

Research Projects

2020 - 2022

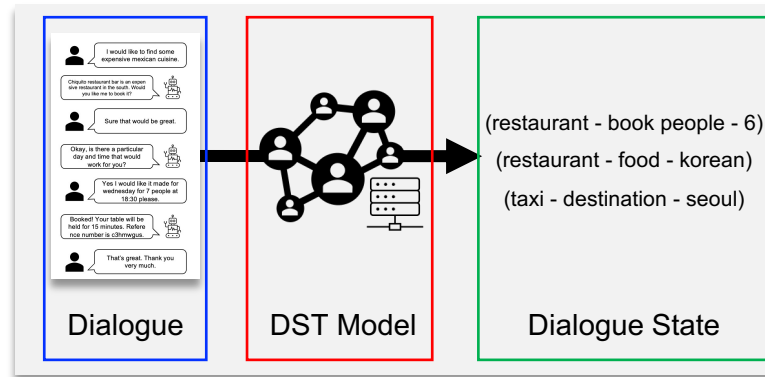
Takyoung Kim

M.S. Student in Industrial & Management Engineering

Korea University

Dialogue State Tracking

(2021.4 – 2021.11, 2022.3 – 2022.8)



Realistic Dataset?

[USR] Saturday please.

[SYS] Reserved. More?

[USR] Sorry, it's Tuesday.

We should consider various **turnback** scenario!

How to Perform Well?

Annotation is too expensive
There are similar situations

Can we exploit information
in the existing dataset?

In Progress

Realistic Evaluation Metric?

<u>Gold State</u>	<u>Pred State</u>
State A	State A
State B	State D
State C	State E

Joint Goal Acc.: 0% (underestimate)

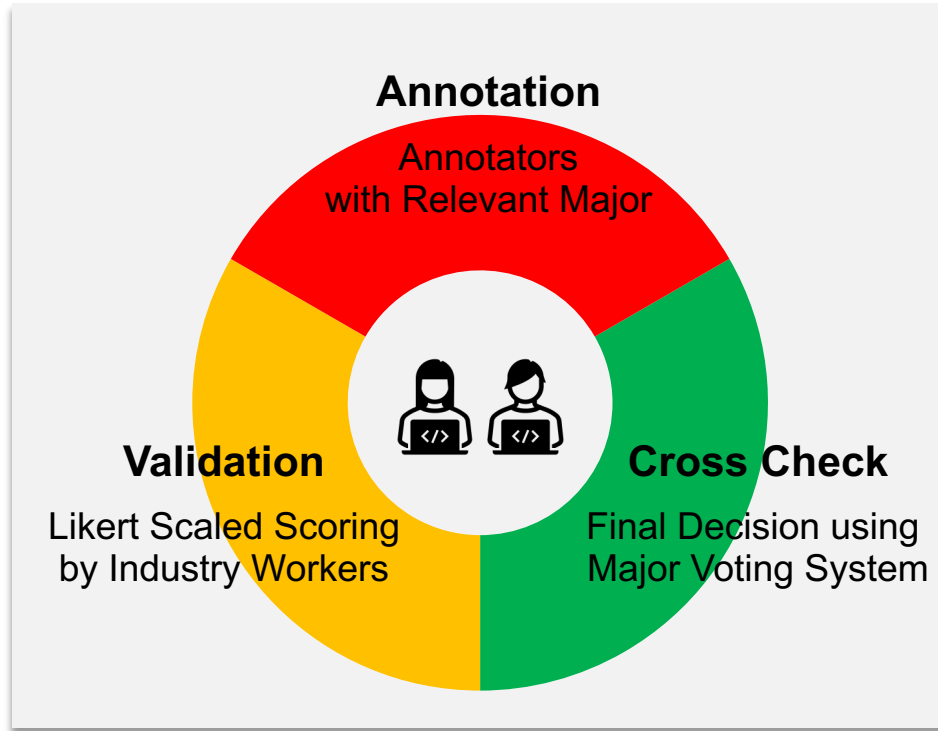
Slot Acc.: 86.67% (overestimate)

Relative Slot Acc.: 20% (proper)

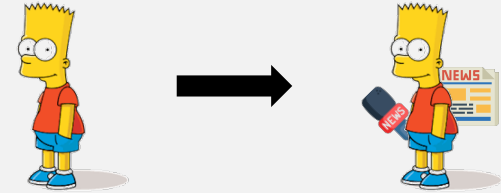
- Independent of predefined slots
- Aligned with human intuition

Korean Document Summarization

(2020.7 – 2020.12, 2021.3 – 2021.6)



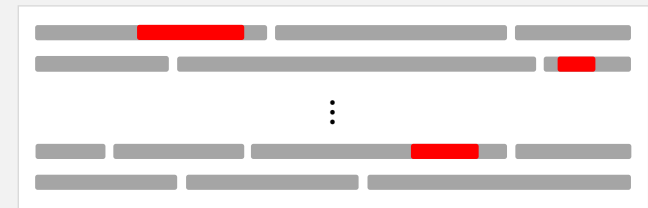
Adaptive KoBART Pretraining



- Task-adaptive pretraining on News domain
- Text-infilling on Korean BART model



News-oriented Summarization



Abstractive
Summary



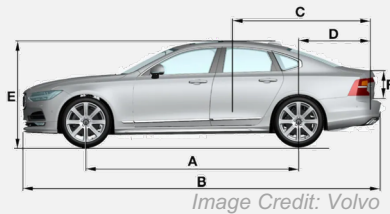
Construct **800,000** Summary Dataset

- Domain: News, Opinion, Magazine, Legal
- Type: Extractive / Abstractive

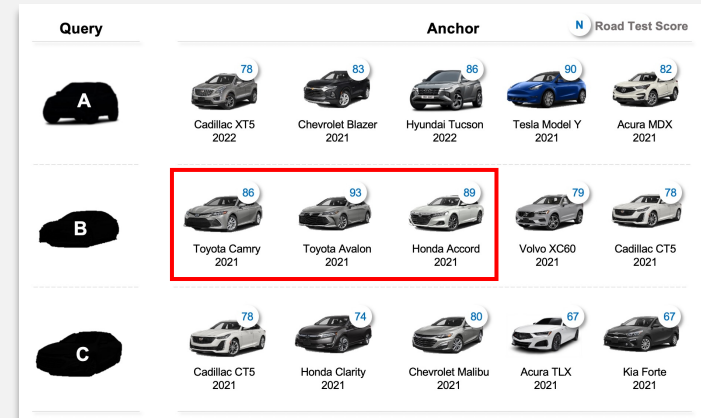
Customer Satisfaction Prediction

(2021.10 – 2022.6)

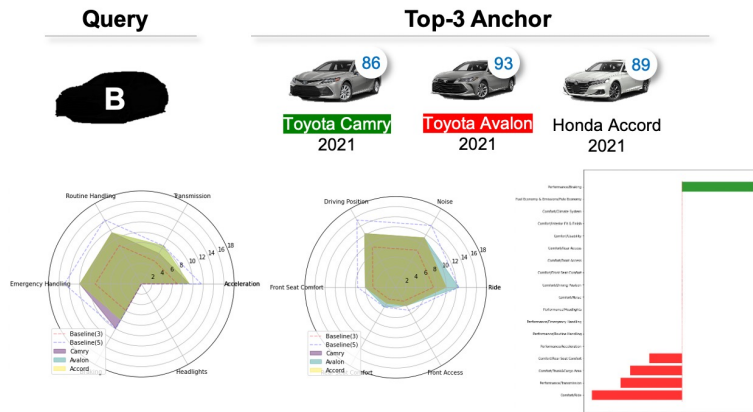
Vectorize Car Dimension



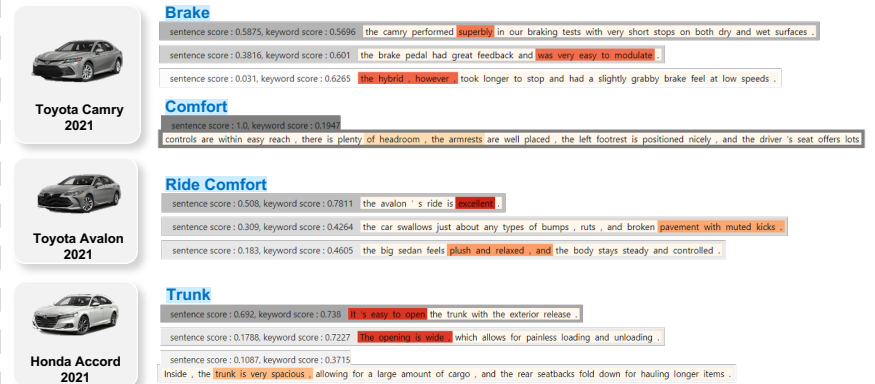
Retrieve Top-k Nearest Car Types



Strength Analysis with Benchmark



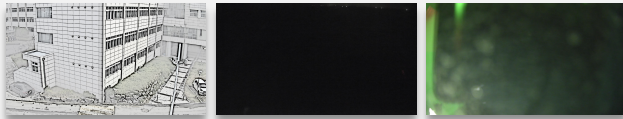
Visualize User Opinion



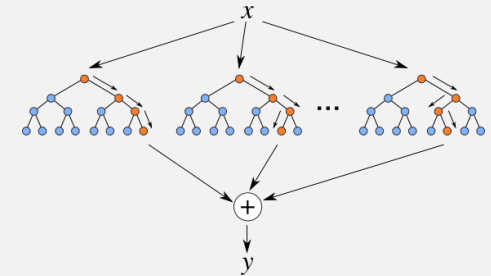
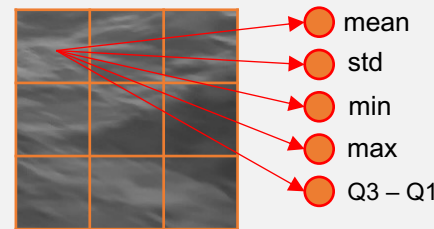
Noisy Wave Height Estimation

(2020.12 – 2021.3)

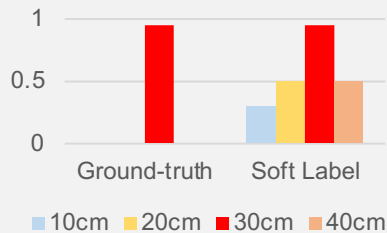
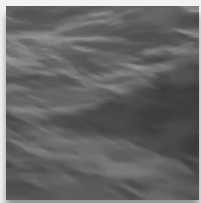
Remove Noisy Data



Regression with Handcrafted Features

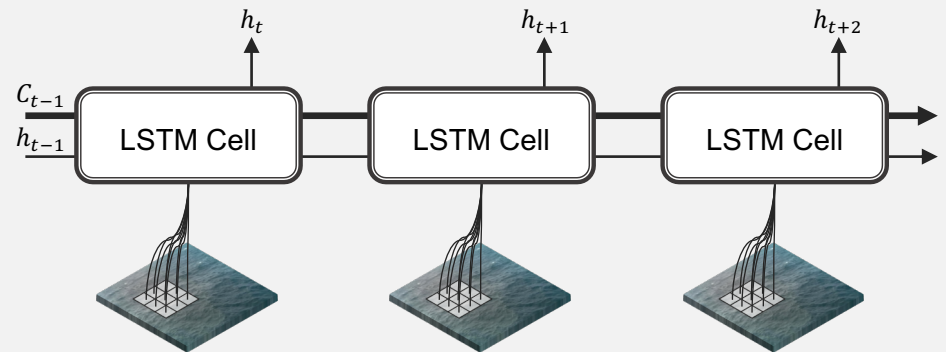


Apply Soft Label



Reflecting similarity among classes
(in classification setting)

Classification Modeling



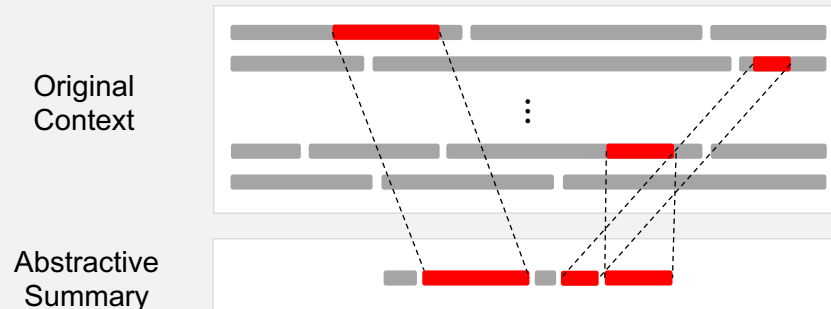
Real-world Constraints: Camera Distribution, Sunlight, Label Credibility, ...

Selective Masking for Task-adaptive Pre-training

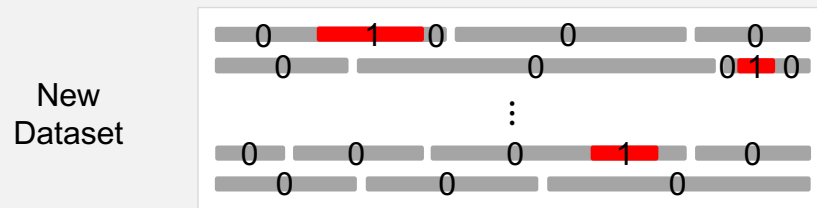
(2021.3 – 2021.7)

Build Content Selector

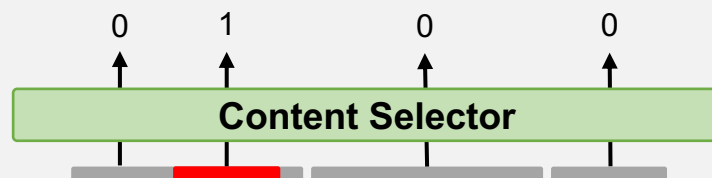
1. Compare context-summary pair



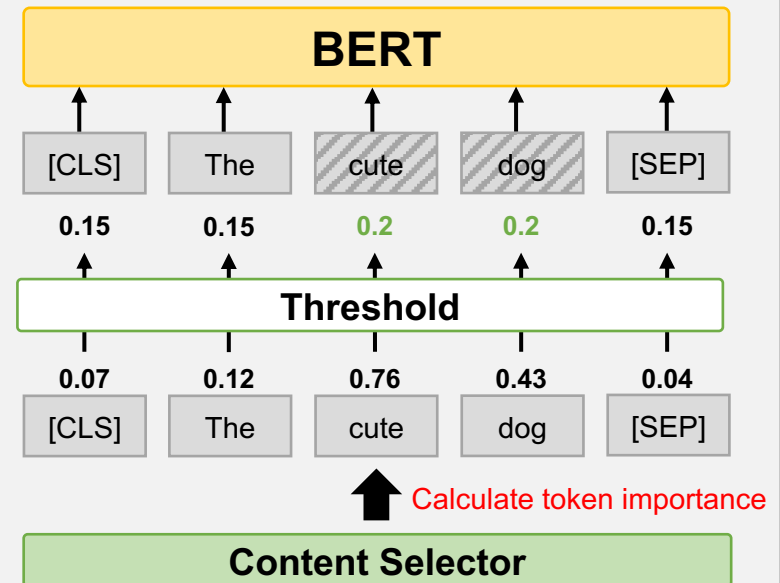
2. Construct dataset for selective masking



3. Train Content Selector



Additional Pre-train w/ Selective Prob.



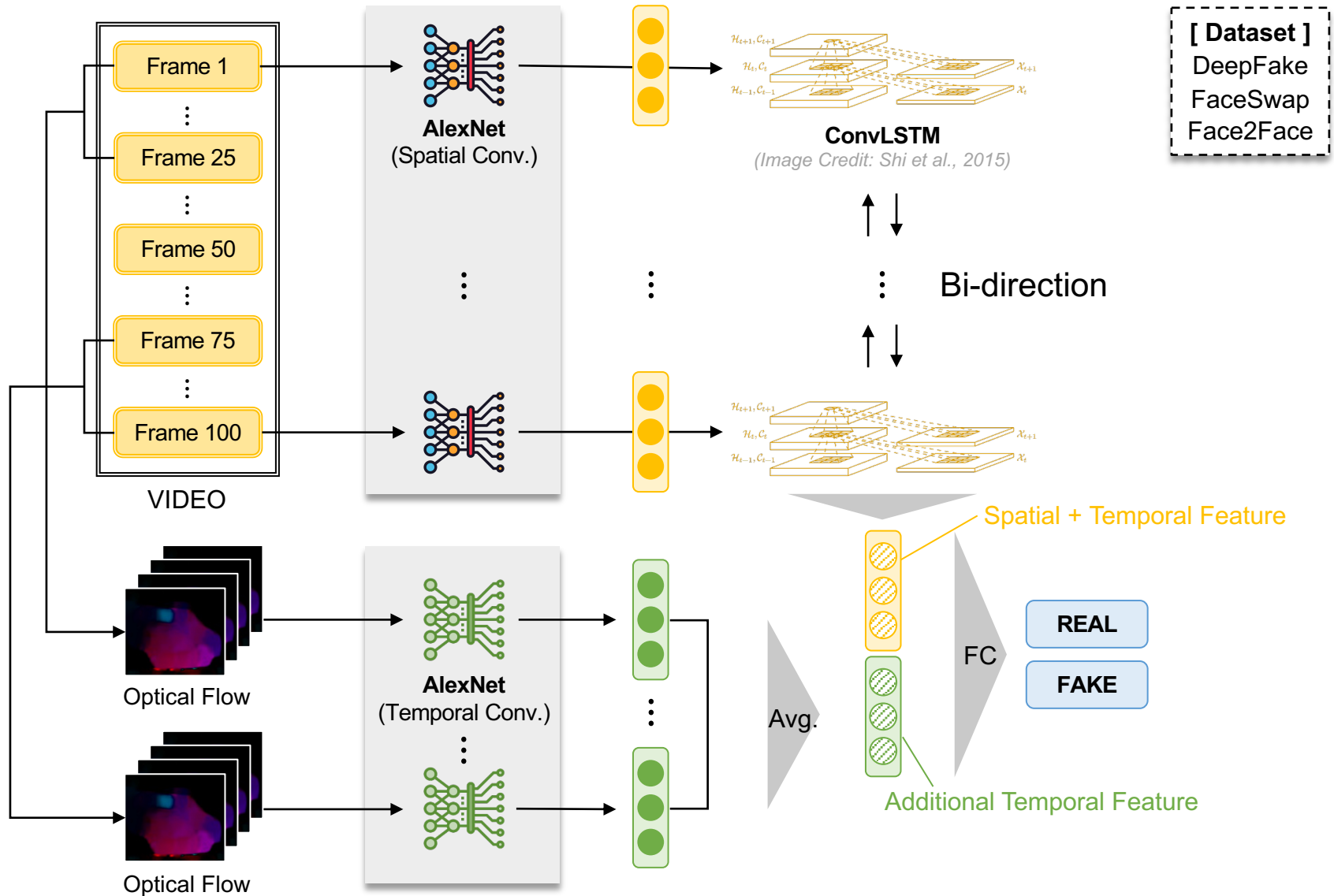
Downstream Task Performance

- Task: Abstractive Summarization
- Performance:
BERT-large < BERT-base

(w/ Selective Masking)

Video DeepFake Detection

(2020.2 – 2020.8)



EOD