FAKULTÄT FÜR WIRTSCHAFTSINFORMATIK
UND WIRTSCHAFTSMATHEMATIK

UNIVERSITÄT MANNHEIM

KOLLOQUIUM

RELATION EXTRACTION WITH MATRIX FACTORIZATION AND UNIVERSAL SCHEMAS

Prof. Dr. Sebastian Riedel

19.05.2014. 17:15.

A5, Raum C 012

The ambiguity and variability of language makes it difficult for computers to analyze, mine, and base decisions on. This has motivated machine reading: automatically converting text into semantic representations. At the heart of machine reading is relation extraction: predicting relations between entities, such as employeeOf (Person, Company). Machine learning approaches to this task require either manual annotation or, for distant supervision, existing databases of the same schema (=set of relations). Yet, for many interesting questions (who criticized whom?) pre-existing databases and schemas are insufficient. For example, there is no criticized (Person, Person) relation in Freebase. Moreover, the incomplete nature of any schema severely limits any global reasoning we could use to improve our extractions.

The need pre-existing datasets can be avoided by using, what we call, a "universal schema": the union of all involved schemas (surface form predicates such as "X-was-criticized-by-Y" as in OpenIE, and relations in the schemas of pre-existing databases). This extended schema allows us to answer new questions not yet supported by any structured schema, and to answer old questions more accurately. For example, if we learn to accurately predict the surface form relation "X-is-scientist-at-Y", this can help us to better predict the Freebase employee(X,Y) relation.

To populate a database of such schema we present a family of matrix factorization models that predict affinity between database tuples and relations. We show that this achieves substantially higher accuracy than the traditional classification approach. More importantly, by operating simultaneously on relations observed in text and in pre-existing structured DBs, we are able to reason about unstructured and structured data in mutually-supporting ways. By doing so our approach outperforms state-of-the-art distant supervision.