

ABSTRACT

The one among the recent activities of the netizens include search for information about individuals of their interest. This kind of information is usually represented in various redundant name formats in different repositories causing the major problem of name ambiguity. Usually, ambiguity in a person's name occurs in two variants, namely (i) referential, when a person uses multiple name variations and (ii) lexical, when there is more than one person with the same name. The goal of name disambiguation is to resolve such ambiguities by linking and merging all the references of the same person together for improving the information quality. It is an essential problem in many applications such as web search, information integration, natural language processing, and digital libraries. The users, in general, perform manual search and find the details of the expected person from the search result.

The research works proposed in this thesis attempt to examine the person's name disambiguation problem for improving the detection of a person's name aliases with techniques like Particle Swarm Optimization (PSO), Extreme Learning Machine (ELM) and Fuzzy Inference System (FIS).

The first work applies PSO to the regularization parameter of logistic regression to improve classification accuracy significantly. It can efficiently handle the unbalanced dataset used to avoid overfitting problem.

The second work proposes the use of ELM, to improve ranking and reduce execution time in web search to detect entity aliases. The results show that the proposed ELM improves ranking and is fast in execution significantly in contrast to Support Vector Machine.

Finally, Mamdani Fuzzy Inference System is proposed to find the closeness between the person's name and his / her alias names for improving the alias detection. It is used as a classifier for finding alias closeness. The fuzzy framework involves the development of fuzzy

inference rules using feature set and defuzzification. The results imply that the developed fuzzy-based framework is accurate and reliable to detect the person's alias names.

The performance of all the proposed techniques, discussed in the thesis have been examined thoroughly on the benchmark alias detection dataset.

The scope for future study is that the proposed approaches can be considered as a multi-objective problem by integrating lexical and referential ambiguities and generalizing them for any type of entity names such as places, things, and other real-world entities.