



Applications of Natural Language Processing in KLM & Elsevier



Assisting Users to Provide Feedback by Asking Smart Follow-up Questions

Hosein Azarbonyad

16 February 2022



Royal Dutch Airlines





We realize value by experimenting with technology and transforming our way of working.



EXPERIMENTING

with new technologies and methodologies

- Blockchain
- Artificial Intelligence
- Robotic process automation
- Virtual reality
- Augmented reality
- Internet of Things



TripReports



Today, Crew and Ground are not able to use feedback sufficiently from **TripReports** to improve **operational excellence** & increase **employee engagement** due to having different systems to report and the difficulty of using them.



Current situation

Problems:

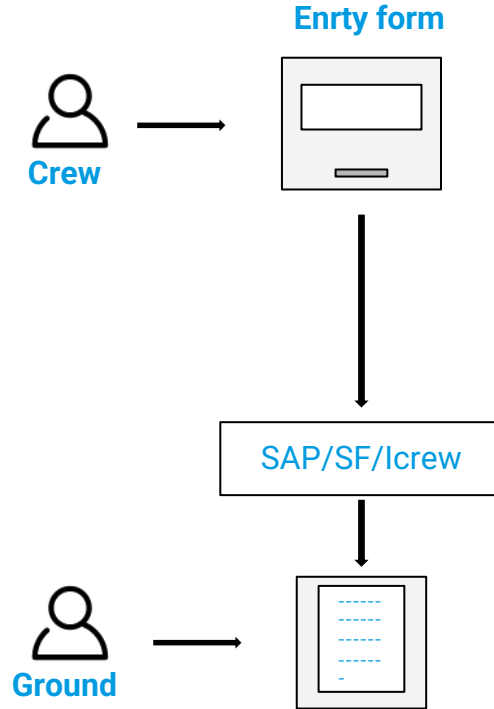
- UX problems
- Categorization
- No feedback

Garbage in

Problems:

- Missing info
- Wrong categories
- UX problems;
(processing, dashboard,
reporting)

Garbage out





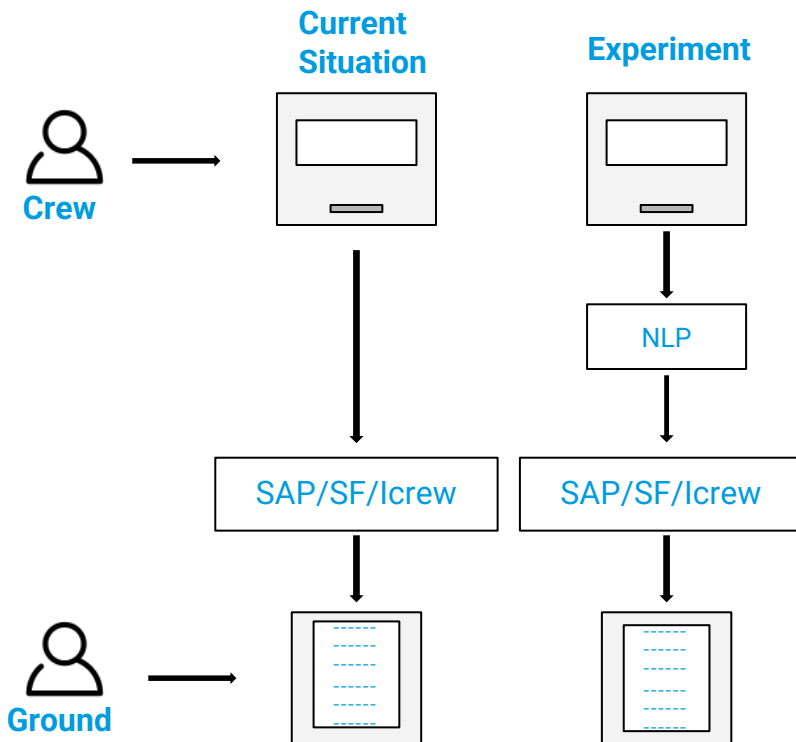
Our solution

Problems:

- UX problems
- Categorise
- No feedback
- **Garbage in**

Problems:

- Missing info
- Wrong categories
- UX problems;
(processing,
dashboard, reporting)
- **Garbage out**



Experiment Improvements:

- ✓ Improve entry form
- ✓ Better and automatic categorization;
 - simplify tree structure (1 tree)
 - NLP / text classification

- ✓ Feedback (only checkbox)

Experiment Improvements:

- ✓ No more missing information
- ✓ No more wrong categories

TripReport classification: challenges

- Large number of labels (more than 7000)
 - Different domains (hotel, catering, passenger, etc) have their own label set
- Labels have hierarchical structure
 - Depths differs per domain
 - Hierarchy implies subset relationship
 - Data skewed towards more generic labels



Tripreport classification: models

- Basic model
 - Flatten labels
 - Text representation: BOW, tf-idf on character n-grams
 - Model: SVM
- Hierarchical model
 - Train local classifiers for each branching point of the label tree
 - Local classifier models are the same as the basic model
- BERT-based model
 - Fine-tune a pre-trained BERT model for classification

Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).



TripReport classification: results

- Experiments are done on hotel related cases
 - Started with 356 labels and cut-it down to 34
 - ~15k reports for training models and hyperparameter search
 - 2k reports annotated for the final evaluation

Model	F1 micro	F1 macro
SVM	0.75	0.62
Hierarchical model	0.74	0.63
Fine-tuned BERT	0.76	0.64
SVM with manual rules	0.83	0.72



Re-designed solution

Enter Trip Report

☒ Complaint ☐ Compliment

 Flight nr. before issue

 Date of occurrence
22-02-2019

 Location

Room number (if applicable)

Describe your issue(s) and where it happened. Please don't include personal information.

 ADD PHOTO(S)

We have detected the categories below, by checking them you help us to get your report to the right person.

Categories
Noise, hygiene and breakfast

☐ I've contacted the hotel reception

☐ Send me feedback on this report

CONTINUE

Auto-saved



Test with users

- Crew satisfaction with POC: 8.3 / 10
- Faster: 4.2 / 5, easier 4.5 / 5
- Classifier results:
 - Accuracy: 0.87
 - Accuracy with user interaction: 0.83
- Identified improvements: auto-fill information



Auto-filling information

KLM Trip Report Log out

Enter Trip Report

☒ Complaint ☐ Compliment

Flight nr. before issue

Date of occurrence
22-02-2019

Location

Room number (if applicable)

Describe your issue(s) and where it happened. Please don't include personal information.

ADD PHOTO(S)

We have detected the categories below, by checking them you help us to get your report to the right person.

Categories
Noise, hygiene and breakfast

☐ I've contacted the hotel reception

☐ Send me feedback on this report

CONTINUE
Auto-saved

Information
extraction

Entailment



How complicated can a report be?

EN

HOSEIN

FILE /

Make up your

[BACK](#)

Step 1/3: Rep

Danger to people

☐ Yes, it caused an in

☐ Yes, there's a chan

Danger to the op

☐ Ground operations

☐ Cargo operations

☐ Aircraft technique

Risk to the envi

☐ Yes

Step 2/3: Describe the situation

Subject *

Send report to *

Add pictures

Location

Station code *
AMS: EHAM Amsterdam/Schiphol

Location type *

Incident or possible risk

Description *

Actions taken

Date and time
19-02-2020 16:22

Authorities present or informed?

Airport authorities

☐ Yes

☐ No

☒ Unknown

FINISH LATER

TO THE LAS

EN

HOSEIN

Saved automatically...

[PREVIOUS](#)

Step 3/3: Fill in the last fields

Fill in the fields below as completely as possible. Your input will be saved automatically, so you can also finish your report at a later stage.

Flight / Equipment

+

Leakage

+

Additional comments or suggestions?

Do you have any ideas to avoid this risky situation / incident?

☐ I agree with the [privacy policy](#) *

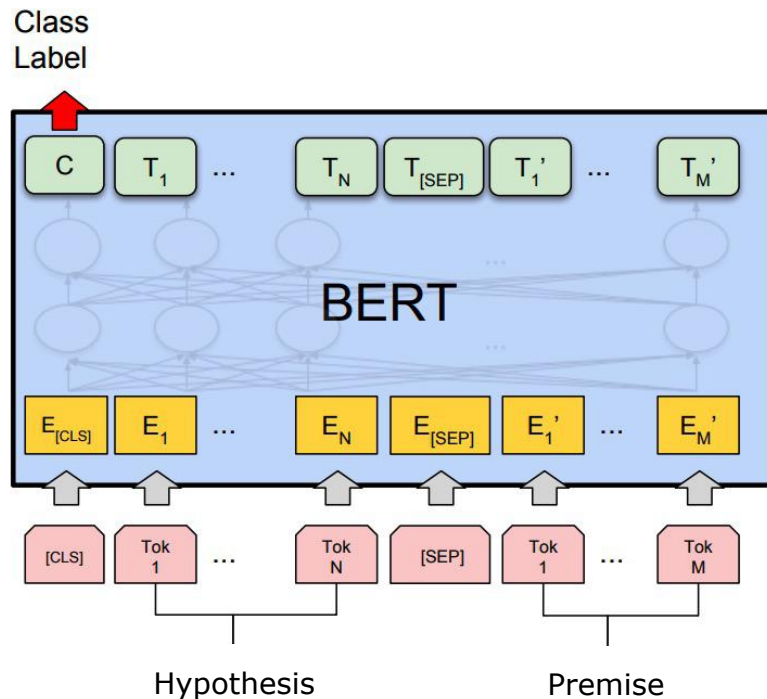
FINISH LATER

SUBMIT NOW



Textual entailment

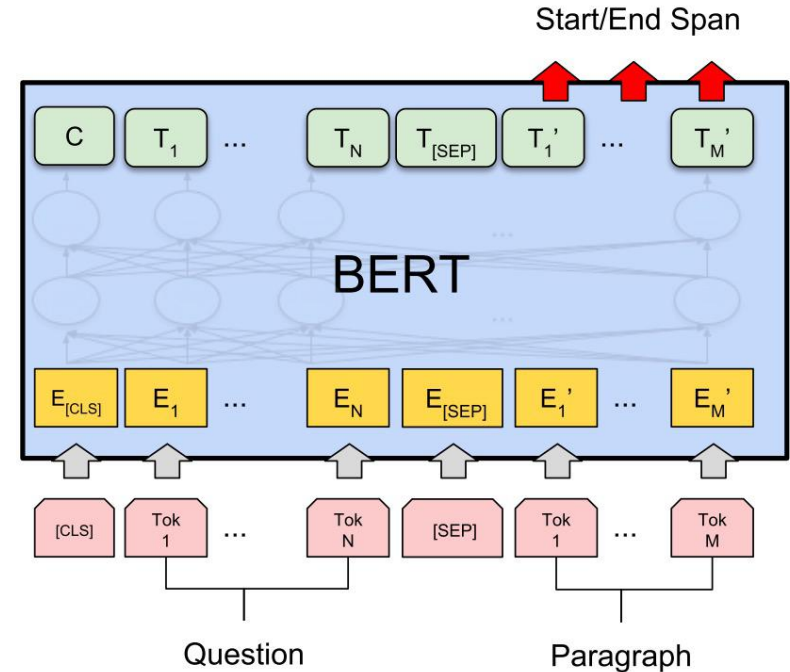
- A fine-tuned BERT model for NLI
 - Trained on XNLI dataset (~112.5k samples)
 - Fine-tuned with 2k annotated samples and evaluated on 1k samples
- Premise: input report
- Hypothesis: “reception is contacted”
- Predict if hypothesis is correct or not (entailment, neutral, contradiction)
 - If ‘neutral’ then ask the user to provide an answer





Fill-in boxes

- A fine-tuned BERT model for question answering
 - Trained on SQuAD v1.1 dataset (100k question-answer pairs)
 - Fine-tuned with 2k annotated pairs and evaluated on 1k pairs
- Paragraph: input report
- Question: “what is the room number”





Results

- Both models can answer questions with a high accuracy
 - There is still a lot of domain specific information
- Checkbox questions are easy to answer

Task	F1 micro	F1 macro
Checkboxes (BERT)	0.93	0.87
Checkboxes (SVM)	0.82	0.75
	F1	Exact Match
Fill-in boxes (BERT)	0.83	0.81
Fill-in boxes (rule-based)	0.61	0.53



Learning

- To really help users with AI, it is important to involve them in the process as early as possible
 - Domain knowledge plays an important role in designing real-world AI models
- Intuitive UX and application design can help users trust the underlying AI model