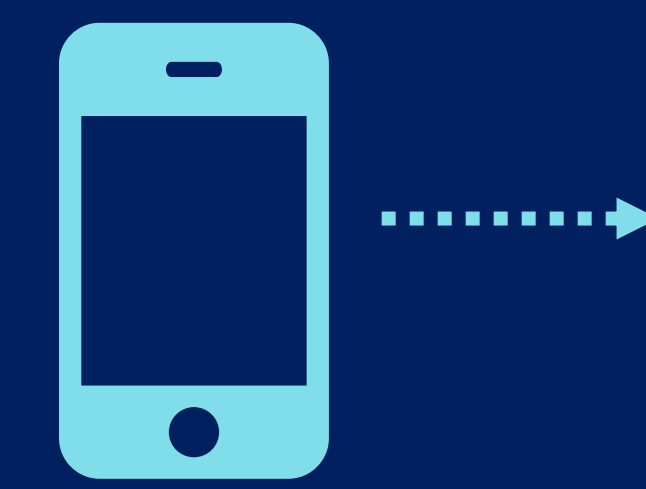


Can rule-based chatbots outperform Neural models without pre-training in Small Data Situations?: A Preliminary Comparison of AIML and Seq2Seq

Md Mabrur Husan Dihyat, Julian Hough
Queen Mary University of London



Take a picture to
download the full paper

TAKEAWAY ANSWER: Yes! Rule-based chatbots using AIML are still safer to use compared to end-to-end Neural model in situations where only small amount of data is available without any pre-training. Much more needs to be done!

INTRO

- While much research has focused on developing and testing chatbot systems within **rule-based** and **deep learning-based** paradigms, few studies compare these two types of system **using the same data**.
- We develop, evaluate and compare the two chatbot paradigms in a real-world scenario: an **IT services chatbot** at Queen Mary University of London which responds to queries relating to computing issues

METHODS

1. A corpus of **200 query-response** pairs was gathered from the university IT services collected between 2018-2020, from which 150 query response pairs were used as training data.
2. The Rule-based chatbot was created using Artificial-Intelligence Markup Language (AIML). Based on the user intents found while analysing the dataset, a total of **ten AIML** files were created where each file addresses a specific issue.
3. For the neural model, we trained an **LSTM (long short-term memory) Sequence-to-Sequence (Seq2Seq)** model on the 150 query-response pairs in the training data, with the responses identical to the AIML templates. The Seq2Seq model has a total of **963,135** trainable parameters.
4. We evaluate the success of the responses to the 50 test set queries automatically using a **ROUGE-1** and **ROUGE-L** comparison to the ground truth response (precision, recall and F-1 measure) and also measure **human-judged task completion success**.

RESULTS

1. Using the first author's judgement, overall, the **AIML** chatbot was found to be approximately **16% more proficient** than the **Seq2Seq** model
2. In terms of **automatic metrics**, in most problem types **the AIML bot outperforms the Seq2Seq model** across the metrics in each category with **two exceptions (Greetings and Agresso issues)**
3. The results suggest that with a small amount of data, both in terms of task success and output quality, it is still **safer** to use a rule-based chatbot with AIML than relying on generalization from an end-to-end neural model, though **no pre-training was used (future work!)**.

TASKS	AIML CHATBOT	SEQ2SEQ CHATBOT
1. GREETINGS	5/7 = 71.4%	5/7 = 71.4%
2. LOGIN ISSUES	3/7 = 42.8%	1/7 = 14.3%
3. MYHR ISSUES	5/11 = 45.5%	3/11 = 27.3%
4. PASSWORD ISSUES	5/8 = 62.5%	1/8 = 12.5%
5. PASSWORD REQUIREMENT	4/6 = 66.7%	4/6 = 66.7%
6. AGRESSO ISSUES	4/6 = 66.7%	5/6 = 83.3%
7. ADDRESS	3/5 = 60%	2/5 = 40%
OVERALL	29/50 = 58%	21/50 = 42%

Figure 1: Task completion rate

Query	Response	Successful
Query 1: I was trying to change the password but I think I have changed the memorable word as well and I forgot them		
Response:		Successful
AIML	To reset your password, please visit the password reset manager at [[Anonymized Website]].	YES
Seq2Seq	It needs to be between 12 to 20 characters long contain upper and lower case characters and numbers.	NO
Query 2: Is it possible to access my [[Anonymous Account]] from outside the [[Anonymised Institution]]		
Response:		Successful
AIML	Please raise a ticket with the IT Service Desk from here: [[Anonymized website]].	NO
Seq2Seq	Yes you can access [[Anonymous Account]] using vdi or direct access. If you do not have access to vdi your line manager centre administrator will need to raise a ticket for your access.	YES
Query 3: How can I find the [[Anonymous Building]] building on [[Anonymous Institution]]		
Response:		Successful
AIML	The [[Anonymous Building]] building is located in our [[Anonymous Location 1]].	YES
Seq2Seq	You can find the building in our [[Anonymous Location 2]].	NO
Query 4: Is it possible to reset the memorable information for myhr		
Response:		Successful
AIML	I suppose to reset the memorable information for myhr is possible, yes.	NO
Seq2Seq	You can call the service desk on [[Anonymous Number]] and ask for a password reset they will email you the new password and the instructions on how to reset it alternatively you can raise a ticket and the service desk will contact you once your password has been reset.	NO

Figure 2: AIML and Seq2Seq chatbot responses

Task	ROUGE-1			ROUGE-L		
	Average precision score	Average recall score	Average f-measure	Average precision score	Average recall score	Average f-measure
Greetings	0.62	0.69	0.68	0.62	0.69	0.68
Login Issues	0.50	0.48	0.48	0.49	0.48	0.48
MYHR Issues	0.49	0.53	0.47	0.46	0.49	0.45
Password Issues	0.61	0.54	0.55	0.59	0.53	0.54
Password Requirement	0.60	0.50	0.54	0.57	0.50	0.51
Agresso Issues	0.67	0.64	0.65	0.66	0.62	0.64
Address	0.54	0.54	0.548	0.54	0.54	0.548

Figure 3: Rouge scores for AIML

Task	ROUGE-1			ROUGE-L		
	Average precision score	Average recall score	Average f-measure	Average precision score	Average recall score	Average f-measure
Greetings	0.79	0.73	0.75	0.79	0.73	0.75
Login Issues	0.33	0.23	0.26	0.32	0.22	0.25
MYHR Issues	0.47	0.35	0.39	0.30	0.19	0.23
Password Issues	0.21	0.27	0.23	0.15	0.18	0.17
Password Requirement	0.64	0.52	0.47	0.60	0.54	0.53
Agresso Issues	0.89	0.87	0.88	0.87	0.86	0.86
Address	0.554	0.45	0.49	0.554	0.45	0.49

Figure 4: Rouge scores for Seq2Seq