

Developing Chatbot for Training Seafarers for better Understanding and Communication by Using Real VTS Data

최성철* · 서원철** · 최승희*** · † 장은규

*가천대학교 산업경영공학과, **부경대학교 시스템경영공학과 ***,† 한국해양수산연수원

Abstract : The advent of the era of data age and advances in artificial intelligence technology has led to innovations in various business areas. In particular, many attempts have been made to improve the stability of the marine accident, which has not previously been applied by a data-drive approach. Most of the marine accidents happen at a time when the vessel is approaching a port and preparing for berthing. Although the cause of the accident has many factors, it is often caused by the difficulties of communication between the ship navigator and the control center. In particular, communication in English makes difficulties for navigators, not English as their first language. To do this, proper English conversation education for sailors is very important. In order to support the issue, this study presents data and framework for the development of a chatbot for ship safety education.

Key words : Maritime Education, Chatbot Framework, Deep learning, LSTM.



Vessel Traffic Service System

- An information system for support information exchange between vessels in the port
- Identify the maritime traffic situation for entering and departing ships and ships
- Ship traffic control to prevent marine accidents from deviations from routes, access to dangerous areas, and risk of collision
- Provides port operation information such as maritime traffic information, port facilities, marina, etc.

Talk

- Communication training between VTS and ship for safety
- Collecting conversation data
- Neural network – RNN and End to End model
- Chatbot framework



† 교신저자 : serius46@daum.net

limitations of communication in VTS

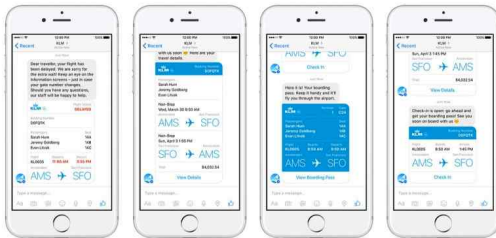
- Not enough trained seamen in English
- Terminology misuse
- To use a lot of words for communication
- Communication misunderstanding

Well training system

Chatbot

- A computer program which conducts a conversation via auditory or textual methods
- From turing test?
- Being widely used with A.I. era
- Generative vs. QA System
- Chatbot is another option for training system

Chatbot



Purpose

- Building a chatbot system to train ship crew members
- Collecting actual communication data
- Question-Answering system
- Neural net based system
- Considering communication-context

Dataset

- Collecting all communication script from ULSAN port
- Between VTS and Shipment
- Total - 21,436 radio communications
- 125 dialogue case, 13 action category (Navigation information, Anchoring, Docking, Departure)
- Conducted by listening and dictation
- Feeding AI machine

Dataset

Dialogue Case	Script	Action Category
LA-PS-01-011	Shore to Ulsan vts	항행 정보
LA-PS-01-012	This is Ulsan	항행 정보
LA-PS-01-013	Ulsan	항행 정보
LA-PS-01-014	Please come in	항행 정보
LA-PS-01-015	over	항행 정보
LA-PS-01-021	Ulsan vts	항행 정보
LA-PS-01-022	get ahead	항행 정보
LA-PS-01-031	Good evening ma'am	항행 정보
LA-PS-01-041	Just want to confirm what time pilot will be boarding on our vessel	항행 정보
LA-PS-01-042	Please check channel 13	항행 정보
LA-PS-01-043	13	항행 정보
LA-PS-01-051	will be contact with pilot	항행 정보
LA-PS-01-052	Ulsan pilot Ulsan pilot	항행 정보
LA-PS-01-053	over	항행 정보
LA-PS-01-061	Ulsan vts	항행 정보
LA-PS-01-062	My good morning Ulsan vts	항행 정보
LA-PS-01-063	This is Ulsan	항행 정보
LA-PS-01-064	Ulsan	항행 정보
LA-PS-01-065	[Ulsan] What is your ETA?	항행 정보
LA-PS-01-066	My ETA is pilot station is 0800	항행 정보
LA-PS-01-067	Copy	항행 정보
LA-PS-01-068	0800 0800	항행 정보
LA-PS-01-069	And what port?	항행 정보
LA-PS-01-071	My last port of call Dokkum Washington USA	항행 정보
LA-PS-01-072	over	항행 정보
LA-PS-01-081	USA	항행 정보
LA-PS-01-082	over	항행 정보
LA-PS-01-083	Your pilot boarding schedule 0800 0800	항행 정보
LA-PS-01-084	Copy	항행 정보
LA-PS-01-085	over	항행 정보
LA-PS-01-086	0800	항행 정보
LA-PS-01-087	0800	항행 정보
LA-PS-01-088	0800	항행 정보
LA-PS-01-089	0800	항행 정보

Preprocessing Dataset

```

1:0 Utan vts Utan vts this is DND NPP Please over is over.
1:1 [DND on ahead].
2: Good evening please just wait to confirm what time pilot will be boarding.
3: We are waiting please check channel 13.
3: Is it in 11 on contact with pilot.
3: Utan vts DND. I am waiting.
3: Yes on ahead.
3: Your heading Pilot heading schedule is 0900 0900.
4: DND. Thank you and please tell me to ***** pilot ladder over? you.
3: contact pilot channel 13.
3: On contact pilot 13 I tried to call but no replies.
3: On contact 1 hour before 1 hour before 1 hour.
3: On contact that thank you.
3: Utan vts Utan vts this is P1234 channel 14 listening.
3: Start on this is Utan vts.
3: DND over.
3: [DND] this is Utan vts on ahead.
4: Can I hear... Do I have any information about pilot over? hour pilot.
3: heading schedule 0900 0900 over.
3: Your time is 0900 and I tried to call pilot but there is no answer Can I.
3: anchor.
3: Your pilot will be boarding at anchor and do you have about the.
3: pilot ladder over? Pilot ladder you contact pilot channel 13 before 1 hour.
3: Roger sir. Before 1 hour contact channel 13 Thank you sir stand by 14 end.
3:
3: Utan control Utan control this is DND.
1: [DND] Utan vts on ahead.
2: confirm this is DND Pilot station at 0900 So we are on only drifting.
3: around pilot station going to drop anchor sir.
3: hour call sign.
3: Listen sir [DND] your anchor call sign.
3: On the call sign is 1802 1802 Do you read well OK your last sign.
3: Last sign is Green China.
3: Your pilot boarding time 10 o'clock today And I give anchor position when.
3: you ready reporting time.
3: OK sir. I receive 0900 pilot boarding. So we are going to drop anchor.
3: Pilot on the ground pilot station over.
3: You signal very weak. you call me back when you ready reporting time.
3: Again sir? Your signal is very weak Call me back 1 hour later please.
3: Do I will call back 1 hour later.
  
```

Choose QA model

- How to give the answer based on the question
- How to represent text information to handle in a model
- Traditional approach vs. Deep learning based approach

Bag of words

- Create a huge word package (corpus) of the dataset
- Give vector index by the orders of words
- Indicates the presence of a word in a document by 1 and 0
- Counter-based
- TF-IDF - Techniques to reduce the weight of high frequency words
- Sparse vector problem – High k dimensions (k – Number of words)

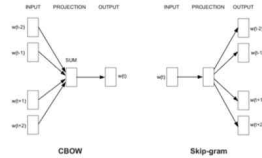
Bag of words

the dog is on the table

0 0 1 1 0 1 1 1
are cat dog is now on table the

Distributed representation

- Learning with neural-net based on probability of occurrence of k-adjacent words
- A dense vector created instead of a sparse vector
- N-dimensional dense vector



Distributed representation

