A STUDY OF CHATBOTS EVALUATING QUALITY INTELLIGENT GUIDED NAVIGATION

AMARDEEP KUMAR

Research Scholar, Dept. of Computer Application,

Sri Satya Sai University of Technology & Medical Sciences, Sehore, Bhopal-Indore Road, Madhya Pradesh, India

Dr. Jitendra Seethlani

Research Guide, Dept. of Computer Application,

Sri Satya Sai University of Technology & Medical Sciences, Sehore, Bhopal Indore Road, Madhya Pradesh, India.

ABSTRACT

Chatbot's are computer application that interacts with users using normal language. This technology started in the 1960 and the aim of technology was to see if chatbot systems could fool users that they were real humans. However, chatbot systems are not only built to copy human conversation, and entertain users. In this paper, we examine other applications where chatbots could be useful such as education, information retrival, business, and e-commerce A chatbot aims to make a conversation between human and machine. The machine has been embedded knowledge to determine the sentences and making a decision itself as response to answer a question. The response purpose is matching the input sentence from user. From input sentence, it will be scored to get the similarity of sentences, the higher score obtained the more similar of reference sentences. The sentence similarity calculation in this paper using web application which divides input sentence as two letters of input sentence. The information of chatbot are stored in the database. The chatbot consists of core and interface. The database has been considered as information storage and interpreter has been considered as stored programs of function and procedure sets for pattern-matching requirement. The interface is standalone which has been built using programming language of Java. In our system Natural Language Processing is used. Our system understands content of conversation based on recent Natural Language Processing (NLP).

KEYWORDS: Chatbot, User, Communication, MySQL, Entertain.

INTRODUCTION

The need of conversational agents has become important with the widespread use of personal machines with the wish to communicate and the desire of their makers to provide natural language interfaces. In the computer age, there are a vast quantity of documents rapidly created to meet user's needs. There are large amount of digital documents on the Internet, which are used in various different ways. All the documents containing the information support many type of decision-making. However, many questions are still unanswered, which brings the idea of constructing a system for questioning and answering doubts. This clearly brings us to see that the obstacle of building such system which is a knowledge provider. Nevertheless, in several condition e.g. in vacation or during a rest, the seller may be unable to communicate with the customer. But for a user, it may take time to answer customer's doubts. As the number of consumers increase the waiting time increases as well, these results in poor customer satisfaction. As a solution, the seller can employ several customer services to answer any questions from customer. However, it involves additional costs for paying customer service officers. As an alternative and better solution, we can develop a system that can receive and understand the question from customer and automatically, give the best response with respect to that question. NLP provide set of human language technology tool .it support the number of major human language. In Chat bot NLP provides Foundational Building Block for higher level and domain specific text under sting Application. Automated live chat support service is the solution for such problem. Chatbot will take care of user questions and problems. It will give the solution for the same. A chatbot is well designed to answer in such a way both general questions and FAQs about a particular domain.

QUALITY OF CHATBOTS AND INTELLIGENT

Chatbot systems originated with programs like ELIZA (Weizenbaum, 1966) which were intended to demonstrate natural language conversation with a computer. An early stated goal of such systems was to pass the Turing Test (Turing, 1950), in which a human interrogator deems a computer sufficiently "intelligent" to pass as a human. However, primitive systems like ELIZA used keyword matching and minimal context identification, and lacked the ability to keep a conversation going. As shown in Figure 1, through interactions with the program, it was easy to guess that ELIZA was a computer. Researchers continued to develop demonstration systems with natural language capabilities, but none were capable of passing the Turing Test. In the early 1980's, ALICE (Wallace, 2003) was created, becoming significant not for its conversational capabilities but because it led to the the development of Artificial Intelligence Markup Language (AIML). AIML is used to declare pattern-matching rules that links user-submitted words and phrases with topic categories. It is eXtensible Markup Language (XML) based, and supports most chatbot platforms and services in use today.

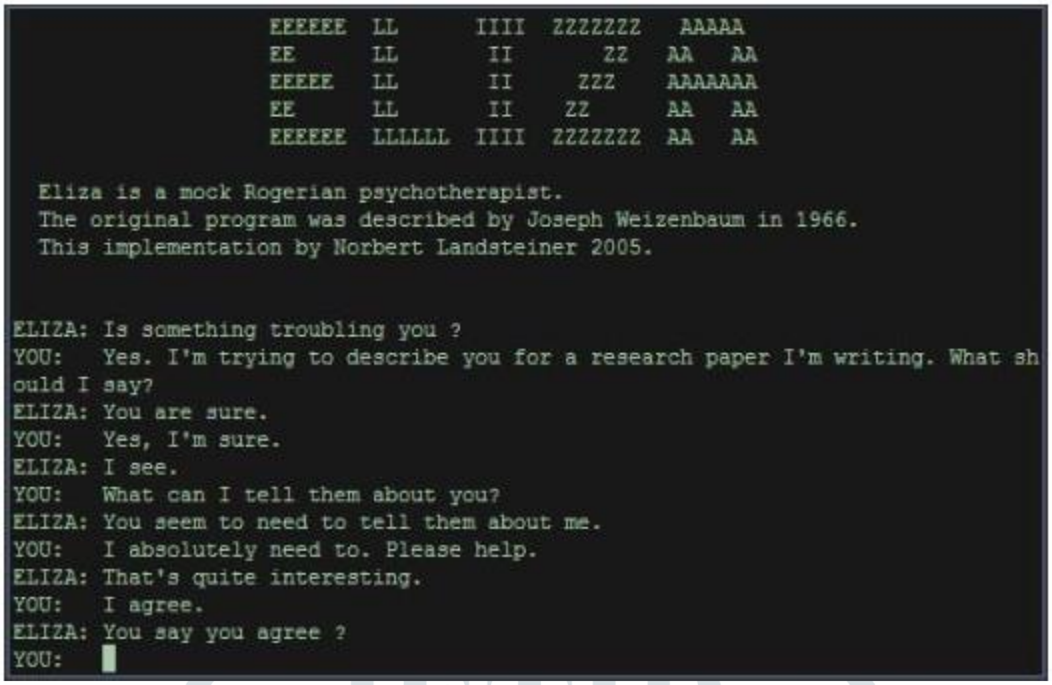

Figure 1. A sample dialog with ELIZA

Chatbots receive natural language input, sometimes interpreted through speech recognition software, and execute one or more related commands to engage in goal-directed behavior (often on behalf of a human user). As intelligent agents, they are usually autonomous, reactive, proactive, and social. The most advanced systems employ machine learning (often Markov chains or deep neural networks) so that they may also adapt to new information or new requests. Chatbots are one category of conversational agents, which are software systems that mimic interactions with real people. They are typically not embodied in the forms of animals, avatars, humans, or humanoid robots (those programs are considered to be "embodied conversational agents"). Conversational agents, a class of dialog systems, have been a subject of research in communications for decades. Interactive Voice Response (IVR) systems (e.g. "Press or Say 1 for English") are also dialog systems, but are not usually considered conversational agents since they implement decision trees. (McTear et al., 2016) These terms are related to each other in Figure 2.

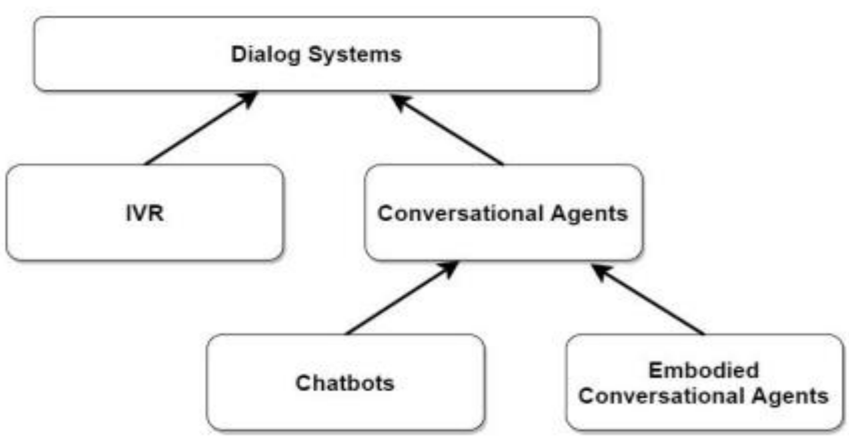

Figure 2. Relationships between classes of software-based dialog systems.

The emphasis in this paper is on the text-based conversational agents found online and in some Internet of Things (IoT) devices, which are sometimes (but not often) embodied. This contrasts with voice-activated conversational agents like Siri, Google Now, Cortana, Samsung S Voice, which are not considered chatbots. Although the earliest software in this genre appeared in Internet Relay Chat (IRC) environments in the mid-1990's, bots have evolved to serve multiple purposes, such as content editing and brokering complex transactions. (Tsvetkova et al., 2016)

These programs serve a range of roles, from personal assistant, to intelligent virtual agent, to companion. There are agents designed to serve as personal university advisors (Ghose & Barua, 2013), educational agents developed to help improve learning outcomes (Kerly et al., 2007), and "Art-Bots" (Vassos et al., 2016) to engage museum visitors in participatory installations. Chatbots are distinct from bots, compromised computers that often run malicious software and can be linked together as botnets to coordinate large-scale denial of service attacks (Thing et al., 2007). However, chatbots can be launched from botnets to shape social perceptions. No one who spends time online is immune from the potential harm of chatbots. Even the director of the annual Loebner Prize Competition in Artificial Intelligence, an event that pits the most sophisticated chatbots against one another, was fooled into thinking a chatbot on a dating service was interested in him romantically. (Epstein, 2007)

Researchers are actively investigating ways to mitigate the harm from chatbots that engage in social engineering. Alarifi et al. (2016), for example, built a corpus of these "sybils" and browsers plug-in to help human users better distinguish between humans and machines. The harm derives from the ease with which these chatbots, especially on social networks like Twitter, pass the Turing Test and convince human participants that their criticisms, abuse, and even rape and death threats are real and originate with other humans. (McElrath, 2017)

The potential for social engineering emphasizes the need to critically examine quality attributes for chatbots, in part to protect the well-being of individuals and societies. Development and implementation of chatbots today is easier, and chatbots themselves are more powerful. Development platforms, some of which implement Software as a Service (SaaS) approaches (e.g. Pandorabots, Chatfuel, Botsify) serve to split the testing responsibility between service provider (who is responsible for testing inputs, execution of actions, and realistic outputs) and client (who evaluates ease of use and effectiveness of task accomplishment). Because of easy integration with social media and developer productivity tools (e.g. Slack, GitHub) chatbots may even be instrumental in improving the work processes in traditional and agile development teams. (Storey & Zagalsky, 2016)

Machine learning approaches are also increasingly integrated to make these agents more adaptive to different input styles and new tasks. Systems are no longer dependent on deterministic responses from rules-based pattern matching, like ELIZA and ALICE. More commonly, systems leverage supervised learning (which requires large training sets), unsupervised learning (like Markov-chain based models), and hybrid intelligence (where humans participate in the training process over time). Supervised learning and hybrid intelligence approaches are more extensive and costly, but can result in systems that are better at on-the-fly problem solving and take less time to achieve goals. (Wilson et al., 2017)

SYNTHESIZED APPROACH:

Analytic Hierarchy Process (AHP) AHP is a structured approach for navigating complex decision-making processes that involve both qualitative and quantitative considerations. First, create a hierarchy of quality attributes and select appropriate metrics to represent each attribute. Second, construct pairwise comparisons between the quality attributes for one or more product options. Next, create comparison matrices and compute the first principal eigenvector of each one to assess relative and global priority. Finally, combine the priorities and compute inconsistency factors to determine which product option best satisfies the hierarchy of quality attributes. (Software is available to make the last two steps easier; code is provided in the Appendix that illustrates how to do this in practice.) For chatbots and conversational agents, the product options might be 1) two or more versions of the conversational system, or 2) the "as-is" system and one or more "to-be" systems under development. Consider an example where you have made improvements to a chatbot system over the past year, and now you want to see if the quality has improved. Have multiple users participate in sessions with your chatbots, and record the metrics you selected. An example is shown in Table 1.

Category	Quality Attribute	Metric	Old	New
Performance	Robustness to unexpected input	% of successes	86-92%	91-93%
	Provides appropriate escalation channels	% of successes	80%	100%
Humanity	Transparent to inspection (known chatbot)	% of users who correctly classify	100%	100%
	Able to maintain themed discussion	0 (low) .. 100 (high)	72 (Avg.) 8 (St. Dev.)	85 (Avg.) 12 (St. Dev.)
	Able to respond to specific questions	% of successes	68-82%	80-85%
Affect	Provides greetings, pleasant personality	0 (low) .. 100 (high)	89 (Avg.) 3 (St. Dev.)	96 (Avg.) 3 (St. Dev.)
	Entertaining, engaging	0 (low) .. 100 (high)	50 (Avg.) 21 (St. Dev.)	66 (Avg.) 4 (St. Dev.)
Accessibility	Can detect meaning and intent	% of successes	85-90%	82-86%
	Responds to social cues appropriately	% of successes	78%	77%

Table 1. Example data to collect for an AHP chatbot quality assessment.

AHP uses pairwise assessments between categories, and within categories. The first step is to set up the attribute hierarchy 1. Notice that the top level displays the goal, the next level shows the quality attribute categories, and the next level from that includes the nine quality attributes that were selected in the prioritization process. At the bottom, the two alternatives (OLD and NEW) are shown. Because the hierarchical structure is central, this approach aligns well with the goal-oriented approach of Kaleem et al. (2016).

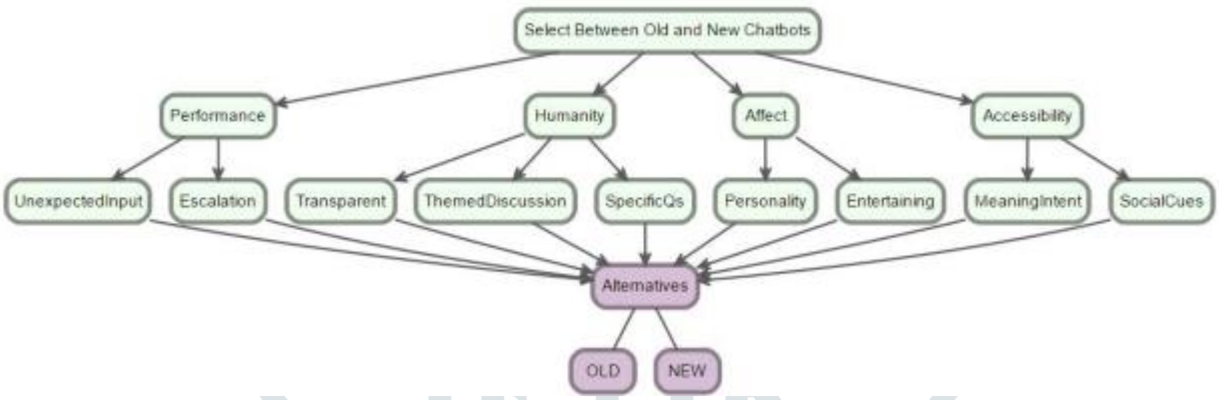

Figure 3. Hierarchical model of prioritized quality attributes.

The first step is to make pairwise comparisons between the categories themselves. Typically in AHP, only the numbers 1, 3, 5, 7, and 9 are used. Create a matrix where each cell indicates how much more important the category in the row is as compared to the category in the column. Record the reciprocal of the value if the category in the column is more important than the category in the row. This means that the number 1 will be in the diagonal, since a category cannot be more or less important to itself. For example, in Figure 4, Performance is much more important than Humanity or Affect (7), but is slightly less important than Accessibility (1/3). Values of 9 and 1/9 indicate that there is a substantial difference between the importance of the two categories being assessed.

Performance Humanity Affect Accessibility

Performance 1 7 7 1/3

Humanity 1/7 1 1/5 1/7

Affect 1/7 5 1 1/7

Accessibility 1/3 7 7 1

Figure 4. Priority matrix for quality categories.

Within each category, create a priority matrix to express the relative merit of each quality attribute as compared to the others. The “Humanity” attribute was selected for display in Figure 5 because it has the most number of attributes for this example. Notice how each measurement has a complementary cell: for example, Transparent is less important than ThemedDiscussion (1/5), so it makes sense that ThemedDiscussion would be more important than Transparent (5, the reciprocal). There will be four priority matrices total at this level (one for each of the four top-level quality attribute categories).

Transparent ThemedDiscussion SpecificQs

Transparent 1 1/5 1/5

ThemedDiscussion 5 1 1

SpecificQs 5 1 1

Figure 5. Priority matrix for selected "Humanity" quality attributes.

Finally, use the measured values (in the "OLD" and "NEW" columns of Table 3) to compare how the different versions of the chatbot perform in terms of each quality attribute. There will be one comparison for each of the nine quality attributes we selected in the original prioritization. Figure 6 shows one of these nine lowest-level priority matrices, showing that the OLD chatbot is much less effective (1/7) in terms of the Escalation attribute - and similarly, the NEW chatbot is much more effective (7).

OLD NEW

OLD 1 1/7

NEW 7 1

Figure 6. Priority matrix for selected "Escalation" quality attribute.

When the first principal eigenvector of each of these matrices are determined and combined, a priority metric for each attribute and category is generated. Although the mechanics of the computations are beyond the scope of this article, instructions are provided in the Appendix for you to use the R Statistical Software (R Core Team, 2016) to perform your own analysis. When the computations are complete, the results help you choose which alternative best satisfies your quality attributes, given that some attributes are more important, and others are less important. Figure 7 shows the results for this example problem. Unfortunately, our efforts have been in vain: the OLD chatbot is weighted much more heavily (66.2% compared to 33.8%) given the priorities that we set between our categories and attributes.

	Weight	OLD	NEW	Consistency
Select Between Old and New Chatbots	100.0%	66.2%	33.8%	18.4%
Accessibility	54.5%	39.1%	15.3%	0.0%
MeaningIntent	47.7%	35.7%	11.9%	0.0%
SocialCues	6.8%	3.4%	3.4%	0.0%
Performance	32.1%	24.6%	7.5%	0.0%
UnexpectedInput	28.1%	21.1%	7.0%	0.0%
Escalation	4.0%	3.5%	0.5%	0.0%
Affect	9.4%	1.6%	7.8%	0.0%
Entertaining	7.8%	1.3%	6.5%	0.0%
Personality	1.6%	0.3%	1.3%	0.0%
Humanity	4.1%	1.0%	3.1%	0.0%
SpecificQs	1.9%	0.3%	1.5%	0.0%
ThemedDiscussion	1.9%	0.5%	1.4%	0.0%
Transparent	0.4%	0.2%	0.2%	0.0%

Figure 7. AHP results for the example problem.

The consistency scores should also be reasonable. Ideally these values should be below 10%, but in practice, numbers below 20% are often acceptable. High values in this column indicate some discrepancy in the individual assessments (for example, prioritizing all elements on one level of your hierarchy low with respect to each other). If there is a problem, inspect all of your assessments to make sure that there have been no data input errors.

CONCLUSION

The device proposed here is an interactive Application, which is capable of answering a questions and Answer. We propose to develop interactive educational software which can run on the desktop. The software helps the user to get answer without reading the file. Initially the software is given input with the file of specific format. Most of the working people in India by using this software will save time and get answer on click. In Future Scope Chat Bot application can be built for the various different fields in different domains along with various domains. For Ex, For communication for Medical, Collages and other important purpose.

REFERENCES

Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003, April). Consolidating the ISO usability models. In Proceedings of 11th international software quality management conference (Vol. 2003, pp. 23-25).

Alarifi, A., Alsaleh, M., & Al-Salman, A. (2016). Twitter turing test: Identifying social machines. Information Sciences, 372, 332-346.

Applin, S. A., & Fischer, M. D. (2015, November). New technologies and mixed-use convergence: How humans and algorithms are adapting to each other. In Technology and Society (ISTAS), 2015 IEEE International Symposium on (pp. 1-6).

IEEE. Beilby, L. J., & Zakos, J. (2014). U.S. Patent No. 8,719,200. Washington, DC: U.S. Patent and Trademark Office.

Bhakta, R., Savin-Baden, M., & Tombs, G. (2014, June). Sharing Secrets with Robots?. In EdMedia: World Conference on Educational Media and Technology (Vol. 2014, No. 1, pp. 2295- 2301).

Bradeško, L., & Mladenić, D. (2012). A survey of chatbot systems through a loebner prize competition. In Proceedings of Slovenian Language Technologies Society Eighth Conference of Language Technologies (pp. 34-37).

Cohen, D., & Lane, I. (2016, February). An oral exam for measuring a dialog system's capabilities. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (pp. 835- 841). AAAI Press.

Coniam, D. (2014). The linguistic accuracy of chatbots: usability from an ESL perspective. Text & Talk, 34(5), 545-567.

Eeuwen, M. (2017). Mobile conversational commerce: messenger chatbots as the next interface between businesses and consumers (Master's thesis, University of Twente). Retrieved on February 1, 2017 from http://essay.utwente.nl/71706/1/van%20Eeuwen_MA_BMS.pdf

Epstein, R. (2007). From Russia, with love. Scientific American Mind, 18(5), 16-17. Fenton, N. E., & Pfleeger, S. L. (1998). Software Metrics: A Rigorous and Practical Approach: Brooks.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104.

Ghose, S., & Barua, J. J. (2013, May). Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor. In Informatics, Electronics & Vision (ICIEV), 2013 International Conference on (pp. 1-5).

IEEE. Goh, O. S., Ardil, C., Wong, W., & Fung, C. C. (2007). A black-box approach for response quality evaluation of conversational agent systems. International Journal of Computational Intelligence, 3(3), 195-203.

Hertzum, M., Andersen, H. H., Andersen, V., & Hansen, C. B. (2002). Trust in information sources: seeking information from people, documents, and virtual agents. Interacting with computers, 14(5), 575-599.

Isaac, A. M., & Bridewell, W. (2014). Mindreading deception in dialog. Cognitive Systems Research, 28, 12-19.

Kaleem, M., Alobadi, O., O’Shea, J., & Crockett, K. Framework for the Formulation of Metrics for Conversational Agent Evaluation. (2016).

In RE-WOCHAT: Workshop on Collecting and Generating Resources for Chatbots and Conversational Agents-Development and Evaluation Workshop Programme, May 28 (p. 20).