

An E-business Chatbot using AIML and LSA

Thomas N T

Amrita Vishwa Vidyapeetham University,
Kollam, Kerala, India
thomasnt13@gmail.com

Abstract—The e-business has completely changed the way of selling products. E-commerce is one of the e-business models which mostly do business over the internet. The major drawback of this field is quality of customer service they provide. In every e-business model, customers have to wait for a long time to get response from the customer service representative. Especially in case of live chat, they talk to multiple customers at a time. The responses may not be relevant as they copy paste pre-written answers. Also, the slow response and the long time wait for the service agent is the biggest headache in this field of online services. As a solution to this problem, we propose a chatbot which automatically gives immediate responses to the users based on the data set of Frequently Answered Questions (FAQs), using Artificial Intelligence Markup Language (AIML) and Latent Semantic Analysis (LSA). Template based questions like greetings and general questions will be answered using AIML and other service related questions use LSA to give responses.

Keywords—E-business; AIML, LSA

I. INTRODUCTION

E-business widely increased their sales by attracting the customers. The number of e-business customers increases daily. So, customer service is a difficult task for them. Customers can buy products by sitting anywhere anytime. But, the quality of customer service is not up to the mark in case of online services. There are different types of customer support services like phone support service, live chat support service etc. In case of live chat support, customers have to wait for their turn. The customer service representative takes time to reply for customer problems. As number of customers increases, the need for customer service also increases.

Automated live chat support service is the only solution for this problem. Chatbot will take care of user queries and problems. It will give the solution for the same. In this paper, a chatbot is designed to answer both general questions and FAQs about a particular domain.

AIML is an artificially intelligent markup language. AIML template is defined with almost all the general queries like how are you?, What's up? etc. It is used to deal general questions and greetings. Queries related to greetings are also handled by AIML. This pattern is a simple language, which can also give random responses for single query or scripts. The main disadvantage of AIML is that the developer has to write patterns for all the questions customers could possibly ask. So, it is a difficult task for developers to design such a system. As a

solution to this, we used the method called LSA[6]. LSA is used to find the semantic similarity between words in vector representation form. For example, vehicle and car has semantic similarity in the vector space according to LSA. So that, the unanswered questions by AIML will be taken up by the LSA part. This technique gives more intelligence to the chatbot. Singular value decomposition reduces the noise in the training data. In this paper a combination of LSA and AIML is used to get customer responses.

II. RELATED WORK

Md. Shahriare Satu and Md. Hasnat Parvez [1] showed a review of integrated applications with AIML based chatbot. The paper explains applications using AIML chatbot instead of human beings to interact with customers and give solution of their problems. They also said about the low cost and configuration of AIML based chatbots.

Wei Yun Gang, Sun Bo, Sun Ming Chen and Zhao Cui Yi [2] introduces various aspects of the Chinese Intelligent Chat Robot Xiao Hui-hui. The lack of dataset and Chinese segmentation system are the most difficult problems they faced during this project.

Salvatore La Bua [3] has proposed an approach of LSA. As most of the chatbots are using AIML, he applied LSA to large amount of documents. It helps to relate words in the vector representation of corpus.

Hadeel Al-Zubaide and Ayman A. Issa [4] proposed an ontology based approach which provides scalability and interoperability properties for seamless support of different domains.

III. PROPOSED SOLUTION

The work flow of the system is shown in the figure 1. The user queries will be first handled by AIML check block. It will check if the entered user query or text is an AIML script or not. Mostly, a user starts conversation by greetings or general questions. AIML is defined with templates of greetings and general queries. Greeting messages by customers like hi, hello, good morning etc will be answered by using AIML templates. It also handles general questions.

Here, we have two types of answers like pattern based answers by AIML and semantic based answers by LSA. The

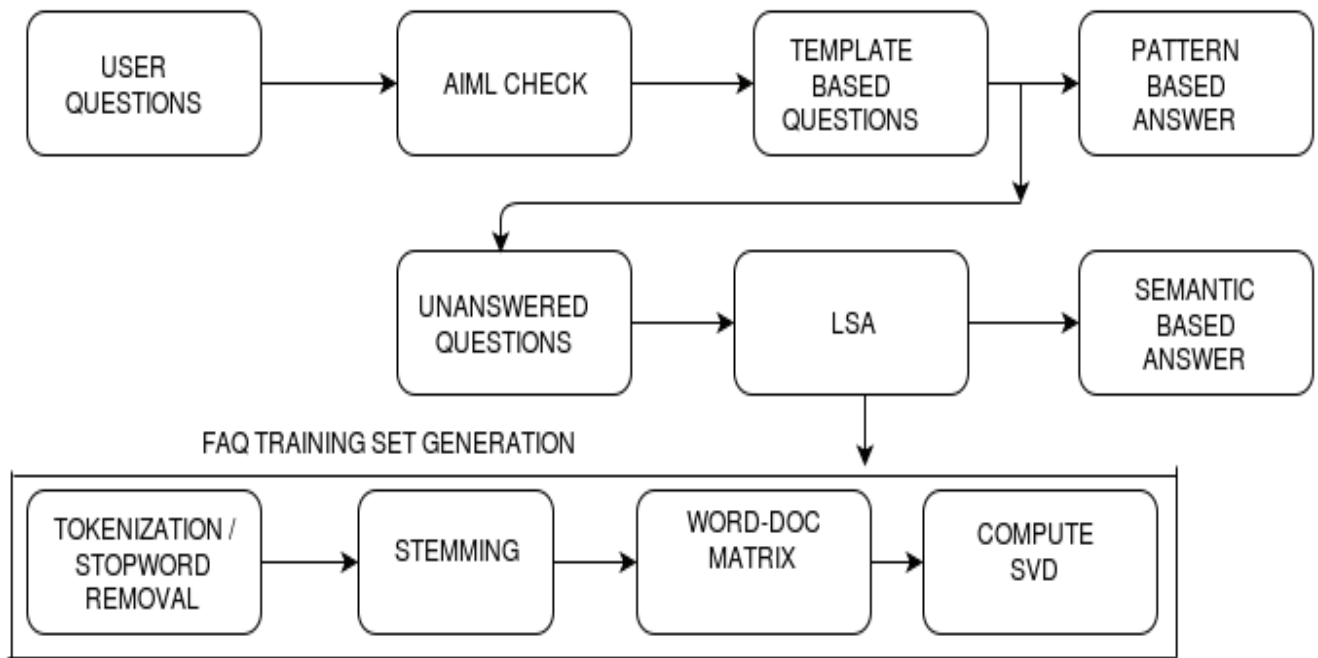


Fig. 1. Workflow of System

system will automatically route to LSA block if AIML is not able to give answer for the particular user input. User can ask any questions. We used the template data of ALICE[7] to run our AIML model.

It's a difficult task for the developer to give all questions user could possibly ask to the chatbot. This is the major drawback of AIML. The unanswered questions from AIML will be automatically routed to the LSA block. The frequently answered questions (FAQs) in the particular e-business domain is used for training the model. The FAQ is made using online data from the internet for various domains.

The FAQ corpus is first tokenized and then the stop words are removed. Next, stemming of words are done using Porter Stemmer algorithm. A word by document matrix is formed where word presence is denoted by 1 and word absence by 0.

The Word by Doc matrix gives relation between a set of concepts arranged as documents and terms. Then singular value decomposition is applied to the word vectors to remove the noise from the dataset. Now, the cosine similarity[5] between user input and FAQ dataset for each questions are checked. The answer from the dataset which is having the minimum distance will be shown to the customer. Here we check the semantic similarity of each words.

All the user conversation is stored in the Hbase. The unanswered questions are stored in a separate table. So that we can track those questions in the Hbase. These questions are then trained. So that, chatbot will be able to give responses to those questions. The similarity score between partially understood questions and FAQ data set is checked for each query. The similarity score below a threshold is considered as irrelevant answers. The irrelevant answers are then, managed by using a 'question back' feature. A set of questions and scripts are defined like 'please rephrase your question', 'can you please be more specific' etc for irrelevant responses. These scripts

will be displayed for partially understood or not understood questions.

IV. CONCLUSION

We tested the model for various scripts written by different users. The model achieved 0.97 precision, which is a significant progress. The chatbot is a combination of AIML and LSA. The combination of technique will help customers to get correct responses. The templates and patterns for general user queries using AIML needs improvement.

The chatbot is designed in such a way that for single template, it gives random responses. LSA based questions are giving correct responses. In future, the chatbot will be able to answer more general questions by defining more templates and patterns.

REFERENCES

- [1] Md. Shahriare Satu; Md. Hasnat Parvez; Shamim-Al-Mamun (2015). "Review of integrated applications with AIML based chatbot". 2015 International Conference on Computer and Information Engineering (ICCI), 87 - 90, DOI: 10.1109/CCIE.2015.7399324
- [2] Wei Yun Gang, Sun Bo, Sun Ming Chen and Zhao Cui Yi. "Chinese Intelligent Chat Robot Based on the AIML Language". Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2014 Sixth International Conference on (Volume:1), DOI:10.1109/IHMSC.2014.96
- [3] "LSA-Bot", [Online]. <http://www.slblabs.com/projects/lisabot>
- [4] Hadeel Al-Zubaide, Ayman A. Issa. "OntBot: Ontology based chatbot". Innovation in Information & Communication Technology (ISIICT), 2011 Fourth International Symposium on, DOI:10.1109/ISIICT.2011.6149594
- [5] "Cosine similarity", [Online]. https://en.wikipedia.org/wiki/Cosine_similarity
- [6] "Latent Semantic Analysis", [Online]. https://en.wikipedia.org/wiki/Latent_semantic_analysis
- [7] A.L.I.C.E Artificial Intelligence Foundation" [Online]. <http://alice.pandorabots.com>

- [8] "Latent Semantic Analysis for Text-based Research" [Online].
<http://www-psych.nmsu.edu/~pfoltz/reprints/BRMIC96.html>
- [9] "An introduction to Latent Semantic Analysis", [Online].
<http://lsa.colorado.edu/papers/dp1.LSAintro.pdf>