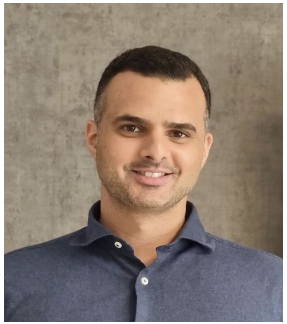
Life is a Circus and We are the Clowns : Automatically Finding Analogies between Situations and Processes



Oren Sultan

@oren_sultan

oren.sultan@mail.huji.ac.il

Prof. Dafna Shahaf

@HyadataLab

dshahaf@cs.huji.ac.il

Analogies in **human cognition**

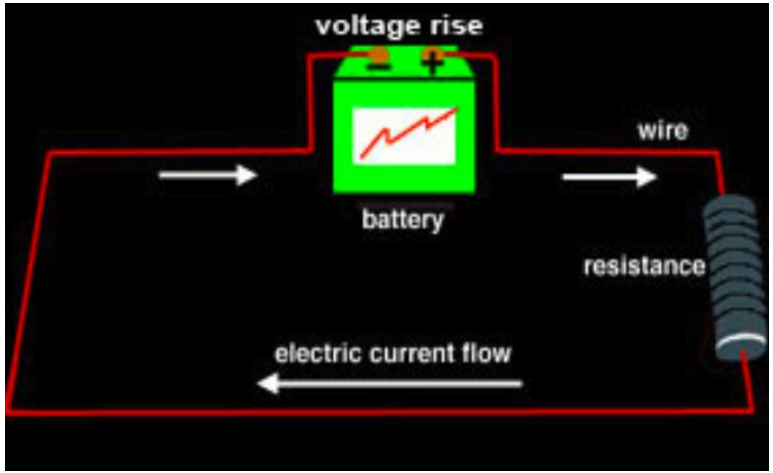
- Analogy-making is a **central** part of **Human Cognition** *(Minsky, 1988; Hofstadter and Sander, 2013; Holyoak, 1984)*
  - Abstract information, adapt to novel situations in terms of familiar ones

driving on the left-hand side of the road in the UK

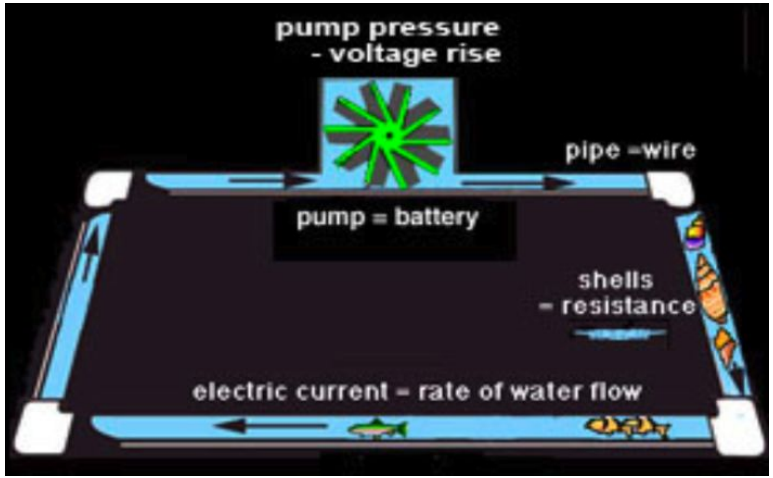
Analogies in **human cognition**

- Analogies play an important role in many areas
  - **Education** – help a teacher to explain a complex subject
  - Politics
  - etc

Electrical circuit

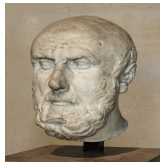


Water pump

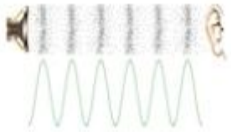


Analogies in **human cognition**

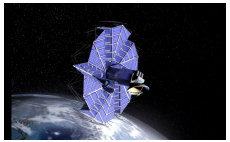
- Analogy-making in innovation
  - Many inventions throughout history are thanks to analogies



[Chrysippus]



[NASA]



[Odon device]

Analogies in **Artificial Intelligence (AI)**

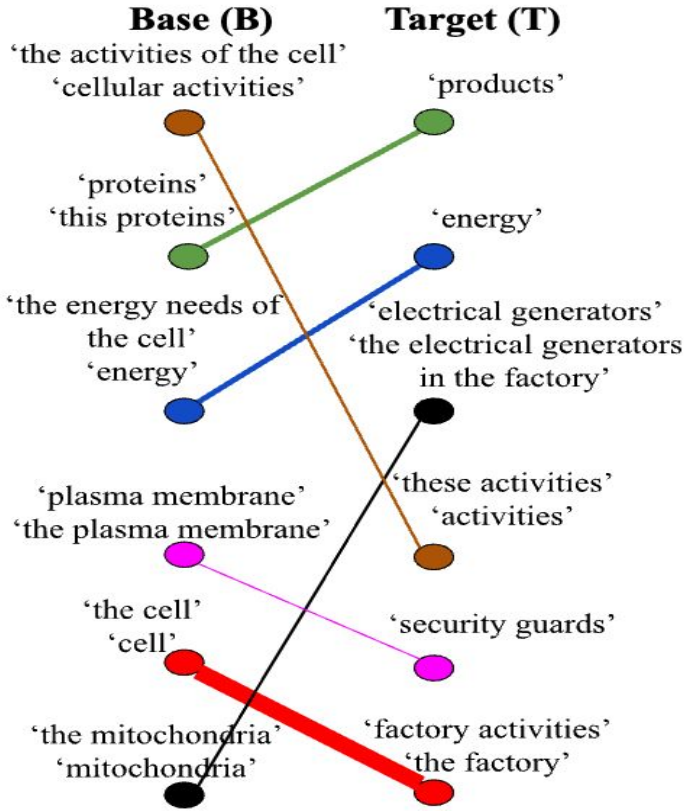
- Analogies in **Natural Language Processing** (**NLP**)
  - Most works focused on word analogies - “a to b is like c to d” *(Mikolov, 2013)*

- **Our focus**: Analogies between **situations** and **processes**
  - **Input**: two domains (e.g., how the heart works / how a pump works).
  - **Goal**: map objects from **base** to **target** according to **relational structure** rather than **object attributes**.

Our work: we tackle a more **realistic** setting – **analogies** between **natural language procedural texts** describing **situations or processes**

Base: Animal Cell
The plasma membrane encloses the animal cell. It controls the movement of materials into and out of the cell. The Nucleus controls the activities of the cell. These cellular activities require energy. The Mitochondria extract energy from food molecules to provide the energy needs of the cell. Animal cells must also synthesize a variety of proteins and other organic molecules necessary for growth and repair. Ribosomes produce these proteins. The cell may use these proteins or move them out of the cell for use in other cells. To move organic molecules, the cell contains a complex system of membranes that create channels within the cell. This system of membranes is called the endoplasmic reticulum.

Target: Factory
Security guards monitor the doors of the factory. They control the movement of people into and out of the factory. Factory activities may be coordinated by a control center. These activities require energy. The electrical generators in the factory provide energy. The factory synthesizes products from raw materials using machines. The factory has hallways to move products through it.



Problem Formulation

Entities: Let $\mathcal{B} = \{b_1, ..., b_n\}, \mathcal{T} = \{t_1, ..., t_m\}$ – entities in the domains (**nouns**).

Relations: Let $\mathcal{R}$ – set of relations – a set of **ordered** entity pairs.

- We focus on **verbs**. (e.g, *"mitochondria* **provides** *energy"*)
- Let $\mathcal{R}(e_1, e_2) \subseteq 2^{\mathcal{R}}$ – set of relations between two entities.

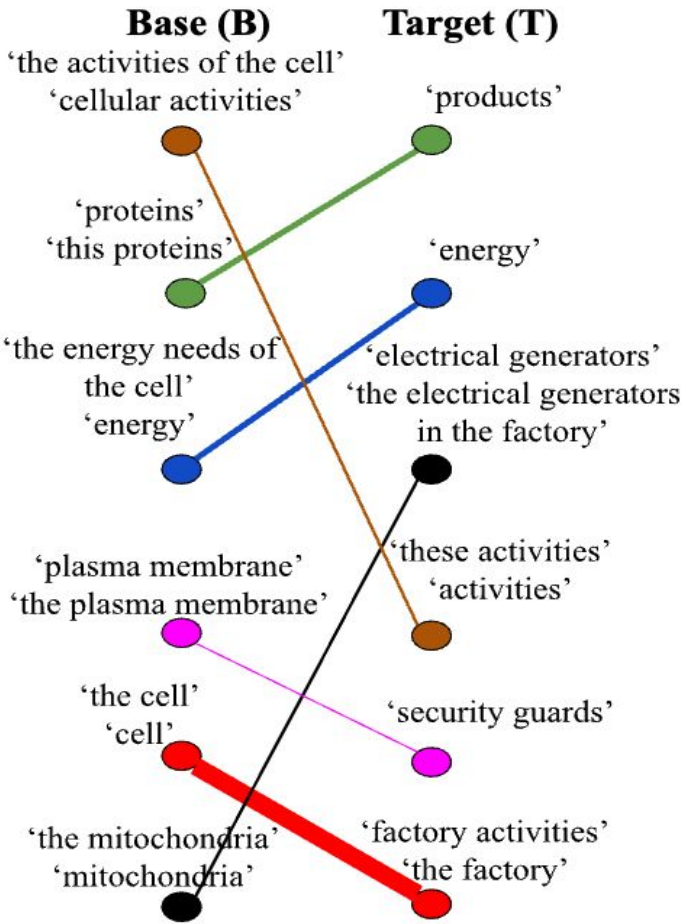
Similarity: Let $sim : 2^{\mathcal{R}} \times 2^{\mathcal{R}} \rightarrow [0, \infty)$– similarity metric between two sets of relations. High **Similarity** ↔ two sets **share many distinct** relations.

$$sim^*(b_i, b_j, t_k, t_l) = sim(\mathcal{R}(b_i, b_j), \mathcal{R}(t_k, t_l)) + sim(\mathcal{R}(b_j, b_i), \mathcal{R}(t_l, t_k)) \quad (1)$$

Objective: find a **consistent mapping** function $\mathcal{M} : \mathcal{B} \rightarrow \mathcal{T} \cup \perp$

- We look for a mapping that maximizes the **relational similarity** between mapped entities:

$$\mathcal{M}^* = \arg\max_{\mathcal{M}} \sum_{\substack{j \in [1, n-1] \\ i \in [j+1, n]}} sim^*(b_j, b_i, \mathcal{M}(b_j), \mathcal{M}(b_i)) \quad (2)$$

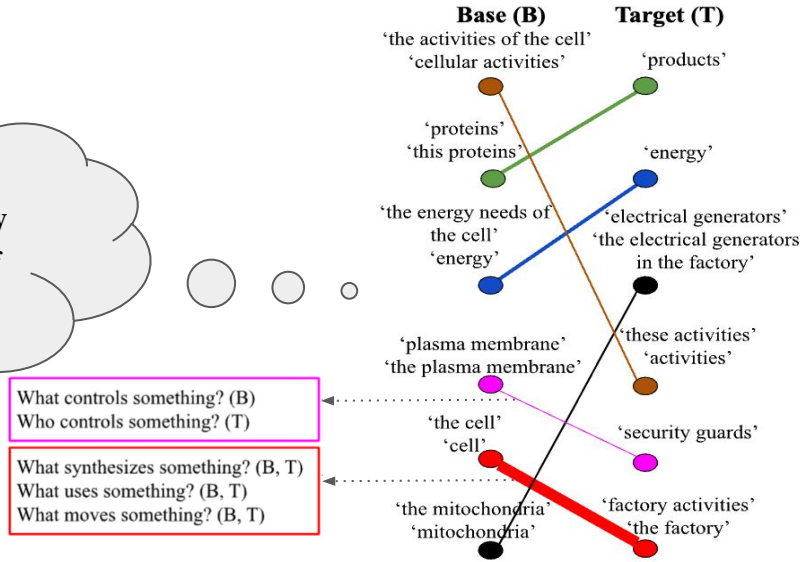


The Main Idea

- **Q**: How can we know that **entities** in the domains **play similar roles**?
- **QA-SRL** model *(FitzGerald, 2018)*
  - **Input:** A sentence. **Output**: questions and answers about the sentence.
  - The **answers** form the **entities**.
  - **Similar questions** between the domains, indicate that the entities may **play similar roles**.

Find Mappings by Questions (FMQ)

Find a mapping that **maximizes** the **similarity** between the **questions** of the mapped **objects.**



Experiments

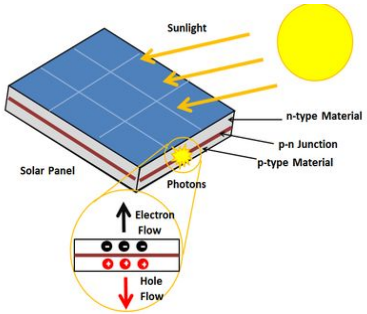
- **RQ1**: Can we leverage our algorithm for **retrieving analogies** from a large dataset of procedural texts?
- **RQ2**: Does our algorithm produce the **correct mapping solution**?
- **RQ3**: Is our algorithm **robust to paraphrasing** the input texts?

- We tested our ideas on **ProPara** dataset *(Dalvi, 2018)* of crowdsourced paragraphs describing **processes**. (e.g, “What happens during photosynthesis?”) were given to 1-6 workers each.

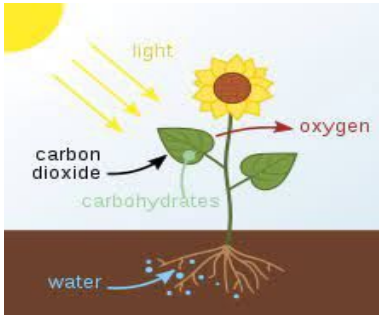
Experiment I: **Mining Analogies** – Examples

Examples for analogies mined by our method (FMQ):

Base: How does a solar panel work?



Target: What happens during photosynthesis?



Experiment I: **Mining Analogies** – Results

Method	Not	Sub	Self	Close	Far
SBERT	0	0	89	11	0
FMV	28	15	26	20	11
FMQ	21	16	29	18	16

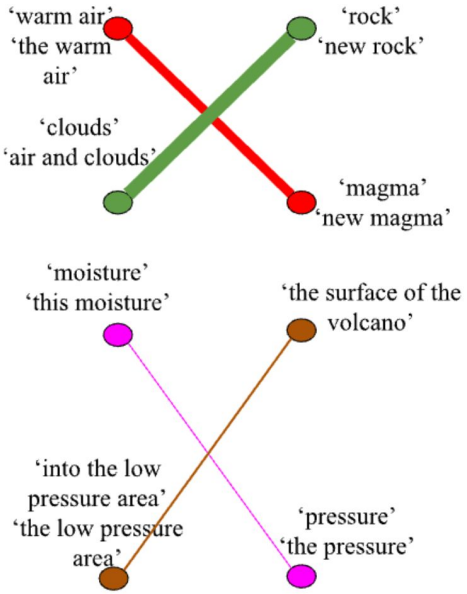
Top-100 of the ranking

Analogies prevalence in data: **~3%**

Experiment II: **Evaluating the Mappings** – Setup

PROMPT: Describe the process by which hurricanes form
Warm water floats up from the ocean.
A pocket of low pressure air is created near the surface of the ocean.
Warm air from outside areas pushes into the low pressure area.
The warm air keeps rising and forms clouds. The wind is getting fed moisture that evaporates from the ocean. This moisture causes the swirling air and clouds to grow bigger.
As it gets bigger the wind rotates faster.

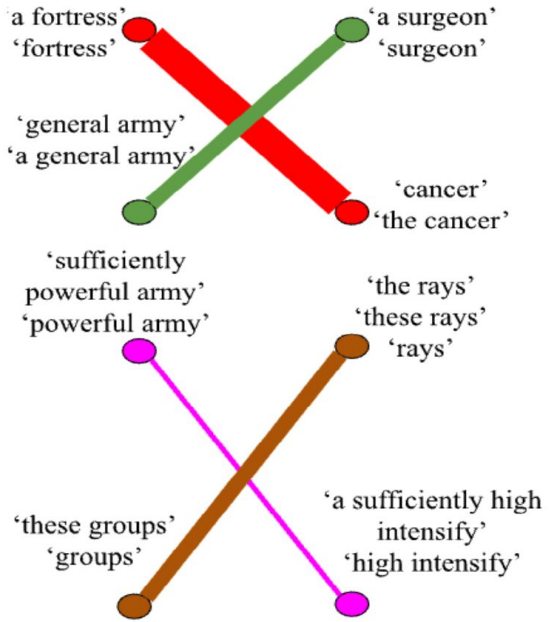
PROMPT: What causes a volcano to erupt?
Magma rises from deep in the earth. The magma goes into volcanos. The volcanos pressure the magma upwards.
The pressure causes the magma to push through the surface of the volcano. The lava cools. The lava forms new rock. New magma is pressured to the surface of the volcano.
The volcano bursts through the rock that formed after the last eruption.



Experiment II: **Evaluating the Mappings** – Setup

Base: The general
A general was trying to destroy a fortress which was situated at the center of a country with roads leading to it, by using his army. He needed to use his army as a complete group in order to destroy the fortress. However, he could not march his army down a road to the fortress because the roads were mined to explode when large groups of men passed over them. After considerable thought, he knew just what to do. He divided his army up into small groups of men, and by sending these groups, simultaneously, from a number of different directions, they converged on the fortress, making up a sufficiently powerful army to destroy it.

Target: The surgeon
A surgeon was trying to destroy a cancer which was situated in the central region of a patient's brain, by using a type of ray.
He needed to use these rays at a high intensity in order to destroy the cancerous tissue. However, at such an intensity the healthy brain tissue will also be destroyed. After considerable thought, he knew just what to do. He divided the rays up into batches of low-intensity rays, and then by sending them, simultaneously, from a number of different directions, they converged on the cancer, making up a sufficiently high intensity to destroy it.



ProPara: **87%** precision
Stories: **94%** precision

Conclusions

Paper

- Analogies are important for **humans** and **AI**.
- We explored analogies between **procedural texts** expressed in **natural language**.
- We develop a **scalable**, **interpretable** method to find **mappings** based on **relational similarity**.

- Our method was able to **mine different type of analogies** (in contrast to **SBERT**).
- Our method produced the correct mappings on both **ProPara** and the **Stories**.
- We showed our method is **robust to paraphrasing**.

Data & Code

Website

Video