CS7637 HW 1

Chenxi Gao
cgao71@gatech.edu

QUESTION 1:

The following is the list that I consider a sandwich:

BLT on white bread; hamburger; turkey and swiss on potato roll; meatball sub; tuna salad on brioche; chicken wrap; chip butty; burrito; grilled cheese; turkey hero; vada pav; cheese quesadilla; veggie burger; egg & cheese biscuit; gyro; patty melt; calzone; sloppy joe.

The following is the list that I do not consider a sandwich:

Ice cream sandwich; ice cream taco; toast; toaster strudel; Klondike bar; buttered biscuit; sushi rolls.

I started with the assumption that BLT on white bread is a sandwich. Therefore, I define the sandwich as **two pieces of bread with something within it**. Hamburger fits well by this definition and is a sandwich so no change. Since potato roll is not bread but I define it as a sandwich, so I generalize the definition: **two pieces of buns with something within it, where bun includes bread and potato roll**. Since meatball sub is a sandwich and has only 1 piece of bread, I generalize the definition: **one or two buns with something within it. Background knowledge: bun includes bread and potato roll**. In a similar manner to how I generalized potato roll, I generalize the definition by extending the background knowledge of bun to **include bread, potato roll, taco shell, tortilla, biscuit.** Then I define ice cream taco not a sandwich by specializing the definition: **something in it must not be cold (below 32 degrees)**. Klondike bar is not a sandwich, so I specialize the definition that **the stuff in it cannot be sweet**. At this moment, I am able to classify sandwiches to the two lists above. The final definition is:

One or two buns with something not cold (below 32 degrees) and not sweet in it. Background knowledge: bun includes bread, potato roll, taco shell, tortilla, biscuit.

Below are two positive and two negative examples:

Burrito has a bun (tortilla), has stuff in it (meat/vegetable), and is neither cold nor sweet. Therefore, it is a sandwich.

grilled cheese has a bun (bread), has stuff in it (cheese), and is neither cold nor sweet. Therefore, it is a sandwich.

Toast does not have stuff in it, so it is not a sandwich.

Sushi does not have bun, so it is not a sandwich. Since they are not sandwich, I do not need to specify the definition.

Although my definition may separate all the candidates above, there might be some sandwich or non-sandwich examples that would make a significant difference to the model. For example, I assume that the stuff in the sandwich is not cold. Imagine a scenario where the stuff in it is sweet but is also considered a sandwich (e.g. bread with strawberry jam). I would need to generalize the definition again to either make jam as an exception or modify the definition, stuff in it cannot be sweet, to any item that cannot be considered a snack. The background knowledge of snack is bar.

Next, I will use a classification approach to define a sandwich. The followings are the parameters and the answers to it that makes the object we are evaluating a sandwich.

1. Is it food?
2. Does it contain staple food (wheat, rice, potatoes, etc.)?
3. Does it contain meat/vegetable?
4. Is it expensive?
5. Is it sweet?
6. Is it served warm?
7. Is there anything inside the staple food?

The following are the sandwiches candidates and their parameters

BLT on white bread: YYYNNNY

Hamburger: YYYNNYY

Turkey and swiss on potato roll: YYYNNYY

Chicken wrap: YYYNNYY

Chip butty: YYYNNYY

Burrito: YYYNNYY

Ice cream sandwich: YYNNYNN

Grilled cheese: YYNNNYY

Turkey hero: YYYNNYY

Ice cream taco: YYNNYNN

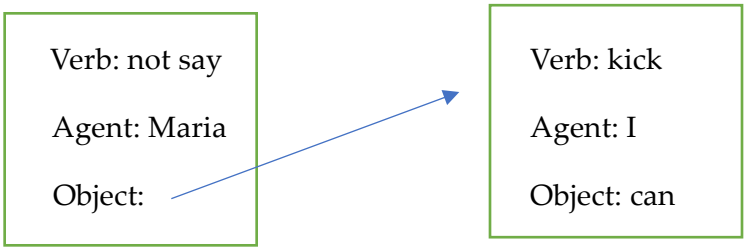
Based on the above parameters, I conclude that sandwiches must be food, containing staple food, not expensive, not sweet, and that there is something inside the staple food. Sandwiches do not have to contain meat/vegetable, neither do it need to be served warm.

Through incremental concept learning, I consider hotdog as a sandwich because it has a bun (bread), has stuff in it (sausage), not cold nor sweet. By using classification, hotdog is a sandwich because it is food, contains staple food, not expensive and not sweet. By using case-based reasoning, I would say that hotdog is most similar to a meatball sub. The bread and their shape are similar; they both contain meat in the center. The only difference is method of cooking the meat inside.

Therefore, I think hotdog is a sandwich.

QUESTION 2:

Below is the frame representation of this sentence: "Maria didn't say I kicked the can."



The AI agent would first locate the first verb and noun in the sentence, “said”, and “Maria”. Because “didn’t” is before “said”, it is served as a negation of the verb, the AI agent would updated “say” to “not say”. Because “say” has many meanings and functions in the sentence, the AI agent would look at the rest of the sentence to know its meaning. It sees another sentence after “say”, so “say” is the beginning of a clause. Otherwise, it is a run-on sentence, and it would not be grammatically correct. If I remember it correctly, a clause is served as an object in the first sentence, “Maria didn’t say”, while it is referring to another sentence, “I kicked the can.” Therefore, the AI agent would separate the sentence to two frames; the frame on the right (clause) is the object of the frame on the left. So far, the AI agent understands that the agent, “Maria”, did “not say” something which is another sentence.

Now the AI agent would try to understand the meaning of the second sentence. It first takes out the first verb and noun, “kick” and “I”, then locates the object, “can”. “Can” has many meanings, but there is only meaning that can be used as a noun, which is a cylindric container usually made of aluminum. Therefore, the AI agent does not need to apply constraints or background knowledge. Now the agent would fully understand the sentence.

When the sentence had a different emphasis placed on it, it usually had a slightly different meaning. The following are the meaning of the same sentence w.r.t. the emphasis.

Maria: it implies that Maria (not someone else) is the person who didn’t say I kicked the can.

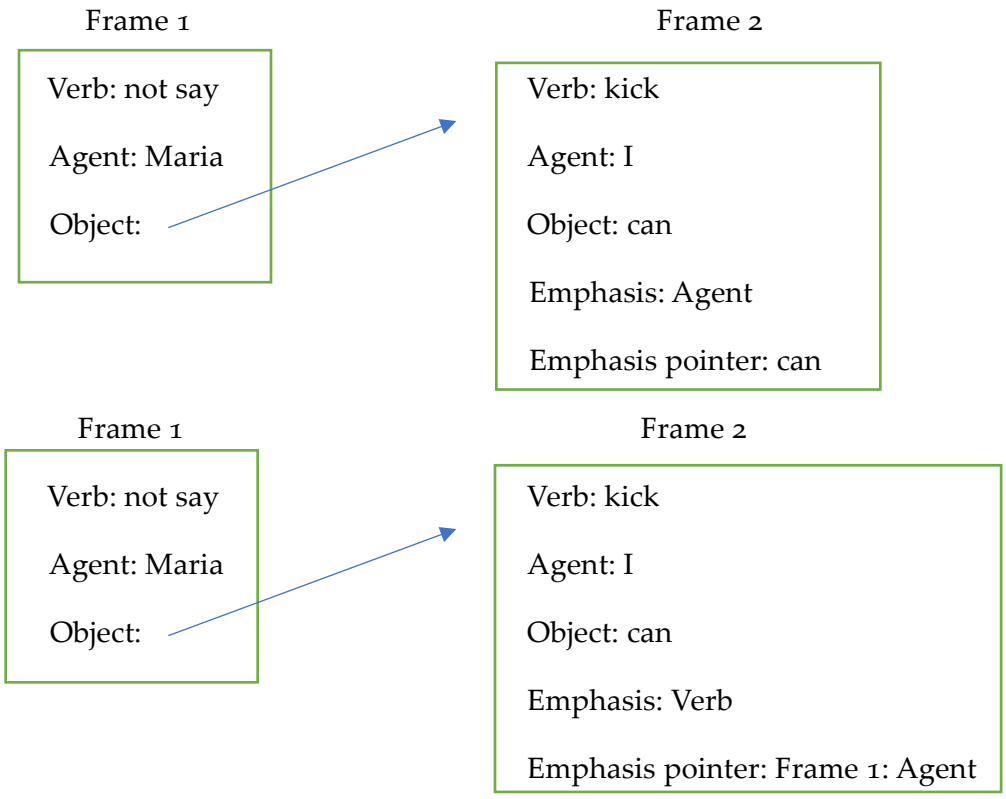
I: it implies that while someone said I kicked the can, Maria wasn’t the one who said that.

Kicked: it implies while Maria said I did something to the can, she didn’t say I kicked it.

Other emphasis does not significantly change the meaning of the sentence.

I guess the AI agent would first pick an emphasis and analyze it based on the background knowledge of the emphasis of the given word and its type. A constraint of emphasis similar to the prepositional constraints would be provided. The constraints would consist of the type of the emphasized word (eg.

Noun, verb, adj., etc.), its location in the sentence (e.g. before a noun, after a clause, etc.), and its corresponding pointer of emphasis. In the below example, a pointer to can means "I did something to the can", and a pointer to Agent in frame 1 implies "someone said".



I think an AI agent can infer whether this sentence whether this sentence is to be taken literally or figuratively by taking out elements, a word or short phrases, in the sentence that corresponds to a literal or figurative meaning, then record it in the frame. There can be a chart of literal word(s) and figurative word(s). For example, when "can" is used as a noun, it is literal. In contrast, the phrase "putting work off" is figurative.

QUESTION 3

Below is a summary of The Toronto Declaration:

Preamble:

This paper illustrates concerns of the machine learning system and artificial intelligence that may cause significant harm to human rights, including "protect individuals from discrimination, to promote inclusion, diversity and equity, and to safeguard equality." (p. 1) The technologies should not harm but indeed protect human rights. The paper primarily focuses on the right to equality and non-discrimination.

Using the framework of international human rights law:

This paper reassures that the states has the obligation to promote and reinforcement human rights, and private sectors should respect human rights. Government, academia, or companies should eliminate the potential risks of those technologies that might violate human rights.

This Declaration focuses on the right to equality and non-discrimination, a critical principle that underpins all human rights. (p. 5)

Governments have responsibilities to prevent discrimination, and a system should be interrogated when discrimination occurs. New and existing technologies should prevent and mitigate against discrimination risks.

Duties of states: human rights obligations:

The paper emphasizes the responsibility of states to promote, protect, respect, and fulfill human rights, and prevent the risk of machine learning systems to violate human rights. States must do the followings steps to mitigate and reduce the harms of discrimination from machine learning in public sector systems: identify risks, ensure transparency and accountability, enforce oversight. States have a duty to take proactive measures to eliminate discrimination and protect human rights.

Responsibilities of private sector actors: human rights due diligence:

Private sector actors have a responsibility to respect human rights. Private sectors actors should follow a human right framework in their systems. There are three core steps to the process of human rights due diligence:

1. Identify potential discriminatory outcomes

2. Take effective action to prevent and mitigate discrimination and track responses
3. Be transparent about efforts to identify, prevent and mitigate against discrimination in machine learning systems. (p. 7)

The right to an effective remedy:

Victims of human rights violations or abuses must have access to prompt and effective remedies, and those responsible for the violations must be held to account. The use of machine learning systems where people's rights are at stake take action to ensure individuals and groups have access to meaningful, effective remedy and redress and pose challenges for ensuring the right to remedy. Additionally, states should supervise the public sectors, evaluate risks of the machine learning systems, give legal guidelines, and provide effective remedies to victims of discriminatory harms linked to machine learning systems.

There are trade-offs inherent to the declaration. While the declaration ensures human rights when the machine learning systems and artificial intelligence technologies completely comply with the declaration, there are harms to the technology firms, the government, and thus might have cause an economic decay in a period of time.

1. In order to comply with the new guidelines, the private sector actors must fully re-evaluate their machine learning models and adopt changes to the them if they violate the human rights. If the changes cannot be made, the private sector actors must discard the existing models and additionally provide remedy to victims. Many companies may face bankruptcy if this happens.
2. The declaration is not welcome to the innovation because the declaration states that states and private sector actors should prevent potential risks of those technologies that might violate human rights. Some innovation and ideas may be stifled because of such risks.
3. States have a duty to proactive measures to eliminate discrimination and protect human rights. In order for the states to identify risks, the officials must closely inspect every machine learning models. This is a very time-consuming task and would be impossible to be completed in a short-term.

4. Because of all the reasons above, there might be a decrease in economic growth, especially for the technology firms.

There are trade-offs if the declaration were totally discarded:

1. Clearly, some of the human rights would be violated. Private sector actors would ignore the consideration of human right when they develop their products. This would raise more human rights concerns.
2. Data privacy (in the last homework) is a good example of human rights that the companies should be protected. If the declaration were totally discarded, companies may maliciously use and collect data from the visitors to their websites without the authorization of the users.

I agree with the declaration that human rights should always be in consideration when the private sector actors develop the machine learning models and when the states inspect the machine learning models. Clearly, machine learning models have to obey the human rights. States and privacy sector actors must provide remedy for those who are harmed. However, I think there are some improvements to actually enforce the declaration in practice on today's machine learning models. First, while the entire declaration is proclaiming to protect, respect, and fulfill human rights, it is still a bit vague about what exactly is human rights in machine learning models, and what kind of actions is a violation of human rights in machine learning models and what kinds are acceptable. The declaration lacks some concrete examples of application of human rights in the field of machine learning models and artificial intelligence. Second, I think the way of fully implementing the declaration where states and private sector actors work together to eliminate risks of human right violation is inefficient given that there are millions of existing machine learning models. This would potentially decrease the innovations or opportunities of emerging technologies. My suggestion is to let the public alongside with states and privacy sector actors decide if a model is acceptable. privacy sector actors should release full documentation of the new model or existing model, including the way it may or may not violate human rights. The model is acceptable if there are no major dissents to the model. For example, if the users are willing to sacrifice some privacy to exchange a better and more reliable service, then the model should not be considered a risk.

QUESTION 4

This is a very interesting question that worth to spend time thinking about it.

I believe that in the end Ethan, the mountain climber, is still the same person as before because his mind was not replaced although all the other parts of his body are replaced. In contrast, Sofia and Akhila are not the same person as before because they lack the presence of mind.

First, replacing a body part except the brain will not make lose your identity. There are thousands of people who have to go through surgery. It is tragic that many of them lose limbs, sensations, or internal organs like gall bladder. Some of them may choose to replace with a robotic limb; some of them may replace a liver or kidney. I know tragedies happens and I am heartbreaking when writing these. But the undeniable the fact is: they did not lose their identities because they have exactly the same memory and experience as before. Sofia and Akhila have an empowered brain. Although they have a portion of their memory that is identical to their past memory, they have a large portion of their memory that does not belong to them. The portion of the memory that does not belong to them makes them to be not the same person as before. Imagine if someone infuse five people's memory and characterization into his/her brain, do you think he/she is the same person who has the same identity as before? Would you have the same interests or strength as before? Therefore, I believe that you are not changed if you lose part of your memory (e.g. Amnesia), but you changed if you have your brain replaced or have something that should not belong to you inside your brain.

Second, I want to talk about the importance of the presence of human brain. The scientists define brain death to be legal death:

> "Brain death (also known as brain stem death) is when a person on an artificial life support machine no longer has any brain functions. This means they will not regain consciousness or be able to breathe without support. A person who's brain dead is legally confirmed as dead. They have no chance of recovery because their body is unable to survive without artificial life support." (NHS)

In other words, by definition you are still alive if you only have your brain present and everything else missing because only brain death is considered legal

death. While the technology of only keeping your brain alive is not there yet, the world where Ethan, Sofia, and Akhila live have a more advanced technology. By fitting the definition of brain death in these scenarios, Ethan is the only person who is alive. It does not matter whether he replaced his body by parts or his entire body at once, He is Ethan and he is still alive. In contrast, Sofia and Akhila are dead already by definition because their brain was dead. It does not matter whether Akhila was built from a backup after a tragic accident because her brain is dead.

Finally, I will briefly illustrate the ideas of personhood and identity in phycological and philosophical view to further supports my argument. What is it to be a person, and what is it to be a nonperson? The most common and mostly accepted answer is "the most common answer is that to be a person at a time is to have certain special mental properties then" (e.g. Baker 2000: ch. 3). I also agree to some proposals of a less direct connection between personhood and mental properties: "for example that to be a person is be capable of acquiring those properties" (Chisholm 1976: 136f.), "or to belong to a kind whose members typically have them when healthy and mature" (Wiggins 1980: ch. 6). Nevertheless,

I acknowledge that there are multiple definitions of personhood and identity, but I believe that existence of mental properties or the presence of mind is most crucial to define a person. A person is no more the same person if he/she has his/her mind replaced.

REFERENCES

NHS Choices, NHS, www.nhs.uk/conditions/brain-death/.

"The Toronto Declaration • Toronto Declaration." Toronto Declaration, 29 Nov. 2019, www.torontodeclaration.org/.

We are material things "constituted by" organisms: a person made of the same matter as a certain animal, but they are different things because what it takes for them to persist is different (Baker 2000, Johnston 2007, Shoemaker 2011).

We are temporal parts of animals: each of us stands to an organism as your childhood stands to your life as a whole (Lewis 1976).

Wiggins, D., 1980, Sameness and Substance, Oxford: Blackwell.