

Extracting Domain Knowledge for Dialogue Model Adaptation

Kuei-Kuang Lin and Hsin-Hsi Chen

Department of Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan
hh_chen@csie.ntu.edu.tw

Abstract. Domain shift is a challenging issue in dialogue management. This paper shows how to extract domain knowledge for dialogue model adaptation. The basic semantic concepts are derived from domain corpus by iterative token combination and contextual clustering. Speech act is identified by using semantic clues within an utterance. Frame states summarize current dialogue condition and state transition captures the mental agreement between users and system. Both Bayesian and machine learning approaches are experimented in identification of speech act and prediction of next state. To test the feasibility of this model adaptation approach, four corpora from domains of hospital registration service, telephone inquiring service, railway information service and air traveling information service are adopted. The experimental results demonstrate good portability in different domains.

1 Introduction

Dialogue management provides a rich human-computer interaction which allows users to convey more complex information than a single utterance. Despite of the recent significant progress in the areas of human language processing, building successful dialogue systems still requires large amounts of development time and human expertise [1]. The major challenging issue is the introduction of the new domain knowledge to the dialogue model when domain is shifted. That usually takes time to handcraft the domain knowledge that a dialogue manager needs. In the past, some papers [2,6] dealt with acquisition and clustering of grammatical fragments for natural language understanding; and some papers [4,9] employed statistical techniques for recognizing speech intentions. This paper emphasizes on how to extract crucial domain knowledge, including semantic concept extraction, speech act identification and formulation of dialogue state transition. Four corpora from different domains are employed to test the feasibility.