CS 520 Final Project 5 Solution

December 20, 2018

1 Solutions

1.1 Pick one kind of bot and make a description

Solution:

I pick LieDetectBot as the research object.

In my opinion, a LieDetectBot is supposed to be able to receive speech from a speaker, transform it into consistent sentences or words, and later then make a decision whether one is telling a lie or not. Now I will try to give a formal representation of the problem. Suppose we have a bot notated as B, in given period of time $1:t$, it receives a piece of sound (speech + outside noise) notated as $sound_{1:t}$, it should output a probability P to indicate how possibly the speaker it records just tells a lie or not. Also, it can have a possibly bonus function, that is to keep track of the speaker's eye sight. I add this component because usually people would unnaturally move their eye sight away from the people they are speaking to as they tell a lie. So eye sight may provide some interesting but probably useful proof to help the bot reach the goal. Specifically, it records a series of continuous images $I_{1:t}$, tries to find out where the speaker's eye is and which possible place the speaker is looking at.

To make the bot functions as we expect, several techniques should be utilized. First of all, after receiving the sound from outside, the bot should be able to recognize it into a most probable paragraph of words. In other words, it should be able to find $\arg\max_{words_{1:t}} P(words_{1:t}|sound_{1:t})$. After that, the bot should be able to extract important critical words from this words set, where it needs to apply Natural Language Processing (NLP) techniques. The NLP module designed for the bot would be able to process $words_{1:t}$ obtained from the last stage and extract important components from it. For example, a set of nouns would be captured and be classified into various categories: time points, people, places, etc. Also verbs and adjectives are captured. We kill keep processing these words, but would also go into a different component in the bot: semantic analysis. In this component, based on the previous work, the semantic relationships among words in the same category or in different categories would be explored and, if it is practical, some entities with proper representation of embeddings would be output. For example, if the speaker mentions that John walked on the Main Street on 4:30 at Saturday

and bought a rose, then a probable embedding for the semantic information can be: [“*name*” : “*John*”, “*time*” : “4 : 30”, “*date*” : “*Saturday*”, “*place* : “*MainStreet*”, “*action*” : {“*bought*”, “*walked*”}, “*object*” : “*rose*”]. What's more, it also analyzes some context information surrounding this sentence and provide embedding for further decision making. At this point, our bot has obtained all extracted information and everything has been represented in a form it can ”understand”. At this stage, the bot can function to distinguish if a lie is hidden by checking out its own Knowledge Base as well as its context information, by applying probability inferences (probably using a Bayesian network). For example, if it finds out that the Main Street is under construction during 1:20 pm to 5:00 pm on Saturday, it may be not likely that John could walk on the pave there. And it may also find out in the context information that John is not a very romantic guy, it queries the Knowledge Base and would find out that buying a rose would have a strong connection with the word “romantic”, then it would infer that buying a rose would not be likely to happen on John, so this increases the probability that if the speaker is telling a lie.

Let's turn to another part. The function of this part is to keep track of the eye speaker's eyes. This involves using computer vision technique, depending on which the bot can recognize the speaker's eyes possibly by facing recognition. By analyzing images recorded continuously in time order, denoted as $I_{1:t}$, together with the prior knowledge that where another speaker is positioned, the bot can keep track of the eye sight of the speaker (for example, where the speaker's eyeballs are directed) and see if his eyesight moves to other places frequently. Also, by using Bayesian network, another probability that if the speaker is telling a lie is obtained.

So far the bot have obtained two kinds of probabilities from two resources that if the speaker is telling a lie, then it puts this information into Bayesian network, which has been trained ahead, and would get a final probability P that how possible the speaker has told a lie just then.

There are some issues, including some solvable but important problems and hard-to-take challenges, that the bot have to face up before completely fulfilling its functions. And these issues are listed below.

Voice recognition: in the real world, there can be a lot of noise int the surroundings. Therefore, it is a practical problem for this part to receive the sounds from outside and filter out those unrelated noises. In addition, a large part of these noises may come from other speakers around, which would make the task harder because it may disturb normal recognition process. And even worse, what about in the situation where there are many foreign language speakers? Also, the speaker near the bot can possibly make it record a wrong order of sentences if it is placed in a conversation where multiple people are involved and there is no obvious order who would say first.

Eyesight tracking: if several people are all standing before the bot (of course they may not even know there is a bot), then it can be possible that the bot captures several eye sights meanwhile. What's more, if several people are in a conversation, the bot may make a mistake because any speaker could look at others in turn (which is regarded as being polite for human beings).

Natural language processing: naturally, people may not always speak in a completely well-structured way, which could be the most satisfied situation for the bot to deal with. Besides, the NLP part should be updated all the time to accommodate new nouns, verbs or even adjectives (for example, 'google' is introduced as a verb now after its widely use).

Semantic extraction: for the same problem as in NLP part, if a naturally not well-structured sentence is to be analyzed, then it can be quite difficult to extract semantic information and make them in the right places accurately. Another challenge comes from context extraction. If someone is switching the topics back and forth frequently in a conversation, then context information extracting can be hard as well.

Knowledge Base: in order to guarantee that the "knowledge" the bot obtains from the "library" is accurate and correct, the Knowledge Base should be maintained every certain period time, when some new entity and relationships may be added or some old ones may be deleted. This means that the bot may go offline every a certain period of time to master new knowledge. What's more, since the knowledge base can be quite large, how to optimize query (like accelerating the query speed) over it is also a practical problem.

Probability calculation and interpretability: to reach the final conclusion that how possible a speaker could tell a lie, we need to combine two probabilities generated from two different sources: language and eyesight tracking. We can use Bayesian network to use these two probabilities to get a final probability. Now everything seems very good. However, I would like to still highlight a new concept here, it is proposed by Prof. Judea Pearl, who is also the father of Bayesian network and the Turing Award winner of 2011. As he proposed at the beginning of 2018, interpretability or explainability of all AIs should be emphasized and caught sufficient attention. Actually he has done much research in enhancing interpretability for AI, especially in the area of causal inference. In order to give user a persuasive result of the final probability, some necessary explainations should be attached as well. From this point of view, plus what I have designed for the bot, I think causal inference is a good technique that can be applied on the Bayesian network and increase the credibility of the final result.

1.2 Bonus

1.2.1 Question i)

Solution:

She wore a costume as a super woman, as shown in the photo.

1.2.2 Question ii)

Solution: I simply tried to make a challenge myself here: draw this painting (using paint program on my laptop) merely with my touch pad and my hands:)