

### Abstract

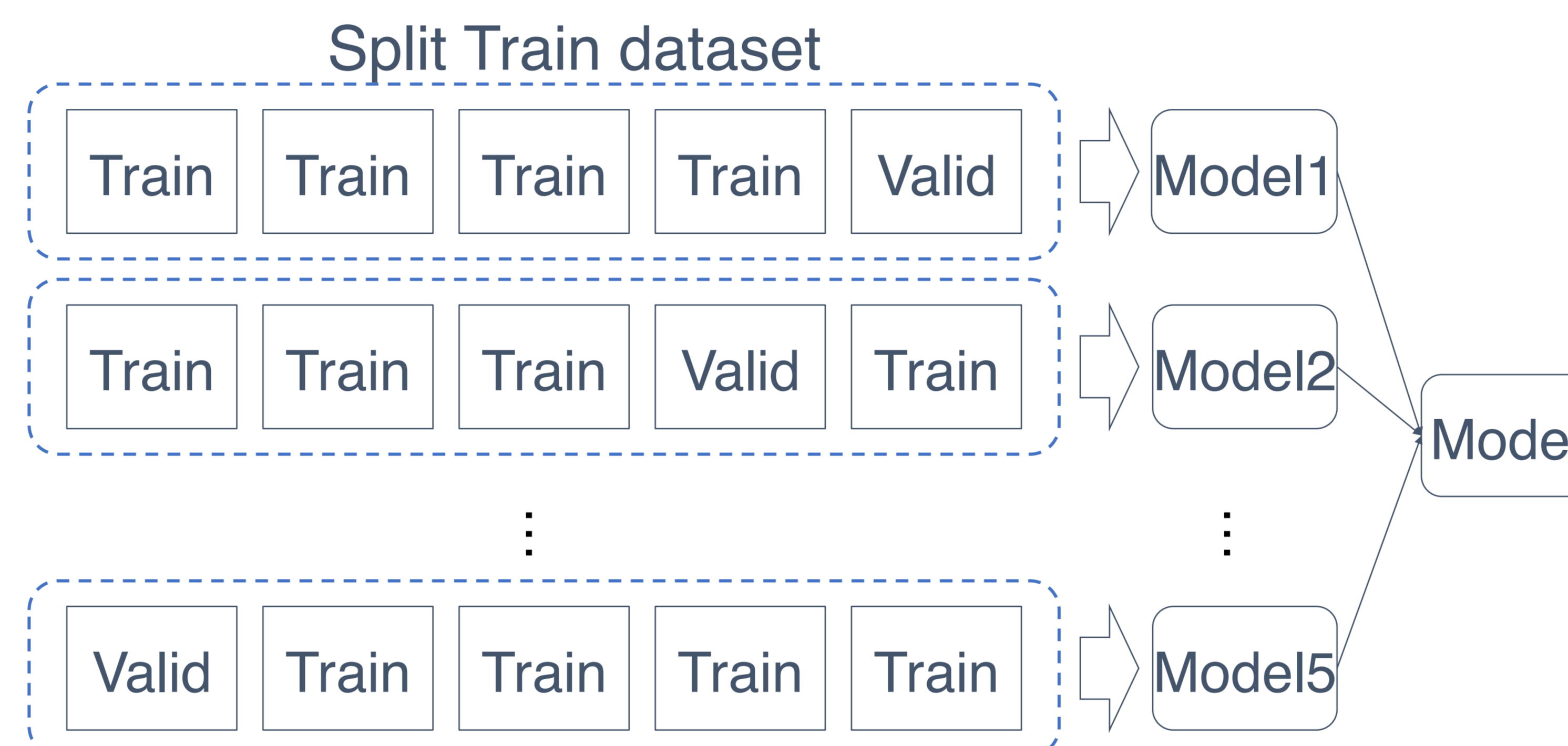
- Investigated the performance of claim detection task **with various numerical formats**
- Two settings for each formats : claim detection only, joint learning (claim detection & category classification)
- Results
  - Best numerical format depended on models and settings
  - Joint learning was effective in some cases

### Numerical Formats

Formats	Example : Year 2018 Fourth Quarter
Mask	Year <b>[MASK]</b> Fourth Quarter : ignores numeral (for comparison)
Marker	Year <b>[NUM] 2018 [NUM]</b> Fourth Quarter : distinguishes target numeral by [NUM]
Digit	Year <b>[NUM] 2 0 1 8 [NUM]</b> Fourth Quarter : avoids subwording numerals
Scientific (sig1)	Year <b>[NUM] 2 [EXP] 3 [NUM]</b> Fourth Quarter : focuses on significant digit & magnitude of numeral
Scientific (sig4)	Year <b>[NUM] 2 . 0 1 8 [EXP] 3 [NUM]</b> Fourth Quarter : more significant digits

### Training Method

- Split train dataset into 5 folds
- Fine-tune a language model for each of 5 train/valid datasets
- Average the predictions from 5 models for final prediction



### Submit Models

- Select submit models in Joint Learning setting
- Best score in each model of BERT (large), RoBERTa and FinBERT

**JRIRD1: BERT (large) with *Marker***  
**JRIRD2: RoBERTa with *Scientific (sig4)***  
**JRIRD3: FinBERT with *Marker***

### Result : Effects of Numerical Formats

Macro-F1 in claim detection task on joint learning setting:

	BERT (base)	BERT (large)	FinBERT	RoBERTa	T5
Mask (1)	0.895	0.899	0.893	<b>0.904</b>	0.896
Marker (2)	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

\* submitted models

- (1) Formats other than Mask were best for each models  
⇒ Numerals are informative (except RoBERTa in joint learning)
- (2) Best formats depend on models

### Result : Effects of Joint Learning

Improvement of macro-F1 in joint learning setting (claim detection):

	BERT (base)	BERT (large)	FinBERT	RoBERTa	T5
Max	0.014	0.014	0.017	0.003	0.005
Min	0.009	<b>-0.004</b>	0.006	<b>-0.013</b>	<b>-0.005</b>

- Not consistent in large models: BERT (large), RoBERTa, T5

### Future Works

- Statistical analysis for the effect of formats
- Investigating optimal setting for joint learning