Neural Computation (VS 265), Problem Set 5 - Holographic Reduced Representation

Due date: November 7, 3:30pm

Fall 2024

General guidelines:

- We are grading problem sets anonymously. **Include your student ID in the submission, but do not include your name.**
- You may work in small groups of 2-3. Note that you are responsible for writing up and submitting your submission individually.
- You are expected to attach any code you used for this assignment but will be evaluated primarily on the writeup.
- If you are including animations as part of your submission, please attach these files separately in your submission (it should not be necessary to run any code to view your animations).

Estimating Analogical Similarity

According to Douglas Hofstadter, our ability to reason by analogy lies at the core of cognition. How does the brain do this? A central element of analogical reasoning is the ability to apply a set of relationships in one domain to another domain. As Tony Plate showed, we can represent such relationships via *binding* of high-dimensional vectors representing roles and fillers, and we can represent concepts, such as might be expressed in a sentence, as a superposition of bindings. Here you will investigate the ability of this framework to capture analogical similarity of sentences via the inner-product between the vectors representing them.

We will be following the experiment as outlined in Chapter 6 of Tony Plate's thesis, available on the course website (or click here).

Here are the steps you should complete:

1. Generate these base vectors and token vectors (Plate uses n=2048, and angle brackets denote unit normalization):

Base vectors		Token vectors
person	**bite**	**jane** = $\langle$**person** + **id**$_{jane}\rangle$
dog	**flee**	**john** = $\langle$**person** + **id**$_{john}\rangle$
cat	**cause**	**fred** = $\langle$**person** + **id**$_{fred}\rangle$
mouse	**stroke**	**spot** = $\langle$**dog** + **id**$_{spot}\rangle$
	lick	**fido** = $\langle$**dog** + **id**$_{fido}\rangle$
bite$_{agt}$	**bite**$_{obj}$	**rover** = $\langle$**dog** + **id**$_{rover}\rangle$
flee$_{agt}$	**flee**$_{from}$	**felix** = $\langle$**cat** + **id**$_{felix}\rangle$
cause$_{antc}$	**cause**$_{cnsq}$	**mort** = $\langle$**mouse** + **id**$_{mort}\rangle$
stroke$_{agt}$	**stroke**$_{obj}$	
lick$_{agt}$	**lick**$_{obj}$	

All base and id vectors are randomly chosen with elements independently distributed as $\mathcal{N}(0, 1/n)$.

2. Generate vectors for the following probe and possible analogical episodes:

Probe: Spot bit Jane, causing Jane to flee from Spot.

Episodes in long-term memory:

E1:	Fido bit John, causing John to flee from Fido.
E2:	Fred bit Rover, causing Rover to flee from Fred.
E3:	Felix bit Mort, causing Mort to flee from Felix.
E4:	Mort bit Felix, causing Felix to flee from Mort.
E5:	Rover bit Fred, causing Rover to flee from Fred.
E6:	John fled from Fido, causing Fido to bite John.
E7:	Mort bit Felix, causing Mort to flee from Felix.
E8:	Mort fled from Felix, causing Felix to bite Mort.
E9:	Fido bit John, John fled from Fido.
E10:	Fred stroked Rover, causing Rover to lick Fred.
E11:	Fred stroked Rover, Rover licked Fred.

The encoding should follow the same structure as in Plate's thesis (note that this example is for "P: Spot bit Jane..." and you will need to modify it for each of the episodes):

$$\mathbf{P}_{bite} = \langle \mathbf{bite} + \mathbf{bite}_{agt} \circledast \mathbf{spot} + \mathbf{bite}_{obj} \circledast \mathbf{jane} \rangle$$
$$\mathbf{P}_{flee} = \langle \mathbf{flee} + \mathbf{flee}_{agt} \circledast \mathbf{jane} + \mathbf{flee}_{from} \circledast \mathbf{spot} \rangle$$
$$\mathbf{P}_{objects} = \langle \mathbf{jane} + \mathbf{spot} \rangle$$
$$\mathbf{P} = \langle \mathbf{cause} + \mathbf{P}_{objects} + \mathbf{P}_{bite} + \mathbf{P}_{flee} + \mathbf{cause}_{antc} \circledast \mathbf{P}_{bite} + \mathbf{cause}_{cnsq} \circledast \mathbf{P}_{flee} \rangle$$

3. Compute the dot products between the target episode P and base episodes E1-11. Run multiple simulations (i.e., generate different random vectors) to get standard deviations of these results. Compare these results to those in Plate's thesis. Do you get similar results?

P: Spot bit Jane, causing Jane to flee from Spot.

Episodes in long-term memory:		Aspects of similarity: OA	FOR	HOR	RFB	HOS	OLI	Type	Dot-products: Avg	Sd
E1:	Fido bit John, causing John to flee from Fido.	✓	✓	✓	✓	✓	✓	LS	0.70	0.016
E2:	Fred bit Rover, causing Rover to flee from Fred.	✓	✓	✓	×	✓	✓	AN^{cm}	0.47	0.022
E3:	Felix bit Mort, causing Mort to flee from Felix.	×	✓	✓	×	✓	✓	AN_1	0.39	0.024
E4:	Mort bit Felix, causing Felix to flee from Mort.	×	✓	✓	×	✓	✓	AN_2	0.39	0.024
E5:	Rover bit Fred, causing Rover to flee from Fred.	✓	✓	✓	$\frac{1}{2}$	✓	×	$SS^{\times I}$	0.58	0.019
E6:	John fled from Fido, causing Fido to bite John.	✓	✓	✓	✓	×	✓	$SS^{\times H}$	0.47	0.018
E7:	Mort bit Felix, causing Mort to flee from Felix.	×	✓	✓	×	✓	×	$FA^{\times I}$	0.39	0.024
E8:	Mort fled from Felix, causing Felix to bite Mort.	×	✓	✓	×	×	✓	$FA^{\times H}$	0.28	0.025
E9:	Fido bit John, John fled from Fido.	✓	✓	×	✓	×	✓	SS^{-H}	0.43	0.019
E10:	Fred stroked Rover, causing Rover to lick Fred.	✓	×	✓	×	×	×	OO_1	0.25	0.024
E11:	Fred stroked Rover, Rover licked Fred.	✓	×	×	×	×	×	OO_2	0.12	0.023

4. Comment on the relative similarities (dot product) between P and different episodes. Is the dot product providing reasonable estimates of analogical similarity? (You may find it helpful to draw upon the "Aspects of similarity" table, and Plate's discussion, in your response.) Why does the similarity in the obtained results reflect the types of analogical similarity in the episodes?