Abstract:

Requirement specification and system design models are the fundamental documents in the software development life cycle. They are the major references to understand user requirements and to guide later system development and maintenance activities. It has been reported that more than 60% of the errors in software products are introduced during the design phase. Errors introduced in the early phases are much harder and more expensive to detect than errors introduced in the coding phase. It is thus highly desirable to improve the quality of software requirement specifications and design models by detecting software defects as early as possible.

In this thesis, we are motivated to provide techniques to improve the quality of software requirement specifications and design models. For software requirement specifications, we propose two works that focus on improving the quality of use cases, which are widely adopted by different software development methodologies to capture user requirements.

First, we propose to find defects in use case descriptions to improve the consistency and integrity aspects of a single use case. We adopt advanced natural language processing techniques to automatically extract action tuples and predicates from use case sentences. We formally define common defects, e.g., inconsistency and incompleteness related defects, in use case documents and propose algorithms to find those defects based on the automatically extracted action tuples, predicates and the control flow related information. The found defects are linked to the original descriptions in use cases to aid improving the quality of the use case document.

Second, we propose to further improve the use case quality by finding missing scenarios and preconditions/postconditions which involve multiple use cases. We adopt the active learning techniques to learn a Deterministic Finite State Automaton (DFA) for each actor/agent in a
use case document. During the learning process, our method finds missing scenarios and missing preconditions/postconditions through interactions with users. The missing scenario is presented as a sequence of actions which is easy to be added to the use case document to improve the integrity of the document.

To find sophisticated errors which may be introduced in the system design phase, we propose to improve the quality of UML state machines models, which are widely adopted to capture the dynamic behaviors of system designs. Our work focuses on finding safety and liveness related defects in UML state machines automatically. We provide an operational semantics for the complete syntax of UML state machines and implement the semantics into the PAT framework, which enables model checking on UML state machines to find liveness and safety related defects.

We evaluated all of our methods with real world documents or models. The evaluation results show that our methods are effective in improving the quality of requirement specifications and design models.