# Research Statement

Bing Liu
School of Computer Science and Engineering, Southeast University, China

I am second year Ph.D. student in Computer Science at Southeast University. My research interests mainly lie in Information Extraction (IE), Natural Language Processing (NLP) and Machine Learning (ML). Recently, I am focusing on IE, especially Relation Extraction (RE).

I have tried to solve the RE problem from two aspects: (1) design novel RE models based on ML, and (2) reduce human participation (for example human annotations) while guarantee the RE performance. I will introduce my current and future work from these two aspects.

**Design RE models based on ML**

- Current Work

➢ RE via Reinforcement Learning (RL)

The lack of sufficient context information is a main obstacle to improve RE performance. In order to alleviate this problem, I have attempted to introduce external evidences of potential facts and human knowledge (low-cost manual patterns) to the RE process. An RL framework was used to incorporate basic relation extractor, relation evidences and low-cost human patterns and a better extraction performance was achieved than the original relation extractor.

- Future Work

In the future, I may explore the other ways to design novel models for RE:

(1) Pay attention to document-level (cross-sentence) RE instead of sentence-level RE. In this case, the context information is more abundant and may be exploited to improve RE performance.
(2) Explore better knowledge representation method. Current mainstream methods are based on DL and word embedding. However, the information provided by word embedding are not disentangled and may be not sufficient for RE.
(3) Address fine-grained and hierarchical RE, which is more applicable in practice.

**Reduce human participation**

- Current Work

➢ Adversarial discriminative denoising for Distance Supervision RE

The lack of labeled data is a common problem for supervised ML methods. Distance Supervision (DS) can generated labeled data for RE automatically via aligning Knowledge Base (KB) with text. However, DS will also bring large amount of noise and severely effects the performance of DS-based RE models. In order to remove the noise from the DS-generated data, I proposed a denoising framework based on adversarial learning. This denoising framework can remove the noise effectively just using handful human annotations and address the non-differentiability problem in existing denoising approaches for DS.

- Future Work

In the future, I may research the other ways of reducing human participation in RE task:

(1) Augment the labeled data via text generation to solve the problems such as the long-tail problem of relation patterns and the low-quality problem of negative samples. The main challenge lies in the control the core semantic in the text generation process.
(2) Reduce the need of training data by exploiting the existing data of other domains, other tasks and other languages.
(3) Explore human-model interaction (human-in-the-loop) modes to improve the efficiency of human participation in RE.