

Relation extraction

Bill MacCartney
CS224u
Stanford University

Overview

Overview

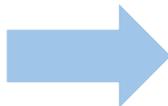
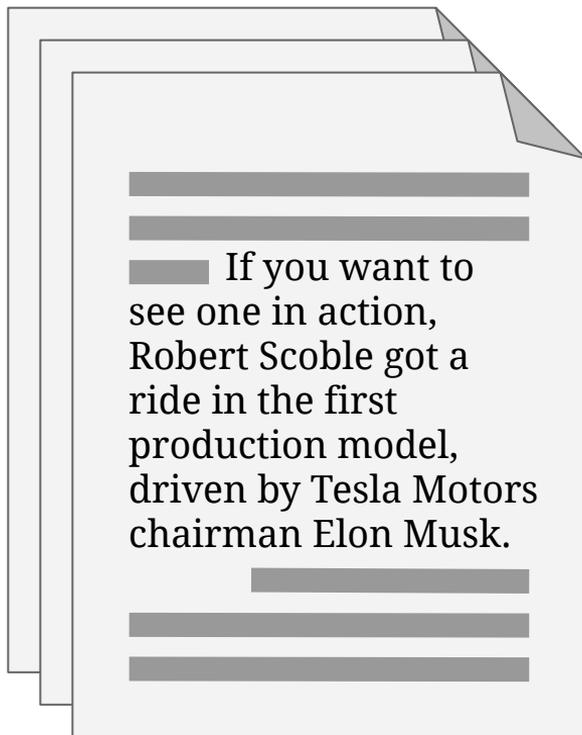
- The task of relation extraction
- Data resources
- Problem formulation
- Evaluation
- Simple baselines
- Directions to explore

The task of relation extraction

- Task definition
- Goal: machine reading
- Practical applications
- Hand-built patterns
- Supervised learning
- Distant supervision

The task of relation extraction

Task definition



relation	subject	object
founders	PayPal	Elon_Musk
founders	SpaceX	Elon_Musk
has_spouse	Elon_Musk	Talulah_Riley
worked_at	Elon_Musk	Tesla_Motors

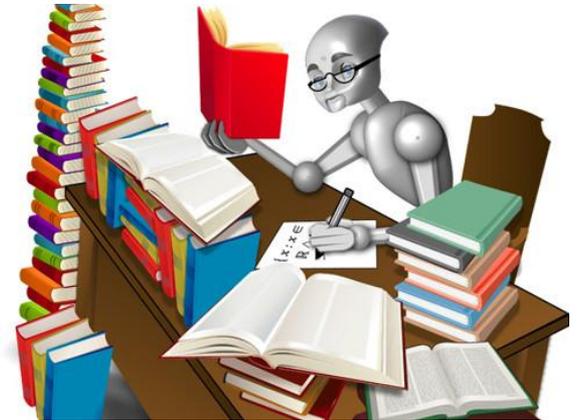
The task of relation extraction

Goal: machine reading

Reading the Web: A Breakthrough Goal for AI

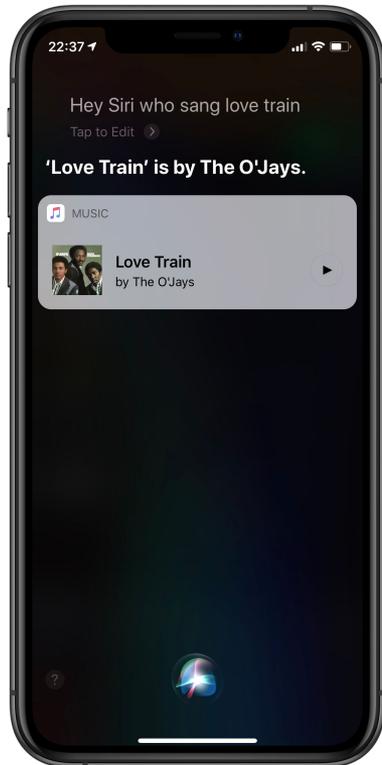
I believe AI has an opportunity to achieve a true breakthrough over the coming decade by at last solving the problem of reading natural language text to extract its factual content. In fact, I hereby offer to bet anyone a lobster dinner that **by 2015 we will have a computer program capable of automatically reading at least 80% of the factual content [on the] web, and placing those facts in a structured knowledge base.** The significance of this AI achievement would be tremendous: it would immediately increase by many orders of magnitude the volume, breadth, and depth of ground facts and general knowledge accessible to knowledge based AI programs. In essence, computers would be harvesting in structured form the huge volume of knowledge that millions of humans are entering daily on the web in the form of unstructured text.

— Tom Mitchell, 2005



The task of relation extraction

Applications: intelligent assistants



/music/artist/track

The O'Jays	Love Train
Cardi B	Bodak Yellow
Selena Gomez	Bad Liar

/film/film/starring

Wonder Woman	Gal Gadot
Dunkirk	Tom Hardy
Tomb Raider	Alicia Vikander

/organization/organization/parent

tbh	Facebook
Kaggle	Google
LinkedIn	Microsoft

/people/person/date_of_death

Barbara Bush	2018-04-17
Milos Forman	2018-04-14
Winnie Mandela	2018-04-11

“Love Train” is a hit single by The O’Jays, written by Kenny Gamble and Leon Huff.

Released in 1972, it reached number one on both the R&B Singles and the Billboard Hot 100, in February and March 1973 respectively, number 9 on the UK Singles Chart and was certified gold by the RIAA. It was The O’Jays’ first and only number-one record on the US pop chart.

Applications: building ontologies

video game
 action game
 ball and paddle game
 Breakout
 platform game
 Donkey Kong
 shooter
 arcade shooter
 Space Invaders
 first-person shooter
 Call of Duty
 third-person shooter
 Tomb Raider
 adventure game
 text adventure
 graphic adventure
 strategy game
 4X game
 Civilization
 tower defense
 Plants vs. Zombies



Mirror ran a headline questioning whether the killer's actions were a result of playing **Call of Duty, a first-person shooter game** ...



Melee, in video game terms, is a style of elbow-drop hand-to-hand combat popular in **first-person shooters and other shooters**.



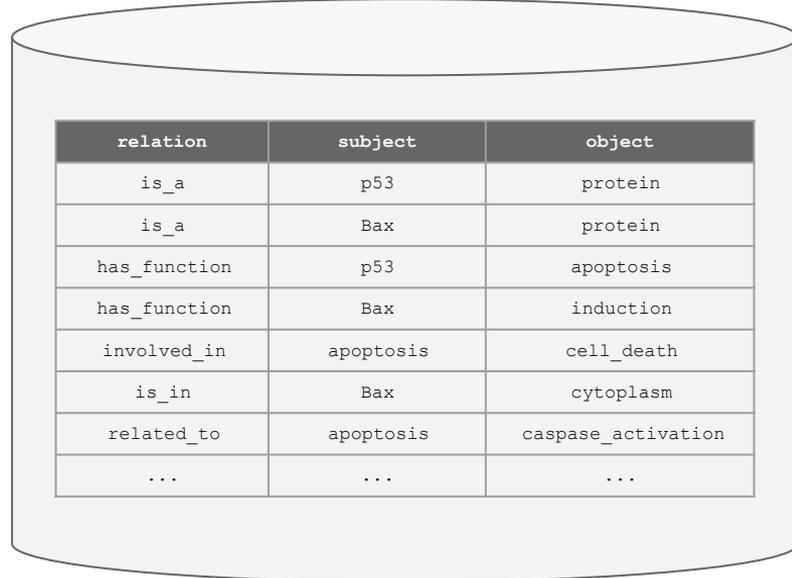
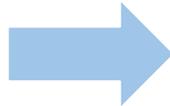
Tower defense is a kind of real-time strategy game in which the goal is to protect an area or place and prevent enemies from reaching ...

The task of relation extraction

Applications: gene regulation



textual abstract:
summary for human



structured knowledge extraction:
summary for machine

The task of relation extraction

Hand-built patterns

Idea: define some extraction patterns



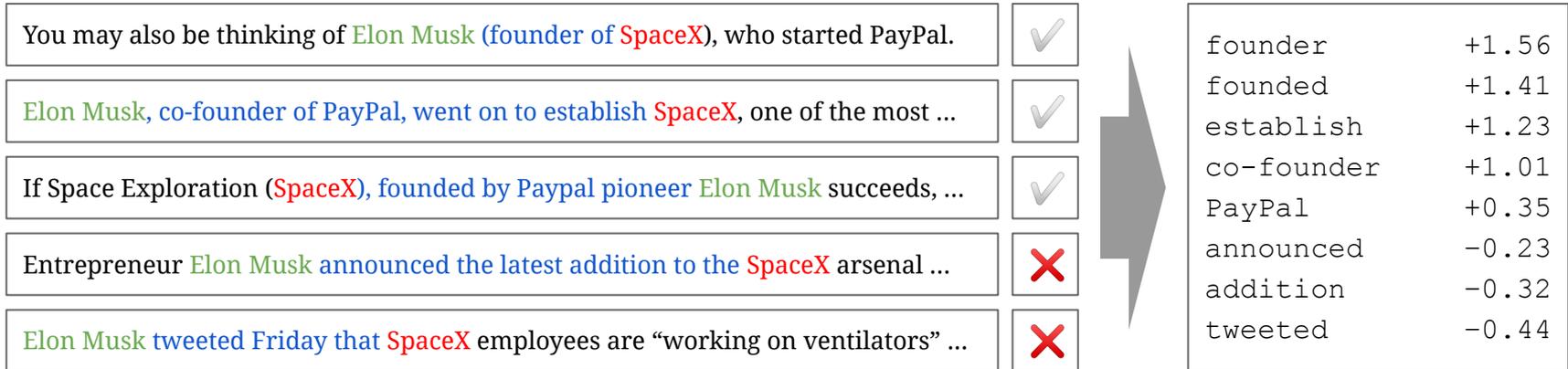
Problem: most occurrences do not fit simple patterns



The task of relation extraction

Supervised learning

Idea: label examples, train a classifier



Success! Better generalizability

Problem: labeling examples is expensive :-)

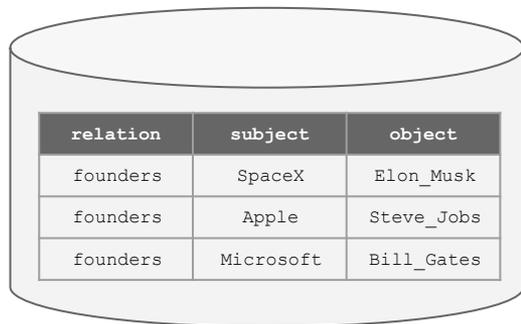
The task of relation extraction

Distant supervision

Idea: derive labels from an existing knowledge base (KB)

Assume sentences with related entities are positive examples

Assume sentences with unrelated entities are negative examples



relation	subject	object
founders	SpaceX	Elon_Musk
founders	Apple	Steve_Jobs
founders	Microsoft	Bill_Gates

Elon Musk, co-founder of PayPal, went on to establish SpaceX, one of the most ...



Entrepreneur Elon Musk announced the latest addition to the SpaceX arsenal ...



Elon Musk dismissed concerns that Apple was poaching the company's talent.



Now we know what Apple would have done with Elon Musk if that deal had ...



Hooray! Massive quantities of training data, practically free!

Qualm: are those assumptions reliable?

Distant supervision: limitations

Distant supervision is a powerful idea — but it has two limitations:

1. Not all sentences with related entities are truly positive examples

Entrepreneur **Elon Musk** announced the latest addition to the **SpaceX** arsenal ...



(but the benefit of *more* data outweighs the harm of noisier data)

2. Need an existing KB to start from — can't start from scratch