Voice-enabled dialogue assistant to for getting information on food nutritions

Authors: Sami Uddin and Ibrokhim Iskandarov

Introduction

In modern days the influence of NLP has become very important in human lives. This branch of artificial intelligence got such a broad spectrum of applications because the algorithms aiding computers to understand people as we do are becoming more powerful and efficient. Despite many breakthroughs in the technology of NLP, it is not a simple task to teach machines the way how we communicate. Some of the common factors which may have a negative impact on the performance of a given NLP can be but not limited to local speech slang, a background noise, how similar the sequence of words is and etc.

Since one of the main roles of using NLP is to simplify to some extent the interaction of humans with computers, we will develop a system which will assist users in gaining information about their requested food nutritions. Also, our system will allow users to assert their interests casually. After making an extensive search on the internet, we concluded that there are no many applications which may perform the above-specified tasks. The overall expected format of the program will be to prompt the user to speak out the food nutrition he wants, and output all relevant information such as calories, energy, nutrition elements etc. One of the main resources to obtain information about the specified food nutrition will be a 3rd party API provided by the United States Department of Agriculture(https://fdc.nal.usda.gov).

Below, we will outline our main research and corresponding conclusions about the literature we found which discusses the theory and algorithms of NLP as well as works which found their direct applications in industry and similar works related to food searches, which we may potentially use for our goals. We will also discuss in some more details the crucial importance of our project in the literature review sections.

State Of Acceptance To Diet Management Applications

That's being said, you are what you eat!

Nowadays, the world is facing an unprecedented level of obesity and numerous diseases such as cancer, diabetes and other chronic illness because of unhealthy food habits. Just in the USA, the percentage of obese people covers 35% of adults and 17% of teenagers[1].

However, diet planning is becoming increasingly popular among people in general. Especially with the help of smartphones and the internet, it's easier than ever to track the food intake and count the calories. Even the older adults are more leaning towards getting virtual assistance [2]. And there have been some applications that focus on helping its consumers by guiding them in managing weights, eating better and effectively.

Studies have shown that people are more likely to use apps if they perceive them effective and useful[3]. And there are quite a number of popular mobile applications out in the market. To name a few, there is an app named Calorie Counter & Food Diary(https://www.mynetdiary.com/). Users can set nutrition and weight goals and can scan barcodes while shopping which helps to make wise decisions. It also lets users track different nutritional values. Another app is *Food Intolerances App*(https://www.baliza.de/en/apps/histamine.html). Not all food is suitable for everybody and this varies from person to person. This app helps people with allergies such as histamine intolerance, mastocytosis, fructose malabsorption, sorbitol intolerance, gluten sensitivity, and lactose intolerance in choosing the right one.

So it's evident that our system will be highly welcome if it serves the purpose effectively and effortlessly and lets users have what they actually want.

Works Towards Designing Dialogue Systems

We are going to develop our system where we take voice input and show results after inferring the command. That's why, In this section, we will briefly discuss the other applications of NLP which found their niche in solving real world problems. The main idea behind all of the products we mention below, are based on a simple receiving an input in terms of speech from the user and respond with the required action based on customer needs. Despite the novelty of NLP technology, it already existed since early 1990th. Back at that time, the functionality of early dialogue assistants were limited to telephone-based call routing systems [4], weather information systems [5], travel planning [6] and etc.

A more recent developments of the dialogue systems include more advanced tasks such as car navigation, entertainment, and communications [7]. These systems are also called a multi-domain which involve several different tasks all at once. Some examples are, telematics, smart home, and intelligent robots. These systems have gradually become capable of supporting multiple tasks and of accessing information from a broad variety of sources and services.

Information Retrieval

Information retrieval is an activity and science that works towards finding relevant information from a collection of resources. These resources can be text, images, audio, videos or metadata that represents data and searches can be full-text or other content based indexing.

IR is also a query based process where the user enters something and the system provides related information. However, the difference between IR and classical SQL queries of a database is, the results from the IR system are typically ranked and it is often expected that they do not match with the user's query. It provides a computed numeric score to each result that matches the user's query and rank according to that. Users get to see the top ranked result and the query can be refined if user wishes[8]

Our plan is to take voice input to our system, convert it into text format and mine the information about food names from that text.

Often it's a case that users have a mental model for their queries, but they lack the knowledge of how an IR system works. Whether the query is short or long, the effectiveness varies. Often the query needs to be further clarified. It's also important to understand which query needs to be clarified. To solve this problem, authors presented a technique that can determine and can avoid unnecessary queries and maintain the performance[9]. From our project's perspective, this will be a very useful technique. Our system will have to determine if it needs further information when a user says "I ate beef" or "I had 1/2 lb beef brisket". When a user says "I ate beef", the IR system of our software will decide if it needs to ask the user about the quantity. Furthermore, a user can ask "Calories in my soup", in which case the IR system will detect that there are many types of soups in the market and it will ask further questions to have specific information about the soup from the user.

Partially Observable Markov Decision Processes For Spoken Dialogue Systems

In this paper, the authors are discussing possible issues when dealing with a Spoken Dialogue System (SDS). It becomes particularly useful when user commands are constrained into digits, place-names, and some short commands. But unfortunately, it becomes error prone when dealing with complex speech domains. In other terms, once the task complexity increases, the accuracy of SDS decreases rapidly.

To improve SDS performance, there are three methods: confidence scores, automated planning and parallel dialogue hypothesis each of these methods are used depending on the type of error we are dealing with. The main problem occurs once we combine these approaches to resolve all types of complexities. This is caused by the lack of an overall statistical framework.

One solution that the authors are proposing is a partially observable Markov decision process (POMDP) that can enable a proper framework [10]. We will see that POMDP is a more general approach and all three methods mentioned above are just special cases of POMDP.
POMDP can be defined as a tuple [11],
{S, A, T, R, O, Z, k, b0}

- S is a set of states
- A is a set of actions
- T is a transition probability from one state to another given the action
- R is expected real values reward
- O is a set of observations
- Z is observation probability
- k is a geometric discount factor
- b0 is an initial belief

The paper describes the definition of each of these arguments and how they are related. The final goal will be to calculate at each time-step, the belief state distribution b. One of the important points in this article covers not only the transition to the regular next state solely depending on the current state but also discusses the situations when the next state may depend not only on the current state but also on few last states. This paper integrates the application of POMDP to all three approaches in resolving the errors in detail.

Even so, despite the clear potential of POMDPs, several key challenges remain. Most crucially, scaling the model to handle real-world problems remains a significant challenge: the complexity of a POMDP grows with the number of user goals, and optimization quickly becomes intractable.

In the current section, we have discussed in some detail the principle of POMDP. We mentioned how important this concept could be when dealing with dialogue systems. We also discussed that the POMDP is a complex structure which unifies all existing methods in identifying various types of errors. Additionally, despite the fact of promising applications of this method, there still some open questions related to this topic.

The relevance of this topic to our work is the following, in order to outline for the user the proper food nutritions, we may also maintain the program to prompt a few questions (sometimes consisting of more) the user.

Here is a simple illustration,

- User: "Cheese"
- Our System: specify which type of cheese
- User: "cheddar cheese"

Next time, when the user wants cheese, the system will show the result for cheddar cheese as a first result.

Conclusion

In conclusion, we will briefly outline the main points we have covered in this literature review. We have discussed the primary reason for diet planning in human's lives these days. The main objective of this was a broad variety of food items which are not always healthy and as a result may lead to illnesses. Also, it's often annoying to go and type anything just to find something related to food nutrition or calorie details. Voice-assistant system to retrieve food nutritions would enable consumers to keep track of the food elements effortlessly. We have also discussed in some detail previous works in the area of voice assistant which found their applications in different areas. Furthermore, what we have also realized was the fact that usually the working principles of voice-assistants sometimes undergo lengthy dialogues between the human and the machine. These dialogues are needed for proper service or information. As it has been mentioned, this sometimes may lead to errors. The various types of the errors and the way how to overcome them has been discussed where we introduce a so called POMDP.

References

[1] Johnson, Nicole Blair, et al. "CDC National Health Report: leading causes of morbidity and mortality and associated behavioral risk and protective factors—United States, 2005–2013." (2014).

[2] Festervand, Troy A., David B. Meinert, and Scott J. Vitell. "Older adults attitudes toward and adoption of personal computers and computer-based lifestyle assistance." *Journal of Applied Business Research (JABR)* 10.2 (1994): 13-22.

[3] Okumus, Bendegul, Anil Bilgihan, and Ahmet Bulent Ozturk. "Factors affecting the acceptance of smartphone diet applications." *Journal of Hospitality Marketing & Management* 25.6 (2016): 726-747.

[4] A.L. Gorin et al.*Speech Communication* 23 18997 113-127. https://scholar.google.com/citations?user=k9pXsKoAAAAJ&hl=en

[5] Zue, Victor, et al. "JUPITER: a telephone-based conversational interface for weather information." *IEEE Transactions on speech and audio processing* 8.1 (2000): 85-96.

[6] Walker, Marilyn A., et al. "DARPA communicator evaluation: Progress from 2000 to 2001." *Seventh International Conference on Spoken Language Processing*. 2002.

[7] Minker, Wolfgang, et al. "The SENECA spoken language dialogue system." *Speech Communication* 43.1-2 (2004): 89-102.

[8] Frankes, William B., and Ricardo Baeza-Yates. "Information retrieval: Data structure & algorithms." *PrenticeHall, Englewood cliffs, NJ* (1992).

[9] Kumaran, Giridhar, and James Allan. "Adapting information retrieval systems to user queries." *Information Processing & Management* 44.6 (2008): 1838-1862.

[10]. Williams, Jason D., and Steve Young. "Partially observable Markov decision processes for spoken dialog systems." *Computer Speech & Language* 21.2 (2007): 393-422.

[11]. Young, Steve, et al. "Pomdp-based statistical spoken dialog systems: A review." *Proceedings of the IEEE* 101.5 (2013): 1160-1179.