Word vectors and Tensor Networks in Ontology and semantic entity mapping

research seminar presentation
Seppo Nyrkkö, UHEL
U40 A112 5.3.2014 klo 10.15-11.45
Motivation

Entities:
Kala
Fisu
Ahven
...
About research

Big question: Combining ontologies

Ontologies vary
- language
- structure: flat, recursive, deep
- relations
  - sub- and superclass
  - functional relations / tags
- labeling
About the experiment

- **Source:** 2 or more ontologies
  - TAP - generic top level ontology
  - PuLS - a medical glossary
- **Generic text material**
  - wikipedia harvester
- **Word vector space model**
  - $D = 10...40$
- **Semi / Unsupervised iterations**
  - user intervention cycle
Target function

Estimate the relation semantic relation $R$ (coverage, overlap)

$g(a, R, b)$

$g \in [0...1]$  

$R$: relation or aspect  
$a, b$: word form representations in vector space
Methods and tools used (SNy, 2014)

R statistical programming env. (r-project.org)
OntoR toolbox (Nyrkkö)
Stanford parser (English dependency syntax)

Reasoning resources: WordNet, FreeBase
Ontologies: PuLS, TAP
Corpora: Collection of Wikipedia top-10 keyword matches
Comparison with ongoing research

Knowledge-base related reasoning with Tensor Networks (Socher et al, 2013)

Model of entity space / optimization problem?

...
“Numbering” the words

Entity numbering

Random word vectors
(Compounds treated as average)

Unsupervised word vectors (clustering based on text corpora)
Word vector space, knowledge base reasoning example

(Socher et al, 2013)
Choosing a Neural Network model (Socher et al, 2013)

Baseline:
Distance model
Single-layer Perceptron

Projection based:
Hadamard (coordinate scaling)
Bilinear (projection and dot product)
Neural Tensor Network (NTN)
Background: Single-layer perceptron neural network

\[ g = U^T f(W [e_1 e_2] + b) \]

\( f = \) nonlinearity function (e.g. tanh)
Figure 2: Visualization of the Neural Tensor Network. Each dashed box represents one slice of the tensor, in this case there are \( k = 2 \) slices.
Model selection boiling down to...

weak / strong entity vector interaction
~ ability to solve the XOR problem

Complexity of parameters / precision tradeoff

Computation time

Scalability
Optimization model

\[ J(\Omega) = \sum_{i=1}^{N} \sum_{c=1}^{C} \max(0, 1 - g(T_i) + g(Q_c)) + \lambda \|\Omega\|^2 \]

g = estimation function
T = Training set (N entries)
Q = Corrupt set (C entries)
Parameter set:
\( \Omega = \{ u, W, V, b, E \} \)
\( \lambda \|\Omega\|^2 = \text{Energy of parameters (Square of euclidean dist.)} \)
Bibliography

Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. 2013. 
Reasoning With Neural Tensor Networks For Knowledge Base Completion. 

Ontology-based Knowledge in Interactive Maintenance Guide. 