Higher-order relations
Higher-order relations

- Represent data on multiple abstraction levels
- Find inherent levels of abstraction
- Discard irrelevant details for task at hand
Higher-order relations

• Represent data on multiple abstraction levels
• Find inherent levels of abstraction
• Discard irrelevant details for task at hand
General principle

• Generalize distributional hypothesis in natural language processing
• Know an object by the company it keeps
• Relate objects in terms of higher-order correlations
• Objects that correlate similarly play similar roles
• Domain-agnostic and interpretable approach
Example: word co-occurrences

She had a cup of coffee in the morning.

Would you like a cup of tea?
Part 1: Batch processing
Batch processing

- Find relations *en masse*
- Internet-scale, preferably
- Resort to cluster computing
Graph transformation

Correlations

Similarities
In (pseudofied) Spark-speak

1: ins = edges.map(((i,j),rij) => (j,(i,rij)))
2: pairs = ins.join(ins).filter((k,((i,rik),(j,rjk))) => i<j)
3: terms = pairs.map((k,((i,rik),(j,rjk))) =>
4:     ((i,j),abs(rik-rjk)-abs(rik)-abs(rjk)))
5:     .reduceByKey((v,w) => v+w)

Scalability

Google Books 5-grams run on Amazon EC2
Twitter hashtags
Wikipedia words
Part 2: Stream processing
Stream processing

- Find relations incrementally
- Speed and low memory usage priorities
- Resource efficient
- Co-occurrence-based
Key observation

Objects that are highly correlated with respect to second-order co-occurrences are highly similar with respect to first-order co-occurrences

Approach

Explicitly count second-order co-occurrences and correlate objects
Buffer and count

(a) Second-order co-occurrences

(b) Co-occurrence buffers
Most correlated words (2nd order)

<table>
<thead>
<tr>
<th>musician</th>
<th>increasing</th>
<th>croatian</th>
<th>wednesday</th>
<th>scholar</th>
<th>hermann</th>
<th>coventry</th>
<th>yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>singer</td>
<td>reducing</td>
<td>yugoslav</td>
<td>thursday</td>
<td>scholars</td>
<td>heinrich</td>
<td>leicester</td>
<td>purple</td>
</tr>
<tr>
<td>pianist</td>
<td>growing</td>
<td>serbian</td>
<td>tuesday</td>
<td>translator</td>
<td>friedrich</td>
<td>norwich</td>
<td>pink</td>
</tr>
<tr>
<td>songwriter</td>
<td>increased</td>
<td>slovenian</td>
<td>monday</td>
<td>playwright</td>
<td>wilhelm</td>
<td>stoke</td>
<td>orange</td>
</tr>
<tr>
<td>guitarist</td>
<td>reduces</td>
<td>croatia</td>
<td>friday</td>
<td>philosopher</td>
<td>georg</td>
<td>swansea</td>
<td>blue</td>
</tr>
<tr>
<td>rapper</td>
<td>reduce</td>
<td>slovak</td>
<td>saturday</td>
<td>poet</td>
<td>wolfgang</td>
<td>cardiff</td>
<td>red</td>
</tr>
</tbody>
</table>
# Accuracy

<table>
<thead>
<tr>
<th></th>
<th>SIMLEX</th>
<th>SIMVERB</th>
<th>MT-287</th>
<th>MT-771</th>
<th>WS-353</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOCO</td>
<td>0.41</td>
<td>0.25</td>
<td>0.59</td>
<td>0.56</td>
<td>0.71</td>
</tr>
<tr>
<td>FOCO</td>
<td>0.35</td>
<td>0.23</td>
<td>0.65</td>
<td>0.59</td>
<td>0.66</td>
</tr>
<tr>
<td>GLOVE</td>
<td>0.31</td>
<td>0.18</td>
<td>0.61</td>
<td>0.57</td>
<td>0.63</td>
</tr>
<tr>
<td>CBOW</td>
<td>0.34</td>
<td>0.22</td>
<td>0.66</td>
<td>0.57</td>
<td>0.69</td>
</tr>
<tr>
<td>SGN</td>
<td>0.41</td>
<td>0.32</td>
<td>0.67</td>
<td>0.60</td>
<td>0.72</td>
</tr>
<tr>
<td>Coverage</td>
<td>0.99</td>
<td>0.96</td>
<td>0.95</td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>
Runtime

Convergence

English Wikipedia run on a Macbook Pro
Part 3: Concept mining
Concept mining

- Concepts as groups of inter-similar objects
- Concepts of concepts
- Levels of abstraction
Concepts as clusters

- Group similar objects using
  - Label propagation (batch)
  - Incremental clustering (streaming)
- A concept is an object in its own right
  - Recalculate relations
Wikipedia words
Concepts

- yale, harvard, duke, oxford, cambridge, school, ucla, stanford, university, college
- arkansas, colorado, jersey, delaware, georgia, kansas, florida, mississippi, minnesota, wisconsin, dakota, massachusetts, indiana, california, maine, pennsylvania, illinois, utah, carolina, louisiana, alabama, connecticut, michigan, virginia, arizona
- grey, white, yellow, gray, blue, pink, red, dark, orange, black, green
- van, convoy, vessel, aircraft, ship, bus, boat, crews, vessels, cycle, bike, vehicle, trains, boats, helicopters, ships, vehicles, jet, helicopter, truck, buses, car, cars, flights, planes, firefighters, motorcycle, trucks, plane
- telegraph, tribune, post, xinhua, times, magazine, newspaper, mirror, observer, herald, guardian
- broadcasting, mining, banking, tech, telecommunications, wholesale, utility, retail, telecom, infrastructure
- main, principal, key, decisive, vital, precious, helpful, valuable, critical, useful, essential, crucial, necessary, important
- appears, sounds, appeared, sound, seemingly, looks, appear, seem, seems, appearing
- soccer, tennis, diving, nba, cycling, boxing, hockey, sailing, basketball, football, baseball, rugby, nfl, golf, cricket, nhl, swimming
Demo time
Olof Görnerup

Intelligent systems unit
olof.gornerup@ri.se
+46 (0)70 252 10 62