Name-entity Disambiguation in Citations by Using Markov Logic Networks

Yichuan Gui

UMBC AN HONORS UNIVERSITY IN MARYLAND — CS computer science, EE electrical engineering

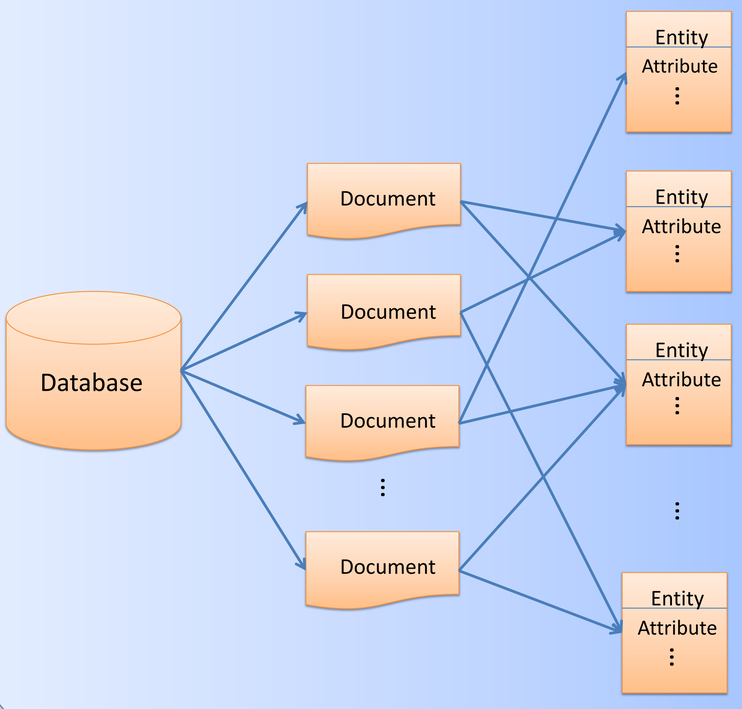
Entity Disambiguation

Entity disambiguation is the process of determining which names, words, or phrases in text correspond to distinct persons, organizations, locations, or other entities.

• What is "*entity*"?
Entity is an object or set of objects in the world.

• What is "*determination*"?

The *determination* is absolutely essential for reasoning, inference, and the examination of social network structures based on information derived from text.



Markov Logic Networks

A *Markov logic network* is a first-order knowledge base with a weight attached to each formula, and can be viewed as a template for constructing Markov networks.

Markov Logic Networks

=

Markov Networks

First-order Logic Formulas with Weight

Logic V.S. Probability

First-order Logic:

$\forall x$ Smokes(x) $\Longrightarrow$ Cancer(x)

P	Q	P->Q
F	F	T
F	T	T
T	F	F
T	T	T

Probability:

Smokes ➡ Cancer ???

Markov Logic Networks:

Smokes(x) => Cancer(x)

+

related weight

Smokes(John) 0.728977
Smokes(Katherine) 0.433007
Smokes(Lars) 0.464004
Smokes(Michael) 0.89796
Cancer(John) 0.375012
Cancer(Katherine) 0.230027
Cancer(Lars) 0.306019
Cancer(Michael) 0.429007

Name-entity Disambiguation in Citations

Citations ⟷ Authors

• An Author has multiple names in different citations

Michael Jordan → M. Jordan, Jordan, M., Michael Jordan, ⋮

• Multiple authors may share the same name

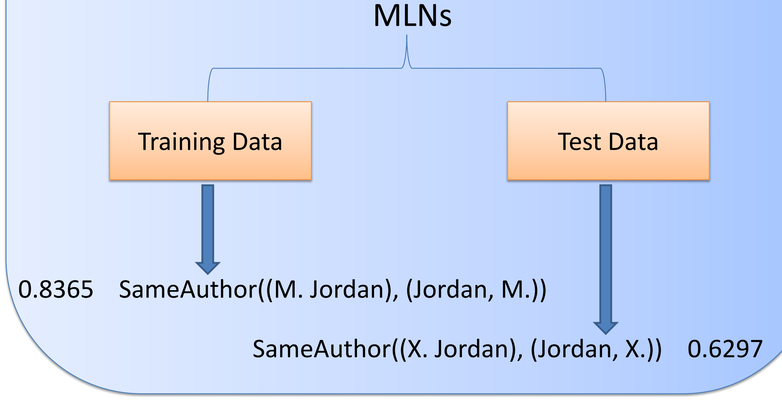
M. Jordan → Michael Jordan, Matthew Jordan, Mark Jordan, ⋮

Implementation with MLNs

Predicates:
HasWordAuthor(author, word)
SameAuthor(Author, Author)

Main Rules:
HasWordAuthor(a1,+w)^HasWordAuthor(a2,+w)
=>SameAuthor(a1,a2)
SameAuthor(a1,a2)^ SameAuthor(a2,a3)
=>SameAuthor(a1,a3)

X. Jordan V.S. Jordan, X.:
a1 = M. Jordan
a2 = Jordan, M.
w = Jordan
SameAuthor((M. Jordan), (Jordan, M.)) √
SameAuthor((X. Jordan), (Jordan, X.)) ?



Future Work

• Adding attributes for entity disambiguation in citations
• Evaluation between MLNs and classical entity disambiguation methods
• Automatic or semi-automatic dataset labeling method

References

• Matthias Blume. Automatic entity disambiguation: Benefits to NER, relation extraction, link analysis, and inference. *International Conference on Intelligence Analysis*, 2005.
• Hui Han, C. Lee Giles, Hongyuan Zha, Cheng Li, and Kostas Tsioutsiouliklis. Two supervised learning approaches for name disambiguation in author citations. *Proceedings of the 4th ACM/IEEE Joint Conference on Digital Libraries*, pages 296-305, 2004.
• Pedro Domingos and Matthew Richardson. Markov logic: A unifying framework for statistical relational learning. *PROCEEDINGS OF THE ICML-2004 WORKSHOP ON STATISTICAL RELATIONAL LEARNING AND ITS CONNECTIONS TO OTHER FIELDS , pages 49-54, 2004*
• http://alchemy.cs.washington.edu