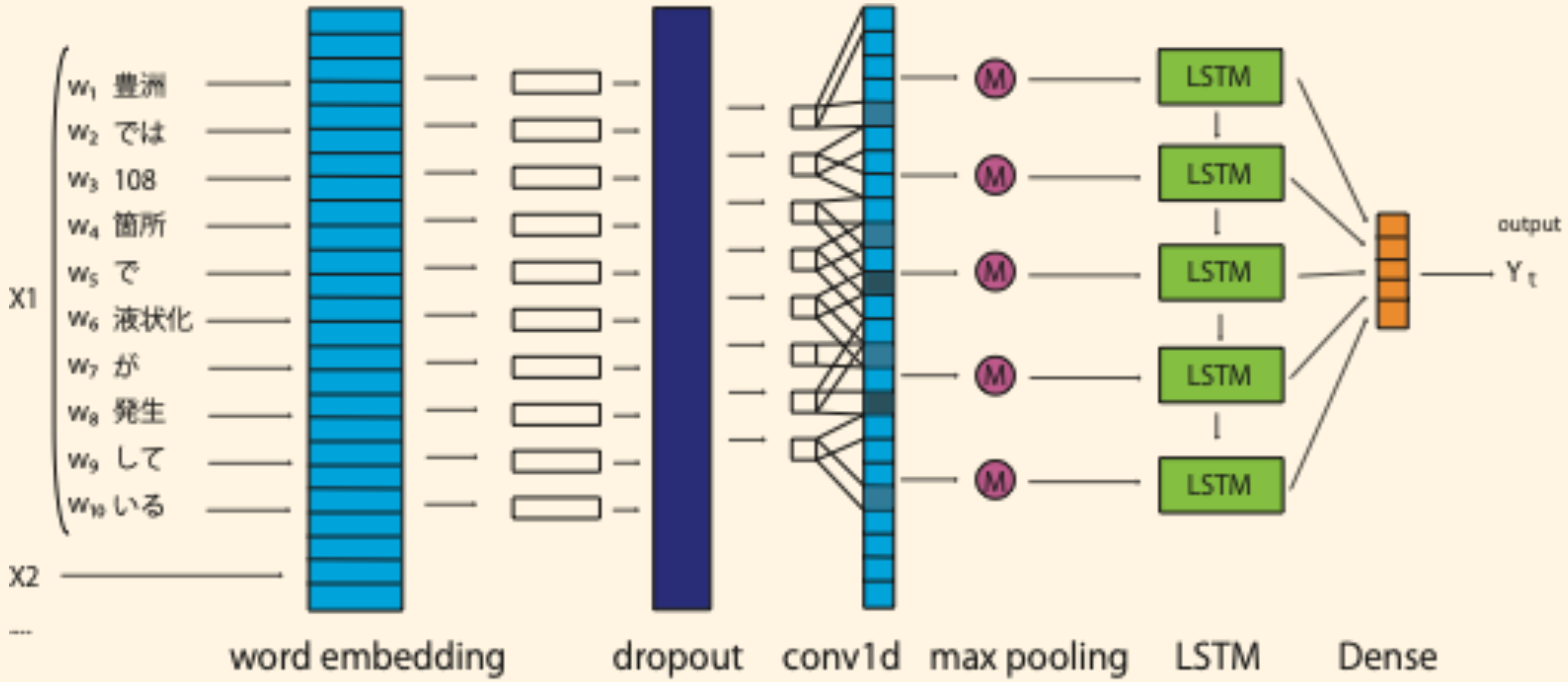
- The relationship between the minutes
  could be taken into consideration as a substitute for evidence.

We compared two models using validation dataset:

- Combine LSTM and CNN models.
- LSTM model only.

Combined models are better!

# Approach: Fact-checkability



# Approach: Relevance

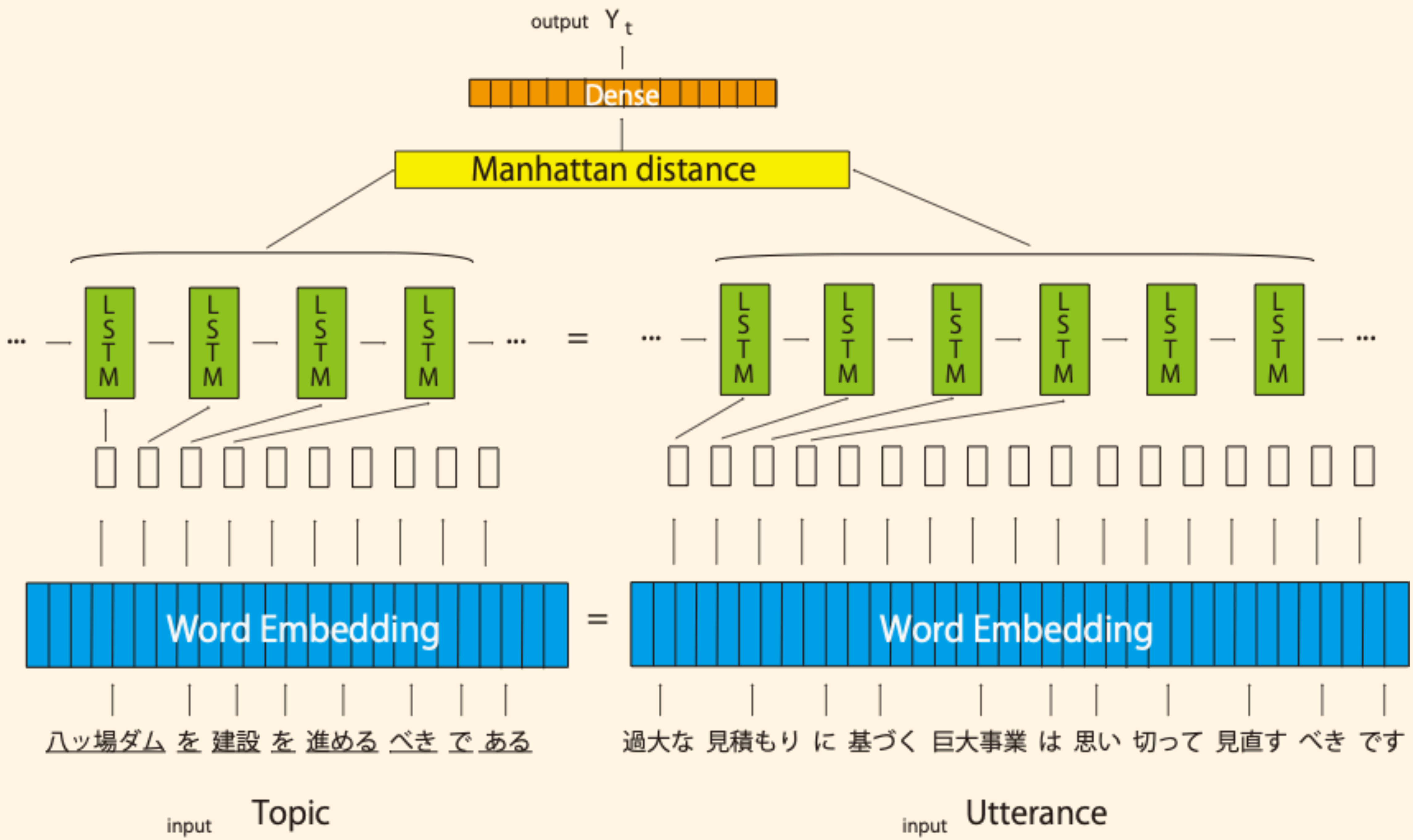
- Binary classification task: “relevance” or “irrelevance”
- Inputs: “Topic” and “Utterance”

We defined optimizer as *Manhattan distance*
between two LSTMs obtained from “Topic” and from “Utterance”.

Manhattan distance

$$\textbf{optimizer} = \exp(\, -||\mathrm{h}^{(left)} - h^{(right)}||_1)$$

# Approach: Relevance

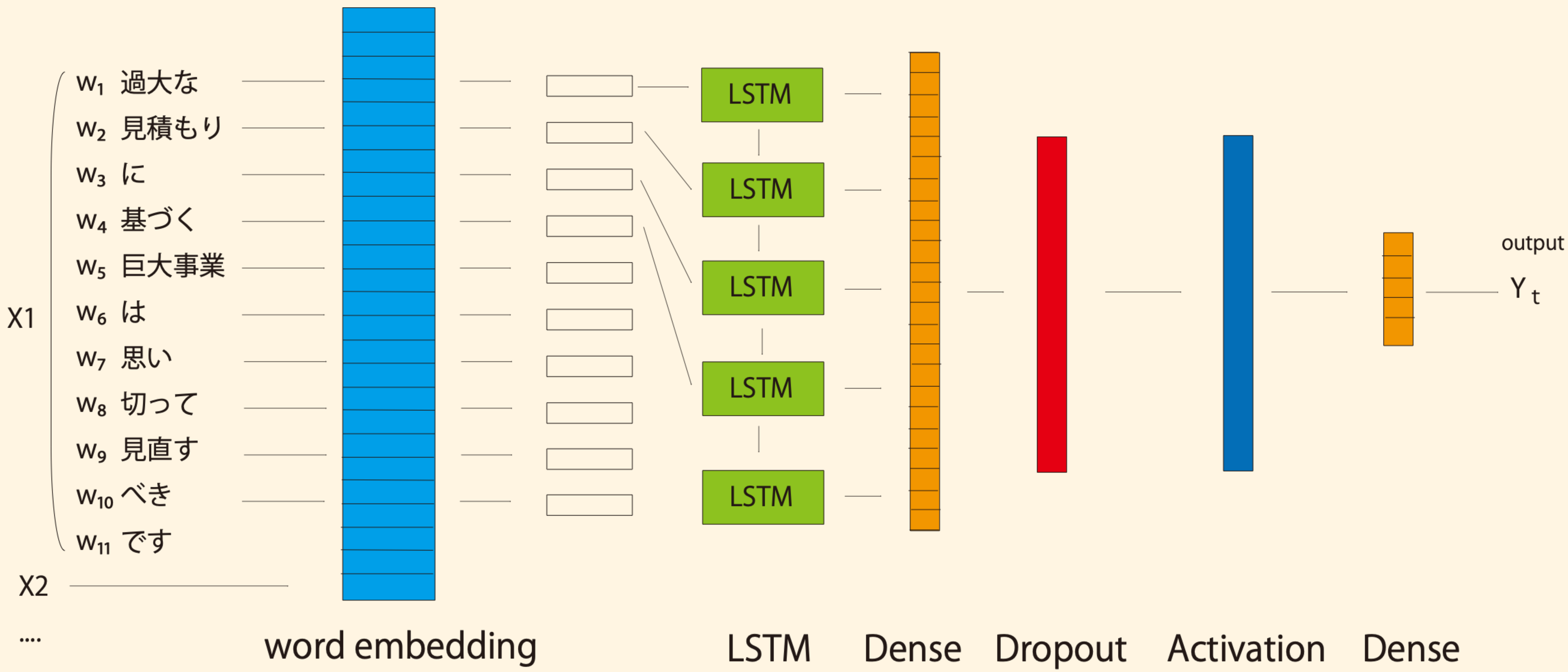


# Approach: Stance

We use simple LSTM model
for classifying “support”, “disapproval”, and “no matter” classes.

- Loss function: sparse categorical cross-entropy
- Activation function: ReLU

# Approach: Stance



# INDEX

# Results: Fact-checkability

The recall & precision scores were higher with the gold standard N3:

- all three assessors gave the common correct answers.

-> The results regardless of people will be identifiable with our approach.

| Gold Standard | | existence | | absence | |
|---|---|---|---|---|---|
| | Accuracy | Recall | Precision | Recall | Precision |
| N1 | 0.966 | 0.782 | 0.978 | 0.406 | 0.938 |
| N2 | 0.810 | 0.863 | 0.865 | 0.660 | 0.673 |
| N3 | 0.918 | 0.944 | 0.945 | 0.841 | 0.839 |
| SC | 0.730 | 0.843 | 0.763 | 0.523 | 0.646 |

N1: one or more; N2: two or more assessors;
N3: three or more;  SC: the weight of the correct score;

# Results: Fact-checkability

The result of Fact-checkability was stably superior.

We confirmed that the model using LSTM and CNN is effective.

Classification results for task participants

| | | existence | | absence | |
|---|---|---|---|---|---|
| team | A | R | P | R | P |
| KSU-08 | 0.735 | 0.407 | 0.722 | 0.914 | 0.738 |
| CUTKB-04 | 0.730 | 0.523 | 0.647 | 0.843 | 0.764 |
| RICT-07 | 0.729 | 0.419 | 0.694 | 0.899 | 0.738 |
| TTECH-10 | 0.719 | 0.176 | 0.500 | 0.931 | 0.743 |
| akbl-01 | 0.708 | 0.438 | 0.626 | 0.857 | 0.736 |
| tmcit-01 | 0.652 | 0.630 | 0.507 | 0.665 | 0.766 |

# Results: Relevance

## Problem

The topic of training data has only a few patterns. →**overfitting**

## Solution in future

Using skip-gram trained with Wikipedia corpus.

| | existence | | absence | |
|---|---|---|---|---|
| Accuracy | Recall | Precision | Recall | Precision |
| 0.865 | 1.000 | 0.865 | 0.000 | NaN |

# Results: Stance

The score is low due to the data shaping problem of the submission data.

| | agree | | disagree | | other | |
|---|---|---|---|---|---|---|
| Accuracy | Recall | Precision | Recall | Precision | Recall | Precision |
| 0.033 | 0.015 | 0.677 | 0.017 | 0.625 | 0.038 | 0.778 |

↓ fixed(not change model)

| | agree | | disagree | | other | |
|---|---|---|---|---|---|---|
| Accuracy | Recall | Precision | Recall | Precision | Recall | Precision |
| 0.807 | 0.269 | 0.627 | 0.212 | 0.610 | 0.963 | 0.824 |

The results improved, but still imbalanced.

# INDEX

# Summary and future work

- It was clarified that both convolution and sequence operations were necessary to estimate the fact-checkability.

- From the data set, we confirmed that the sentences including the fact checkable information shared similar facts with the target sentences provided in the task.

- We need to adjust the models for Relevance and Stance tasks in future.