CS7637 Fall 2018 Assignment 2

Zheng Fu
zfu66@gatech.edu

Question 1

Positive example list: BLT on white bread, veggie burger, egg & cheese biscuit, patty melt, hamburger, sloppy joe, meatball sub, tuna salad on brioche, chip butty, grilled cheese, turkey hero, vada pav. Negative example list: turkey and swiss on potato roll, toaster strudel, Klondike bar, buttered biscuit, gyro, sushi rolls, calzone, ice cream taco, ice cream sandwich, chicken wrap, burrito, toast, Klondike bar.

According to the New Webster Third International Dictionary, a sandwich was defined as[1]:

> "two thin pieces of bread, usually buttered, with a thin layer (as of meat, cheese, or savory mixture) spread between them."

So based on the above definition, I designed a concept diagram of sandwich as Figure 1A. Then I used the algorithm of incremental concept learning introduced in the lecture (see Figure 2) to step through the learning process along with the following four examples:

Positive example 1: BLT on white bread fits the current definition of the sandwich concept portrayed in Figure 1A. So I do nothing.

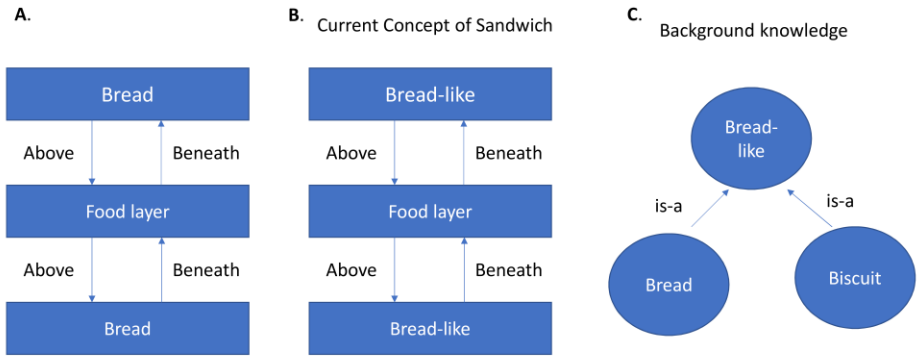

Figure 1. Concept diagrams of sandwich

Positive example 2: Egg & cheese biscuit does not fit the current definition of the sandwich concept portrayed in Figure 1A. Because it uses biscuit instead of bread, I added background knowledge to the current concept and modified it accordingly (see Figure 1B and 1C).

Negative example 1: Toast does not fit the current definition of the sandwich in Figure 1. So I do nothing.

Negative example 2: Burrito fits the current definition of sandwich in Figure 1. However the food layer is wrapped within the bread-like items instead of between them, and I added forbid-link[2] as shown in Figure 3.

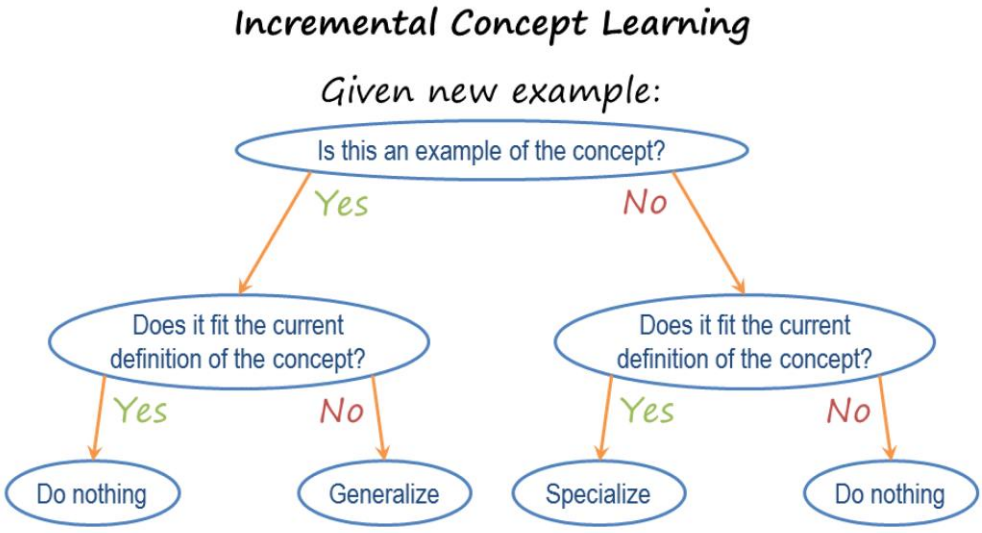

Figure 2. The algorithm of incremental concept learning.

Moreover some sandwiches that I did not include in the learning process would make a significant difference to the model. For instance, ice cream taco and ice cream sandwich do fit the model in Figure 3, yet in human common-sense a sandwich is a meal with at least room temperature. So both ice cream taco and ice cream sandwich are not sandwich, and such concepts as 'temperature', 'function' etc. are not included in the current model.

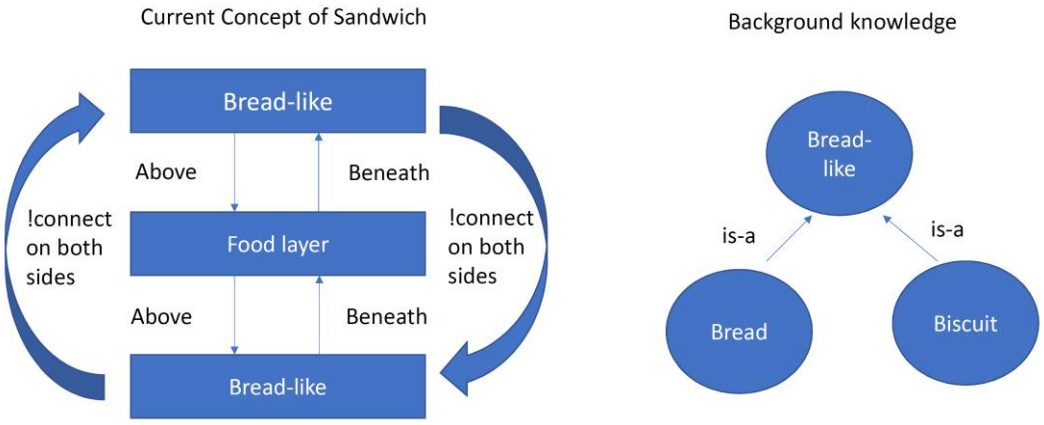

Figure 3. A concept of sandwich with forbid-link.

Classification

Here I use the following three features to classify different kinds of sandwich:

1) Has two slices of bread-like? 2) Has a split roll? 3)Has meat as filling?

And I defined the values for those parameters for the following six sandwiches:

Hamburger: 1) Yes; 2) No; 3) Yes; Veggie burger: 1) Yes; 2) No; 3) No; BLT on white bread: 1) Yes; 2) No; 3) Yes; Chip butty: 1) Yes; 2) No; 3) No; Meatball sub: 1) No; 2) Yes; 3) Yes; Vada pav: 1) No; 2) No; 3) No.

So based on these values, an abstracted classification was made as Figure 4.

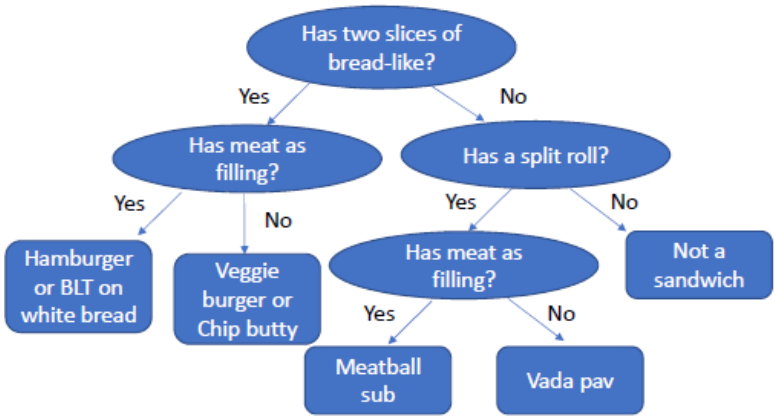

Figure 4. An abstracted classification of six sandwiches.

Is a hot dog a sandwich?

1. Incremental concept learning: A hot dog fits the model shown in Figure 3 so it is a sandwich.
2. Classification: By using the abstracted classification model in Figure 4, a hot dog is a meatball sub-like sandwich.
3. Case-based reasoning: Again in the retrieving step meatball sub could be drawn as most "similar" to a hot dog for adaptation, evaluation and storage.

Question 2

The AI agent could use the principles of Understanding to make sense of sentence "I never said Amy planted that seed" as shown in Figure 5A. Here I used a thematic role frame[2] to help AI agent to understand this sentence, in which the agent is 'I', the co-agent is 'Amy', the verb 'said' belongs to primitive 'Speak', the verb 'planted' belongs to primitive 'Move-object' and the object is 'that seed'.

Different emphases will alter the meaning of this sentence, and we could change the values in the thematic role frame to help AI agent to understand these different

meanings. Here are few different interpretations of the sentence with different emphases:

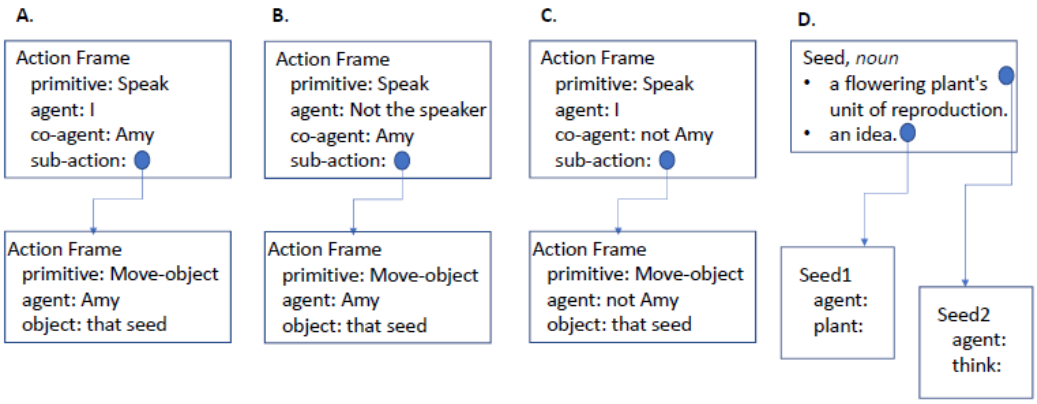

Figure 5. Thematic role frames for AI to understand sentence "I never said Amy planted that seed".

1. ***I*** never said Amy planted that seed. That means 'Amy planted that seed' was spoken, yet it was not by the current speaker. So the value of 'Agent' was changed from 'I' to 'Not the speaker' (see Figure 5B).
2. I never ***said*** Amy planted that seed. That means the agent may imply 'Amy planted that seed' through other way instead of speaking.
3. I never said ***Amy*** planted that seed. That means the seed was planted, but not by Amy. So the value of 'co-agent' was changed from 'Amy' to 'Not Amy' (see Figure 5C).
4. I never said Amy planted ***that*** seed. That means a seed was planted by Amy, yet it was not the given seed that the speaker mentioned.

In order to help AI agent to decide if the sentence describes a literal plant seed or a figurative seed representing an idea, I mimic the lecture example of clarifying ambiguous verb 'take' and introduce background knowledge to AI agent as Figure 5D. First, the background knowledge tells AI agent that the word 'seed' has two meanings. Second, the background knowledge tells AI agent when 'seed' means a literal plant seed, it will typically have an agent with verb 'plant'. Also the background knowledge

tells AI agent when 'seed' means a figurative seed representing an idea, it will have an agent with verb 'think'.

Question 3

Preamble

With the fast development of machine learning technologies, the potential discriminative risk of using machine learning system may increase as well. The primary goal of this declaration is to discuss and investigate the harm of machine learning technologies to human rights, especially the right of equity[3].

Using the framework of international human rights law

Private sector actors who create machine learning system should ensure that system is designed and applied in a proper way to protect and respect human rights, and they are responsible to prevent the occurrence of any kinds of discriminations. If such preventions were failed, the issues must be addressed immediately and the system should be interrogated. Government measures, academia, legal and civil law experts should be able to adequately participate the discussion of conceiving, developing and deploying machine learning system, so as to identify and address the potential harm to human rights of all individuals and groups[3].

Duties of states: human rights obligations

Using machine learning technologies in public sector systems or through public-private relationships may cause high risk of discriminatory or human rights-harming outcomes. Therefore it is states' responsibilities to make sure the current evaluation approaches to protect and promote the human rights are updated to consider and address the potential discrimination posed by machine learning technologies. By doing so[3]:

1. Prior to deploying and using machine learning system in public context, the states should ensure the harms of discrimination are fully investigated.
2. Applying machine learning system in public sector must be explainable and intelligible.
3. Mechanism of independent oversight should be established and all machine learning made or supported decision should meet the international accepted standards for due process.

Human rights due diligence

Human rights due diligence was defined as a process that the private sector actors take steps to ensure human rights abuse did not emerge. Such process has three key steps[3]:

1. While developing and deploying new machine learning technologies, the private sector actors should evaluate the risk that the system will lead to direct discriminations as well as differential treatment.
2. Inadvertent discrimination should be prevented while designing and training machine learning models. Regular, ongoing quality assurances checks and real-time auditing is required for eliminating human rights harm in the life-cycle of machine learning system implementation.
3. Risk and instance of discrimination should be disclosed, and mechanisms should be established to inform the negative impact of such disclosed information on decision making and outcomes.

The right of an effective remedy

Meaningful and effective remedy and redress should be accessible for individuals and groups and provided by companies and private sector actors who develop and deploy machine learning technologies. Meanwhile the states should make a clear line in terms of which parties and individuals are legally responsible for the outcomes and decisions made by using machine learning technologies[3].

Tradeoff inherent to the declaration

On one hand, the declaration clearly deemed that "avoid using 'black box systems' that cannot be subjected to meaningful standards of accountability and transparency"[3]. And innovations or opportunities rooted in deep learning models may be lost due to this declaration, since the outcomes of such models highly rely on optimizing weighting factors within complicated neural networks. It is hard for even experienced machine learning engineers to explain how the model was trained and the relationship between the training process and the outcomes.

On the other hand, if the declaration were discarded, marginalised groups who are vulnerable to discrimination may be exposed to the risk of human rights harm. Because the currently machine learning technologies could only conclude "correlation" relationships instead of 'inference", and the outcomes may be interpreted as causation for certain marginalised groups due to the insufficient datasets.

My opinions

As far as I am concerned, I partially agree with the sections of the declaration in terms of protecting human rights and privacy. Yet I did not support its attitudes toward the machine learning technologies. I always believe the values of technologies are neutral and if we feed the machine learning system non-representative or biased dataset, we will harvest discriminative results for sure. Thus I believe the declaration should pay more attention on making rules of collecting and sharing unbiased datasets, especially for those tiny companies with limited budgets.

Question 4

Positive light:

100% Literacy: Or What if AI Could Fix Our Broken Educational System?[4]

Summary:

We encounter an increasing challenge in our daily life and we are not sufficiently aware of: more than one one-third of children in the United States lack fundamental reading skills[4]. Yet using Learning Ovations, an AI-based professional support system, could fix it and produce even better outcomes.

Description:

Learning Ovations could analyze tremendous student assessment data and determine an individualistic approach to each student in the class. It means each kid could have his/her own personalized activities, material, and instruction that enable him/her to read at the appropriate grade level. More than that, Learning Ovations is able to map the individualistic curriculum to the existed school curriculum, therefore teachers could respond to their students individually as well.

Evaluation:

I did not believe the authors oversold the significance of Learning Ovations, since it cited few randomized trials outcomes conducted by Department of Education (E.D.) and the National Institutes of Health (N.I.H.) to approve their assertions. Kids with personalized instructions have 94% literacy achievement rate, whereas such rate in the control groups are 78%[5].

Negative light:

Artificial Intelligence: The Robots Are Now Hiring.[6]

Summary:

A revolution occurs in the hiring process of more and more companies. AI and machine learning technologies were heavily applied in evaluating potential employees.

Description:

The authors used DeepSense as an example to portray the results to audiences unfamiliar with AI. DeepSense is a company focuses on searching people's social media accounts and extracting their personality traits. Such information will be integrated and introduced to a scientifically based personality test. Traditionally employers accessed candidates based on their skill sets, now with the help of AI, hiring managers is prone to make decisions based on people's personality.

Evaluation:

The authors quoted Ifeoma Ajunwa, a sociology and law professor of Cornell, to speak of their concerns about the fairness, transparency and accuracy in the hiring process while using AI as an assistance. Overall I believe the AI development was portrayed fairly, and the missing part is further discussion on protecting human privacy while using such technologies.

References

1. 5 Ways to Define a Sandwich, According to the Law. http://mentalfloss.com/article/501011/5-ways-define-sandwich-according-law.
2. Winston, P. H. (1993). Artificial intelligence (3rd ed.). Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.
3. The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems. https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/.
4. 100% Literacy: Or What if AI Could Fix Our Broken Educational System? https://www.forbes.com/sites/cognitiveworld/2018/10/08/100-literacy-or-what-if-ai-could-fix-our-broken-educational-system/#666b1ecb77b4.
5. http://www.united2read.org/learningovations/.

6. Artificial Intelligence: The Robots Are Now Hiring. https://www.wsj.com/articles/artificial-intelligence-the-robots-are-now-hiring-moving-upstream-1537435820.