A Structured Prediction Approach for Entity Coreference Resolution
Janardhan Rao Doppa, J. Walker Orr, Mohammad NasresFahani, Jed Irvine
Thomas G. Dietterich, Xiaoli Fern and Prasad Tadepalli

Introduction

Coreference resolution problem is encountered in various forms across CS
• NLP – Noun-phrase coreference resolution
  “The dog put the ball in his mouth”
• Databases - database cleaning and database integration

Problem formulation

Given
• \( x \): set of incomplete and noisy mentions (records) extracted from a single document
• \( I \): set of integrity (hard) constraints
• \( R \): set of rules (soft constraints)

Predict
• \( y \): set of clean and complete version of the true entities as a result of clustering entities in \( x \), such that every entity in \( y \) satisfies all the integrity constraints
• \( y \) violates as few rules as possible

An example from NFL domain

Joint feature function \( \phi(x, h, y) \)
• syntactic rules that capture our beliefs on extraction errors
• integrity constraints and rules
• \# entities in the output (MDL principle)
• \# of ignored facts (don’t want to drop any information)
• \# of missing facts

Cost function learning
• linear cost function \( f = w \cdot \phi(x, h, y) \)
• goal of learning: \( \argmin_{(h,y) \in Y} w \cdot \phi(x, h, y) = (h^*, y^*) \)
• Structured prediction with latent variables
  1. optimize for \( h \)
  2. fix \( h^* \) and optimize for \( y \)

Prediction (greedy “argmin” procedure)
• repeatedly merge the pair of entities \((e_i, e_j) \in x \) that maximally improve the cost

Weight learning (“averaged” perceptron)
• for each training example \((x, h^*, y^*)\)
  Predict: \((h, y) = \argmin_w f(x, h, y)\)
  If \((h, y) \neq (h^*, y^*)\)
  Learn: \( w = w + \alpha(\phi(x, h, y) - \phi(x, h^*, y^*)) \)
• repeat until convergence or max. iterations

Experiments and results

Real data
• Manually annotated the data extracted from 25 NFL documents (LDC training corpus) by UW system
• 5 documents for training and 20 documents for testing

Synthetic data
• Took a subset of the real data and added some additional facts, e.g., homeTeam and awayTeam, in a consistent way
• 5 documents for training and 10 documents for testing

Structured prediction approach

Structural prediction problem \((X, Y, \phi, \Delta)\)
• \( x \) is an input from the input space \( X \)
• \( (h, y) \) is an output from the output space \( Y \)
• \( \phi: X \times Y \mapsto \mathbb{R}^d\) is the joint feature function
• \( \Delta: X \times Y \times Y \mapsto \mathbb{R} \) is the loss function

Conclusions
• Integrity constraints are more effective in improving accuracy than rules
  1. most rules in the domain are deterministic
  2. there are only a small number of relationships