

CHATBOT USING AIML AND LSA

M. Sumithra^{1a}, Roshan Azaruddin G^{2*}, Nilesh Narayanan^{3*}, Vimal G A^{4*}, Saravanan S^{5*}

^aAssociate Professor, Department of Information Technology, Panimalar Engineering College

^{*}Student, B.Tech Information Technology, Panimalar Engineering College

-----***-----

ABSTRACT

Chatbots are artificial intelligence-based systems that mimic human conversation (AI). It is intended to be the ultimate virtual assistant for entertainment purposes, assisting with things such as answering queries, getting driving directions, turning up the thermostat in a smart home, and playing one's favourite music, among others. Chatbots are currently playing a pivotal role in business groups because of their ability to multitask and by reducing the cost involved in customer service. As there is a dearth in task accomplishment chatbots need to be as efficient as possible. Using this prototype a paradigm of a chatbot is conceived to resolve problems in an inch-perfect manner using AIML(Artificial Intelligence Markup Language and LSA(Latent Sematic Analysis) with python. Chatbot is a program that runs on artificial intelligence (AI) and machine learning platforms. Chatbots are now becoming more and more popular with business groups as they reduce customer service costs and allow to manage multiple users at the same time. However, to perform many tasks, you need to design your chatbot as efficiently as possible.

1. INTRODUCTION

Chatbots are at the forefront of the software business. Thousands of chatbots are created by startups and employed by enterprises to assist them in their daily operations. An artificial neural network inspired by the neural nodes of the human brain powers a chatbot that uses AI and machine learning to function. Chatbots are computer programmes that can have natural-sounding chats. Facebook, for example, provides a machine learning chatbot that allows businesses to communicate with their customers via the Facebook Messenger app. In the world of web services, there has been a significant development, with everything now being linked to the internet. Everything is delivered to your house, which is a highly user-friendly method. Various sorts of customer service are offered, including live chat assistance and phone (telephone) services. Many web-based services, such as Ebusiness, entertainment, virtual support, and others, are available in today's period. In the world of web services, there has been a significant development, with everything now being linked to the internet. The definition of natural and instinctive connection modes is one of the most significant tasks in the field of Human Computer Interaction (HCI). Numerous initiatives have been dedicated to this goal in particular.

2. EXISTING SYSTEM

The program that converses with humans using natural language are called chatbots. Plethora of organizational domains use chatbots to fill the void of humans.

2.1 ELIZA:

ELIZA is the main chatbot created by Joseph Weizenbaum with a keyword adjustment strategy. It was thought to persuade the customer to search for information safely. Keyword, if a keyword is found, the appropriate one. The response has been restored. If there is no catchphrase present, ELIZA will continue according to the stated principles. Get more data from the client to continue the discussion walk.

2.2 ALICE:

Richard Wallace came up with Alice's idea in 1998. Need help with design adjustments and store the data in artificial intelligence markup language (AIML) records. AIML records are similar to XML documents that help chatbots store design information. AIML records are similar to XML documents created to store chatbot design information. Code 2 shows an example of the discussion

3. PROPOSED SYSTEM

The process of user discussing generally start with a simple welcome or general questions ,the ones which are first taken care In AIML is user inquiries ,it is used to check whether the inquiry which is entered is in the form of a AIML script. User dialogue mainly starts off evolved with welcome or widespread questions. AIML is characterised with widespread inquiries and welcome which is spoke back via way of means of using AIML formats. This operation is divided into 3 parts:

- ◆ User put up the question on chatbot
- ◆ Processing is achieved at the customers question to in shape the predefined layout via way of means of the developer
- ◆ Pattern matching is executed among person entered question and knowledge (sample).

Here in our system we some additional feature like adding faqs for the bot even by a non programmer AIML is characterized by general inquiries, requests, and greetings that are answered using the AIML format.

4. CONCEPT

4.1 AIML:

- ◆ AIML stands for Artificial Intelligence Modeling Language. AIML is an XML-based markup language used to create artificial intelligence applications. AIML makes implementations easy to program, understand, maintainable, and create human interfaces. In this tutorial, you will learn the basics of AIML. This tutorial has covered all the basic components of AIML, including good examples. AIML-based chatbots fall into the category of rule-based chatbots, but some self-learning is feasible.
- ◆ AIML is a language for creating the brains of chatbots.
- ◆ The chatbot's NLU processes AIML, and chat behavior is controlled by AIML rules.
- ◆ Chatbot applications have multiple sets of AIML and may behave differently.
- ◆ The flow chart below shows how an AIML-based chatbot can handle a set of inputs that represent essentially the same text. AIML was originally developed by Dr. Richard Wallace and the global free software community from 1995 to 2002. It formed the first significantly expanded Eliza's idea called "A.L.I.C.E." ("Artificial Language Internet Computer Entity") and won numerous awards. The
- ◆ AIML interpreter AliceBot program is on the market under the GNU GPL. This could be the usual testing and development of AIML-based intelligent bots. The latest version of is AIML 2.0 with powerful NLP processing.

4.1.1 OVERALL ARCHITECTURE DIAGRAM:

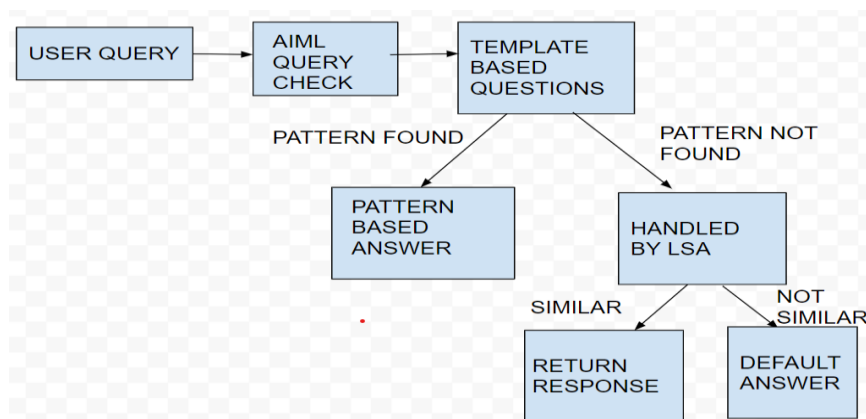


Figure1: Overall Architecture Diagram

4.2 LSA:

- Latent Semantic Analysis (LSA) is a method of natural language processing that use a statistical approach to determine the relationship between words in a document. LSA is tasked with dealing with the following types of problems:
- For example, mobile, phone, cell phone, and telephone are all similar, but when we ask "The cell phone has been ringing," only papers including "cell phone" are returned, but documents containing mobile, phone, and telephone are not.

4.2.1 LSA ARCHITECTURE DIAGRAM:

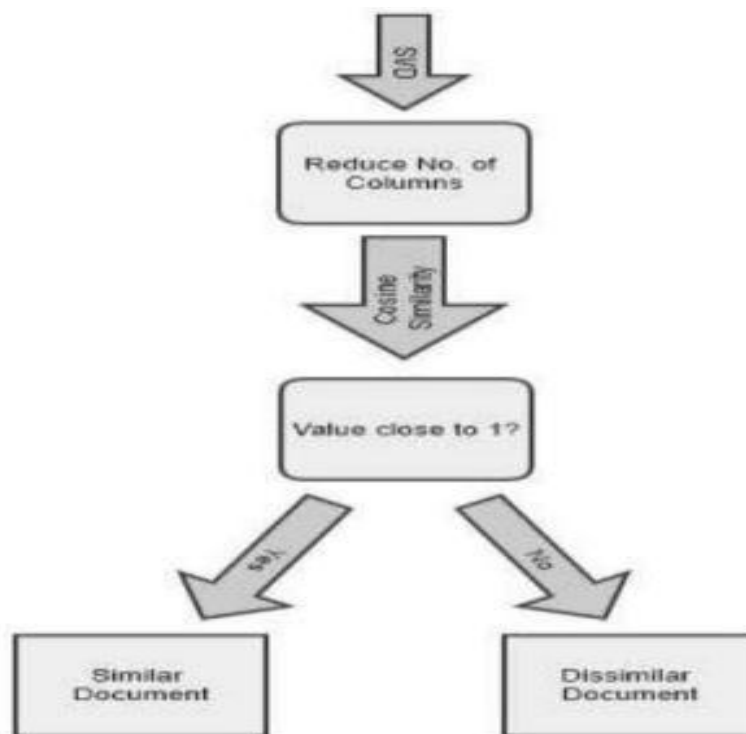


Figure 2: Architecture Diagram of LSA

4.2.2 LEMMATIZATION:

Lemmatization usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma .

4.2.3 REMOVING STOP WORDS:

Stop words are a set of commonly used words in any language. For example, in English, “the”, “is” and “and”, would easily qualify as stop words. In NLP and text mining applications, stop words are used to eliminate unimportant words, allowing applications to focus on the important words instead.

4.2.4 DOCUMENT TERM MATRIX:

A document-term matrix is a mathematical matrix that describes the frequency of terms that occur in a collection of documents. In a document-term matrix, rows correspond to documents in the collection and columns correspond to terms.

4.2.5 TERM FREQUENCY:

In the case of the term Frequency, the weights represent the frequency of the term in a specific document. The underlying assumption is that the higher the term frequency in a document, the more important it is for that document.

$$TF(t) = c(t,d)$$

4.2.6 INVERSE DOCUMENT FREQUENCY:

In the case of IDF, the underlying idea is to assign higher weights to unusual terms, i.e., to terms that are not so common in the corpus. IDF is computed at the corpus level, and thus describes corpus as a whole, not individual documents. It is computed in the following way:

$$IDF(t) = 1 + \log(N/df(t))$$

4.2.7 SINGULAR VALUE DECOMPOSITION:

Singular Value Decomposition is the statistical method that is used to find the latent(hidden) semantic structure of words spread across the document.

Let

C = collection of documents.

d = number of documents.

n = number of unique words in the whole collection.

$$M = d \times n$$

The SVD decomposes the M matrix i.e word to document matrix into three matrices as follows

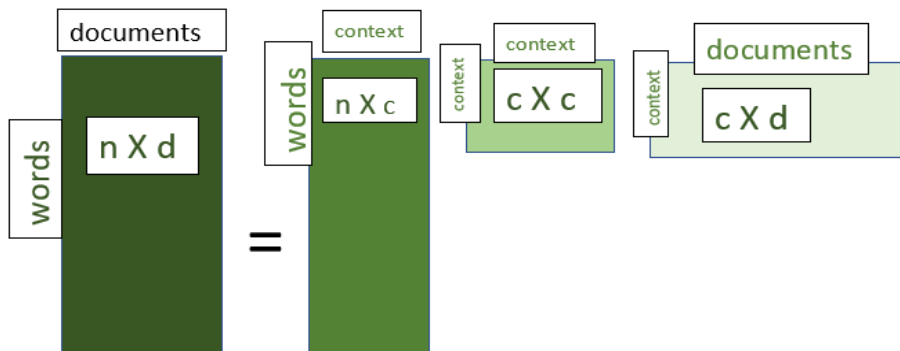


Figure 3: Decomposition of document term matrix

where

U = distribution of words across the different contexts

Σ = diagonal matrix of the association among the contexts

VT = distribution of contexts across the different documents

A very significant feature of SVD is that it allows us to truncate few contexts which are not necessarily required by us. The Σ matrix provides us with the diagonal values which represent the significance of the context from highest to the lowest. By using these values we can reduce the dimensions and hence this can be used as a dimensionality reduction technique too. If we select the k the largest diagonal values in Σ a matrix we obtain

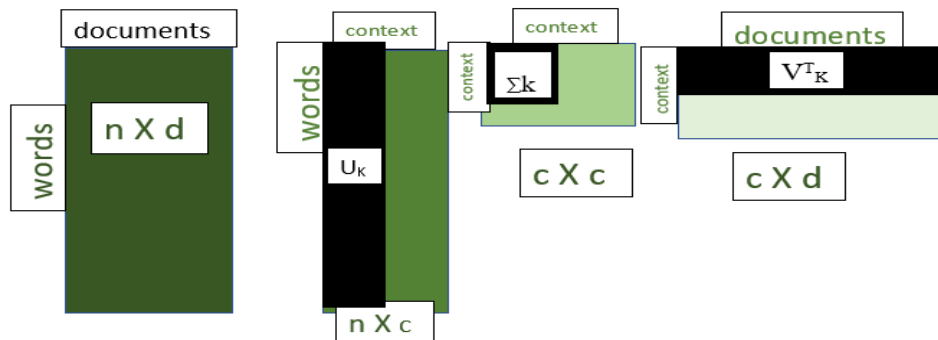


Figure 4: Reduced matrix using Singular value decomposition preserving similarity

$$M_k = U_k \Sigma_k V_k^T$$

where

M_k = approximated matrix of M

U_k, Σ_k, V_k^T are the matrices containing only the k contexts from U, Σ, V^T respectively

4.2.8 COSINE SIMILARITY:

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.

Similarity = $(A \cdot B) / (||A|| \cdot ||B||)$ where A and B are vectors.

5. CONCLUSION

Artificial Intelligence conversational agents are becoming popular for web services and systems like scientific, entertainment and commercial systems, and academia. But more effective human computer interaction will takes place by querying missing data by the user to provide satisfactory answer. In this paper we have proposed and implemented an interactive chatbot for University environment using AIML. The next step towards building chatbots involves helping people to facilitate their work and interact with computers using natural language or using their set of rules. Future Such chatbots, backed by machine learning technology, will be able to remember past conversations and learn from them to answer new ones. The challenge is to talk to different bot users and multiple users. For future work, you can create a mixed chatbot According to AIML and LSA. This allows the client to interact Use chatbots in a more natural way. We can improve it Discussion by changing including patterns and templates General customer inquiries and correct answers regarding AIML Frequently given in LSA.

REFERENCES

- [1] J. Weizenbaum, "Elizaa computer program for the study of natural language communication between man and machine," Communications of the ACM, vol. 9, no. 1, pp. 36-45, 1966.
- [2] A. M. Neves, F. A. Barros, and C. Hodges, "Iaiml: A mechanism to treat intentionality in aiml chatterbots," in Tools with Artificial Intelligence, 2006. ICTAI'06. 18th IEEE International Conference on. IEEE, 2006, pp. 225-231.
- [3] B. A. Shawar and E. Atwell, "Chatbots: are they really useful?" in LDV Forum, vol. 22, no. 1, 2007, pp. 29-49.

- [4] N. Thomas, "An e-business chatbot using aiml and lsa," in *Advances in Computing, Communications and Informatics (ICACCI)*, 2016 International Conference on. IEEE, 2016, pp. 2740–2742.
- [5] M. Sumithra and Dr. S. Malathi, "A Novel Distributed Matching Global and Local Fuzzy Clustering (DMGLFC) FOR 3D Brain Image Segmentation for Tumor Detection", *IETE Journal of Research*, doi.org/10.1080/03772063.2022.2027284, 2021
- [6] B.Buvanswari and T.Kalpalatha Reddy, "A Review of EEG Based Human Facial Expression Recognition Systems in Cognitive Sciences" International Conference on Energy, Communication, Data analytics and Soft Computing (ICECDS), CFP17M55-PRJ:978-1-5386-1886-8, August 2017.
- [7] Chethana, C., Subbiah Swaminathan, S. Sharanyaa, E. Sathish, R. Prathipa, and Anuradha Thakare. "Application Of Reverse Engineering in the Process of Utilization of Human Brain in Artificial Intelligence." *Journal of Optoelectronics Laser* 41, no. 3 (2022): 89-93.
- [8] M. Sumithra and Dr. S. Malathi, "Modified Global Flower Pollination Algorithm-based image fusion for medical diagnosis using computed tomography and magnetic resonance imaging", *International Journal of Imaging Systems and Technology*, Vol. 31, Issue No.1, pp. 223-235, 2021
- [9] K. Sridharan, and Dr. M. Chitra "SBPE: A paradigm Approach for proficient Information Retrieval", *Jokull Journal*, Vol 63, No. 7; Jul 2013
- [10] Sharanyaa, S., P. N. Renjith, and K. Ramesh. "Classification of Parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques." 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2020.
- [11] M. Sumithra and Dr. S. Malathi, "3D Denselex NET Model with Back Propagation for Brain Tumor Segmentation", *International Journal Of Current Research and Review*, Vol. 13, Issue 12, 2021.
- [12] B.Buvaneswari and Dr.T. Kalpalatha Reddy, "EEG signal classification using soft computing techniques for brain disease diagnosis", *Journal of International Pharmaceutical Research*, ISSN : 1674-0440, Vol.46, No.1, Pp.525-528, 2019.
- [13] Sharanyaa, S., P. N. Renjith, and K. Ramesh. "An Exploration on Feature Extraction and Classification Techniques for Dysphonic Speech Disorder in Parkinson's Disease." In *Inventive Communication and Computational Technologies*, pp. 33-48. Springer, Singapore, 2022.
- [14] K. Sridharan, and Dr. M. Chitra "Web Based Agent And Assertion Passive Grading For Information Retrieval", *ARPN Journal of Engineering and Applied Sciences*, VOL. 10, NO. 16, September 2015 pp:7043-7048
- [15] M. Sumithra and Dr. S. Malathi, "Segmentation Of Different Modalities Using Fuzzy K-Means And Wavelet ROI", *International Journal Of Scientific & Technology Research*, Vol. 8, Issue 11, pp. 996-1002, November 2019.
- [16] Sharanyaa, S., S. Lavanya, M. R. Chandhini, R. Bharathi, and K. Madhulekha. "Hybrid Machine Learning Techniques for Heart Disease Prediction." *International Journal of Advanced Engineering Research and Science* 7, no. 3 (2020).
- [17] M. Sumithra and S. Malathi, "A Survey of Brain Tumor Segmentation Methods with Different Image Modalities", *International Journal of Computer Science Trends and Technology (IJCST)* – Vol. 5 Issue 2, Mar – Apr 2017
- [18] B.Buvaneswari and Dr.T. Kalpalatha Reddy, "High Performance Hybrid Cognitive Framework for Bio-Facial Signal Fusion Processing for the Disease Diagnosis", *Measurement*, ISSN: 0263-2241, Vol. 140, Pp.89-99, 2019.
- [19] Sharanyaa, S., and M. Shubin Aldo. "Explore places you travel using Android." In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 4796-4799. IEEE, 2016.
- [20] M. Sumithra and Dr. S. Malathi, "A Brief Survey on Multi Modalities Fusion", *Lecture Notes on Data Engineering and Communications Technologies*, Springer, 35, pp. 1031-1041, 2020.
- [21] Sharanyaa, S., and Madhumitha RP. "Eyeball Cursor Movement Detection Using Deep Learning." RP, Madhumitha and Rani. B, Yamuna, *Eyeball Cursor Movement Detection Using Deep Learning (July 12, 2021)* (2021).

- [22] M. Sumithra and S. Malathi, "A survey on Medical Image Segmentation Methods with Different Modalities", International Journal of Engineering Research and Technology (IJERT) – Vol. 6 Issue 2, Mar 2018.
- [23] B.Buvaneswari and Dr.T. KalpalathaReddy,"ELSA- A Novel Technique to Predict Parkinson's Disease in Bio-Facial",International Journal of Advanced Trends in Computer Science and Engineering, ISSN 2278-3091,Vol.8,No.1,Pp. 12-17,2019
- [24] K. Sridharan , and Dr. M. Chitra, Proficient Information Retrieval Using Trust Based Search On Expert And Knowledge Users Query Formulation System, Australian Journal of Basic and Applied Sciences, 9(23) July 2015, Pages: 755-765.
- [25] Sharanyaa, S., and K. Sangeetha. "Blocking adult account in osn's using iterative social based classifier algorithm." International Journal of Scientific Engineering and Science 2, no. 1 (2018): 33-36.
- [26] B.Buvaneswari and Dr.T. Kalpalatha Reddy, "ACPT- An Intelligent Methodology for Disease Diagnosis",Journal of Advanced Research in Dynamical and Control Systems,ISSN : 0974-5572,Vol.11,No.4,Pp.2187-2194,2019.
- [27] Sumithra, M., Shruthi, S., Ram, S., Swathi, S., Deepika, T., "MRI image classification of brain tumor using deep neural network and deployment using web framework", Advances in Parallel Computing, 2021, 38, pp. 614–617.
- [28] K. Sridharan , and Dr. M. Chitra"RSSE: A Paradigm for Proficient Information Retrieval using Semantic Web" , Life Science Journal 2013;10(7s), pp: 418-425
- [29] Sharanyaa, S., S. Vijayalakshmi, M. Therasa, U. Kumaran, and R. Deepika. "DCNET: A Novel Implementation of Gastric Cancer Detection System through Deep Learning Convolution Networks." In 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), pp. 1-5. IEEE, 2022.fs