RICT at the NTCIR-14 QALab-PoliInfo Task

RICOH
imagine. change.

Jiawei Yong, Shintaro Kawamura, Katsumi Kanasaki, Shoichi Naitoh, and Kiyohiko Shinomiya
Ricoh Company, Ltd.

Index

RICOH
imagine. change.

Segmentation subtask

- Overall thought for segmentation
- Cue-phrase-based idea
  - ◎ Semi-supervised segmentation
- Results and conclusion

Classification subtask

- Research challenges
- Research methods
- Results and conclusion

RICOH
imagine. change.

Segmentation subtask

Segmentation subtask in 2 steps

RICOH
imagine. change.

input: Date, Speaker, Summary

minutes

Date:
Speaker:
xxxxxxxxxxx
xxxxxxxxxxx
xxxxxxx
xxxxxxxxxxx
xxxxxxxxxxx
xxx
xxxxxxxxxxx
xxxxxxxxxxx
xxxxxxxx
xxxxxxxxxxx

segmentation

segments

初めに、xxx
xxxxxxxxxxxxxxx
xxx見解を求めます。

次に、xxx
xxxxxxxxxxxxxxx
xxx見解を求めます。

最後に、xxx
xxxxxxxxxxxxxxx
xxx質問を終わります。

search

contiguous segments that correspond to the input

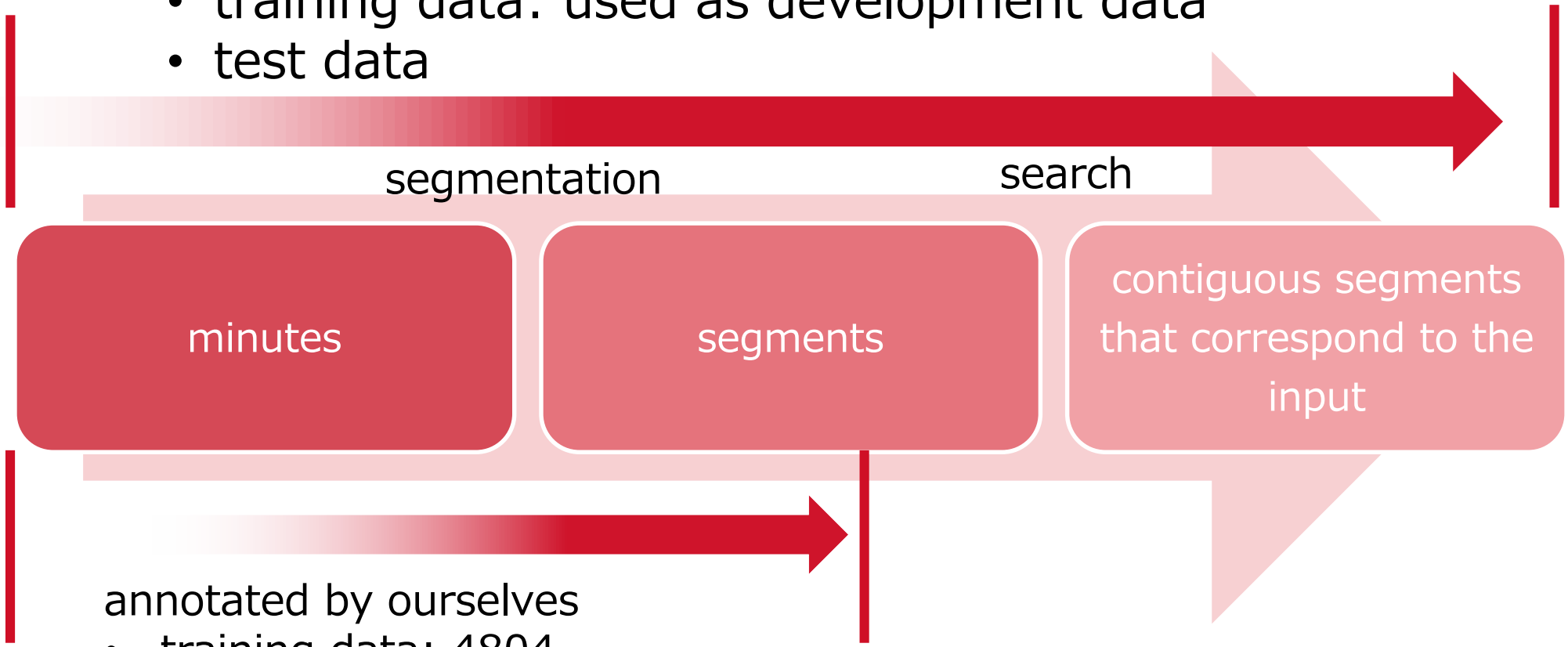
次に、xxx
xxxxxxxxxxxxxxx
xxx見解を求めます。

最後に、xxx
xxxxxxxxxxxxxxx
xxx質問を終わります。

Data sets for the segmentation subtask

RICOH
imagine. change.

data sets provided by the task organizer
- training data: used as development data
- test data



annotated by ourselves
- training data: 4804 utterances, 995 segments
- development data: 3438 utterances, 683 segments

Cue-phrase-based idea (segmentation step)

RICOH
imagine. change.

- **Hints for topical segmentation**

  ☐ Lexical cohesion
  TextTiling was tried in the dry run
  not reliable

  ☑ Cue phrases
  used in the formal run
  effective for speech in the assembly

Models for segmentation step (formal run)

RICOH
imagine. change.

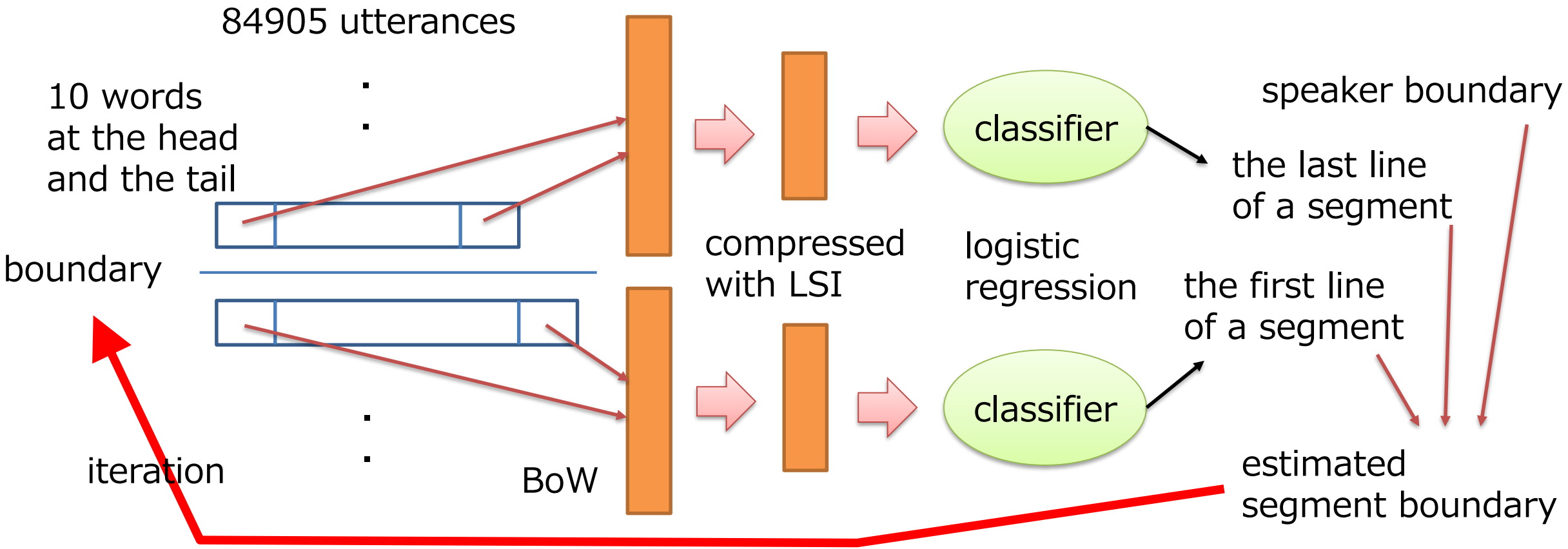
Submitted 5 Runs

- Rule-based Model (string pattern matching) … Run 1
- Supervised Model
  - BoW ⇒ SVM … Run 2
  - pre-trained word2vec ⇒ LSTM … Run 5
  - *word embeddings ⇒ HAN (unsubmitted)
- Semi-supervised Model (Original method)… Run 3
- No segmentation Model (each utterance is a segment) … Run 4

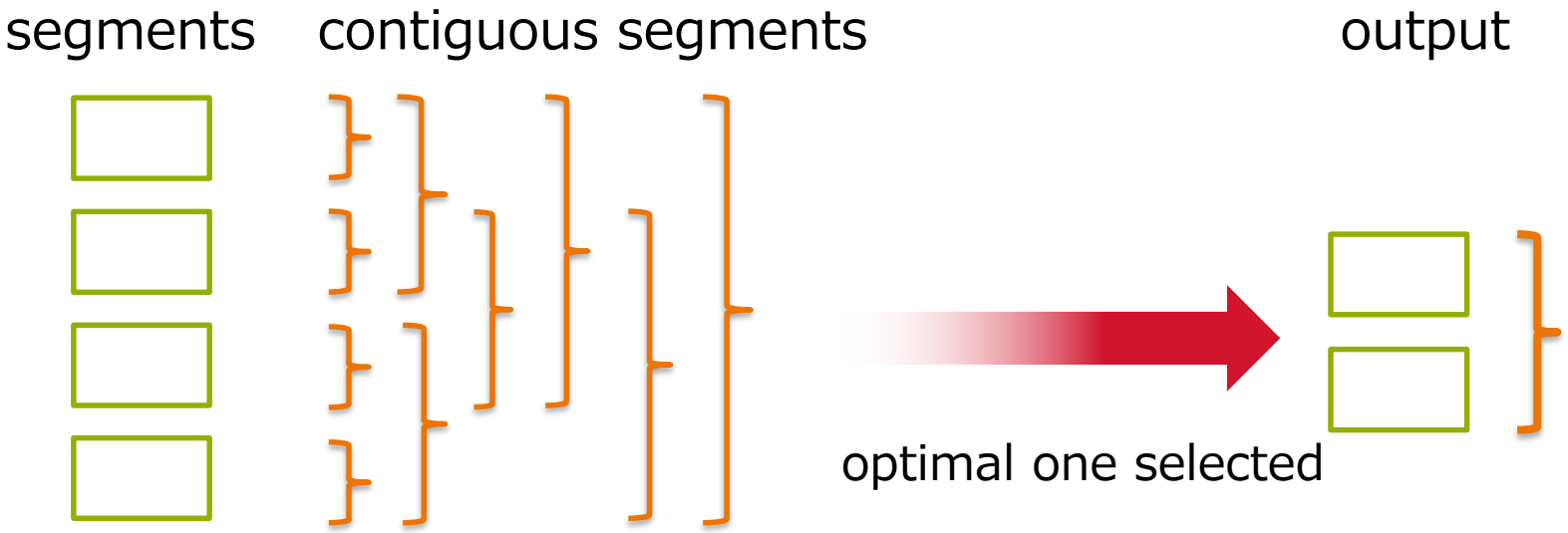
Semi-supervised model (Segmentation step)

RICOH
imagine. change.

- Segment boundaries are learned through bootstrapping.



Search step



- maximize $\sum_{i=1}^{k} idf(t_i) - \lambda k \log(n)$

Coverage of weighted words $t_i (i = 1, \ldots, k)$ in the summary

Penalty for the length (n utterances)

Hyperparameter λ is tuned by the development data. (0.4 for questions and 0.7 for answers)

Evaluation results

RICOH
imagine. change.

The performance of the methods when applied to the test data set (mean values of 5 runs)

Segmentation method	Question			Answer		
	Recall	Precision	F1	Recall	Precision	F1
rule-based	0.851	0.913	0.881	0.949	0.903	0.925
SVM	0.819	0.851	0.834	0.913	0.939	0.925
LSTM	0.916	0.690	0.780	0.909	0.925	0.914
HAN	0.871	0.874	0.873	0.949	0.921	0.934
semi-supervised	0.836	0.760	0.796	0.907	0.814	0.858
no segmentation	0.828	0.715	0.767	0.680	0.839	0.751

- The rule-based segmentation was the best during the formal run (Top 1 in F1). The method using a hierarchical attention network (unsubmitted one) also shows good performance.

Conclusions on segmentation subtask

RICOH
imagine. change.

- Assembly speeches can be effectively segmented by cue phrases.
- A rule-based segmentation and a neural network segmentation combined with a simple search model give good results. They can be baselines for more advanced methods that take syntactic or semantic features into account.
- A semi-supervised segmentation that does not require training data is also feasible.

RICOH
imagine. change.

Classification subtask

Research challenges in classification

RICOH
imagine. change.

Training Data

01 ·Quality

- The kappa statistics among annotators are quite low to the same sentence labelling.

Challenge1: Low Kappa Statistic

02 ·Quantity

- The quantity of labelled utterances for each topic are insufficient.

Challenge2: Underfitting

03 ·Imbalance

- The volume of different labels in different topics are in a great disparity.

Challenge3: Imbalanced Learning

We have submitted 7 runs by challenge combinations

Research methods in classification

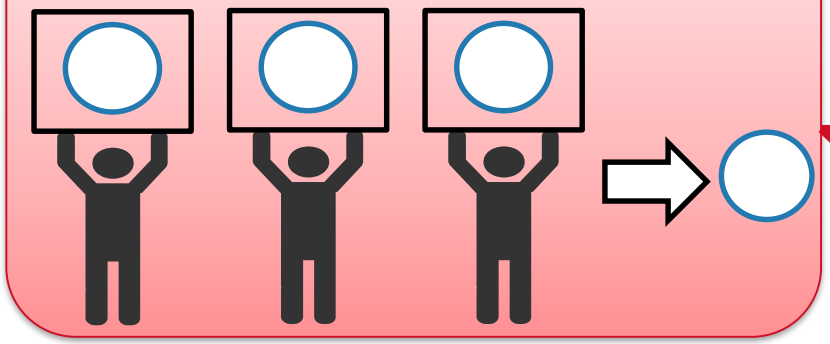
RICOH
imagine. change.

01 **Challenge1: Low Kappa Statistic**

Fact Checkability Subtask

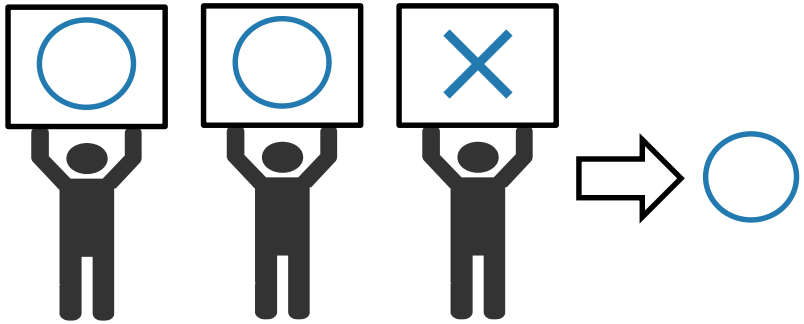
News Detection Support for Fact Check（NLP2018）

① **Unanimous training data** (4710)



LSTM
①F1 score:0.91

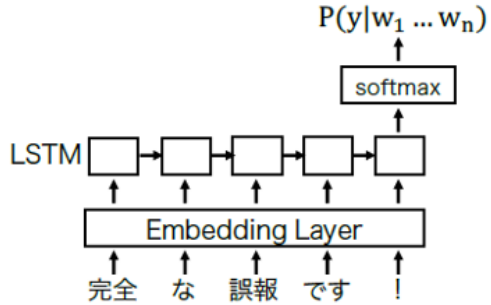
② **Majority training data** (10291)



LSTM
②F1 score:0.81

input

input



LSTM Classifier

SVM Classifier

Suspicious News Detection Using Micro Blog Text (2018)

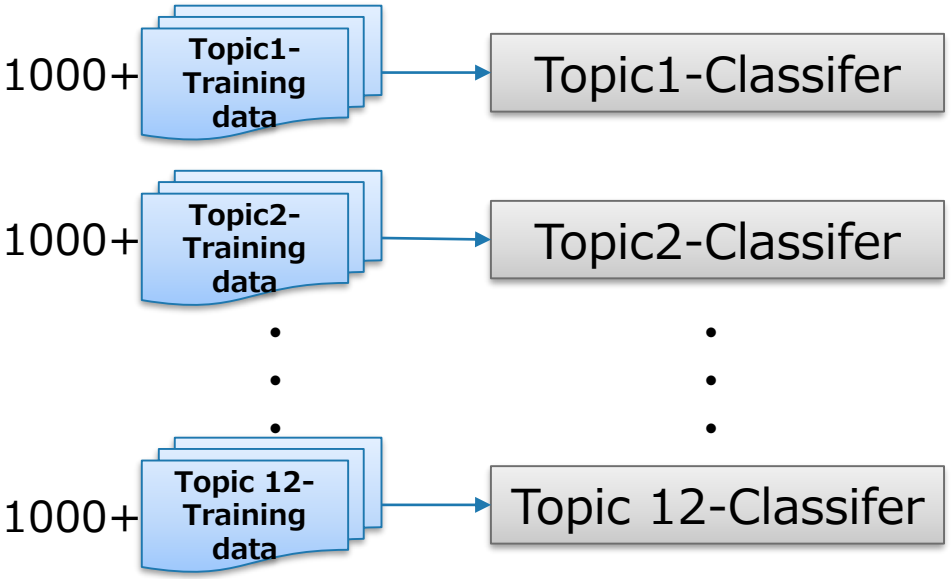
Research methods in classification

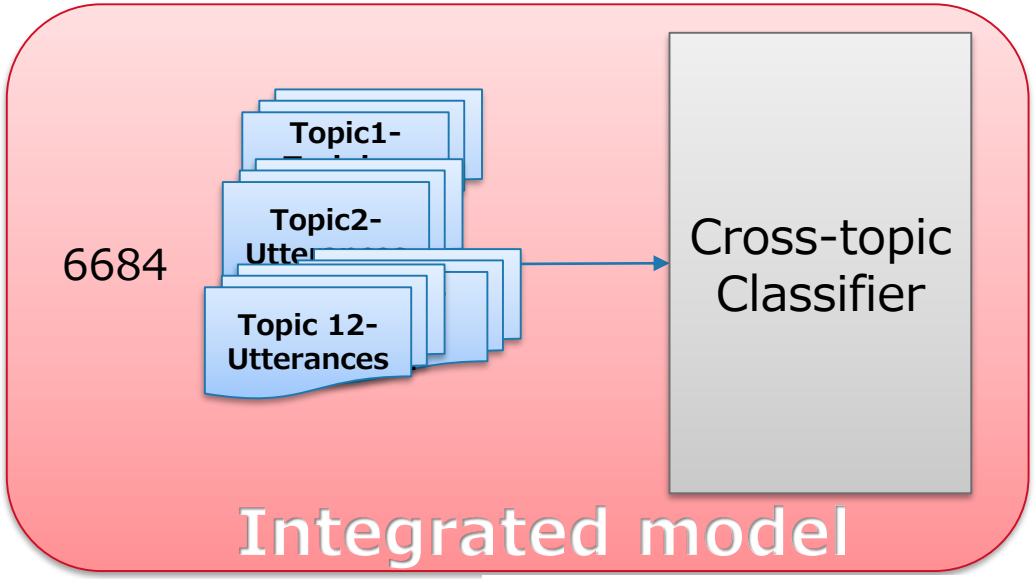
RICOH
imagine. change.

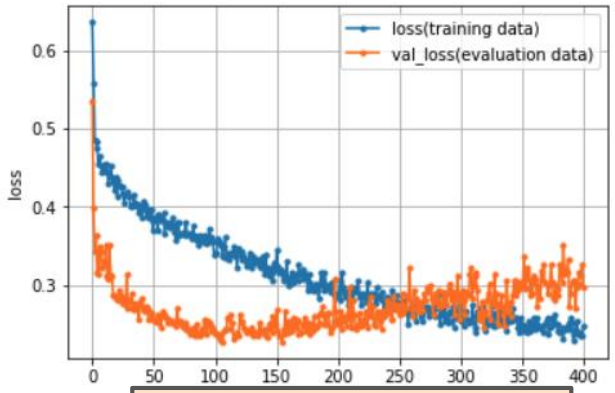
02

Challenge2: Underfitting

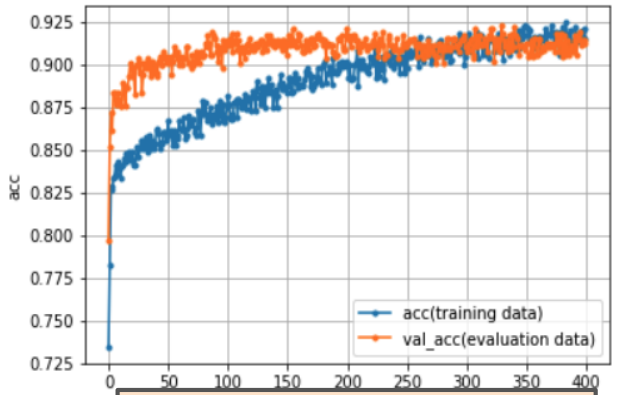
Stance Classification Subtask







The variation of Loss rate



The variation of accuracy rate

The underfitting problem has been alleviated.

Research methods in classification

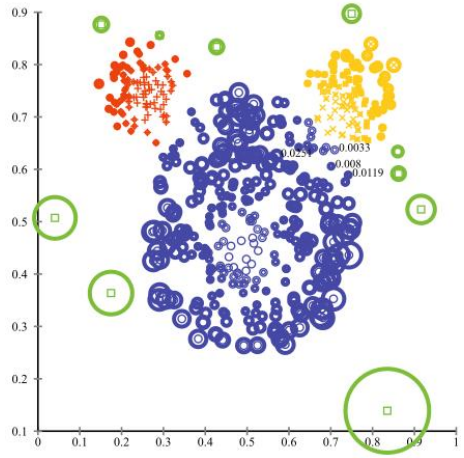
RICOH
imagine. change.

03

Challenge3: Imbalanced Learning

Relevance & Stance Classification Subtask

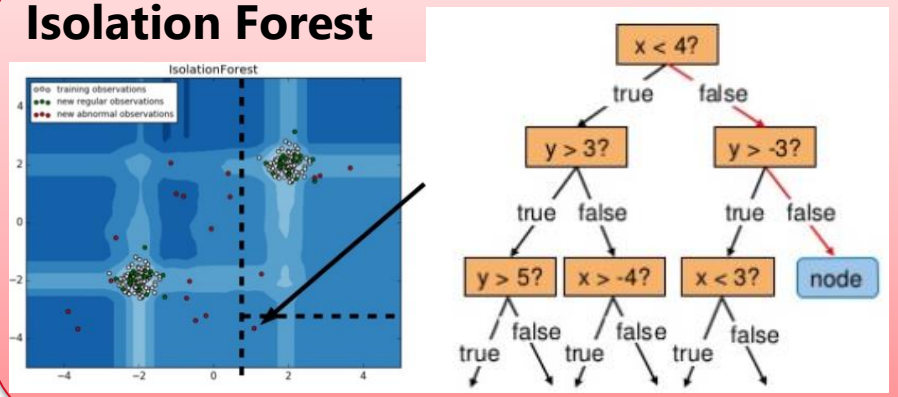
Relevance（"1"） : irrelevance（"0"） = 9390 : 901 ≒ 10 : 1



outlier detection

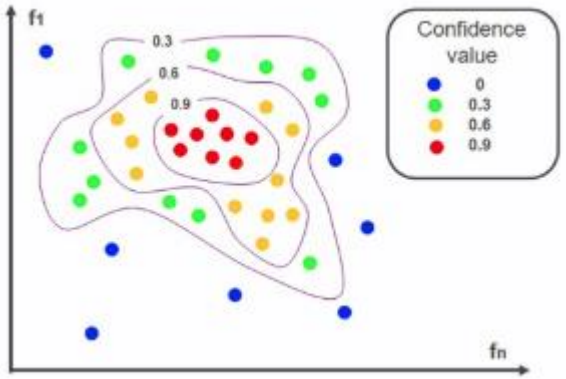
We regard Majority class as **normal data**, minority class as **outlier value**.

Isolation Forest



The F1 score of Minority Class

One class SVM



Evaluation results

RICOH
imagine. change.

The performance of the methods when applied to the test data set for classification

Classification Subtasks	Top Values of RICT Runs for each criteria						
	Accuracy	**1-Recall**	**1-Precision**	**1-F1**	**0-Recall**	**0-Precision**	**0-F1**
1. Relevance	0.857 (rank 7)	0.99	0.865	0.923 (rank 7)	0.524	0.332	0.406 (rank 2) Imbalanced Learn
2. Fact-checkability	0.729 (rank 3)	0.693	0.476	0.564 (rank 3) Low kappa	0.899	0.738	0.811 (rank 3) Low kappa
3. Stance	0.808 (rank 1)	0.295	0.63	0.40 (rank 3) underfitting	0.962	0.827	0.889 (rank 2) underfitting
		2-Recall	**2-Precision**	**2-F1**			
		0.194	0.579	0.290 (rank 4) underfitting			

Conclusions on classification subtask

RICOH
imagine. change.

- We have showed the assembly utterances can be classified by supervised learning methods with a high accuracy.

- The selection of training data acts an important role for supervised learning method. We shall select out the training data in consideration of quality quantity and balance.

①Low Kappa Statistic Challenge②Underfitting Challenge③Imbalanced Learn Challenge

Unanimous training data **Integrated model** **Isolation Forest**

RICOH
imagine. change.

Thank you for your attention.

RICOH

imagine. change.