SemEval-2012 Task 2: Measuring Degrees of Relational Similarity

David Jurgens  
Department of Computer Science  
University of California, Los Angeles

Saif Mohammad  
Emerging Technologies  
National Research Council Canada

Peter Turney  
Emerging Technologies  
National Research Council Canada

Keith Holyoak  
Department of Psychology  
University of California, Los Angeles
Talk Outline

- Motivating Example
- Task Description
- Data Annotation Gathering
- Systems and Performance
- Discussion
The relational search engine

List all things that are part of a ... car
List all things that are part of a ... car

Antenna
Hubcaps
Seats
Roof
Wheel
Engine
Tires
Windows
The relational search engine

List all things that are part of a ... car

Antenna
Hubcaps
Seats
Roof
Wheel
Engine
Tires
Windows

How might we rank these items?
The relational search engine

List all things that are part of a ... car

Car:Antenna
Car:Hubcaps
Car:Seats
Car:Roof
Car:Wheel
Car:Engine
Car:Tires
Car:Windows

These are all analogous pairs, but vary in how strong the relation is
The relational search engine

List all things that are part of a ... car

Car: Antenna
Car: Hubcaps
Car: Seats
Car: Roof
Car: Wheel
Car: Engine
Car: Tires
Car: Windows

What is the most prototypical example of the shared relation?
Talk Outline

• Motivating Example
• Task Description
• Data Annotation Gathering
• Systems and Performance
• Discussion
Task 2: Measuring Degrees of Relational Similarity

Given example pairs having approximately the same relation

Car: Antenna
Car: Hubcaps
Car: Seats
Car: Roof
Car: Wheel
Car: Engine
Car: Tires
Car: Windows

1. Identify what the relation is
2. Rate each pair according to the degree that it expresses that relation
Task 2: Measuring Degrees of Relational Similarity

bouquet: flower
army: soldiers
library: book
arsenal: weapons
herd: cow
troop: soldier
paragraph: word
album: photos
class: student
beach: sand
garden: plot

1. Identify what the relation is
Task 2: Measuring Degrees of Relational Similarity

1. Identify what the relation is

A X is made from a collection of Y

bouquet:flower
army:soldiers
library:book
arsenal:weapons
herd:cow
troop:soldier
paragraph:word
album:photos
class:student
beach:sand
garden:plot
Task 2: Measuring Degrees of Relational Similarity

1. Identify what the relation is

   A $X$ is made from a collection of $Y$

2. Rate each pair according to the degree that it expresses that relation

bouquet:flower
army:soldiers
library:book
arsenal:weapons
herd:cow
troop:soldier
paragraph:word
album:photos
class:student
beach:sand
garden:plot
Task 2: Measuring Degrees of Relational Similarity

1. Identify what the relation is

   A X is made from a collection of Y

2. Rate each pair according to the degree that it expresses that relation

   51.7 bouquet:flower
   50.0 army:soldiers
   37.3 library:book
   35.7 arsenal:weapons
   23.6 herd:cow
   21.1 troop:soldier
   20.7 paragraph:word
   18.2 album:photos
   10.5 class:student
   -7.5 beach:sand
   -32.8 garden:plot
Task 2: Relation Taxonomy

10 Relation Categories, Divided into 79 subcategories

**Class Inclusion**
- Taxonomic - flower:tulip
- Function - weapon:knife

**Cause-Purpose**
- Cause:Effect - joke:laughter
- Agent:Goal - climber:peak

Task 2: Relation Taxonomy

Includes some more challenging subcategories...

**Similar**
Dimensional Naughty - copy:plagiarize

**Contrast**
Asymmetric Contrary - hot:cool

**Space-Time**
Contiguity - ocean:coast
Task Data

• Lists of example pairs for all 79 subcategories
  • Pairs vary in quality
• Prototypicality ratings for 10 subcategories
• All materials used to crowdsource the ratings
  • Includes example description of each relation, “An X is a kind of Y”
Talk Outline

• Motivating Example
• Task Description
• Data Annotation Gathering
• Systems and Performance
• Discussion
Crowdsourcing Graded Relational Annotations

Phase 1
Seed Pairs and example relation → Generate Examples

Phase 2
List of pairs → Prototypicality Rating → Ratings
Gathering Relation Examples

Consider the following word pairs:
flower:tulip, emotion:rage, poem:sonnet

What relation best describes these X:Y word pairs?
to X is to have a Y receive some object/service/idea
Y is an unacceptable form of X
a Y is a part of an X
Y is a kind/type/instance of X

• Question 1 asked Turkers to pick the relation shared by 3 seed pairs

• Question 2 asked Turkers to provide four additional examples with the same relation
Rating Prototypicality

• Question 1 same as Phase 1
• Question 2 used the MaxDiff format

Given prototypical examples of a subcategory:
  flower: tulip, emotion: rage, poem: sonnet

weapon: spear
bird: swan
automobile: van
hair: brown

Select which pair is the **best** example of the relation and which is the **worst** example.
Talk Outline

• Motivating Example
• Task Description
• Data Annotation Gathering
• Systems and Performance
• Discussion
Participants

- University of Texas, Dallas
  - two systems
- University of Minnesota, Duluth
  - three systems
- Benemérita Universidad Autónoma de Puebla (México)
Evaluation Metrics

Systems provide numerical ratings for each pair

- Use the ratings to answer MaxDiff questions
  - weapon:spear
  - bird:swan
  - automobile:van
  - hair:brown

- Compare system ranking with Turker ranking using Spearman’s rank correlation

  [Highest scoring is best example]
Baselines

• Generate a random ordering of pairs
• Score pairs according to the pair’s words’ Point-wise Mutual Information (PMI)
  • a measure of statistical association of the pairs’ words
Average Correlation Performance

Spearman's Rank Correlation

- BUAP
- UMD-V2
- UMD-V0
- PMI
- UTD-SVM
- Random
- UTD-NB
Correlation Performance per Subcategory

- Number of Significant Correlations
- p < 0.05
- p < 0.01

- BUAP
- UMD-V2
- UMD-V0
- UMD-V1
- PMI
- UTD-SVM
- UTD-NB
- Random
MaxDiff Performance

% MaxDiff Questions Answered Correctly

- BUAP
- UMD-V2
- UMD-V0
- UMD-V1
- PMI
- UTD-SVM
- Random
- UTD-NB
Talk Outline

• Motivating Example
• Task Description
• Data Annotation Gathering
• Systems and Performance
• Discussion
Categorical Performance

• Were some subcategories harder than others?
Measuring the impact of pair reversals

Spearman's Rank Correlation

With Reversals
Without Reversals

BUAP
UMD-V2
UMD-V0
PMI
UTD-SVM
Random
UMD-V1
PMI
UTD-NB
Future Work

• Relations aren’t simply binary
  • Especially when relational reasoning comes into play

• Future SemEval task

• Dataset has many uses in psychology as well as computational linguistics

• Spark more interest
Thank you!

https://sites.google.com/site/semeval2012task2/

David Jurgens  
Department of Computer Science  
University of California, Los Angeles

Saif Mohammad  
Emerging Technologies  
National Research Council Canada

Peter Turney  
Emerging Technologies  
National Research Council Canada

Keith Holyoak  
Department of Psychology  
University of California, Los Angeles

Questions? jurgens@cs.ucla.edu