

# **JRIRD at the NTCIR-16 FinNum-3 Task: Investigating the Effect of Numerical Representations in Manager's Claim Detection**

---

Shunsuke Onuma and Kazuma Kadowaki  
The Japan Research Institute, Limited

## ■ Abstract

---

- Participate in Manager's Claim Detection (English subtask) of FinNum-3
  - Claim detection: judges whether a target numeral is in a manager's claim or not
  - Numerical category classification: classifies a target numeral into one of 12 categories
- Investigate the performance of the claim detection task with **various numerical representations**
- Experiment on two task settings
  - Claim detection only
  - Joint learning
    - claim detection & numerical category classification

## Our Approach for Manager's Claim Detection

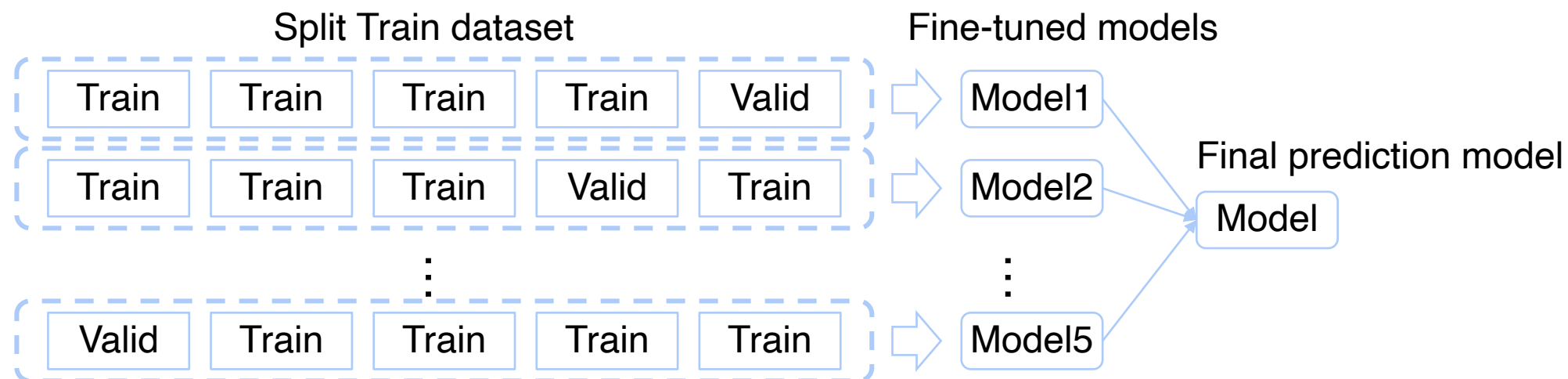
- Use five pre-trained language models and fine-tuned them
  - BERT (base), BERT (large), FinBERT, RoBERTa (large), T5 (large)
- Preprocess the input texts with the following numeral formats:

Format	Example: Fiscal Year 2018 Fourth Quarter
Mask	Fiscal Year [MASK] Fourth Quarter
Marker	Fiscal Year [NUM] 2018 [NUM] Fourth Quarter
Digit	Fiscal Year [NUM] 2 0 1 8 [NUM] Fourth Quarter
Scientific (sig1)	Fiscal Year [NUM] 2 [EXP] 3 [NUM] Fourth Quarter
Scientific (sig4)	Fiscal Year [NUM] 2 . 0 1 8 [EXP] 3 [NUM] Fourth Quarter

- Expect: *Digit* and *Scientific* help language models better recognize numerals
  - *Digit* splits numerals into each digit (avoids subwording numerals)
  - *Scientific* indicate significant digit(s) and magnitude of each numeral

## Training Method

- Split train dataset into 5 folds
  - train dataset : valid dataset = 4 : 1  $\rightarrow$  5 train/valid datasets
- Fine-tune a language model for each of 5 train/valid datasets
  - Grid search for best hyperparameters
- Average the predictions from 5 models for final prediction
  - Voting for T5 and soft average for other models



## Select Models

- Select models for submission
  - joint learning setting
  - Best score in each model of BERT (large), RoBERTa and FinBERT
    - Macro-F1 score (dev) of the claim detection task
    - Experiment using T5 is not conducted before submitting
- Submit models
  1. BERT (large) with *Marker*
  2. RoBERTa with *Scientific (sig4)*
  3. FinBERT with *Marker*

## Results : Effect of Numerical Formats

Macro-F1 (test) for the claim detection task on joint learning:

	BERT (base)	BERT (large)	FinBERT	RoBERTa	T5
Mask	0.895	0.899	0.893	<b>0.904</b>	0.896
Marker	0.903	<b>0.908</b> <sup>*1</sup>	0.910 <sup>*3</sup>	0.904	0.893
Digit	<b>0.911</b>	0.902	0.901	0.897	0.900
Scientific (sig1)	0.900	0.897	0.899	0.901	<b>0.903</b>
Scientific (sig4)	0.904	0.903	<b>0.911</b>	0.895 <sup>*2</sup>	0.901

### Results

Score: best score in each pretrained model \* : submitted models

- Numerals are informative
  - **Formats other than *Mask* were best for each models (except RoBERTa)**
- Best formats depend on models
  - We need further experiment to investigate the effect of formats

## Results : Effect of Joint Learning

Improvement of macro-F1 for the claim detection task by joint learning:

	BERT (base)	BERT (large)	FinBERT	RoBERTa	T5
Mask	0.011	0.014	0.006	0.001	-0.002
Marker	0.011	0.013	0.017	0.003	-0.005
Digit	0.009	0.003	0.008	-0.005	-0.002
Scientific (sig1)	0.014	-0.004	0.008	-0.008	0.005
Scientific (sig4)	0.009	0.002	0.017	-0.013	0.004

Red: negative effect

### Results

- Improve constantly in small models: BERT (base) and FinBERT
- **Not consistent** in large models: BERT (large), RoBERTa and T5
  - Our setting of joint learning might not be optimal

## ■ Conclusion

---

- Investigate the performance of the claim detection task with various numerical formats in the FinNum-3
- Results
  - Numerals are informative in the claim detection task
  - Best numerical formats depends on the models and settings
  - Joint learning is effective in some cases
- Future works
  - Statistical analysis for the effect of formats
  - Investigating optimal setting of joint learning