T- CHATBOT PROJECT REPORT

Gülin Tunca, Yasemin Turhan and Süleyman Yiğit Tosun

Department of Software Engineering, Istanbul Aydin University, Istanbul, Turkey

Abstract

Using technology in globalization word, we are going to examine data mining a way with help to need of person's requirements to get easier. In these facilities, we are determining to project requirement to use data analysis with applying to usage of chat bot that used in every area daily life to make need data analysis and taking hand to requirements of customer and company owners with resolve most efficacy and clear way. In this project, our target is get resolve to problems that gets by even customer or employees with most effective way and with that increase customer satisfaction and decrease working load.

1.INTRODUCTION

A Chatbot is a conversational agent that interacts with users in a certain domain or on a certain topic with natural language sentences. Many chatbots have been deployed on the Internet for the purpose of seeking information, site guidance. Most existing chatbots consist of dialog management modules to control the conversation process and chatbot knowledge bases to response to user input .Typical implementation of chatbot knowledge bases contains a set of templates that match user inputs and generate responses. Templates currently used in chatbots, however, are hand coded. Therefore, The construction of chatbot knowledge bases is time consuming and difficult to adapt to new domains. We get faced a lot of problem as we make analysis in sector. These problems are not only effect company, they affect employees and customers as a big problem as well. After we saw this opportunity as a serve to people, we decide to make this project. Most of companies gives real time chat support nowadays. With this chat support customers can get help in real time from companies. Most of time it gets very complicated job for companies cause of not enough number of employee to get solution customer in real time. That's why most of time customer waits to connect an employee to get help. This causes to customer less satisfaction and loss of time even if they have a simple problem to solve. In the same time this causes performance loses and emotional distraction as employee's side. As the companies side their profit margin ratio decreases to get more employee. We are taking hand to this problem and creating a chat bot with using big data of companies. With chat bot, we can solve problems with no need of any person. It is being also much more because it has no contain real person in solving step. That why we can eliminate problems as waiting time etc. in solving step and get most efficacy and fast way to resolve.

To sum up, chat bot supply companies and customers to get solution in most efficacy and fast way. In these facilities chat bot's aim is get data from customer and companies and resolve quickly, and decrease companies work load and increase customer satisfaction.

2. RELATED WORK

Competitions such as the Alexa Prize , ConvAI and WOCHAT, rank submitted chatbots by having humans converse with them and then rate the quality of the conversation. However, asking for absolute assessments of quality yields less discriminative results than soliciting direct comparisons of quality. In the dataset introduced for the ConvAI2 competition, nearly all the proposed algorithms were evaluated to be within one standard deviation of each other (Zhang et al., 2018). Therefore, for our human evaluation task, we ask humans to directly compare the responses of two models given the previous utterances in the conversation. Both Facebook and Amazon have developed evaluation systems that allow humans to converse with (and then rate) a chatbot (Venkatesh et al., 2018; Miller et al., 2017). Facebook's ParlAI 4 is the most comparable system for a unified framework for sharing, training, and evaluating chatbots; however, ChatEval is different in that it entirely focuses on the evaluation and warehousing of models. Our infrastructure relies only on output text files, and does not require any code base integration .

3. Technology

3.1.Python (latest version prefered):

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Why Python is useful?

• Beginner Friendliness: Python was designed to be easy to understand and fun to use (its name came from Monty Python so a lot of its beginner tutorials reference it). Fun is a great motivator, and since you'll be able to build prototypes and tools quickly with Python, many find coding in Python a satisfying experience. Thus, Python has gained popularity for being a beginner-friendly language, and it has replaced Java.

• Easy to Understand: Being a very high level language, Python reads like English, which takes a lot of syntax-learning stress off coding beginners. Python handles a lot of complexity for you, so it is very beginner-friendly in that it allows beginners to focus on learning programming concepts and not have to worry about too much details.

• Very Flexible: As a dynamically typed language, Python is really flexible. This means there are no hard rules on how to build features, and you'll have more flexibility solving problems using different methods (though the Python philosophy encourages using the obvious way to solve things). Furthermore, Python is also more forgiving of errors, so you'll still be able to compile and run your program until you hit the problematic part.

• Many Open Source Frameworks and Tools

As an open source programming language, Python helps you to curtail software development cost significantly. You can even use several open source Python frameworks, libraries and development tools to curtail development time without increasing development cost. You even have option to choose from a wide range of open source Python frameworks and development tools according to your precise needs. For instance, you can simplify and speedup web application development by using robust Python web frameworks like Django, Flask, Pyramid, Bottle and Cherrypy. Likewise, you can accelerate desktop GUI application development

3.2 Pycharm

PyCharm is an IDE for Python developed by JetBrains. PyCharm is built for professional Python developers, and comes with many features to deal with large code bases: code navigation, automatic refactoring, and other productivity tools, in a single unified interface.

•Why Pycharm is useful?

•Code Completion: PyCharm has great code completion, whether it’s for a built-in or an external package.

•Python Type Hinting: If code completion fails you, type hints (as in episode_id: int) will turn it right back on. Notice the dropdown lists methods from the int class.

•Package Management: we can all use pip. But a nice visual representation of what’s installed, whether it’s current, and ability to search and add new packages is sweet.

3.3 SQLite

SQLite is not directly comparable to client/server SQL database engines such as MySQL, Oracle, PostgreSQL, or SQL Server since SQLite is trying to solve a different problem.
Client/server SQL database engines strive to implement a shared repository of enterprise data. They emphasize scalability, concurrency, centralization, and control. SQLite strives to provide local data storage for individual applications and devices. SQLite emphasizes economy, efficiency, reliability, independence, and simplicity.
SQLite does not compete with client/server databases. SQLite competes with fopen().

Why it is useful?
•Client/Server Applications
If there are many client programs sending SQL to the same database over a network, then use a client/server database engine instead of SQLite. SQLite will work over a network filesystem, but because of the latency associated with most network filesystems, performance will not be great. Also, file locking logic is buggy in many network filesystem implementations (on both Unix and Windows). If file locking does not work correctly, two or more clients might try to modify the same part of the same database at the same time, resulting in corruption. Because this problem results from bugs in the underlying filesystem implementation, there is nothing SQLite can do to prevent it.
A good rule of thumb is to avoid using SQLite in situations where the same database will be accessed directly (without an intervening application server) and simultaneously from many computers over a network.

•High-volume Websites
SQLite will normally work fine as the database backend to a website. But if the website is write-intensive or is so busy that it requires multiple servers, then consider using an enterprise-class client/server database engine instead of SQLite.

•Very large datasets
An SQLite database is limited in size to 140 terabytes (247 bytes, 128 tibibytes). And even if it could handle larger databases, SQLite stores the entire database in a single disk file and many filesystems limit the maximum size of files to something less than this. So if you are contemplating databases of this magnitude, you would do well to consider using a client/server database engine that spreads its content across multiple disk files, and perhaps across multiple volumes.

•For device-local storage with low writer concurrency and less than a terabyte of content, SQLite is almost always a better solution. SQLite is fast and reliable and it requires no configuration or maintenance. It keeps thing simple. SQLite "just works".

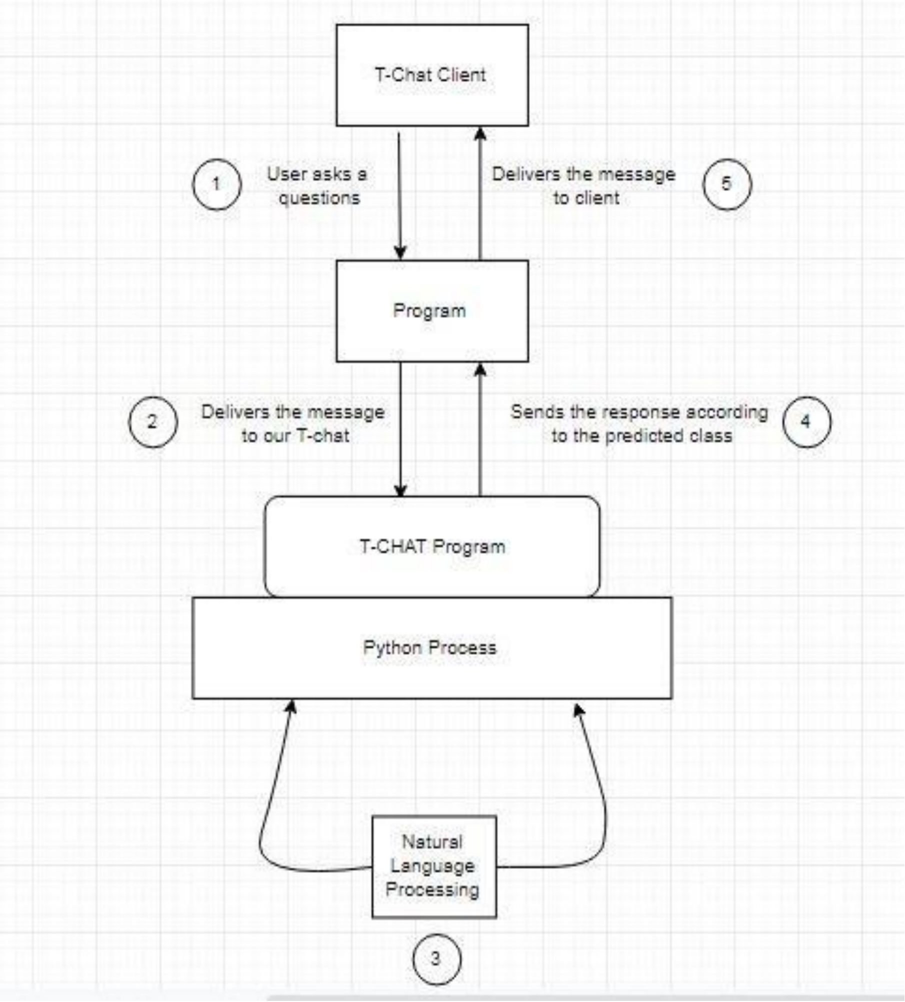
3.4 JSON

JSON is a syntax for storing and exchanging data.JSON is text, written with JavaScript object notation.
When exchanging data between a browser and a server, the data can only be text.JSON is text, and we can convert any JavaScript object into JSON, and send JSON to the server.We can also convert any JSON received from the server into JavaScript objects.This way we can work with the data as JavaScript objects, with no complicated parsing and translations.

Why we use?

Since the JSON format is text only, it can easily be sent to and from a server, and used as a data format by any programming language.JavaScript has a built in function to convert a string, written in JSON format, into native JavaScript objects:JSON.parse()

4. Chatbot Working Logic


T-Chat Client
1
User asks a questions
Delivers the message to client
5
Program
2
Delivers the message to our T-chat
Sends the response according to the predicted class
4
T-CHAT Program
Python Process
Natural Language Processing
3

Fiqure 1.1

4.1 Natural Language Processing

Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken. NLP is a component of artificial intelligence .
The development of NLP applications is challenging because computers traditionally require humans to "speak" to them in a programming language that is precise, unambiguous and highly structured, or through a limited number of clearly enunciated voice commands. Human speech, however, is not always precise -- it is often ambiguous and the linguistic structure can depend on many complex variables, including slang, regional dialects and social context.

•How natural language processing works: techniques and tools;

Syntax and semantic analysis are two main techniques used with natural language processing. Syntax is the arrangement of words in a sentence to make grammatical sense. NLP uses syntax to assess meaning from a language based on grammatical rules. Syntax techniques used include parsing (grammatical analysis for a sentence), word segmentation (which divides a large piece of text to units), sentence breaking (which places sentence boundaries in large texts), morphological segmentation (which divides words into groups) and stemming (which divides words with inflection in them to root forms).

Semantics involves the use and meaning behind words. NLP applies algorithms to understand the meaning and structure of sentences. Techniques that NLP uses with semantics include word sense disambiguation (which derives meaning of a word based on context), named entity recognition which determines words that can be categorized into groups), and natural language generation (which will use a database to determine semantics behind words).

Current approaches to NLP are based on deep learning, a type of AI that examines and uses patterns in data to improve a program's understanding. Deep learning models require massive amounts of labeled data to train on and identify relevant correlations, and assembling this kind of big data set is one of the main hurdles to NLP currently.

Earlier approaches to NLP involved a more rules-based approach, where simpler machine learning algorithms were told what words and phrases to look for in text and given specific responses when those phrases appeared. But deep learning is a more flexible, intuitive approach in which algorithms learn to identify speakers' intent from many examples, almost like how a child would learn human language.

Three tools used commonly for NLP include NLTK, Gensim, and Intel NLP Architect. NTLK, Natural Language Toolkit, is an open source python modules with data sets and tutorials. Gensim is a Python library for topic modeling and document indexing. Intel NLP Architect is also another Python library for deep learning topologies and techniques.

4.2 Artificial Neural Networks

Neural Networks are a way of calculating the output from the input weighted connections which are calculated from preapred iterations while training the data. Each step through the training data amends the weights resulting in the output with accuracy.

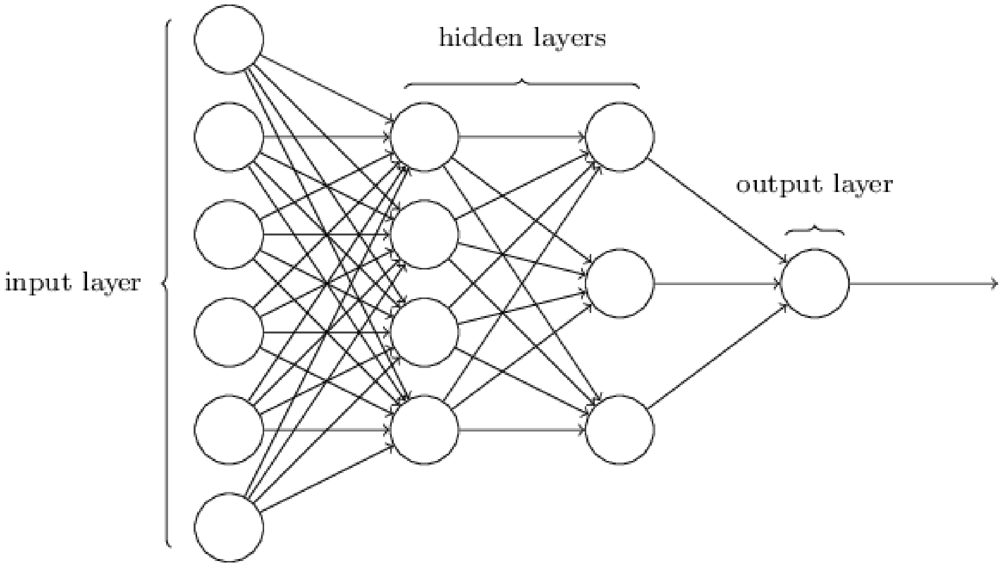


Figure 1.2

4.3 Chatbots As Virtual Assistants Logic

Allowing users to have natural conversations with a system opens up new possibilities for accessing content and services. Users can make direct inquiries in a natural way rather than rely on traditional navigational or search capabilities. A chatbot is a technology that allows users to have natural conversations to access content and services. Chatbots typically take the form of a chat client, leveraging natural language processing to conduct a conversation with the user. Chatbots control conversation flow based on the context of the user's requests and respond with natural language phrases to provide direct answers, request additional information or recommend actions that can be taken. The diagram below provides a high level description of how a chat client could be used to leverage natural language processing to assist with access to content or perform data queries.

4.4 Chatbot Response

This will make your chatbot more real, human-like and more reasonable. Each file has two databases: Special and Matrix. Each database also contains two fields:
1st field: "Request" is where the user inputs are compared with data recorded in this field. 2nd field: "Respond" is where the chatbot outputs are taken from data recorded in this field.

Data: We used Question-Answer data set. Searched whether NLP is used to answer questions.
We read the article "Customer service chatbot, Lei Cui,Shaohan Huang, Microsoft Research Asia" . We searched datasets through Kaagle website. We have decided that the most appropriate dataset is the Question-Answer dataset. We read the article "Question Generation as a Competitive Undergraduate Course Project" and got information about the Question-Answer dataset. Also, we used prepared dataset 2006-3 on reddits for commands.

5. Experiments, Analysis and Performance

Discuss the experiments that you performed to demonstrate that your approach solves the problem. The exact experiments will vary depending on the project, but you might compare with previously published methods, perform an ablation study to determine the impact of various components of your system, experiment with different hyperparameters or architectural choices, use visualization techniques to gain insight into how your model works, discuss common failure modes of your model, etc. You should include graphs, tables, or other figures to illustrate your experimental results.

6. Conclusions

T-Chat is systematic evaluation of chatbots. Specifically, it is a repository of model code. and parameters, evaluation sets, model comparisons, and a standard human evaluation setup. ChatEval seemlessly allows researchers to make systematic and consistent comparisons of conversational agents. We hope that future researchers–and the entire field–will benefit from T-Chat.

7. References

[1] D. Parnas, "On the Criteria to be Used in Decomposing Systems into Modules", *Communications of ACM*, vol. 15, no. 12, pp. 1053-1058, 1972.

[2] S. McConnell, *Rapid Development*, Microsoft Press, 1996.

[3] The CHAOS Reports, http://www.standishgroup.com.

[4] ISO/IEC 9126-1, *Software Engineering - Product Quality - Part 1: Quality Model*, 2001.

[5] J. Heaton, Generating Faces with a Generative Adversarial Networks (GAN) in Keras/Tensorflow 2.0 (7.2), 11 December 2019, https://www.youtube.com/watch?v=Nrsy6vF7rSw

[6] J. Louis, Image Processing and Neural Networks Intuition: Part, January 16, 2018. https://www.datasciencecentral.com/profiles/blogs/image-processing-and-neural-networks-intuition-part-1

[7]: Sample Chatbot Report in Turkey

https://www.iabturkiye.org/UploadFiles/Reports/Chatbot-iab912018131741.pdf

[8]: Python Chat Bot Tutorial - Chatbot with Deep Learning

https://www.youtube.com/watch?v=wypVcNIH6D4

[9]: About What is NLP? And how is train

https://lechatbot.com/chatbotunu-nas%C4%B1l-e%C4%9Fitirsin-e56c2c283ceb

[10]Question Generation as a Competitive Undergraduate Course Project

Noah A. Smith

[11] Chatbot youtube course

https://www.youtube.com/watch?v=dvOnYLDg8_Y

[12] Natural Language Process

https://searchbusinessanalytics.techtarget.com/definition/natural-language-processing-NLP

[13] Pycharm

https://www.g2.com/products/pycharm/reviews

https://blog.michaelckennedy.net/2015/11/19/9-reasons-you-should-be-using-pycharm/

[14] SQLite for Database

https://www.sqlite.org/whentouse.html

[15]Python

https://www.python.org/doc/essays/blurb/

http://www.bestprogramminglanguagefor.me/why-learn-python

https://medium.com/@mindfiresolutions.usa/python-7-important-reasons-why-you-should-use-python-5801a98a0d0b

http://www.bestprogramminglanguagefor.me/why-learn-python

[16] Chatbot Datasets

https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/?st=j9udbxta&sh=69e4fee7

http://files.pushshift.io/reddit/comments/