

## **Abstract**

Discovery of attributes associated with various entities from text is important for building a large database of facts. Traditionally, systems focused on only one task or adopted an incremental pipeline of the two tasks: Extracting ⟨entity type, attribute name⟩ pairs (by ontology construction fed with query logs) and extracting ⟨entity, attribute name, attribute value⟩ tuples (by open information extraction systems). In this paper, we investigate joint extraction of the type-level ⟨entity type, attribute name, attribute value type⟩ tuples and the entity-level tuples by replacing the entity mentions in the text with their types and generalizing frequent informative segments, called meta patterns. The typed entities can help find precise attribute names associated with the types; and the frequency of type-level tuples can help group the entity-level tuples from numerous kinds of expressions. Mining meta patterns enables not only their mutual enhancement but also automatic and domain-independent discovery of attribute features from massive text corpora, relying on no linguistic assumptions nor query logs nor human annotations. We propose a novel framework, called MetaPAD, that runs a data-driven text segmentation algorithm with meta-pattern quality assessment, detects synonymous meta patterns to share the same type-level tuple and merge the entity-level tuples, and adjusts entity types in meta patterns into appropriate granularity for precise associations with attribute names. Empirical results show that MetaPAD significantly outperforms the state-of-the-art methods on both attribute name extractions for entity types and tuple extractions for facts from news, tweets and biomedical text.

## **Bio**

Dr. Meng Jiang is a postdoctoral researcher in University of Illinois at Urbana-Champaign. He received his Ph.D. from the Department of Computer Science at Tsinghua University, Beijing in 2015. He received his bachelor from the same department in 2010. He visited Carnegie Mellon University in 2013 and University of Maryland, College Park in 2016. You can find more about him here:

<http://www.meng-jiang.com>.

His research lies in the field of data mining, focusing on user behavior modeling. He has given two tutorials in major conferences. His Ph.D. thesis won the Dissertation Award at Tsinghua University. His work on "Suspicious Behavior Detection" was selected as one of the Best Paper Finalists in KDD'14. His work on "Social Contextual Recommendation" has been deployed by the Tencent social network. The package of his work on "Automatic Attribute Discovery" is now transferring to U.S. Army Research Lab.