Natural Language Understanding Across Application Domains and Languages

Tsui Wai-Ching
徐 惠 貞

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Systems Engineering and Engineering Management

摘要

本論文主要是探討如何將信念網絡 (Belief Networks) 應用於自然語言的理解。我們利用了信念網絡 (BNs) 去擷取在特定範疇的說話當中，一些不同語意概念的配合與所帶出的意向之間的因果關係。基於所得的結果，我們就能推斷出一句說話的意向。其中一項研究是這個方法的範疇 (domain) 可移植性及語言 (language) 可移植性。我們做過由英文航空範疇移植到兩種語言(中文及英文) 的股票範疇的測試，都有不錯的表現。但是從一個範疇跳到另一個範疇或從一種語言跳到另一種語言往往會面對一個問題: 缺乏訓練的數據。在這種情況下，我們必須先成立有限度能力的自然語言理解系統，以幫助數據的收集。針對這個傳統的雞與卵問題，我們提出了一系列的法則爲信念網絡選取語意概念及爲其或然率賦值。我們亦証實這個語言理解模型與語音識別系統的可給合性。於這綜合系統內，我們的語言理解部件學習耦合的語音識別系統的識別樣式。利用語音識別器所提供的資訊，我們的語言理解模型可接收語音識別器就每個輸入回覆的多個最佳假設，並於話音理解過程中編入置信指標。我們的綜合模型的應用系統，在話音理解方面，比基準測試更有合理的表現。

Abstract

This thesis describes methodologies for porting a natural language understanding (NLU) framework across domains and languages, and integrating the framework with speech recognition component in a conversational system. Approaches towards language understanding usually involve much handcrafting, for instance, in writing grammars or annotating corpora, hence portability is a desirable trait in the development of language understanding systems. Belief Networks (BNs) have been previously explored to be used in the field of natural language understanding. In this thesis, we devise a methodology for porting a BN-based NLU framework across domains and languages. Our approach involves parsing the user's input query with a semantic grammar, followed by informational goal inferencing using BNs. The approach has been proven to be effective in understanding English queries in the ATIS (Air Travel Information Systems) domain. We apply the approach to the English ISIS (stocks) domain and extend to process the Chinese stocks queries and has shown promising results. With the high proportion of common concept categories sharing across the two languages, a large amount of grammar rules are reusable as we port from English to Chinese. Porting an NLU framework to new domains or languages often implies another problem – the lack of domain-specific training data. To enhance the NLU development with mini-

mum data requirement, a set of principles are devised for semantic concept selection and probability assignment in the BNs. We have also illustrated the ease of integration of our NLU framework with a speech recognition system. In the integrated framework, our NLU framework is extended to receive the N-best hypotheses generated by the speech recognizer and incorporate the speech recognition confidence scores into the spoken language understanding process. Application of our integrated framework on the spoken language understanding achieved statistically significant improvement over the benchmark algorithm.

Acknowledgments

I would like to take this opportunity to thank all those who have helped in the production of this thesis.

First of all, I would like to express my deepest gratitude to my thesis advisor, Prof. Helen Meng for her constructive criticism, guidance and support. Without her patient help, this work cannot be finished. Also, I would like to thank for my ex-advisor Prof. Boon-Toh Low. It has been a great pleasure for me to work with him.

I wish to thank the members of my thesis committee, Prof. Li-de Wu from Fu Dan University; Prof. C. H. Cheng and Prof. Jerome Yen at the Department of Systems Engineering and Engineering Management from CUHK, for their careful reading of my thesis and their precious comments given on this thesis.

My special thanks go to Carmen and Kon Fan for sharing their experience and expertise on the use of Belief Networks. Thanks to my friends Sally, Tony, Kin, Fat Yuk, Tiffany, Brenda and Ada, for sharing with me when I am frustrated by the research. Thanks also go to Bunny, Patrick, Philip, Ah Fan and the rest of the HCCL members for making my postgraduate studies joyful and unforgettable.

Finally, I would like to express my most heartfelt gratitude to Ka Fu, my parents and my sisters for their endless support, understanding, sharing, patience, encouragement and love. Without their warm care, the thesis would not reach its current stage.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

With the expanding population of computer users over the world, we have progressed into an information age. Previously, transaction operations and information inquiries used to be performed via menu-driven interfaces. For example, online applications through web browsers, touch-tone telephones, touchscreen interfaces on computers, etc.. The primary disadvantage of navigating menus is the time it takes by users to go through all the options and the difficulty of matching the users' goals to the options. In recent years, the popularity of such automatic services enabling human-computer interactions has been increasing dramatically and their usage has penetrated different application areas. But at the same time, people also have higher expectations on the service quality. They pursue more efficient and comfortable interaction modes with the machines. Natural language is one of the most natural and flexible ways that people can communicate with computers. Since natural language is an indispensable means of human-human communication, minimal training is required for using natural language human-computer interfaces. The ease-of-use feature of the natural language interface motivates researchers' interest in developing more and more effective language understanding technologies. Nowadays, a large number of conversational systems were emerged [17, 64]. Conversational systems are sometimes referred as spoken language systems. They provide conversational interfaces to humans. With the use of conversational interfaces, people can easily interact with the computers as if they are interacting with a human operator.

The natural language understanding (NLU) component is one of the prominent technologies in a conversational system. Within a restricted domain of expertise, NLU allows users to freely communicate with computers in their own languages, in both spoken and typed forms. NLU not only spots the keywords or key phrases in the user's request, but also classifies it into an appropriate domain-specific informational goal according to the underlying intention of the user's information-seeking request. One of the well-known applications of NLU technology is the call routing services. In these services, NLU determines the informational goal of the caller's request. The call is then routed to the appropriate destination as directed by the informational goal. Examples include the AT&T's "How May I help you?" task [2, 21] and Bell Laboratories' "How may I direct your call?" financial services for a call center [7].

NLU technologies are usually restricted to domain-specific and language-dependent tasks. Approaches towards language understanding generally require a parsing procedure to transform a query into semantic concepts. To map the transformed concepts to the appropriate informational goal, one may make use of some domain-specific heuristic rules. They are usually hand-designed by domain and linguistic experts. However, the task of writing heuristic rules is daunting and expensive, and often forms a major bottleneck in the development of language understanding systems.

In view of the disadvantages of applying a rule-based approach in NLU system development, automatic techniques for NLU development are desired by many. Previously, a NLU framework which involves semantic tagging and informational goal inference using BNs[1] (BNs) [48] was developed [37] for the English Air Travel Information Systems (ATIS) domain [50], and the approach showed promising results. The BN uses machine-learning techniques to automatically capture domain-specific knowledge and offers a number of advantages in handling NLU problems. In Section 1.2, the merits of using BN in performing NLU tasks will be fully discussed. In our NLU framework, we use BNs to capture the causal relationships between the semantic concepts parsed from the user's query and its underlying informational goal(s). By observing the presence and absence of the semantic concepts in the query, the BNs are able to infer the underlying informational goal(s) of the query. In this thesis, we explore the portability of this BN-based NLU framework across both application domains and languages, from the English ATIS domain to the English ISIS[2] stocks domain [34, 35], and from the English language to the Chinese language. The ISIS stocks domain offers additional complexities that were not encountered in the ATIS domain. One is semantic disambiguation, and another is the accommodation of out-of-vocabulary (OOV)

[1] Belief Networks is also known as Bayesian Networks.

[2] ISIS abbreviates Intelligent Speech for Information Systems. It is the name of a trilingual spoken dialog system (SDS) for the financial domain.

words, i.e. words that are absent from our grammars. Verbalized numeric expressions occur frequently in the stocks domain, and they may correspond to a number of possible specifications, for example, number of shares, lot size and price, etc. It is necessary to disambiguate among these specifications for accurate interpretation of the user's query. In order to handle these two issues, transformation-based parsing technique that utilizes local context as well as constraints for semantic disambiguation and OOV interpretation has been adopted in this thesis.

Usually, porting an NLU framework to a new domain or a new language implies the problem of the lack of training data. However, data collection is time-consuming and expensive. It hinders NLU development. To reduce the data requirement, another focus in our thesis is to investigate the alternative approach for hand assigning the probabilities in the BNs of our framework.

Moreover, with the widespread use of conversational systems, integration between the speech recognition and language understanding technologies becomes another research interest. In Section 1.3, the architecture of a typical conversational system will be introduced. Besides, we will describe the challenges faced by the researchers in spoken language understanding. In this thesis, our BN framework will be extended for handling spoken inputs. Under the extended framework, our BNs can process N-best outputs from speech recognizers and possess higher tolerance against speech recognition errors.

1.2 Natural Language Understanding Using Belief Networks

In the last section, we have pointed out the deficiencies of rule-based approaches in handling natural language understanding problems, such as extensive handcrafting for grammar rules. Our work in this thesis is motivated by the idea of automatically capturing the semantic / syntactic constraints from the real data for natural language understanding. Real data contains realistic samples that reflect the actual input that the NLU system will receive and be asked by the users. Capturing domain-specific constraints from real data is supposed to better model real world situations. Moreover, as the captured domain constraints generalize the real world cases as covered in the training data, when compare with those rigid rule-based approaches, machine learning approaches are more robust in handling unseen cases.

Machine-learning (or data-driven) approaches for natural language understanding is desirable as it often provides a reasonable coverage of the domain being modeled. With the availability of training data, data-driven approaches can be easily ported across different restricted domains and languages. BN is an example of machine-learning approaches and can be applied in natural language understanding with the following advantages [24]:

1. The dependencies between a query's informational goal(s) and the relevant semantic concepts can be effectively captured in the BN topology.

2. BNs identify the informational goal by means of probability inferencing. When massive data is available, this provides an attractive alternative to handcrafting the heuristics between parses and their interpretations.

3. BNs can handle incomplete information, which may be the consequence of imperfect speech recognition. Thus, BNs can be applied to handle spoken queries.

4. Optional incorporation of prior knowledge is enabled in the BN framework to aid the inference process.

With the merits of BNs, a BN-based NLU framework was developed previously [37]. In this work, we will investigate the ease of portability of such BN-based NLU framework across both application domains and languages.

Statistical approaches to natural language understanding require much data for development, while heuristic approaches have high demand on the completeness of domain knowledge handcrafted in the heuristic rules. BNs in conjunction with Bayesian statistical techniques facilitate the combination of domain knowledge and data. Anyone who has experienced statistical approach developments understands the importance of prior domain-specific knowledge, especially when the data is scarce and expensive. As a matter of fact, data collection for system development is a catch-22 problem [19]. In order to collect data that reflect the actual usage, one needs to have a real system that users can speak to. Therefore, for a new domain or language, a NLU system with limited capabilities must first be developed to support an "experimenter-in-the-loop", or wizard data collection paradigm to collect data for the further development. Once the system begins to be mature, it can then be switched to the "system-in-the-loop", or wizardless paradigm to collect more data to enhance system refinement (Figure 1.1). Prior knowledge is an essence for developing the limited capabilities in a prototype system. In BNs, causal semantics make the encoding of causal prior

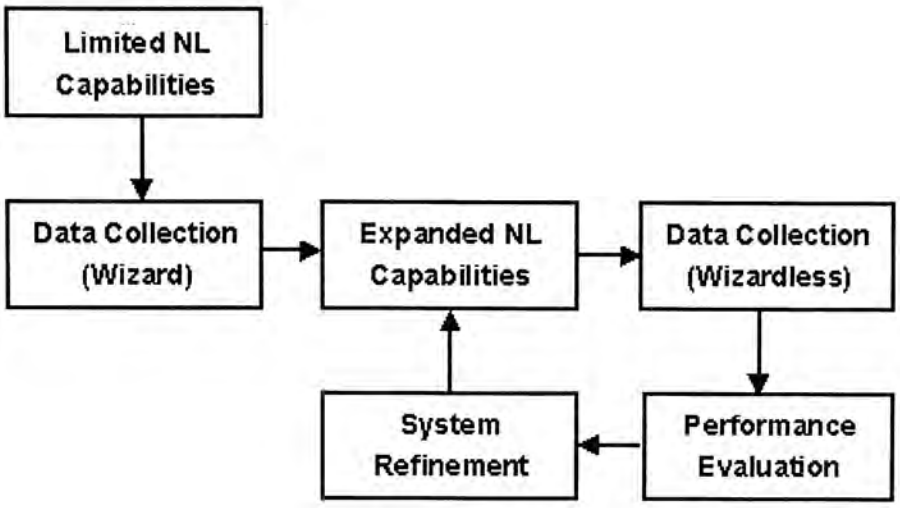

Figure 1.1: Illustration of data collection procedure.

knowledge straightforward. It drives us to devise a set of principles for the incorporation of prior knowledge into the BNs. Following these principles, we can hand-assign the statistical probabilities in the BNs.

1.3 Integrating Speech Recognition with Natural Language Understanding

Conversational interface combines several human language technologies to enable information access and transactional processing using spoken dialogue. The most popular technologies include speech recognition, natural language understanding and speech generation. A generic block diagram of a typical conversational interface is depicted in Figure 1.2. It illustrates the collaborations among the different speech technologies in a single conversational system.

In a conversational interface, many interesting research issues can be found. They are not only restricted to one specific component, but can

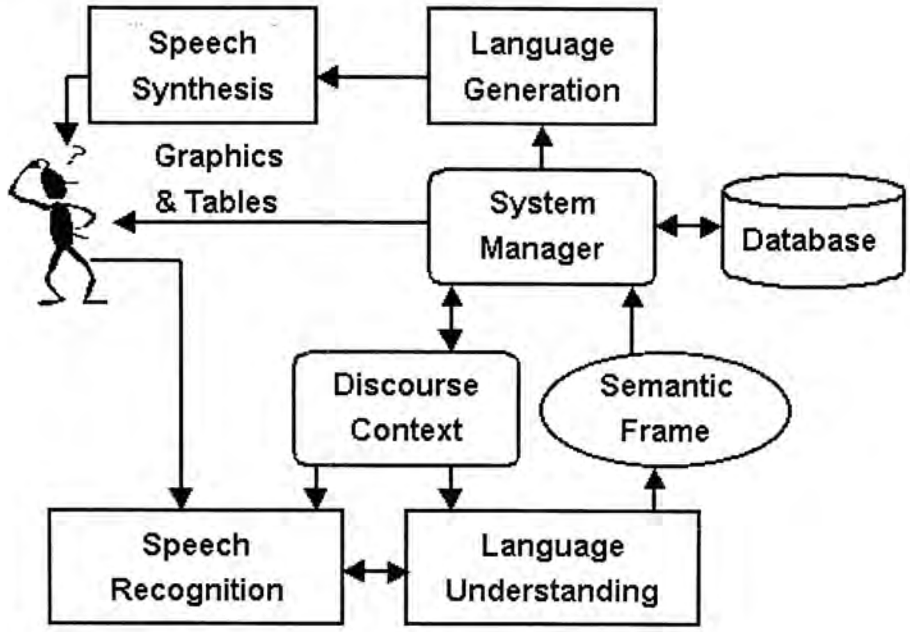

Figure 1.2: A generic block diagram of a typical conversational interface [64].

also be the integration between any pair of the technologies, or even the overall system architecture. As another focus in this thesis, we will investigate the integration between the speech recognition and the natural language understanding technologies.

In an integrated framework for speech recognition and natural language understanding, the NLU component is responsible for inferring the underlying meaning of the recognized output from the speech recognition component. One major concern in such an integrated framework is how the NLU component integrates with the speech recognition component. As a matter of fact, speech recognition technology is still far from perfect. Although many speech recognizers nowadays are able to output N-best hypotheses for each spoken input, the correct hypothesis cannot be guaranteed to receive the highest scoring and so be put at the toppest ranking. Therefore, many simple approaches that work only on the single best hypothesis may not be suitable all the times and so have room for improvement. One possible extension for spoken language understanding is to utilize the semantic / syntactic knowledge acquired during the natural language understanding process for re-ranking the recognition hypotheses. By integrating the knowledge from both the speech recognition and natural language understanding components, the hypotheses are rescored. Our work aims to tighten the relationship between the speech recognition and the natural language understanding in a conversational system. The BNs in our NLU framework are customized with the speech recognizer under investigation. We utilize the speech recognition information provided by the recognizer, such as the hypotheses that captures the recognizer's recognition behavior and the recognition confidence scores generated for each of the hypotheses, in the training and informational goal inference process of the NLU. By doing so, we strive to improve the NLU performance for spoken inputs.

1.4 Thesis Goals

The success in previous work [37] is a testament to the feasibility of using BNs for natural language understanding. In this thesis, we will explore the portability of this BN-based NLU framework across both application domains and languages. The ease of portability is illustrated by porting our NLU framework from the English ATIS (Air Travel Information Services) domain to the English ISIS (Intelligent Speech for Information Systems) stocks domain, and also from the English language to the Chinese language. After developing an appropriate and portable framework, unavailability of data is often a bottleneck that hinders the NLU system development. But with the merits offered by the causal semantics of BNs, development of NLU framework from prior knowledge is supported. A set of principles are devised for incorporating the prior knowledge into BNs through probability assignments. Following the principles, a NLU system is developed with the reduced data requirements and the NLU performance is proved to be comparable to the trained framework. Our BN-based NLU framework is also extensible for spoken language understanding. We will show that our proposed approach for integrating a speech recognition system with our NLU framework can drive to the better performance in spoken language understanding.

1.5 Thesis Organization

The rest of this thesis is organized as follows. Chapter 2 reports on previous work related to the NLU, portability issues of NLU across both application domains and languages and also, the integration of speech recognition system with NLU system.

In Chapter 3, we detail our NLU framework. Our framework for natural language understanding couples semantic tagging with BNs for informational goal inference. Then, we move to illustrate the portability of our NLU framework across both application domains and languages. The complexities caused by lexical ambiguity, semantic ambiguity [29] as well as the presence of out-of-vocabulary words encountered in the ISIS stocks domain will be discussed.

In Chapter 4, we propose a solution to the problem of the data sparseness often faced by NLU system developers when they develop systems for new domains or new languages. We devise guidelines for concept selection and probability assignments in the BNs. Benchmark experiments and the error analysis for our approach will be included in this chapter.

Chapter 5 describes our proposed algorithm to integrate a speech recognition system that outputs N-best hypotheses with our NLU framework. Experiments showing the improvement obtained by our approach over the simplest approach will be reported. A detailed analysis on our approach will also be included.

In the last chapter, we will give a summary on our research findings. Some possible research directions on the issues of NLU, and its integration with speech recognition systems will be covered.

Chapter 2

Background

The ultimate goal of natural language understanding is to generate the meaning representation encoding the full meaning of users' requests from both spoken and typed inputs. With the development of different approaches for natural language understanding, the portability ease of language understanding algorithms becomes another main concern. Moreover, being integrated with a speech recognition component in a conversational system, the handling of speech recognition output has challenged NLU developers for many years. Speech recognition outputs may often contain disfluencies, unknown words as well as recognition errors. Therefore, NLU approaches towards spoken language should be robust.

Natural language understanding for both text and spoken inputs and the portability of different language understanding frameworks are popular research topics in the area of speech and language technologies. Different approaches have been proposed for this area in the past decades. In this chapter, we will describe the background information regarding this research area. In Section 2.1, the previous work on natural language understanding will be reviewed. In Section 2.2, some examples on porting different NLU frameworks across application domains and languages will be covered. In Section 2.3, several approaches on spoken language understanding will be presented. Since we are going to demonstrate the portability of a BN-based NLU framework across application domains and languages in this thesis, the background information of Belief Networks will be provided in Section 2.4. As introduced in the previous chapter, the transformation-based parsing technique is adopted in our work to disambiguate complexities abounded in the ISIS stocks domain. In Section 2.5, we will present the general idea of the transformation-based parsing technique.

2.1 Natural Language Understanding Approaches

Natural language understanding determines the meaning of users' inputs and builds the appropriate semantic representations. To perform a natural language understanding task, researchers need to address the choice of a grammar formalism that is well adapted to the application and the practical issue of how to implement the grammar. Grammar formalism serves as a basis for the extraction of sentence constituents. Nowadays, the two most commonly used grammar formalisms are namely the *Syntactic Formalism* and *Semantic Formalism*.

Natural language understanding has been traditionally dominated by syntactic approaches. Some examples can be found in [5, 9, 12, 16, 25, 33]. In these approaches, each sentence is parsed into a sequence of syntactic constituents, such as np (noun phrase), vp (verb phrase), det (determiner), etc.. The parsed syntactic structure provides interpretation of the meaning of the sentence. However, a complete syntactic analysis needs to account for all words in an utterance. When working with spontaneous speech, such an approach may break down dramatically in the presence of recognition errors and various spontaneous events such as extragrammaticalities, interjections, false starts, hesitations, repetitions, self-corrections, ellipsis and interruptions. Besides, writing and modifying syntactic grammar rules are usually hard for non-specialists and take a lot of time. As a result, many researchers exclude pure syntactic algorithms, in favor of semantic-driven approaches [1, 13, 40, 42, 52, 64].

Semantic-driven approaches give direct explanations of the meaning of the sentence in terms of the words spoken and the relations between the words. Previous examples adopting semantic-driven approaches in natural language understanding include [8, 13, 27, 30, 39, 40, 41, 42, 43, 47, 51, 52, 61, 62, 63]. In these examples, semantic approaches analyze the sentence structure at a semantic level rather than a syntactic level. They define primarily semantic constituents by using semantic labels to abstract words and using syntax only to a minor degree. It results in semantic approaches can get rid of the constraints provided by syntax, and so are more adaptable to spontaneous human-machine interaction. Moreover, semantic approaches enhance the construction of semantic representations, which are essential elements for database retrievals.

Sentence parsing refers to the analysis of a sentence with respect to grammar rules in order to dissect the sentence structure [22]. The "parsed query" can be the output of a grammar-based parser [8, 53], probabilistic recursive transition networks [51] or an ergodic HMM [13, 41, 42], etc.. To interpret the meaning of the sentences, two methodologies may be applied. They are the *rule-based* approaches and the *stochastic* approaches and will be discussed in Section 2.1.1 and 2.1.2 respectively.

2.1.1 Rule-based Approaches

Previous work adopting the rule-based approaches can be found in [3, 13, 27, 28, 40, 52, 53, 63]. The simplest approach involves direct mapping of the parsed utterance into a pre-defined template for meaning interpretation [27, 28, 40, 52, 53, 63]. The semantic sequence can also be mapped to the semantic frame with domain independent rules [13]. In this approach, by knowing which template is best matched to the query, the topic of the utterance can then be automatically determined. But the two steps may be swapped in some approaches. For instance, the utterance theme is determined prior to the mapping of the semantic frames, and the theme is then used to decide which template should be used for the meaning representation [3].

A pure rule-based approach requires domain and linguistic experts to define all the necessary rules. It tends to be time-consuming and error prone, because understanding results are very often inconsistent if the rules have been made by two different persons or even by the same person in different periods. Besides, when the domain coverage increases, the rules become more complex and non-manageable and thus, the maintenability of the rules will be in question. However, rule-based approaches can be adopted without the need of large amounts of training data and it would be an advantage for new applications, when large amounts of annotated training data are unavailable.

2.1.2 Stochastic Approaches

Stochastic approaches have been explored in order to incorporate automatic learning methods for NLU. The relations between the semantic information encoded in the corpus are automatically learned from large amounts of training data and stored in form of parameters. There exists a number of different techniques modeling the stochastic approaches. They include *Hidden Markov Models (HMM)*, *Hidden Understanding Models (HUM)*, *Decision Trees* and *Bayesian Belief Networks (BNs)*.

Ergodic Hidden Markov Models have been adopted in the work of [13, 41, 42]. In the HMM models, the hidden HMM states are modeled by the semantic labels s_j. Since all states can follow each other, the model is ergodic. The modeling consists of maximizing the conditional probability $P(S|O)$ of some state sequence S given the observation sequence O. The pre-processed words are defined as the observation o_m. To understand the observed word sequence O, the semantic concept sequence S that maximizes the posterior probability $P(S|O)$ is found.

A similar formulation called Hidden Understanding Models (HUM) is proposed in [39, 51]. Given a string of input word W and a discourse history H, the task of the statistical language understanding system is to search among the many possible discourse-dependent meaning M_D for the most likely meaning M_O as in Equation 2.1.

$$M_o = \arg\max_{M_D} P(M_D|W, H) \tag{2.1}$$

The application of decision trees in natural language understanding has been explored in [33]. In this approach, decision trees function as decision-making devices that assign a probability to each of the possible choices based on the context of the decision $P(f|h)$, where f is the future vocabulary representing the set of choices, and h is the history as the context of the decision. The probability $P(f|h)$ is determined by asking a sequence of questions about the context, each is presented as a tree node in the decision tree. Each of the possible answers to the questions is associated with a branch extending from the node corresponding to the question. By assigning a probability distribution to the possible choices, decision trees provide a ranking system which not only specifies the order of the choices, but also gives a measure of the relative likelihood that each choice is the one should be selected.

BNs are applied in natural language understanding to capture the relationships between the semantic concepts and the informational goals of the queries. With the acquired knowledge on the restricted domain, the BNs are able to classify the underlying intention encoded in each of the users' queries. This framework has been applied to the English ATIS domain and reported promising results [37]. In this thesis, one of the focus is to assess the ease of portability of the statistical BN-based NLU framework. In Section 2.4, the background information on the BNs is provided.

Statistical approaches are more robust in dealing with spontaneous speech input. The automatic learning capability of statistical NLU frameworks facilitate their portability to both new domains and languages. However, statistical approaches are rather expensive because a lot of labeled data are needed. It usually forms the bottleneck for the NLU development for the new applications.

In this thesis, we will work on the statistical natural language understanding with the use of BNs. Similar to other statistical approaches, a large amount of training data is required for the training of the BN probabilities. In an attempt to ameliorate the reliance of training data, we have also devised a set of principles for the BN probabilities estimation in this work.

2.1.3 Mixed Approaches

Based on the pros and cons of each of the rule-based and stochastic approaches, a hybrid approach has been proposed in [8]. In this approach, an understanding module, which is composed of a specialized decision tree called Binary Classification Trees (BCTs) and a Rule-based Module (RBM), is adopted. BCT is a statistical algorithm. It is a binary tree in which each internal node has a label with a YES/NO question and two children, one for a positive answer to the question and one for the negative one. A path starting from the root and ending in a leaf is defined for each input sentence by answering the question for the node applied to the sentence. A class is associated to each leaf: the output is given by the class associated to the end of the path. Similar to other statistical approaches, BCT requires a lot of labeled data for development. However, data are not always available. In view of this, RBM is responsible for dealing all the phenomena for which the data are insufficient to properly train the statistical part. It in turns that the amount of knowledge to be put into the RBM is in fact decided on the basis of the amount of data available.

2.2 Portability of Natural Language Understanding Frameworks

NLU technology is usually restricted to domain-specific and language-dependent tasks. With the more NLU approaches developed, the applicability of a NLU framework across both the application domains and languages is desirable. With the brought on of the NLU technology, portability, which is measured as the efforts to move from one application to another, has gained more and more attentions from the researchers. In this section, we will review some of the previous work done in this research area.

2.2.1 Portability across Domains

The earlier work addressing portability issues of NLU systems across domains mainly focus on the portability of grammar rules. Most probably, syntactic grammar rules are ported from one domain to another, owing to the generality of the syntactic grammars [25, 53]. Examples include the TINA NLU system described in [53] and the PUNDIT natural language processing system reported in [25]. TINA has been ported from the TIMIT sentences to the Resource Management (RM) task [45]; while PUNDIT has been ported from a range of messages to the Resource Management domain. During the grammar conversion process, new sentences are parsed one by one with the grammar of the original domain. Whenever a parse fails, new grammar rules are added. Once the grammar has been expanded to accommodate the new set of sentences, a subset grammar can be created automatically that only contains the rules needed in the new domain, eliminating any rules that were particular to the original domain.

The automatic learning capability of a statistical approach facilitates the portability of NLU framework. One example is the HMM-based NLU framework described in [13]. With the semantic lexicon and the annotated training corpus prepared for the new domain, the HMM-based NLU framework was easily ported from the hotel reservation domain to the travel information accessing domain.

2.2.2 Portability across Languages

With the fast growing of multilingual human-computer interfaces, information can be accessed globally by allowing people to communicate with the computers in their native tongues. In view of this, more focus is put on the research work regarding the portability of NLU across different languages. Some related work are reported in [18, 40, 59, 60, 62].

Porting an understanding framework across languages usually involves the translation of the grammar rules from one language to another language. By doing so, Minker [40] has successfully ported the NLU system from French ATIS to American English ATIS. TINA [53] was originally designed for the English language, but has been incorporated into many systems to support different languages. For example, the GALAXY system [20] was ported from its English version to its Mandarin version named as "YINHE" [59, 60]. Also, the MIT VOYAGER system has been ported to Japanese and Italian [18].

To minimize the effort involved in mapping the semantic frame for different languages, Ward [62] has proposed the use of unified formal language. The statistical parsers that are trained automatically from multi-lingual natural language sentences, along with a single formal representation for the meaning of the sentences. Though the statistical NLU models are trained separately for each language to capture the different language constructions, the outputs are language independent. This framework has been ported from the English language to the French language.

2.2.3 Portability across both Domains and Languages

Some previous work on porting the NLU frameworks across both domains and languages can also be found in [41, 54].

In [41], the statistical HMM-based NLU component was migrated from the English ATIS system to the French MASK (Multimodal-Multimedia Automated Service Kisosk) system. Due to the similarity of the ATIS and MASK domains, equivalent query preprocessing and semantic decoding strategies were applied to the ported MASK domain. Correspondences between the labels and the words are automatically learned from a large annotated training corpus and memorized in the form of statistical model parameters. The TINA NLU system [53] was adopted to support the GALAXY system development [54]. The system can field spoken and typed questions in four domains - weather information, flight information, air travel planning and navigation, and in multiple languages.

2.3 Spoken Language Understanding

A conversational system should be able to recognize the words that are spoken by a user and give appropriate response to the user's request. To achieve high accuracy of the conversational system, the most ideal case is to have a perfect speech recognizer and a robust NLU component in the system. However, unfortunately, today's speech recognition technology is far from perfect and error-free. Speech recognition errors cause the NLU component to fail or give an incorrect interpretation. To alleviate the understanding problem originated from mis-recognition, different strategies have been developed to detect when recognition errors have occurred and take appropriate response to recover from these errors.

2.3.1 Integration of Speech Recognition Confidence into Natural Language Understanding

One of the essential approaches is to integrate the speech recognition information into natural language understanding process. Speech recognizers output their recognized results in form of word graphs [44]. When examining the recognition results, confidence scores for the recognition output can be computed on various levels, including the phonetic level, the word level and the utterance level. N-best hypotheses [11] may be devised from the word graph, for example, by the maximization of the posterior probability (MAP) method [4].

2.3.2.1 Word Graph Recognition Output

Word graphs provide a compact representation for the recognition results. Each node in the word lattice, associated with a confidence score, refers to a hypothesized word. Every path in the word lattice represents a probable hypothesis for the input query. Fetter [14] has suggested an approach to re-score the word confidences in the word graph. Each word confidence is re-scored as the weighted sum of the original acoustic score and the probability of the word to have that acoustic score.

The advantage of word graphs is complete coverage on the possible hypotheses. However, the large coverage dramatically increases the operation time in searching across the word lattice to collect the embedded hypotheses.

2.3.2.2 N-best Recognition Outputs

Due to the complexity imposed by the word graphs in searching the hypotheses, N-best interfaces are preferred by the most researchers. Previous work that involve the integration of N-best hypotheses with language understanding include [23, 26, 58]. In these approaches, the utterance-level acoustic score generated by the recognizer was linearly combined with the score given by the NLU system, for example, the TINA parser in [26] and the BNs in [58]. Word-level confidence scores generated for each word in the N-best hypotheses are used for detecting potential recognition errors [23]. The word with confidence score lower than a certain threshold is either rejected by the system or confirmed by the user, depending on the rejection strategies adopted.

N-best interface narrows the coverage of possible hypotheses, but can highly speed up the searching process in the word graph. However, the number of hypotheses (N) can be flexibly tuned to each application so as to achieve the optimal performance.

2.3.2 Integration of Other Potential Confidence Features into Natural Language Understanding

Other than the speech recognition confidence scores, other confidence features have been adopted by the researchers. For instance, linguistic and application specific features [46], expected word error rates [56], semantic class language models [15] and exemplary concept graphs [32]. Combined with the acoustic score, the hypothesis receiving the highest overall confidence score is selected and passed to the language understanding system for further process.

2.4 Belief Networks

2.4.1 Overview

Belief Network (BN)[1] is one of the popular choices for uncertain reasoning [49]. BN is a directed acyclic graph consisting nodes and directed arcs as shown in Figure 2.1. The nodes in a BN represent propositional variables of interest, while the directed arcs represent informational or causal dependencies among the variables. BNs allow the causal dependencies between the variables to be easily interpreted. The dependencies are quantified by conditional probabilities for each node given its parents in the network. For instance, for any variable v with parents p_1, p_2 ... p_n, the conditional probability is defined as $P(v|p_1, p_2 \ldots p_n)$. The prior probability $p(v)$ is obtained if v has no parents.

Consider the example extracted from [49] as demonstrated in Figure 2.1.

[1]Belief Networks are also known as Bayesian Networks, Bayesian Belief Networks and Causal Networks

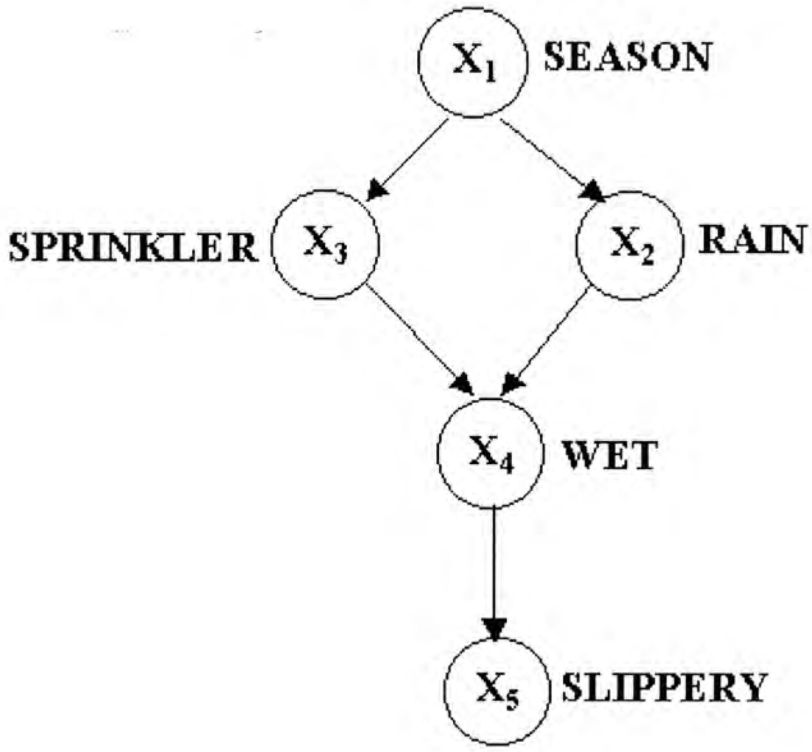

Figure 2.1: A Belief Network depicting the causal relationship among the 5 variables [49].

The BN illustrates the causal relationship among five variables as represented by the five nodes. They are the season of the year (X_1), whether it is raining (X_2), whether the sprinkler is on (X_3), whether the road is wet (X_4), and whether the road is slippery (X_5). Each node exhibits a Boolean value, the value is true if the represented event takes place. Directed arcs are found to link the different variable pairs. For example, the directed arc linking from X_2 to X_4, depicts the fact that "raining makes the road wet". Moreover, we observe that there is no direct linking between X_1 and X_5. The absence of the linking illustrates the fact that the season cannot have direct influence on the slipperiness of the road.

As one of the powerful features of BNs, they give the direct representations of the world, but not the reasoning processes. The directed arrows in the diagram represent only the causal relationship, but not the flow of information. Reasoning processes can operate on the BNs by propagating information in any direction. For example, if it rains, then the road is probably wet; if someone slips on the road, it provides evidence for the road is wet.

Any complete probabilistic model of a domain must represent the joint distribution. By applying the conditional independencies in the BNs, the joint probabilities can be obtained by factoring into local, conditional distributions for each variable given its parents:

$$P(X) = \prod_{i=1}^{n} P(x_i|parents(x_i)) \tag{2.2}$$

So, the joint probability for the network shown in Figure 2.1 can be derived from the conditional probabilities as illustrated in Equation 2.3.

$$P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4) \tag{2.3}$$

where x_i is the value of variable X_i for i=1,2 ... 5.

2.4.2 Bayesian Inference

Many different algorithms are developed for Bayesian inference, many of them can be found in [48]. In this work, we focus on the simple usage of BNs, which is denoted as the naive Bayesian inference. Naive Bayesian Networks consist only of one parent and a number of child nodes as depicted in Figure 2.2. The theoretical foundation for Bayesian Networks is Bayes' rule. Consider the example shown in Figure 2.2, for the hypothesis node H, we have three evidence nodes namely E_1, E_2 and E_3, with the corresponding values e_1, e_2

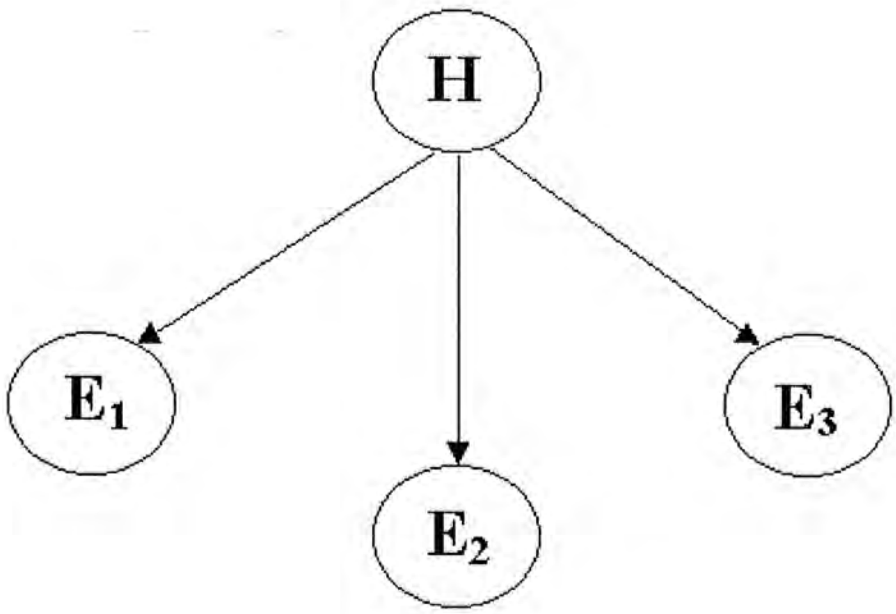

Figure 2.2: A Naive Belief Network.

and e_3. $P(h|e_1, e_2, e_3)$ is computed according to Bayes' Rule as shown in Equation 2.4.

$$P(h|e_1, e_2, e_3) = \frac{P(h, e_1, e_2, e_3)}{P(e_1, e_2, e_3)} \tag{2.4}$$

Since the value of h is mutually exclusive and exhaustive and following the Equation 2.2, Equation 2.4 can be further derived as:

$$P(h|e_1, e_2, e_3) = \frac{P(e_1|h)P(e_2|h)P(e_3|h)P(h)}{\sum_{h'=0,1} P(e_1|h')P(e_2|h')P(e_3|h')P(h')} \tag{2.5}$$

By substituting both the prior and the conditional probabilities into Equation 2.5, the a posteriori probability $P(h|\vec{e})$ can be evaluated.

2.5 Transformation-based Parsing Technique

In recent years, transformations have been applied to a diverse set of problems, including part of speech tagging, pronunciation network creation, prepositional phrase attachment disambiguation, and parsing.

Transformation-based systems are typically deterministic. Transformation rules are in form of an ordered list. Each rule is checked with the condition as specified and is applied whenever matched. Then the next rule is investigated until the last rule in the list has been processed. This algorithm allows the intermediate results from classifying one object to be available in classifying other objects. Transformation-based parsing (TBP) was first introduced in [5] for part-of-speech tagging problem. An initial state annotator assigns each word its most likely tag as indicated in the training corpus. An ordered list of transformations is then learned to improve tagging accuracy based on contextual cues. The transformation rules should be conform to any of the non-lexicalized and lexicalized templates as shown in Table 2.1 and Table 2.2 respectively.

In this thesis, the transformation-based parsing technique is used for disambiguating complexities and inferring the category of unseen words during the natural language understanding process.

2.6 Chapter Summary

In this chapter, we have given the background information for the research in this thesis. Previous work on natural language understanding are reviewed, including the natural language understanding approaches and the portability issue on the language understanding frameworks. We have also presented the general approaches for spoken language understanding. Next, we have given a brief introduction on the BNs, with which, our NLU framework was developed. Also, the idea of transformation-based parsing technique is covered at the end of this chapter.

Action	Change tag a to tag b
Template 1	The preceding word is tagged z.
Template 2	The following word is tagged z.
Template 3	The word two before is tagged z.
Template 4	The word two after is tagged z.
Template 5	One of the two preceding words is tagged z.
Template 6	One of the two following words is tagged z.
Template 7	One of the three preceding words is tagged z.
Template 8	One of the three following words is tagged z.
Template 9	The preceding word is tagged z and the following word is tagged w.
Template 10	The preceding word is tagged z and the word two before is tagged w.
Template 11	The preceding word is tagged z and the word two after is tagged w.
Template 12	The following word is tagged z and the word two before is tagged w.
Template 13	The following word is tagged z and the word two after is tagged w.

Table 2.1: Non-lexicalized transformation templates for the action "Change tag a to tag b". a, b, z and w are variables over the set of parts of speech.

Action	Change tag a to tag b
Template 1	The preceding word is w.
Template 2	The following word is w.
Template 3	The word two before is w.
Template 4	The word two after is w.
Template 5	One of the two preceding words is w.
Template 6	One of the two following words is w.
Template 7	The current word is w and the preceding word is x.
Template 8	The current word is w and the following word is x.
Template 9	The current word is w and the preceding word is tagged z.
Template 10	The current word is w and the following word is tagged z.
Template 11	The current word is w.
Template 12	The preceding word is w and the preceding tag is t.
Template 13	The preceding word is w and the following tag is t.
Template 14	The following word is w and the preceding tag is t.
Template 15	The following word is w and the following tag is t.
Template 16	The current word is w, the preceding word is w_2 and the preceding tag is t.
Template 17	The current word is w, the preceding word is w_2 and the following tag is t.
Template 18	The current word is w, the following word is w_2 and the preceding tag is t.
Template 19	The current word is w, the following word is w_2 and the following tag is t.

Table 2.2: Lexicalized transformation templates for the action "Change tag a to tag b". w, w_2 and x are variables over all words in the training corpus, while a, b, t and z are variables over all parts of speech.

Chapter 3

Portability of the Natural Language Understanding Framework across Application Domains and Languages

In this chapter, we will explore the portability of a Belief Network based natural language understanding framework across both application domains and languages. A framework for natural language understanding which involves semantic tagging and informational goal inference using Belief Networks (BNs) has been developed previously [37]. The approach has been applied to the ATIS (Air Travel Information Systems) domain in English, and has shown promising results. In the coming sections, we will first introduce this BN-based NLU framework. We will investigate the portability of this framework across both application domains and languages: from the

ATIS domain [50] to the ISIS (Intelligent Speech for Information Systems) stocks domain [34, 35] and from the English language to the Chinese language. Ambiguous entities such as the various numeric expressions and new words abound in the stocks domain, and present a challenge to the portability of our framework. We have analyzed the contextual information of the ambiguous entities, and proposed the transformation-based parsing technique for solving the problem. We found that this technique is effective in disambiguating among the various kinds of numeric expressions prevalent in the stocks domain, as well as infer possible semantic categories for out-of-vocabulary (OOV) words. Moreover, in order to port our NLU framework to the Chinese language, some special pre-processing mechanisms are described. Finally, experiments on the goal identification of our ported NLU framework will be evaluated.

3.1 Natural Language Understanding Framework

Figure 3.1 outlines our NLU framework. Our NLU system couples semantic tagging with BNs for informational goal inference. We use a semantic tagger to transform the input query into a sequence of semantic concepts. This forms the input of a BNs suite to infer the informational goal of the query. Each BN in the suite works as a classifier for an unique goal and outputs the a posteriori probability for each input. By comparing with the pre-defined threshold value, each BN is able to make a binary decision regarding the presence or absence of the goal. Combing the N-binary decisions from each of the BNs in the suite, we can identify multiple goal queries and reject out-of-domain (OOD) sentences.

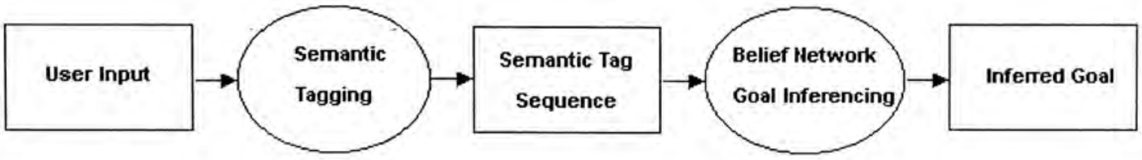

Figure 3.1: The natural language understanding framework

3.1.1 Semantic Tagging

Semantic tagging is the process to abstract the input queries into a set of semantic concepts based on some pre-defined grammar rules. We have written an English grammar and a Chinese grammar to support biliteral / trilingual input. This includes English and two Chinese dialects – Putonghua and Cantonese. Here are some examples of the English grammar rule and the Chinese grammar rule for the pre-terminal <StockNameHK> referring to the class of Stock Names found in Hong Kong:

<StockNameHK> → HSBC, Cheung Kong, Hopewell ...

<StockNameHK> → 匯豐,匯豐銀行,匯豐控股,長實,長江實業,合和 ...

The grammar rules tell the semantic tagger that when the words in the query match any of the listed terminals on the right hand side of the grammar rules, such words in the query should be tagged as the pre-terminal tag written on the left hand side. So, in our example, if the terms HSBC, Cheung Kong or Hopewell, etc. appear in an English query or 匯豐, 長實 or 合和, etc. appear in a Chinese query, then the tagger would tag these items in the query as <StockNameHK>.

Semantic tagging renders our approach robust towards extra-grammaticalities and spoken disfluencies, which are ignored at an early stage. The concept tags aim to match the attribute labels for database access, though some tags tend to be syntactic in nature (for example, <Article>). Sentences are automatically tagged using a transformational procedure. This allows the intermediate results from one grammar rule to be available in applying the later rules. The semantic tagger produces a concept sequence for every input query string. Some examples are shown in Table 3.1. Such kind of semantic concept sequences forms the input to our BNs.

GOAL	SELL
QUERY	放 五手 匯豐 係 一百蚊
TAG	<SellAction> <LotNumber> <StockNameHK> <At> <PriceValue>

GOAL	REQUEST_REALTIME_QUOTES
QUERY	What is the bid price of HSBC?
TAG	<Quest> <CompanyStat> <Of> <StockNameHK>

Table 3.1: Examples of semantic concept sequences output by the semantic tagger.

3.1.2 Informational Goal Inference with Belief Networks

Training queries that are tagged semantically are used to train a suite of BNs. A BN is developed for each domain-specific goal, to make a binary decision regarding whether the goal is present or absent in the query. The pre-defined (basic) topology, which is equivalent to the Naive Bayesian Networks, is shown in Figure 3.2.

For a network corresponding to goal G_i, we select the M concepts C_1, C_2 ... C_M which have the highest Information Gain (IG) with the goal G_i. IG measures the degree of association between a concept and a goal, taking into

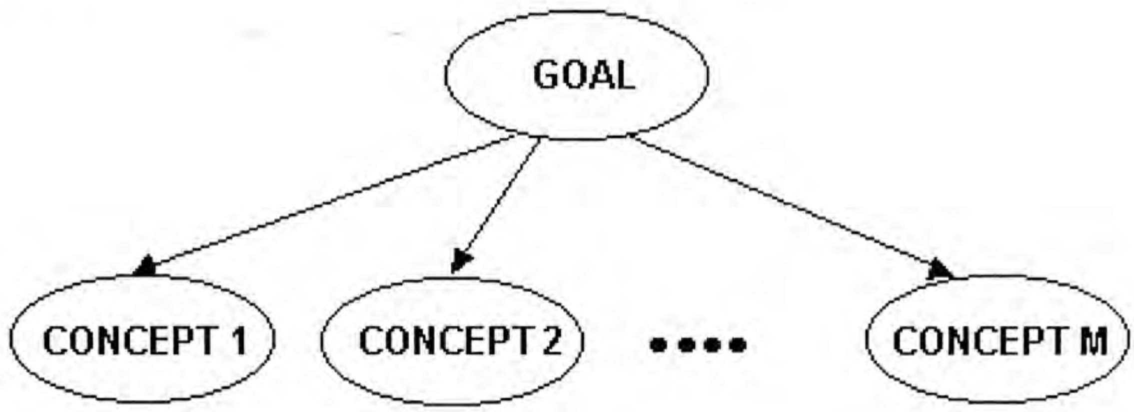

Figure 3.2: The pre-defined structure of our Belief Network. The arrows of the acyclic graph are drawn from cause to effect.

consideration both the presence and absence of the entities. The selected concepts form the input of the BN corresponding to the goal G_i. Equation 3.1 is the formula for the IG between a concept C_k and the goal G_i.

$$IG(C_k, G_i) = \sum_{c=0,1} \sum_{g=0,1} P(C_k = c, G_i = g) log \frac{P(C_k = c, G_i = g)}{P(C_k = c)P(G_i = g)} \tag{3.1}$$

The selected concepts are regarded as most indicative of the goal. Each BN then applies Bayesian inferencing as defined in Equation 3.2 to derive P($G_i|\vec{C}$).

$$P(G_i = 1|\vec{C}) = \frac{P(\vec{C}|G_i = 1)P(G_i = 1)}{P(\vec{C})} \tag{3.2}$$

This computation is simplified by assuming independent concepts as are captured in our basic BN topology.[1]

We develop one BN per informational goal. So, we have in total N BNs to represent a finite set of goals in the restricted domain. Each BN outputs its own a posteriori probability P($G_i|\vec{C}$) reflecting the relevance of the input query. The decisions across all the BNs are combined to identify the output

[1] A more complex topology is described in [36].

goal(s) of an input query. We may label the query with the (single) goal giving the highest value of P($G_i|\vec{C}$) across all BNs. Alternatively, we may label the query with all the goals for which the BNs voted positive – this achieves multiple-goal identification. In the case when all BNs vote negative, the input query is rejected as out-of-domain. In the latter scheme, the a posteriori probability is compared with a pre-set threshold θ_i in order to make the binary decision regarding the presence or absence of the goal G_i. The decision rule for each particular goal G_i is:

$$P(G_i|\vec{C}) < \theta_i \Rightarrow G_i = 0 \qquad \text{(Goal } G_i \text{ is absent)}$$

$$P(G_i|\vec{C}) \geq \theta_i \Rightarrow G_i = 1 \qquad \text{(Goal } G_i \text{ is present)}$$

We strive to achieve high values for precision and recall for G_i. Precision (P) is the percentage of queries in the test set correctly classified to belong to G_i, out of all the queries which are classified as G_i. Recall (R) is the percentage of queries in the test set correctly classified to belong to G_i, out of all the queries, which truly belong to G_i. We define the threshold value θ_i of goal G_i by combining Precision and Recall into a single score which optimizes with the F-measure [57] as shown in Equation (3.3).

$$F_\beta(\theta_i) = \frac{(1+\beta^2)RP}{\beta R + P} \tag{3.3}$$

where β=1 in our experiments to treat P and R with equal importance.

3.2 The ISIS Stocks Domain

ISIS abbreviates Intelligent Speech for Information Systems. It is a spoken dialog system in the stocks domain [34, 35]. It can handle typed queries in

Chinese or English, or spoken queries in Putonghua, Cantonese or English to support biliteral / trilingual input. We have chosen the ISIS stocks domain, in both English and Chinese, to demonstrate the ease of portability of our NLU framework previously developed for the English ATIS domain.

To aid the task of domain definition and grammar development, we collected some sample queries in both English and Chinese. We requested that our subjects compose questions that they will ask of a stock broker, such as questions on real-time stock quotes, or simulated investor accounts. In this manner, we collected 2048 Chinese queries and 2157 English queries and have split them into disjoint training and testing sets. Table 3.2 shows the distribution of our collected corpora.

	Number of Chinese Queries	Number of English Queries
Training sentences	1547	1614
Testing sentences	501	543

Table 3.2: Distribution of our training and testing corpora.

Here are some of our collected examples:

- *Amend my purchase order for HSBC from three to six lots please.*

- 如果和記黃埔升了兩塊錢,請給我賣出兩千股

  (Translation: If Hutchison Whampoa rises another two dollars, please sell two thousand shares.)

The stocks domain presents new complexities for NLU. Verbalized numbers abound in the domain-specific queries, and they can refer to stock codes (commonly used in Hong Kong), stock prices, number of lots, number of shares, etc. Consider the query:

"I would like to purchase Cheung Kong at ninety five a share."

Verbalized numbers are parsed to obtain their numeric values, and the numeric expressions are classified into the appropriate semantic category with considerations of both left and right contexts. We shall elaborate on this in Section 3.3.2. Another subtlety in this example is that "a share" really means "per share", and the user is not trying to purchase a single share of Cheung Kong.

Additionally, we have found that in the ISIS domain, contextual information may strongly influence semantic interpretation. Consider the following example pair in Table 3.3. The two examples are similar in that "two dollars"

Query 1 (Incremental example)
If Hutchison rises another $\boxed{\text{two dollars}}$, sell three lots for me.
Query 2 (Decremental example)
If Hutchison drops another $\boxed{\text{two dollars}}$, buy three lots for me.

Table 3.3: A pair of example queries illustrating the semantic ambiguity with the phrase "two dollars". The phrase refers to an incremental share price in Query 1 but a decremental share price in Query 2.

need to be decoded semantically as an incremental share price. However, it refers to an increment in the former example, and a decrement in the latter.

3.3 A Unified Framework for English and Chinese

In this chapter, we have extended the NLU framework to a unified NLU framework that is applicable to both the Chinese and English languages. There are three main components in our framework, namely the semantic tagging, transformation-based parsing and informational goal inference with BNs.

3.3.1 Semantic Tagging for the ISIS domain

We hand-designed a set of semantic tags (or concept tags) based on our English sentences (for example, <StockNameHK>, <ShareNumber>). Some syntactic tags are included as well (for example, <Article>). Had there been more data collected, we believe we could have applied a semi-automatic procedure for acquiring such structures from unannotated corpora [55]. The set of concept tags forms the pre-terminal categories of our English grammar, and the tags are designed and processed to match the attribute labels for subsequent database access. A real-time data capture component continuously updates a relational database based on a dedicated Reuters satellite feed. As an example, consider the stock "HSBC" with concept tag <StockNameHK>. According to invocations specified in our grammar, this tag automatically invokes a procedure that converts it into 0005.HK (which signifies that it is the stock 0005 from the Stock Exchange of Hong Kong). The new tag 0005.HK matches the RIC[2] code, and can be used directly for database access. Examples of other concept tags include <PriceValue>, <LotNumber> and <ShareNumber>.

Input queries in Chinese are first tokenized into words by means of a greedy algorithm together with a 1100-word lexicon. The lexicon currently covers the 33 constituent stocks in the Hang Seng Index. We maximized the reuse of English concept tags (hand-designed with reference to the English

[2]RIC stands for Reuters Instrument Code.

queries) as pre-terminal categories for processing the Chinese sentences we have collected. At this initial stage we are using a single Chinese grammar for both Putonghua and Cantonese queries. For example:

QUERY: 請問你有沒有長實的成交量

(Translation: do you have the trading volume of Cheung Kong.)

TOKENIZE WORDS: 請問你 有沒有 長實 的 成交量

TAG: <ShowWord> <HasOrNot> <StockNameHK> <Dummy>
<CompanyStat>

Our English and Chinese grammars have 164 and 177 preterminal tags respectively. Of these, 148 tags are common between the two grammars, achieving about 84% sharing thus far. Examples of language-specific tags include: English preposition tags like <At> (for English preposition "at"); <TeenDigit> (for English numbers like "eleven", "twelve", etc.) and <Twenty> and <Thirty> (for the Chinese numbers "廿" and "卅").

3.3.2 Transformation-based Parsing

As mentioned previously, the stocks domain presents new complexities for natural language understanding. Much ambiguity exists for numeric expressions, and we need to identify precisely to what they refer, be it share price, lot number, stock code, share number, etc. We can conceive of other complexities - for example, the request "I'd like to close my position on Intel" may be either a buy or sell request depending on the investor's holdings.

In order to handle these complexities in the ISIS domain, we applied the technique of *Transformation-based parsing (TBP)* [5]. Transformation-based parsing was proposed by Brill and has been applied to corpora such as the Penn Treebank. TBP has also been shown to achieve better performance than a corpus-derived PCFG [6]. In TBP, the grammar consists of a sequence of precedence-ordered transformation rules. To the best of our knowledge, this work is one of the first attempts to apply TBP towards natural language understanding. We found that the context-sensitive transformations offer a powerful solution for handling the ambiguities in our current domain. We illustrate with the following examples.

3.3.2.1 Handling Numeric Expressions

QUERY: Sell two more lots of Cheung Kong when it gains another two dollars per share.

TAG: <SellAction> <NumberValue> <More> <Lot> <Of> <StockNameHK> <Quest> <Dummy> <Gain> <More> <PriceValue>

In this query, "two dollars" is an incremental share price relative to the current share price. Our intermediary grammar rules will be able to parse "two dollars per share" to be a <PriceValue>. However the transformation rule:

RULE: <PriceValue> <+PriceValue> PREVBIGRAM <Gain> <More>

which specifies that we should change the concept tag <PriceValue> to <+PriceValue> (i.e. an incremental share price) if its previous concept bigram is <Gain> <More>.

A similar technique can be used to handle the phrase "two more lots" in the example query above. Here a pair of precedence-ordered transformation rules are applied:

RULE: <NumberValue> <LotNumber> NEXT1OR2TAG <Lot>

RULE: <LotNumber> <+LotNumber> NEXTTAG <More>

where the first rule indicates that the numeric expression should be a specified lot number if its following one or two tags is <Lot>; and the second rule changes the concept tag <LotNumber> to <+LotNumber> (i.e. the incremental lot number) if its following tag is <More>.

3.3.2.2 Handling Out-of-Vocabulary (OOV) Words

Language understanding is hampered by words that lie outside of the vocabulary specified in the grammar's terminal categories. However, we have developed a technique that makes use of the context-sensitive transductions to infer a possible concept tag for the <OOV>. To illustrate with an example, assume that PCCW (Pacific Century CyberWorks Limited) is a stock name which is an <OOV>.

QUERY: I'd like to check the news about PCCW over the past two weeks.
TAG: <ShowWord> <News> <RegardingWord> <OOV> <RelativeDate>

Here our semantic tagger has tagged "PCCW" as <OOV>. However we have specified the transformation:

RULE: <OOV> <StockNameOOV> PREVBIGRAM <News> <RegardingWord>

which states that an <OOV> tag may be inferred as a new stock name if it is preceded by the tag bigram <News> <RegardingWord>. Hence we have a conjecture regarding the semantic category of the <OOV>. Within the context of a dialog system, this partial information may contribute towards the generation of an intelligent response such as "Sorry, I do not know about the stock PCCW."

A similar technique is used to handle OOV words in Chinese queries. OOV words are often tokenized as a sequence of mono-character or bi-character units, for example,

QUERY: 我想查詢 桊電國際 最新價格
(Translation: I would like to ask for the latest share price of Global Link.)
TAG: <ShowWord> <OOV> <CompanyStat>

Since the stock name is not known by the system, its characters are tagged as <OOV>. We perform "n-gram grouping" whereby the entire sequence is grouped into a single unit to form "桊電國際", which is in turn given a single <OOV> tag. Thereafter, transformation rules for <OOV> tags (similar to the one presented in the English example above) are applied to transform the tag from <OOV> to <StockNameOOV>.

These transformation rules further modify our set of concept tags, and at this point there are 80 concept tags in our English grammar and 78 concept tags in our Chinese grammar. Among them only two are language-specific tags, which shows a high degree of sharing across languages. Henceforth, all our queries are represented as a sequence of such concept tags.

3.3.3 Informational Goal Inference with Belief Networks for the ISIS domain

We have modeled the ISIS domain with 10 domain-specific goals, they are as listed in Table 3.4. Then we went through our 1614 English and 1547 Chinese

REQUEST_CHART	REQUEST_NEWS
ASK_TRENDS	AMENDMENT
CANCELLATION	REQUEST_ACCOUNT_INFO
REQUEST_REALTIME_QUOTES	BUY
SELL	SYSTEM_META_COMMANDS

Table 3.4: The ten informational goals defined in the ISIS domain.

queries and annotated each with the most appropriate informational goal to form our English and Chinese training corpora. A small number of queries are labeled with multiple goals, examples are shown in Table 3.5.

QUERY	Please show me the daily chart of HSBC and the closing price of Hang Lung.
GOAL(S)	REQUEST_CHART, REQUEST_REALTIME_QUOTES
QUERY	請告訴我匯豐的現價還有給我在一百塊的時候購入二千股 (Translation: Please give me the latest share price of HSBC and buy two thousand shares when it hits a hundred dollars per share.)
GOAL(S)	REQUEST_REALTIME_QUOTES, BUY

Table 3.5: Example queries with multiple goals.

In order to acquire a disjoint test set, we collected another 501 Chinese queries and 543 English queries. These were also annotated with informational goals. Table 3.6 summarizes the statistics of our corpora. While out-of-domain queries are not used in training, a few are included in our testing corpora for investigating the capability of rejection. Examples of the OOD queries include:

- 請告訴我星加坡元兌港元的匯率

- *Give me the exchange rate for the Singapore Dollar against the Hong Kong Dollar. (Translation of the above)*

	Number of Chinese Queries	Number of English Queries
Training sentences	1547	1614
Single goal sentences (train)	*1530*	*1593*
Multiple goal sentences (train)	*17*	*21*
Testing sentences	501	543
Single goal sentences (test)	*483*	*532*
Multiple goal sentences (test)	6	6
Out-of-domain (test)	*12*	*5*

Table 3.6: Statistics of our training and testing corpora. Out-of-domain queries are not used in training.

3.4 Experiments

Our experiments are performed with our collected ISIS data sets so as to evaluate the performance of our NLU framework to a ported domain. As mentioned previously, we have in total defined 10 domain specific goals from the data sets. Each of the sentence can be classified into any of the three categories regarding the goals identified: the single goal queries, the mutliple goal queries and the OOD sentences.

3.4.1 Goal Identification Experiments

We have developed 10 BNs, one for each goal that we have defined for the ISIS domain. For each goal, we computed the Information Gain for all goal-concept pairs to select the subset of the concepts as input to the corresponding BN. Goal inference proceeds as specified in Equation 3.2. The threshold values that optimize the F-measure based on the training data are used for the BNs. The BNs with optimized thresholds are used to perform goal identification on our test set queries. In cases where multiple goals are hypothesized for single-goal queries, we penalize the insertion errors. Within-domain queries that are wrongly rejected as OOD, or OOD queries that are wrongly identified as within-domain are all penalized as errors.

A series of experiments were conducted in which we varied the BN input dimensionality, which is equivalent to the number of selected concepts per goal. Variation covered the whole range of concepts up to the full concept set. The goal classification accuracies on both the Chinese and English training sets, calculated according to our scoring scheme mentioned above, are shown in Figure 3.3 and Figure 3.4 respectively.

Performance accuracies in the plots are normalized based on the full size of the training data sets. As observed in the figures, training accuracies increase with input dimensionality and converged at 26 concepts per goal in Chinese and 28 concepts per goal in English. This suggests that 26 and 28 concepts per goal are the suitable parameter settings for the Chinese and English languages respectively. Performance with the optimal parameter settings for the Chinese and English are summarized in Table 3.7.

The Chinese test set yielded a goal identification accuracy of 93.7%. 11 of the 12 OOD queries were correctly rejected. But only 3 of the 6 multiple-goal queries have both of the goals correctly identified, and another 3 of them have their goals partially identified by getting one of the two goals in the query correctly identified.

In English, the test set yielded 92.0% goal identification accuracy. Among all the testing queries, all the OOD queries were correctly rejected. For the multiple goal queries, only 3 out of the 6 have identified both of the goals.

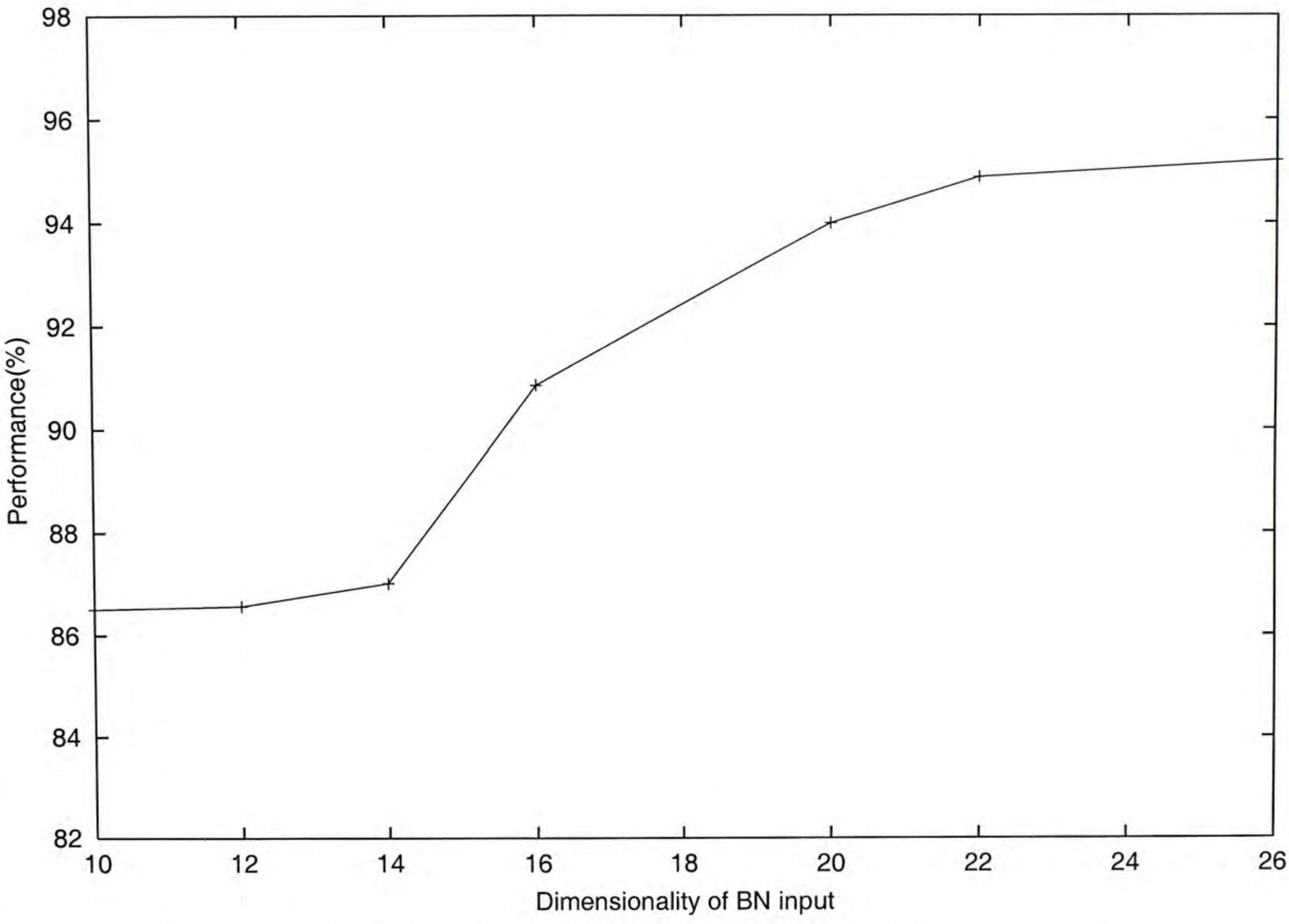

Figure 3.3: Goal classification performance on the training set for various input dimensionalities in the Chinese BN.

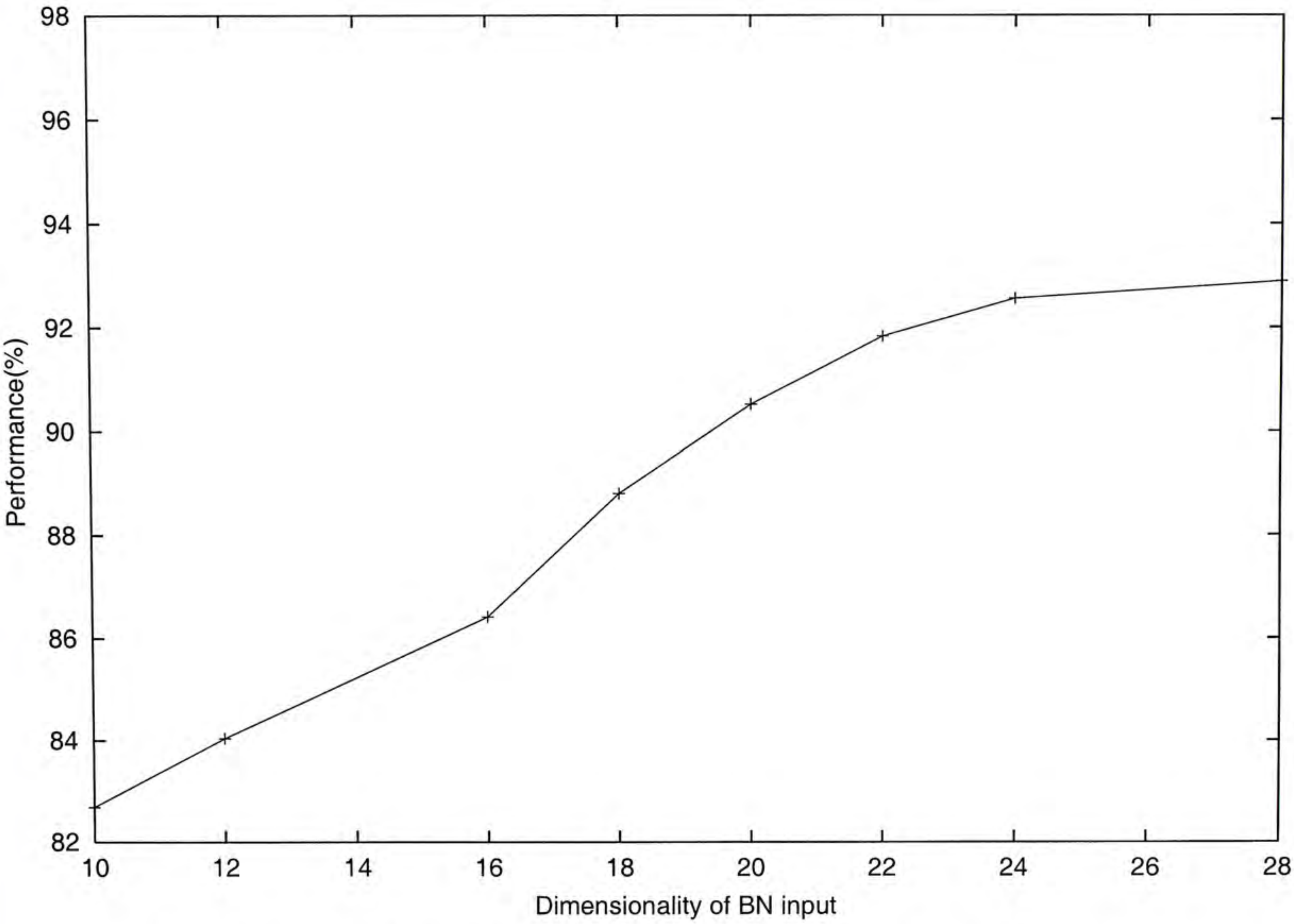

Figure 3.4: Goal classification performance on the training set for various input dimensionalities in the English BN.

	Chinese Testing Data Set	English Testing Data Set
Total Number of Test Queries	501	543
Multiple Goal Identification Performance	75.0% (9/12)	66.7% (8/12)
Single Goal Identification Performance	94.2% (455/483)	92.5% (492/532)
OOD Rejection Performance	91.7% (11/12)	100% (5/5)
Overall Performance	93.7% (475/507)	92.0% (505/549)

Table 3.7: Experimental results of the Chinese and English goal identification performance on the testing sets.

and only another 2 of them have got partially correct by having one out of two goals correctly identified in each of them.

The identified goal(s) together with the concepts and their values form the semantic frame that represents the meaning of the query. This is used to formulate a SQL expression for database access.

With the satisfactory performance obtained, we have demonstrated that our NLU framework is applicable across languages and domains.

3.4.2 A Cross-language Experiment

We have also conducted a side experiment where the BNs trained on Chinese were used for goal identification on the English queries. Both English training and testing sets were used since they are disjoint from the Chinese training corpus. The goal identification accuracy was 80.0% by using the 26 concepts optimized for the Chinese in Section 3.4.1. Similarly, the BNs trained on the

English queries were also used for goal identification of the Chinese training and testing sets. The goal identification accuracy was 84.0% with the use of 28 concepts optimized for the English language in Section 3.4.1.

To find out the main cause of the degradation suffered by the cross testing, we have extracted the accuracies of the queries under each goal and are organized as in Table 3.8 and Table 3.9. The discrepancies of each goal's identification performance are also included.

	English Trained Belief Networks (%) (1)	Chinese Trained Belief Networks (%) (2)	Net changes obtained from cross language testing (%) (3)=(2)-(1)
REQUEST_CHART	97.9	96.4	-1.5
REQUEST_NEWS	87.0	85.0	-2.0
ASK_TRENDS	90.7	47.2	**-43.5**
AMENDMENT	94.3	34.5	**-59.8**
CANCELLATION	85.7	90.5	+4.8
REQUEST_ACCOUNT_INFO	96.2	78.4	**-17.8**
REQUEST_REALTIME_QUOTES	96.0	91.9	-4.1
BUY	92.3	90.4	-1.9
SELL	84.4	91.7	+7.3
SYSTEM_META_COMMANDS	95.5	98.4	+2.9

Table 3.8: Comparison between the results of testing the English test sets on the English trained BNs and the Chinese trained BNs (cross-language testing). Column (3) refers to the net changes obtained from the cross testing. The main degradations in cross-language testing originate from a few particular goals.

When compared with the monolingual experiments, the main degrada-

	Chinese Trained Belief Networks (%) (1)	English Trained Belief Networks (%) (2)	Net changes obtained from cross language testing (%) (3)=(2)-(1)
REQUEST_CHART	100	95.2	-4.8
REQUEST_NEWS	93.3	78.9	-14.4
ASK_TRENDS	96.8	56.2	**-40.6**
AMENDMENT	100	47.1	**-52.9**
CANCELLATION	90.9	92.1	+1.2
REQUEST_ACCOUNT_INFO	90.4	88.2	-2.2
REQUEST_REALTIME_QUOTES	92.4	75.5	-16.9
BUY	93.6	96.6	+3.0
SELL	94.1	99.1	+5.0
SYSTEM_META_COMMANDS	100	100	0

Table 3.9: Comparison between the results of testing the Chinese test sets on the Chinese trained BNs and the English trained BNs (cross-language testing). Column (3) refers to the net changes obtained from the cross testing. The main degradations in cross-language testing originate from a few particular goals.

tions were mainly originated from a few particular goals. After investigating the discrepancies, two possible factors are suggested for explaining the observations.

3.4.2.1 Language-specific Elements

As mentioned in the earlier section, there are two language-specific elements defined in our English domain. One of the English-specific tags is <Instead>. The <Instead> tag abstracts the English word phrase "instead of". Although <Instead> is very similar to the English concept tag <AmendAction> which abstracts the words like "change" in meaning, they have very different behavior. For example, consider the pair "change X to Y" and "Y instead of X". To identify the passive / active roles of the objects X and Y, different transformation rules are used for disambiguation by checking the grammatical structure. However, in Chinese, the case is much simpler as we do not have such synonym. It only has the tag <AmendAction>. Thus, when test the English queries on the Chinese BNs, the concept <Instead> in those English sentences (mainly of the goal AMENDMENT) are obsolete to the Chinese BNs. As a result of bypassing the important evidence, the Chinese BNs suite fails to identify the correct informational goal(s) of such English sentences. For instance:

GOAL: AMENDMENT

QUERY: *Please sell Hopewell instead of Henderson Land.*

TAG: <Dummy> <SellAction> <StockNameHK> <Instead> <CancelledStockNameHK>

INFERRED GOAL: OOD

Similar explanation can be applied on the queries embedding another English language-specific concept <OrderName>. The concept <OrderName> can often be found in the REQUEST_ACCOUNT_INFO sentences. For example:

GOAL: REQUEST_ACCOUNT_INFO

QUERY: *Help me to check the sell order on Hang Seng Bank.*

TAG: <Dummy> <ShowWord> <Article> <OrderName> <StockNameHK>

INFERRED GOAL: REQUEST_REALTIME_QUOTES

The concept <OrderName> is ignored by the Chinese BNs. By considering the rest concepts, the sentence looks like a query requesting for the real-time quotes. Therefore, the Chinese BNs misclassified the English REQUEST_ACCOUNT_INFO sentence as a REQUEST_REALTIME_QUOTES query.

3.4.2.2 Overfitting to Sparse Training Data

We have collected our English and Chinese data sets separately, and so the two data sets are disjoint and independent of one another. There is no guarantee that all types of sentences (more exactly is the combination of semantic concepts) in one language have been trained sufficiently in the other languages. Therefore, owing to the overfitting of the small training data sets of both languages, certain degree of degradation is expected when we perform the cross-testing on our BNs.

For example, for the goal ASK_TRENDS, there are certain training queries in Chinese that ask for the prediction of the trend of a whole stock category.

But in our collected English training data set, users ask for the prediction on particular stocks. Both of these two kinds of query are very common in one language, but have no instance found in their corresponding counterpart language. It caused the BNs fail to identify the correct goal of such particular types of sentences during cross-testing. As described previously, in the first example listed below, the concept <StockCat>, referring to the stock categories, is commonly seen in the Chinese sentences of the goal ASK_TRENDS. However, it is quite rare for the English sentences under the same goal. Therefore, this Chinese sentence was rejected by the English BNs and was classified as an out-of-domain sentence. Moreover, in the second example, the concept <–PriceValue> found in the English query is also unexpected by the Chinese BN of the goal ASK_TRENDS. Bypassing the <–PriceValue> concept, the rest concepts are well matched with that of a Chinese REQUEST_REALTIME_QUOTES sentence. As a result, this sentence is wrongly classified as the goal REQUEST_REALTIME_QUOTES.

GOAL: ASK_TRENDS

QUERY: 預算一下藍籌股的走勢

TAG: <AskTrend> <StockCat> <Trend>

INFERRED GOAL: OOD

GOAL: ASK_TRENDS

QUERY: *Can you tell me when Cheung kong drops five dollars.*

TAG: <ShowWord> <Quest> <StockNameHK> <MarketMovement> <–PriceValue>

INFERRED GOAL: REQUEST_REALTIME_QUOTES

Additionally, when we investigate our experimental results, we found that there exists a number of sentences of the Chinese goal AMENDMENT with similar patterns failed to be correctly classified when it is cross-tested by the English BNs. For example:

GOAL: AMENDMENT
QUERY: 我想改買四手恆隆
TAG: <Dummy> <AmendAction> <LotNumber> <StockNameHK>
INFERRED GOAL: OOD

We observed from the data sets that the English sentences corresponding to the AMENDMENT actions usually provide also some information about the amended order. However, such kind of information is missed in that sorts of Chinese AMENDMENT queries. By checking the output of the BNs on this sentence, the a posteriori probability for the goal AMENDMENT is quite high. Nonetheless, when compare with the threshold value of this goal, it fails and so, this sentence was not identified as the goal AMENDMENT.

3.5 Chapter Summary

In this chapter, we have described our BN-based NLU framework and demonstrated that our NLU framework is applicable across both the application domains and languages. Our framework consists of a semantic tagging procedure followed by informational goal inference using BNs. This framework was previously used for English in the ATIS (air travel) domain, and is currently extended to handle both English and Chinese for the ISIS (stocks) domain. The stocks domain presents new complexities for NLU which were not observed in the air travel domain – the prominent ones being the disambiguation of various kinds of numeric expressions, and new words (OOV). We applied a transformation-based parsing technique, which utilizes both the left and right contexts for disambiguating the numeric expressions, and infer a possible category of the OOV word. The concept tags resulting from parsing are fed into BNs previously trained from the training sets of collected English and Chinese queries. Experiments with disjoint test sets gave goal identification accuracies of 92.0% for English queries and 93.7% for Chinese queries. These results demonstrate that our NLU framework is applicable to Chinese and the ISIS domain. A side experiment with cross-language configurations gave goal-identification accuracies of 80.0% (testing English queries on BNs trained on Chinese queries) and 84.0% (testing Chinese queries on BNs trained on English queries), which suggest that much commonality is captured across the two languages.

Chapter 4

Enhancement in the Belief Networks for Informational Goal Inference

In Chapter 3, we have introduced our unified natural language understanding (NLU) framework to process ISIS stocks domain sentences in both Chinese and English. As a data-driven approach, our BN-based NLU framework offers us the ease of portability. However, the training procedure generally requires a large amount of data, which may form a bottleneck for system development. To alleviate the reliance of training data, we devised principles for concept selection and the probability assignment in each BN. We also benchmarked the goal identification performance of the BNs developed with our principles.

4.1 Semantic Concept Selection in Belief Networks

Information Gain (IG) captures the degree of association between a concept and a goal, taking into consideration both the presence and absence of the entities. In our NLU framework, IG is applied to select the subset of concepts having the highest IG values from the concept pool. Concepts that are selected to indicate the presence of the goal are regarded as the positive evidence for the goal, while the concepts that are selected to indicate the absence of the goal are deemed as the negative evidence. In Chapter 3, BNs using concepts selected by IG have shown promising results on goal identification. However, without sufficient training data, IG cannot be applied efficiently. Hence we devise a set of concept selection principles to replace the IG concept selection scheme. These principles are intended for application when training data are scarce.

4.1.1 Selection of Positive Evidence

The ultimate goal of a NLU system is to generate the meaning representation of the input sentence. Meaning representation encodes the full meaning of the input and that may be in form of semantic frames. In a semantic frame, different types of information that are indispensable for the accomplishment of requests are defined as slots. NLU system extracts information from the input to fill in these slots. As a result, each semantic frame stores a significant portion of positive evidence for the corresponding goal. Figure 4.1 illustrates an example semantic frame defined for the goal SELL in English.[1] In the

[1] In Appendices A and B, semantic frames of all the goals defined for both Chinese and English are included.

GOAL:SELL

{

stock := <StockRicHK>|<StockName>|<All>
price := <PriceValue>|<MarketPrice>
+price := <+PriceValue>
–price := <–PriceValue>
lot number := <LotNumber>
share number := <ShareNumber>|<AllOfStock>

}

Figure 4.1: The semantic frame for the goal SELL in English.

semantic frame, we can find the slot names residing on the left-hand side. These slot names represent the categories of information required for the completion of SELL requests. On the right-hand side of each slot, the values for the slots are listed, as they are extracted from the query. We can simply select these concepts as the positive evidence of the goal. Following this strategy, the positive evidence for the English goal SELL are selected as in Table 4.1.

StockRicHK	StockName	All
PriceValue	MarketPrice	+PriceValue
-PriceValue	LotNumber	ShareNumber
AllOfStock		

Table 4.1: The positive evidence extracted from a semantic frame for the goal SELL in English.

Positive evidence selected from the semantic frames are relatively informative in nature and can be common across multiple goals. For example, the concept <ShareNumber> has been selected as the positive evidence of the goals BUY, SELL, AMENDMENT, etc.. Such kind of concept sharing across multiple goals weaken the classification power of the BNs. Therefore, we have to further select some concepts which are more goal-specific to the BNs. One obvious example is the action concepts, for example <SellAction>. To express the selling intention of a sentence of the goal SELL, it is impossible to miss the selling concept <SellAction> abstracting the words like "sell", "selling", etc. So, in this case, the concept <SellAction> is extracted as another piece of positive evidence for the goal SELL.

Additionally, a more complete positive evidence concept set for the goals can be captured by including the concepts that usually coexist with the selected concepts in the sentences. For example, the concepts <All> and <AllOfStock> are previously selected for the English goal SELL, but we found that they often occur with the concept <Portfolio> in the sentences under the goal SELL. Here are the examples:

- QUERY : *Sell me all of the HSBC stocks in my portfolio.*
  TAG : <SellAction> <AllOfStock> <StockRicHK> <Stock> <Portfolio>
- QUERY : *Help me sell all the stocks in my account.*
  TAG : <Dummy> <SellAction> <All> <Stock> <Portfolio>

The concept <Portfolio> is therefore included as another piece of positive evidence for the goal SELL. Applying this principle, the concept <CompanyStat> abstracting quote items like "ask price" and the concept <MarketMovement> abstracting the market descriptors such as "drops", are also selected for the goal SELL. Moreover, extended from the concepts <CompanyStat> and

<MarketMovement>, the concepts <PriceValue>, <+PriceValue> and <-PriceValue> are also included by referencing the following examples:

- QUERY : *Please sell three lots of HSBC at the ask price of one hundred dollars.*

  TAG : <Dummy> <SellAction> <LotNumber> <Of> <StockRicHK> <At> <CompanyStat> <Of> <PriceValue>

- QUERY : *Sell three lots of Hang Lung when it drops two more dollars.*

  TAG : <SellAction> <LotNumber> <Of> <StockRicHK> <When> <Dummy> <MarketMovement> <-PriceValue>

Lastly, the positive evidence of the English goal SELL are finalized as in Table 4.2. By applying again the concept selection principles on the rest

GOAL:SELL		
StockRicHK	StockName	All
Portfolio	PriceValue	MarketPrice
+PriceValue	-PriceValue	LotNumber
ShareNumber	AllOfStock	SellAction
MarketMovement	CompanyStat	

Table 4.2: The concept set of positive evidence selected for the goal SELL in English.

of the nine goals in English, the positive evidence are determined and have been put under Appendix C. Moreover, our concept selection principles are supposed to be applicable across languages. We have gone through the positive evidence selection procedure for the 10 goals in Chinese. The selected positive evidence for the Chinese goals are listed in Appendix D.

4.1.2 Selection of Negative Evidence

Comparing the positive evidence selected for the goals in last section, we found that some of the goals have high degree of overlap among their selected concepts. In our domain, there exists some cases that a concept sequence of one goal contains the complete concept sequence of another goal. Consider the following sentence pair:

- GOAL : SELL
  QUERY : *Sell three lots of HSBC.*
  TAG : <SellAction> <LotNumber> <Of> <StockRicHK>

- GOAL : AMENDMENT
  QUERY : *Sell three lots of HSBC instead of Hang Seng Bank.*
  TAG : <SellAction> <LotNumber> <Of> <StockRicHK> <Instead> <CancelledStockRicHK>

We can clearly see that the entire concept sequence abstracting the former sentence of the goal SELL is completely embedded in the latter concept sequence of the goal AMENDMENT. If the BNs classify the sentence goal solely by positive evidence, the BN of the goal SELL will mis-fire for the sentence of the goal AMENDMENT. It results that the sentence of the goal AMENDMENT will not only be classified as the goal AMENDMENT, but also of the goal SELL. It raises the importance of the consideration of negative evidence in the BNs so as to increase their rejection capabilities.

To determine the negative evidence, we can make use of our knowledge in the domain. For our example, in each of the AMENDMENT sentences, either of the two concepts <AmendAction> and <Instead> should be found. But on the contrary, neither of these two concepts should occur in the sentences of the goal SELL. As a result, both of them are injected into the BN of the goal SELL as negative evidence. Similarly for the BN of the goal BUY, AMENDMENT sentences that involve buying action are also misclassified as the sentences of the goal BUY. So, the concepts <AmendAction> and <Instead> are also included to the BN of the goal BUY as negative evidence. Moreover, another confusion is commonly found in the transactional goals like BUY, SELL, CANCELLATION and AMENDMENT with the goal REQUEST_ACCOUNT_INFO. The BNs of the transactional goals mis-fire for the sentences regarding the enquires on the progress of placed order. For example:

- GOAL : BUY
  QUERY : *Buy three lots of HSBC at one hundred dollars.*
  TAG : <BuyAction> <LotNumber> <Of> <StockRicHK> *<At> <PriceValue>*
- GOAL : REQUEST_ACCOUNT_INFO
  QUERY : *Has my order of buying three lots of HSBC at one hundred dollars processed?*
  TAG : <AskAcc> <Order> <Of> <BuyAction> <LotNumber> *<Of> <StockRicHK> <At> <PriceValue> <Processed>*

The BN of the goal BUY misclassifies the sentence of the goal REQUEST-_ACCOUNT_INFO as the goal BUY. To enhance those transactional BNs to reject the sentences of the goal REQUEST_ACCOUNT_INFO, we have injected some negative evidence to each of them. They include the concepts <Processed>, <NotYet> and <AskAcc>. All of these concepts are often found in the sentences of the goal REQUEST_ACCOUNT_INFO, but rare to be seen in other sentences.

We have checked across the rest goal pairs and have completed the selection of negative evidence for the goals in both English and Chinese. The complete concept sets for all the 10 goals in English and Chinese are listed in Appendices E and F respectively.

4.2 Estimation of Statistical Probabilities in the Enhanced Belief Networks

Now, we have come to the second step of our NLU framework enhancement. With the concept sets selected for the BNs, we have to estimate the statistical probabilities modeling the relationship between the goals and the selected concepts. For each BN of a particular goal, we need to estimate the prior probability of the goal and the posterior probabilities regarding the presence / absence of each of the selected concepts given the presence / absence of the goal. For the sake of simplicity, the pre-defined topology will be applied as before. Under this topology, we assumed conditional independence is held in the BNs.

1. Prior Probabilities

   The prior probabilities of a goal G_i is represented in terms of:

   - $P(G_i = 1)$
   - $P(G_i = 0)$

where $P(G_i = 1) + P(G_i = 0) = 1$

2. Posterior Probabilities

   For each of the concepts selected for a BN, we have to define two pairs of posterior probabilities, they are:

   - $P(C_j = 0|G_i = 1)$ and $P(C_j = 1|G_i = 1)$
   - $P(C_j = 0|G_i = 0)$ and $P(C_j = 1|G_i = 0)$

   where $P(C_j{=}0|G_i{=}1){+}P(C_j{=}1|G_i{=}1){=}1$

   $P(C_j{=}0|G_i{=}0){+}P(C_j{=}1|G_i{=}0){=}1$

In the coming section, the guidelines for the hand assignment of these statistical probabilities are discussed.

4.2.1 Estimation of Prior Probabilities

In the ISIS stocks domain, we have defined 10 domain-specific goals in total. Without any prior knowledge on the goal distribution, we chose to evenly assign the prior probability P(G_i=1) for the goal G_i where i=1,2 ... 10 as 0.1 for the sake of fairness. Henceforth, by the principle of marginalization, the value of $P(G_i = 0)$ for the goal G_i is then determined as follows:

$$\begin{aligned} P(G_i = 0) &= 1 - P(G_i = 1) \\ &= 1 - 0.1 \\ &= 0.9 \qquad\qquad for\ i{=}1,2 \ldots 10 \end{aligned} \tag{4.1}$$

The hand-assigned prior probabilities for the 10 ISIS domain specific goals in both Chinese and English are illustrated as in Table 4.3.

GOAL(G_i)	P(G_i=1)	P(G_i=0)
REQUEST_CHART	0.1	0.9
REQUEST_NEWS	0.1	0.9
ASK_TRENDS	0.1	0.9
AMENDMENT	0.1	0.9
CANCELLATION	0.1	0.9
REQUEST_ACCOUNT_INFO	0.1	0.9
REQUEST_REALTIME_QUOTES	0.1	0.9
BUY	0.1	0.9
SELL	0.1	0.9
SYSTEM_META_COMMANDS	0.1	0.9

Table 4.3: The estimated prior probabilities of the 10 goals.

4.2.2 Estimation of Posterior Probabilities

For each goal-concept pair in a BN, two pairs / four individual posterior probabilities are estimated. They are P(C_j=0|G_i=1), P(C_j=1|G_i=1), P(C_j=0|G_i=0) and P(C_j=1|G_i=0). In the section, we describe the general principles for assigning P(C_j=1|G_i=1) and P(C_j=1|G_i=0). The remaining P(C_j=0|G_i=1) and P(C_j=0|G_i=0) are then derived as the complement of the former two probabilities.

4.2.2.1 Posterior Probabilities Given $G_i = 1$

We assign the probability P(C_j=1|G_i=1) based on the importance, function, characteristic and occurrence of the concept C_j in the sentences of the goal G_i. To enhance the hand assignment of probability, we have divided the concepts into 5 categories:

Category 1 : Indispensable Concepts

Indispensable concepts refer to the concepts that are mandatory for a sentence in order to exhibit the intention of the goal G_i. Examples in our domain include the concept <SellAction> in the sentences of the goal SELL (Table 4.4, E.g.1), the concept <BuyAction> in the sentences of the goal BUY (Table 4.4, E.g.2) and the concept <CancelAction> in the sentences of the goal CANCELLATION (Table 4.4, E.g.3), etc.. For each of the concepts C_j in this category, we hand-assign the posterior probability $P(C_j=1|G_i=1)$ to 1. However,

E.g. 1	GOAL	SELL
	QUERY	Sell three lots of HSBC.
	TAG	$\boxed{\text{<SellAction>}}$ <LotNumber> <Of> <StockRicHK>
E.g. 2	GOAL	BUY
	QUERY	I want to buy two thousand shares of Hopewell.
	TAG	<Dummy> $\boxed{\text{<BuyAction>}}$ <ShareNumber> <Of> <StockRicHK>
E.g. 3	GOAL	CANCELLATION
	QUERY	Please cancel my order of Cheung Kong.
	TAG	<Dummy> $\boxed{\text{<CancelAction>}}$ <Order> <Of> <StockRicHK>

Table 4.4: Examples of indispensable semantic concepts (Category 1) found in ISIS domain sentences.

there is a special case found in our domain. In Chinese, the concept <AmendAction> (for the words like "改","轉", etc) is a dispensable concept that should be found in all the sentences of the goal AMENDMENT. But, it is not the case of English. In each of the English sentences of the goal AMENDMENT, either the concepts <AmendAction> or <Instead> should be found. The two concepts are in fact an alternative of one another. Although they are indispensable for sentences of the goal AMENDMENT in nature, there is no guarantee which of them may occur. Therefore, a slight penalty is given to the probabilities and finally, 0.95 is assigned to each of the posterior probabilities P(AmendAction=1|AMENDMENT=1) and P(Instead=1|AMENDMENT=1).

Category 2 : Important Concepts

Concepts that abstract important information for the completion of requests are categorized as *important concepts*. So, all the concepts that provide information to fill in the slots of semantic frames are belonging to this category. For instance, the concepts <LotNumber> and <ShareNumber> are important concepts for the goals BUY and SELL. Although these concepts are necessary for the completion of requests, there are no guarantee that they must occur in the regarded sentences. As a result, we lower the posterior probability P(C_j=1|G_i=1) to 0.75.

Category 3 : Auxiliary Concepts

Auxiliary concepts are those concepts that do not carry informative content, but are usually found in the sentences of particular goals. A common example is the concept <ShowWord> (for the phrases like "Show me", "Have a look", etc). It can be usually found in the sentences of the goals CHART and REQUEST_REALTIME_QUOTES. Consider the following sentences pair:

- QUERY : *Show me the bid price of HSBC, please*
  TAG : <ShowWord> <CompanyStat> <Of> <StockRicHK> *<Dummy>*

- QUERY : *The bid price of HSBC, please*

  TAG : <CompanyStat> <Of> <StockRicHK> <Dummy>

Both of the two sentences are meaningful and so, the word phrase "Show me" in the first sentence is in fact optional in nature. For the concepts under this category, we assign the posterior probabilities P(C_j=1|G_i=1) as 0.5.

Category 4 : Irrational Concepts

Irrational concepts are concepts that are informative, but are rare to be found in the sentences of the particular goal. For example, under usual and normal situations, people sell their stocks when the stock prices rise. However, sometimes, people may put a stop-loss[2] order like:

QUERY : *Help me to sell HSBC when it drops two more dollars.*

TAG : <Dummy> <SellAction> <StockRicHK> <Dummy> <MarketMovement> <-PriceValue>

The concept <-PriceValue>, abstracting the decremental price information "two more dollars", is captured as a concept for the goal SELL. However, we believe that the probability P(-PriceValue=1|SELL=1) would not be high. For this case, the posterior probability P(C_j=1|G_i=1) of 0.25 is suggested.

[2]An order placed with a broker to buy or sell when a certain price is reached. It is designed to limit an investor's loss (or lock in profit) on a security position. This is sometimes called stop market order.

Category 5 : Negative Concepts

As introduced in Section 4.1.2, *negative concepts* are selected to provide negative evidence for the goals. The occurrence of such negative concepts diminishes the chance of the presence of the goal. So, it is straightforward to assign the posterior probability P(C_j=1|G_i=1) of all of the negative concepts to 0.

After defining the posterior probabilities P(C_j=1|G_i=1) of each goal-concept pair, by applying the rule of conditional marginalization, the corresponding value of P(C_j=0|G_i=1) is simply computed as follows:

$$P(C_j = 0|G_i = 1) = 1 - P(C_j = 1|G_i = 1) \tag{4.2}$$

Now, we can summarize our hand-assigned probabilities P(C_j=1|G_i=1) and P(C_j=0|G_i=1) as shown in Table 4.5.

	Category	P(C_j=1|G_i=1)	P(C_j=0|G_i=1)
1(a)	Indispensable concepts	1	0
1(b)	Indispensable concept alternatives	0.95	0.05
2	Important concepts	0.75	0.25
3	Auxiliary concepts	0.5	0.5
4	Irrational concepts	0.25	0.75
5	Negative concepts	0	1

Table 4.5: Summary of the hand-assignment of posterior probabilities P(C_j=1|G_i=1) and P(C_j=0|G_i=1).

4.2.2.2 Posterior Probabilities Given $G_i = 0$

To assign the posterior probability of P(C_j=1|G_i=0) for the goal G_i, we need to estimate the probability that the concept C_j being present while the goal G_i is absent. More exactly, we have to estimate the occurrences of the concept C_j for all sentences except those of the goal G_i. We have designed a scheme for this posterior probability estimation. Under this scheme, we first evaluate the number of goals in whose sentences that the concept C_j would be found. For this, we can make use of the clues given by the positive evidence that have been selected for each goal. Henceforth, for each goal G_i, the posterior probability P(C_j=1|G_i=0) is evaluated as the proportion of the rest 9 goals from whose sentences the concept C_j can be found. Generally, we can classify the concepts into two categories:

Category 1 : Concepts Occurring for Multiple Goals

As discussed in earlier sections, there are some concepts shared by sentences from multiple / different goals. For instance, referring to Section 4.1.1, there are in total 4 goals declaring the concept <ShowWord> as one of their positive concepts. The 4 goals are REQUEST_CHART, REQUEST_NEWS, REQUEST_REALTIME_QUOTES and REQUEST_ACCOUNT_INFO. To estimate the posterior probability P(ShowWord=1 | G_i=0) for any of the 4 goals, let say P(ShowWord=1|REQUEST_CHART=0), 3 goals out of the rest 9 goals are counted as follows:

$$\begin{aligned} P(C_j = 1|G_i = 0) &= (4-1)/9 \\ &= 0.33 \end{aligned} \tag{4.3}$$

The corresponding posterior probability P(C_j=0|G_i=0) is estimated as the complement of P(C_j=1|G_i=0):

$$\begin{aligned} P(C_j = 0|G_i = 0) &= 1 - P(C_j = 1|G_i = 0) \\ &= 1 - 0.33 \\ &= 0.67 \end{aligned} \tag{4.4}$$

But, for those posterior probability P(C_j=1|G_i=0) of the rest 6 goals, for example, P(ShowWord=1|SELL=0), it becomes:

$$\begin{aligned} P(C_j = 1|G_i = 0) &= 4/9 \\ &= 0.44 \end{aligned} \tag{4.5}$$

The corresponding posterior probability P(C_j=0|G_i=0) would then be:

$$\begin{aligned} P(C_j = 0|G_i = 0) &= 1 - P(C_j = 1|G_i = 0) \\ &= 1 - 0.44 \\ &= 0.56 \end{aligned} \tag{4.6}$$

Category 2 : Concepts Occurring Solely for a Particular Goal

For a concept C_j that occurs only in the sentences of a particular goal G_i, we can simply assign the posterior probability P(C_j=1|G_i=0) to 0:

$$P(C_j = 1|G_i = 0) = 0 \tag{4.7}$$

By marginalization, the probability of P(C_j=0|G_i=0) is computed as:

$$\begin{aligned} P(C_j = 0|G_i = 0) &\quad 1 - P(C_j = 1|G_i = 0) \\ &= 1 - 0 \\ &= 1 \end{aligned} \tag{4.8}$$

The zero probability implies that the occurrence of the concept C_j should not be expected in the sentences of all the goals except G_i.

One example is the concept <Chart>. This concept occurs only in the sentences of the goal REQUEST_CHART and so, the probability P(Chart=1|REQUEST_CHART=0) is hand-assigned as 0, while P(Chart=0|REQUEST_CHART=0) is assigned as 1. Similarly, for any of the rest 9 goals, let say BUY, the posterior probabilities would be assigned as:

$$\begin{aligned} P(Chart = 1|BUY = 0) &= 1/9 \\ &= 0.11 \end{aligned} \tag{4.9}$$

$$\begin{aligned} P(Chart = 0|BUY = 0) &= 1 - P(Chart = 1|BUY = 0) \\ &= 1 - 0.11 \\ &= 0.89 \end{aligned} \tag{4.10}$$

Base on the above categorization, the hand assignment of the statistical probabilities for each selected concept under the 10 goals in English and Chinese are shown in the Appendices E and F respectively.

4.3 Experiments

So far, we have introduced the principles to develop BNs with only a handful of training data. The principles cover the positive and negative evidence selection as well as the hand-assignment of probabilities in each BN. A series of experiments have been carried out to evaluate how well these principles apply to our BNs. In each experiment, both the English and Chinese data sets for the ISIS stocks domain applied in Chapter 3 are reused.

4.3.1 Belief Networks Developed with Positive Evidence

Based on the positive concepts defined in Section 4.1.1 and the probability assignment scheme described in Section 4.2, two suites of BNs (one for each language) are developed. We have tested the BNs with the testing sets of the corresponding language, 80.0% and 80.2% goal identification accuracies are achieved for the Chinese and English languages respectively. To benchmark our results, the goal identification performance tested on the trained BNs are reported in Table 4.6 and Table 4.7. We see that only slight degradations are obtained in test set performance when compared with that are tested on the BNs automatically trained with the training data.

	Trained BNs		Hand-assigned BNs	
	Training Set	Testing Set	Training Set	Testing Set
Performance(%)	86.1	82.8	80.7	80.0

Table 4.6: The overall goal identification performance of the Chinese BNs. These results are generated with Chinese BNs containing only concept nodes that serve as positive evidence of the corresponding goal node.

	Trained BNs		Hand-assigned BNs	
	Training Set	Testing Set	Training Set	Testing Set
Performance(%)	85.7	85.6	78.1	80.2

Table 4.7: The overall goal identification performance of the English BNs. These results are generated with English BNs containing only concept nodes that serve as positive evidence of the corresponding goal node.

4.3.1.1 Error Analysis

Investigating the performance for each goal in Chinese and English as shown in Table 4.8 and Table 4.9, we observed that there exists a few goals in each language that obtain relatively poor performance in both the hand-assigned and trained BNs. In Chinese, the goals include REQUEST_ACCUONT_INFO, REQUEST_REALTIME_QUOTES, ASK_TRENDS and AMENDMENT. But in English, we have the goals of ASK_TRENDS, AMENDMENT and REQUEST_ACCUONT_INFO. Since the performance of these goals are poor (just close or below the average accuracy rate) in both the hand-assigned and trained BNs suites, we believe that is not caused by either of the two probability estimation schemes. Instead, the poor performance are suggested to be the result of the incomplete concept sets. The examples in Table 4.10 explain our argument.

In each of the illustrated examples, apart from the reference goal, a spurious goal is mis-identified for the sentence. To investigate the misclassification, we have compared the concept sequences of the query with the selected positive concept set for the corresponding spurious hypothesis goal. We found that the matched concepts are those that can be commonly found in the queries of the spurious hypothesis goal. More surprisingly, we may even find some queries of the spurious goal that contain all of the matched concepts. Since all the matched concepts in the queries are considered to be positive evidence throughout the goal inference process while the rest unmatched concepts will be ignored by the BNs, it is understandable that why the spurious hypothesis goal has been mis-fired. So, to improve the goal identification accuracy, the rejection capability of the BNs is enhanced by incorporating some negative evidence in the BNs.

Goal	**Performance(%)**	
	Trained BNs suite	**Assigned BNs suite**
REQUEST_CHART	97.7	93.2
REQUEST_NEWS	86.7	93.3
ASK_TRENDS	61.3	80.7
AMENDMENT	42.4	51.5
CANCELLATION	90.9	81.8
REQUEST_ACCOUNT_INFO	82.7	76.9
REQUEST_REALTIME_QUOTES	74.4	69.7
BUY	98.4	74.2
SELL	88.2	92.2
SYSTEM_META_COMMANDS	100	100

Table 4.8: Goal identification performance of each goal based on the test set. These results are generated with Chinese BNs containing only concept nodes that serve as positive evidence of the corresponding goal node.

4.3.2 Belief Networks with the Injection of Negative Evidence

In this experiment, to develop the BNs suites, the complete concept sets that contain both the positive and negative evidence are used. The goal identification performance of the newly developed Chinese and English BNs are reported in Table 4.11 and Table 4.12 respectively. As in the previous experiment, to benchmark the BNs suite developed with our principles, the goal identification performance tested on the trained BNs suites are included.

Goal	Performance(%)	
	Trained BNs suite	**Assigned BNs suite**
REQUEST_CHART	97.9	91.5
REQUEST_NEWS	92.6	83.3
ASK_TRENDS	85.2	70.4
AMENDMENT	45.7	60.0
CANCELLATION	76.2	91.5
REQUEST_ACCOUNT_INFO	76.4	68.9
REQUEST_REALTIME_QUOTES	94.7	86.7
BUY	94.2	80.8
SELL	91.1	82.2
SYSTEM_META_COMMANDS	95.5	95.5

Table 4.9: Goal identification performance of each goal based on the test set. These results are generated with English BNs containing only concept nodes that serve as positive evidence of the corresponding goal node.

REFERENCE GOAL	ASK_TRENDS
QUERY	我想知道長實今日開市價預計會係幾多
HYPOTHESIS GOAL(S)	ASK_TRENDS, REQUEST_REALTIME_QUOTES
MATCHED POSITIVE CONCEPTS OF THE GOAL REQUEST_REALTIME_QUOTES	<ShowWord>, <StockRicHK>, <RelativeDate>, <CompanyStat>, <Quest>
EXAMPLE QUERY OF THE GOAL REQUEST_REALTIME_QUOTES FORMED BY THE MATCHED CONCEPTS	我想知道長實今日開市價係幾多

REFERENCE GOAL	AMENDMENT
QUERY	轉十六塊四毛沽長江基建
HYPOTHESIS GOAL(S)	AMENDMENT, SELL
MATCHED POSITIVE CONCEPTS OF THE GOAL SELL	<PriceValue>, <SellAction>, <StockRicHK>
EXAMPLE QUERY OF THE GOAL SELL FORMED BY THE MATCHED CONCEPTS	十六塊四毛沽長江基建

REFERENCE GOAL	AMENDMENT
QUERY	Please change my order to buy two lots of HSBC from one lot of Hysan.
HYPOTHESIS GOAL(S)	AMENDMENT, BUY
MATCHED POSITIVE CONCEPTS OF THE GOAL BUY	<BuyAction>, <LotNumber>, <StockRicHK>
EXAMPLE QUERY OF THE GOAL BUY FORMED BY THE MATCHED CONCEPTS	Please help me to buy two lots of HSBC.

REFERENCE GOAL	REQUEST_ACCOUNT_INFO
QUERY	Have my order of buying three thousand shares of HSBC completed?
HYPOTHESIS GOAL(S)	REQUEST_ACCOUNT_INFO, BUY
MATCHED POSITIVE CONCEPTS OF THE GOAL BUY	<BuyAction>, <ShareNumber>, <StockRicHK>
EXAMPLE QUERY OF THE GOAL BUY FORMED BY THE MATCHED CONCEPTS	Please help me to buy three thousand shares of HSBC.

Table 4.10: Examples suffering from the insufficiency of positive evidence for goal inference. Spurious hypothesis goals are mis-fired for the queries due to the use of incomplete concept set. Valid example queries of the spurious hypothesis goals can be formed by the matched positive concepts.

	Trained BNs suite		Hand-assigned BNs suite	
	Training Set	**Testing Set**	**Training Set**	**Testing Set**
Performance(%)	93.2	90.7	88.0	88.4
Net Change(%)	+7.1	+7.9	+7.3	+8.4

Table 4.11: The overall goal identification performance of the Chinese BNs. These results are generated with Chinese BNs containing the complete concept sets (including both the concept nodes that serve as positive evidence and negative evidence of the corresponding goal node). Net changes refer to the performance difference obtained over the BNs suites developed with the positive evidence.

	Trained BNs suite		Hand-assigned BNs suite	
	Training Set	**Testing Set**	**Training Set**	**Testing Set**
Performance(%)	90.4	90.2	83.8	85.6
Net Change(%)	+4.7	+4.6	+5.7	+5.4

Table 4.12: The overall goal identification performance of the English BNs. These results are generated with English BNs containing the complete concept sets (including both the concept nodes that serve as positive evidence and negative evidence of the corresponding goal node). Net changes refer to the performance difference obtained over the BNs suites developed with the positive evidence.

Goal	Trained BNs suite		Hand-assigned BNs suite	
	Accuracy (%)	Net change (%)	Accuracy (%)	Net change (%)
REQUEST_CHART	97.7	0	100	+6.8
REQUEST_NEWS	86.7	0	93.3	0
ASK_TRENDS	100	+38.7	100	+19.3
AMENDMENT	100	+57.6	100	+48.5
CANCELLATION	90.9	0	81.8	0
REQUEST_ACCOUNT_INFO	88.5	+5.8	77.9	+1.0
REQUEST_REALTIME_QUOTES	77.3	+2.9	90.9	+21.2
BUY	98.4	0	87.1	+12.9
SELL	94.1	+5.9	84.3	-7.9
SYSTEM_META_COMMANDS	100	0	100	0

Table 4.13: Goal identification performance of each goal based on the test set. These results are generated with Chinese BNs containing the complete concept sets (including both the concept nodes that serve as positive evidence and negative evidence of the corresponding goal node). The net changes refer to the performance difference obtained over the BNs suites developed with positive evidence.

Goal	Trained BNs suite		Hand-assigned BNs suite	
	Accuracy (%)	Net changes (%)	Accuracy (%)	Net changes (%)
REQUEST_CHART	100	+2.1	100	+8.5
REQUEST_NEWS	92.6	0	83.3	0
ASK_TRENDS	83.3	-1.9	79.6	+9.2
AMENDMENT	85.7	+40	97.1	+37.1
CANCELLATION	81.0	+4.8	95.2	+3.7
REQUEST_ACCOUNT_INFO	84.0	+7.6	77.4	+8.5
REQUEST_REALTIME_QUOTES	94.7	0	86.7	0
BUY	94.2	0	78.9	-1.9
SELL	93.3	+2.2	82.2	0
SYSTEM_META_COMMANDS	95.5	0	95.5	0

Table 4.14: Goal identification performance of each goal based on the test set. These results are generated with English BNs containing the complete concept sets (including both the concept nodes that serve as positive evidence and negative evidence of the corresponding goal node). The net changes refer to the performance difference obtained over the BNs suites developed with positive evidence.

As shown in the Table 4.11 and Table 4.12, improvements in goal identification accuracy ranging from 4.6% to 8.4% are obtained after injecting the negative evidence to the BNs. Table 4.13 and Table 4.14 depict the goal identification accuracy for each goal. From these tables, we observed that the largest improvements are found in the goals like AMENDMENT, ASK_TRENDS, etc. out of the ten goals in both languages. Moreover, after re-checking the examples shown in Table 4.10, all the examples are found to be correctly identified by the BNs injected with the negative evidence. Such satisfactory results justify the use of negative evidence in the BNs development.

4.4 Chapter Summary

Portability of NLU framework across domains and languages often implies the lack of training data. To enhance portability of our NLU framework, we have devised some principles to reduce data requirement during the NLU development. Our proposed approach offers an efficient way to develop a NLU system with limited training data. The devised principles cover the guidelines for concept selection as well as the hand-assignment of probabilities embedded in the BNs. The concept selection principles specify the selection criteria for both the positive and negative concepts, while the probability assignment guidelines govern the hand-assignment for both the prior and posterior probabilities in the BNs. We have performed a series of experiments in which our devised principles are applied to the BN development. The incorporation of negative concepts has been proven to be favorable by achieving between 4.6% and 8.4% improvement when compared with only using positive concepts in both Chinese and English data sets. Besides, less than 5% degradation is obtained for the goal identification when compared with the performance tested on the trained BNs. These results justify the use of our devised principles in BN development.

Chapter 5

Integration between Speech Recognition and Natural Language Understanding

In chapter 3, we have illustrated the feasibility of using Belief Networks (BNs) for natural language understanding (NLU) on perfect sentences. However, with the expanding use of conversational systems, natural language understanding on error-free sentences is not enough. As a matter of fact, a spoken language system should be able to recognize the words that are spoken by a user, understand his / her request recognized by the speech recognizer and give the appropriate response. To achieve high accuracy in a conversational system, the most ideal case is to have a perfect speech recognizer and a robust NLU component in the system. However speech recognition (SR) technologies nowadays are still far from perfect, recognition errors are often unavoidable. Speech recognition errors may cause the NLU component to fail and misunderstand the user's input. As a result, unexpected responses will be generated by the system and the user will be frustrated.

To alleviate the problem resulted from imperfect speech recognition, integration of speech recognition information output by a speech recognizer with natural language understanding is highly desirable. In this chapter, we describe our work that merges semantic and informational goal decoding with speech recognition. To demonstrate the feasibility of our approach, we have coupled our NLU framework [38] devised in Chapter 3 with a Chinese speech recognizer developed for the ISIS stocks domain [10]. When a user speaks a sentence to the SR-NLU integrated framework, a list of N-best hypotheses,[1] each associated with a confidence score is generated by the speech recognizer. To accommodate the N-best hypotheses output from the speech recognizer, we have extended our BN-based NLU framework. In the extended framework, the speech recognition information is utilized in both of the BNs training and informational goal inference processes.

Previously, our BNs are trained with perfect error free sentences. To enable the BNs in our extended NLU framework to learn the recognition behavior of the investigated speech recognizer, uncertain hypotheses output from the speech recognizer are used for training instead. Integrated scores are calculated by combining the speech recognition confidence scores with the a posteriori probabilities output by the BNs for each hypothesis. The N-best hypotheses are re-prioritized with the integrated scores and the informational goal of the sentence is then determined.

[1] N can be any number from 1 to 10.

5.1 The Speech Corpus for the Chinese ISIS Stocks Domain

We investigate our approach for spoken language understanding within the ISIS stocks domain. To collect the speech data, a set of typed reference sentences[2] regarding the ISIS stocks domain was prepared. Our collected corpus covers 9 domain-specific goals in total. They are REQUEST_CHART, REQUEST_NEWS, ASK_TRENDS, AMENDMENT, CANCELLATION, REQUEST_ACCOUNT_INFO, REQUEST_REALTIME_QUOTES, SELL and BUY.[3] According to the underlying intention, each reference sentence was tagged as one of the 9 defined goals. Speakers were asked to read the reference sentences in Chinese (Cantonese) in a natural way. The collected speech data were then passed to the speech recognizer. In Table 5.1, we show an example of a reference sentence belonging to the goal REQUEST_CHART, with its N-best recognition hypotheses generated by the speech recognizer. The numbers in the last column refer to the confidence scores output by the speech recognizer. Each of the confidence score is obtained as the result of the linear combination of the acoustic score and the language score output by the speech recognizer. A semantic tagger transformed each of the recognized N-best hypotheses into semantic concept sequences using the predefined grammar rules. The generated semantic concept sequences are then used for the training of BNs and the evaluation of our extended NLU

[2]The typed reference sentences are error-free. We refer to these as perfect reference sentences.

[3]The goal SYSTEM_META_COMMANDS was included in Chapter 3 but not here because no data were collected for this goal.

Reference	我想要晒科技股禽日既走勢圖	
Hypothesis 1	我想要晒科技股禽日既走勢圖	-24774.4
Hypothesis 2	我想要晒科技股禽日既走勢	-25307.2
Hypothesis 3	我想要晒科技股禽日既週	-26373.5
Hypothesis 4	我想要晒科技股禽日既就	-26381.5

Table 5.1: The N-best hypotheses of the reference sentence "我想要晒科技股禽日既走勢圖" of the goal REQUEST_CHART. The N-best hypotheses are prioritized with the confidence scores as listed in the right column.

framework. For experimental purposes, the whole data set has been split into disjoint training and testing sets. With the perfect reference sentences, each data set can be further split into a perfect reference set and an imperfect hypotheses set. Table 5.2 captures the distributions of the four data sets. We have in total 5967 hypotheses generated from 978 reference sentences and 2944 hypotheses generated from 495 reference sentences in our training and testing sets respectively. Table 5.3 evaluates the accuracy rates for the

	Reference Sentence #	**Hypothesis Sentence #**
Training	978	5967
Testing	495	2944

Table 5.2: Data distribution of the training and testing data sets. Reference sentences refer to those perfect typed sentences while the hypothesis sentences are those uncertain N-best hypotheses generated by the speech recognizer.

first-best hypotheses of our corpus in terms of sentence accuracy, character accuracy, concept sequence accuracy and concept accuracy.[4] Among the

	Speech Recognition		Natural Language Understanding	
	Sentence Accuracy(%)	Character Accuracy(%)	Concept Sequence Accuracy(%)	Concept Accuracy(%)
Training	45.7	86.1	53.2	79.0
Testing	47.0	86.3	52.3	79.1

Table 5.3: Accuracy rates of the training and testing spoken data sets from the perspectives of speech recognition and natural language understanding.

four accuracy rates, sentence accuracy rate and character accuracy rate were measured from the speech recognition perspective while the concept sequence accuracy rate as well as the concept accuracy rate were evaluated from the viewpoint of natural language understanding. As observed, concept sequence accuracy rate is higher than the sentence accuracy rate, and the character accuracy rate is higher than the concept accuracy rate. Such observations can be explained as follows.

In our BNs, all input queries are first abstracted into semantic concept sequences. The abstraction relaxes the exact context matching constraint as applied in the evaluation of sentence accuracy. Some examples are provided in Table 5.4.

In each of the example pairs, the indicated word pairs cannot be exactly matched in context, but their conceptual meaning are coherent and so have been abstracted into the same concept tag. Besides, since concepts usually

[4]The accuracy rates obtained for the training set are generally poorer than that of the testing set because we had 300 training sentences recorded in a noisy environment that mismatched with the quiet studio training environment of the rest 678 training sentences for the recognizer.

Concept	<OpenPrice>
Example 1a	估下 長實 聽日 開市價 幾多 <AskTrend> <StockRicHK> <RelativeDate> <OpenPrice> <Quest>
Example 1b	估下 長實 聽日 開市 幾多 <AskTrend> <StockRicHK> <RelativeDate> <OpenPrice> <Quest>
Concept	<StockRicHK>
Example 2a	宜家 中電控股 息收 幾多 <RelativeDate> <StockRicHK> <CompanyStat> <Quest>
Example 2b	宜家 中電 息收 幾多 <RelativeDate> <StockRicHK> <CompanyStat> <Quest>
Concept	<BuyAction>
Example 3a	盡入 和黃 <BuyAction> <StockRicHK>
Example 3b	盡掃 和黃 <BuyAction> <StockRicHK>

Table 5.4: Examples illustrating the relaxation of context matching evaluation scheme by concept abstraction. The indicated word pairs cannot be exactly matched in context, but their conceptual meaning are coherent and so have been abstracted into the same concept tag.

contain more than one characters, in most of the cases, the number of characters in one single sentence can be multiples of the number of concepts in that sentence. Therefore, it is easier to have a sentence with all concepts correct rather than a sentence with all characters correct. It explains why we obtain higher concept sequence accuracy rate than the sentence accuracy rate.

Concept abstraction can relax the strict context matching constraint as applied in the calculation of character / sentence error rate. However, once there are incorrect concepts caused by speech recognition errors, even for only one character, the degradation in concept error rate is often higher than that of the character error rate. For instance, in Table 5.5, the word phrase "五手", meaning the concept "<LotNumber>", was mis-recognized as the word phrase "五股", which is referring to the concept "<ShareNumber>". Such recognition error on the character "手" contributes to 20% character error rate, but 33.3% concept error rate for this sentence. It reflects the fact that usually, concept accuracy rate suffers more significantly from mis-recognition than character accuracy rate does.

5.2 Our Extended Natural Language Understanding Framework for Spoken Language Understanding

Our speech recognizer generates N-best hypotheses accompanied with confidence scores for each input utterance. Each of the BNs is trained on N-best recognition hypotheses. Since we have in total 9 BNs, for a sentence with 10-best hypotheses generated by the speech recognizer, we will have 90

Reference Sentence	入五手和黄 <BuyAction> <LotNumber> <StockRicHK>
Hypothesis Sentence	入五 股 和黄 <BuyAction> <ShareNumber> <StockRicHK>

Table 5.5: An example illustrating the recognition result for the sentence "入五手和黄". The hypothesis contains one mis-recognized character that lowers the concept accuracy rate by 33.3% but only 20% in the character accuracy rate. This explains why the character accuracy rate is higher than the concept accuracy rate.

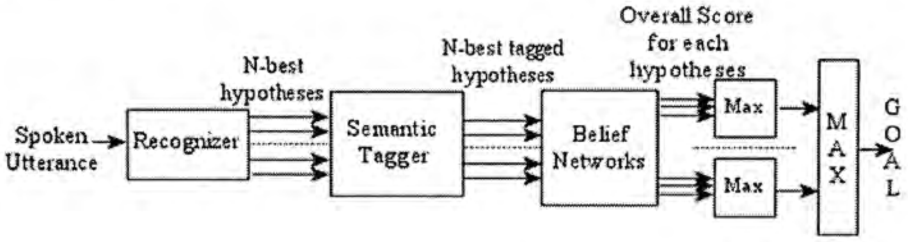

Figure 5.1: Graphical representation of informational goal identification scheme for a spoken utterance with N-best recognition outputs.

hypothesis-BN pairs. For each hypothesis-BN pair, we compute an overall score that combines the recognition confidence score with the a posteriori probability output from the BN. The strategy for combination will be described later. Then, the N-best hypotheses are re-ranked according to the integrated score. To accommodate the use of N-best recognition hypotheses, we have revised the goal identification scheme of our NLU framework. The scheme is a max-of-max strategy and is depicted in Figure 5.1. We first identify the hypothesis that scores highest for each BN. The nine hypotheses thus obtained are compared with one another, and the best-scoring among the nine represents the most probable goal for the sentence.

5.2.1 Integrated Scoring for Chinese Speech Recognition and Natural Language Understanding

To perform the max-of-max evaluation for single-goal identification, an integration score (I) is calculated for each hypothesis-BN pair. In Equation 5.1, we have stated the definition of the integration score.

$$I_{ij} = w \times R_i + (1 - w) \times \ln P(G_j = 1|\vec{\mathbf{C}}_{\mathrm{i}}) \tag{5.1}$$

where I_{ij} *is the integration score of the hypothesis* i *for the goal* G_j.

$\ln P(G_i = 1|\vec{C}_j)$ *is natural log of a posteriori probability output for the hypothesis* i *from the BN of the goal* G_j.

R_i *is the confidence score output from the speech recognizer for the hypothesis* i.

w *is a free parameter for linear weighting.*

5.3 Experiments

Our experiments are conducted on the collected speech data sets for the Chinese ISIS stocks domain. As mentioned previously, each sentence in our data sets is composed of a perfect reference sentence and a list of imperfect N-best hypotheses. Therefore, we have in total defined 4 data sets, namely the perfect reference training set, perfect reference testing set, imperfect hypothesis training set and imperfect hypothesis testing set. Each sentence was assigned with a single domain-specific goal. We have in total 9 goals defined by the data sets. It results in having 9 BNs developed in each experiment.

5.3.1 Training and Testing on the Perfect Reference Data Sets

We train the BNs with the training set of perfect reference sentences. We perform informational goal inference on the reference sentences of both the training and testing sets. Applying our max-of-max goal identification scheme (Figure 5.1), we obtain the goal identification accuracies as shown in Table 5.6. We observed that over 96% goal identification accuracies are obtained.

	Goal Identification Accuracy
Training	96.9%
Testing	96.4%

Table 5.6: Goal identification accuracies of Belief Networks using perfect reference sentences for both training and testing.

This result is a little bit higher than the 94.2% single goal identification performance under the multiple goal identification scheme as obtained for the text corpus in Chapter 3. Such satisfactory results suggest that our BN-based NLU framework is applicable to the reference set of our collected corpus.

5.3.2 Mismatched Training and Testing Conditions – Perfect Reference versus Imperfect Hypotheses

In the experiment, we alternate between the perfect reference sentences and the imperfect hypothesis sentences for both the training and testing processes. Table 5.7 tabulates the goal identification accuracies for the training and testing sets under each environment combination. We observed that under the same training environment, if we test on perfect reference sentences,

Combination #	Environment		Goal Identification Accuracies(%)	
	Training	Testing	Training	Testing
1	Perfect Reference	Perfect Reference	96.9	96.4
2	Perfect Reference	Imperfect Hypotheses	84.2	81.4
3	Imperfect Hypotheses	Perfect Reference	94.5	95.0
4	Imperfect Hypotheses	Imperfect Hypotheses	86.2	83.8

Table 5.7: Goal identification accuracies under different combinations of training and testing environments of BNs.

the goal identification performance is better than testing on imperfect hypothesis sentences. This is because the reference sentences have the complete set of correct concepts. Examples can be found from the pairs of perfect reference sentences training (Table 5.7, the first combination outperforms the second combination) and also the pairs of imperfect hypothesis sentences training (Table 5.7, the third combination outperforms the last combination).

But with the same testing environments, higher goal identification accuracies are obtained for those cases in which the training environments are compatible with the testing environments. In these cases, the BNs can better model the understanding task. For instance, to test the perfect testing reference sentences, BNs can provide a better model by training with the training reference sentences and so better performance can be obtained (Table 5.7, the first combination) when compared to the BNs trained with imperfect training hypothesis sentences (Table 5.7, the third combination). Similarly, to test on imperfect hypothesis sentences, BNs trained with imperfect training hypothesis sentences (Table 5.7, the last combination) can provide a better model than the BNs trained with perfect training reference sentences (Table 5.7, the second combination). So, for goal identification based on imperfect hypothesis sentences output from the speech recognizer, it is more desirable to use BNs trained with imperfect hypothesis sentences.

5.3.3 Comparing Goal Identification between the Use of Single-best versus N-best Recognition Hypotheses

When the N-best hypotheses are passed to a NLU component, two popular strategies may be adopted. One is to work on the most likely hypotheses only, i.e. the first-best hypotheses generated by the speech recognizer for each spoken input. This is the simplest strategy for spoken language understanding. However, this strategy may miss the more plausible candidates residing in the lower ranks in the N-best hypothesis list. In view of this, the idea that utilizes the whole N-best hypothesis lists is implemented. The proposed approach is believed to be more flexible and effective. Therefore, we have implemented our algorithm with the N-best strategy. Table 5.8 shows the performance of goal identification evaluated on the N-best hypothesis lists. The results tested on the first-best hypotheses are also included for comparison. Under each suite of BNs, the goal identification accuracies testing on the whole N-best hypothesis lists are higher than those based only on the first-best hypotheses. It implies that, apart from the first-best hypothe-

Combination #	Environment		Goal Identification Accuracies(%)	
	Training	Testing	Training	Testing
1	**Perfect Reference**	**First Best hypotheses**	83.2	79.2
2	**Perfect Reference**	**Whole N-best hypothesis lists**	84.2	81.4
3	**Imperfect Hypotheses**	**First Best hypotheses**	85.8	81.8
4	**Imperfect Hypotheses**	**Whole N-best hypothesis lists**	86.2	83.8

Table 5.8: Goal identification accuracies testing on the first-best hypotheses and the whole N-best hypothesis lists.

ses, the hypotheses residing beyond the first rank are also valuable assets for the goal identification process. In the N-best approach, the operation on each input utterance needs to consider up to 10 times the number of hypothesis when compared to only using the first-best. However, the use of N-best hypotheses can be justified by having satisfactory improvements. With reference to Table 5.8, although we only have around 2% improvement under both the perfect and imperfect training environments, the changes are in fact significant. Referring to the testing performance under the perfect training environment (see rows 1 and 2 of Table 5.8), the goal identification performance is improved from 79.2% to 81.4% when using the N-best approach instead of the first-best approach. The 2.2% improvement refers to 11 sentences that can only be identified by the N-best approach out of the 103 sentences mishandled by the first-best approach. These 11 sentences amounts to 10.7% out of the 103 that can be improved. Similarly, for the testing performance under the imperfect training environment (see rows 3 and 4 of Table 5.8), the goal identification performance is improved from 81.8% to 83.8% when using the N-best approach instead of the first-best approach. Here, the 2.0% improvement refers to 10 sentences out of the 90 sentences mishandled by the first-best approach. These 10 sentences amounts to 11.1% out of the 90 that can be improved. Moreover, since the operation time on each utterance is only few milliseconds, the increase in the operation time on processing up to 10 times of hypotheses will not have significant impact on the system performance.

5.3.4 Integration of Speech Recognition Confidence Scores into Natural Language Understanding

So far, the single-goal identification scheme determines the informational goal of the input utterance by maximizing the a posteriori probabilities of the N-best hypotheses output from the BNs. In this section, we investigate the effect of the use of integration score that integrates both the speech recognition confidence with the NLU score.

5.3.4.1 Optimization of the Linear Weighting w in the Integration Score

A series of experiments were performed in which we varied the linear weighting w in the integration score. The linear weighting balances the scores of speech recognition and natural language understanding. Variations covered the range from 0 to 1. With the use of integration score, the goal identifi-

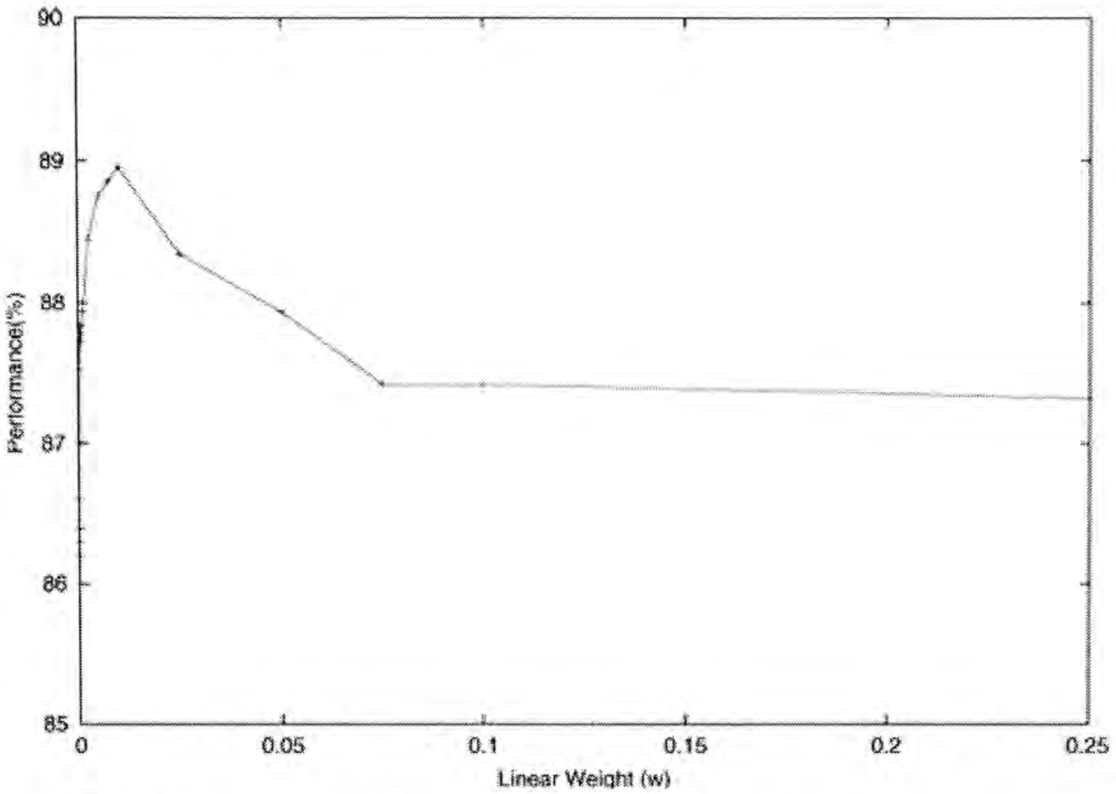

Figure 5.2: Optimization of the linear weighting w for the integration score with our training hypothesis data set (the N-best recognition hypothesis is used).

cation accuracies testing the N-best hypothesis lists of the training set on the BNs developed with hypothesis sentences are plotted in Figure 5.2. As observed in Figure 5.2, the goal identification accuracy increases with the weighting w from 0 to 0.01. Beyond the optimal weighting 0.01, the goal identification accuracy drops. Because of a scaling problem, we have only shown the results up to the weighting w equals 0.25 in Figure 5.2. The value of the optimized weighting is rather small as it serves to balance the recognition confidence levels which have a large range (a negative number with magnitude 10^5 in our collected corpus), and the small natural log value of the a posteriori probability (a negative number with magnitude 10^2 in our corpus). Applying the optimized weight in our integration score, the goal identification accuracies tested on both the training and testing N-best hypotheses are shown in Table 5.9.

	Training	Testing
Performance (%)	89.0	85.9

Table 5.9: Goal identification accuracies testing the training and testing N-best hypotheses on the BNs developed with the training hypothesis sentences using the optimized integration scores.

5.3.4.2 Investigating the Effect of Incorporating Integration Score in Goal Identification

With the use of the N-best hypotheses for testing, we compare the performance of incorporating the integration score with that using only the a posteriori probabilities output from the BNs in the goal identification process. We used w=0.01 in Equation 5.1. The goal identification accuracies are reported in Table 5.10. Under both of our training environments, the use of integration score can further improve the goal identification accuracies by 2.1% to 2.4%.

5.3.5 Feasibility of Our Approach for Spoken Language Understanding

Concluding the previous experiments, we find the following features that may boost up the performance on spoken language understanding:

1. Learning the recognition behavior of the coupled speech recognizer by training the NLU framework with speech recognition hypotheses.

Combination #	Training Environment	Evaluation Measure	Goal Identification Accuracy(%)	
			Training	Testing
1	Perfect Reference	BN probabilities	84.2	81.4
2	Perfect Reference	Integration Scores	86.9	83.8
3	Imperfect Hypotheses	BN probabilities	86.2	83.8
4	Imperfect Hypotheses	Integration Scores	89.0	85.9

Table 5.10: Goal identification accuracies tested on the N-best hypothesis list evaluated with the BN probabilities and integration scores. The use of integration score can further improve the goal identification accuracies by using only the BN probabilities.

2. Utilizing the whole N-best recognition hypothesis list by performing understanding on each of them.

3. Integrating the recognition confidence scores generated by the speech recognizer into natural language understanding.

Compiling these features, our approach on spoken language understanding is devised. The example shown in Table 5.11 helps to illustrate how our approach works. Table 5.12 showed the natural log values of the a posteriori

BUY: 係十二個半買晒所有長實		
Hyp. 1	係 十二個半 買 晒所有 長實 <At> <PriceValue> <BuyAction> <AllOfStock> <StockRicHK>	-26373.49
Hyp. 2	係 十二個半 賣 晒所有 長實 <At> <PriceValue> <SellAction> <AllOfStock> <StockRicHK>	-26385.23

Table 5.11: The speech recognition hypotheses output for the example "係十二個半買晒所有長實".

probabilities output by the BNs representing the two confusable goals BUY and SELL.

Maximizing the natural log values of the aposteriori probabilities output

	GOAL	
	BUY	SELL
Hypothesis 1	-0.0212	-0.3392
Hypothesis 2	-0.3324	**-0.0070**

Table 5.12: Natural log values of the a posteriori probabilities output by the BNs corresponding to the goals BUY and SELL for the example shown in Table 5.11.

from the BNs, Hypothesis 2 is selected and the underlying goal of the sentence is hence identified as SELL. On the contrary, if we evaluate the hypotheses with the integration scores as shown in Table 5.13, Hypothesis 1, inferring the goal BUY should be selected. As the reference goal of the sentence is BUY, it suggests that our integration score approach has a better performance in this example.

	GOAL	
	BUY	SELL
Hyp. 1	$I=$ **0.01× (-26373.49)+0.99 × (-0.0212)** $=$ **-263.7558**	$I=$ 0.01 × (-26373.49)+0.99 × (-0.3992) $=$ -264.0708
Hyp. 2	$I=$ 0.01× (-26385.23)+0.99 × (-0.3324) $=$ -264.1812	$I=$ 0.01 × (-26385.23)+0.99 × (-0.0070) $=$ -263.8593

Table 5.13: The integration scores calculated for each hypothesis-BN pair in the example shown in Table 5.11.

5.3.5.1 Evaluation of Our Extended Natural Language Understanding Framework

In this section, we develop a BNs suite with the training N-best hypothesis sentences. To perform the informational goal inference, for each input utterance, the whole N-best list is passed to the BNs. Integration scores for each hypothesis-BN pair are calculated and used to determine the informational goal of the sentence. Table 5.14 tabulates the goal identification accuracies under this approach. For benchmark purpose, we have also included the performance obtained from the traditional approach. In the traditional approach, the BNs are trained with the perfect reference sentences. To determine the informational goal of a sentence, only the a posteriori probabilities of the First-best hypotheses output from the BNs are compared.

Environment		**Confidence Measure**	**Goal Identification Accuracy(%)**	
Training	**Testing**		**Training**	**Testing**
Perfect Reference	**First Best hypotheses**	**BN probability**	83.2	79.2
Imperfect Hypotheses	**Whole N-best hypothesis lists**	**Integration Score**	89.0	85.9

Table 5.14: Goal identification accuracies obtained from the approaches of using integration score and BN probability.

5.3.5.2 Significance Testing

From Table 5.14, the goal identification accuracy of our approach is 85.9%. When compared with the 79.2% performance of our benchmark approach, it refers to 6.7% improvement achieved by our approach. To test the significance of the achieved improvement, we have conducted a significance test using sign test [31]. In Table 5.15, we have included the analytical figures used in the significance test.

	Sentence #
Correctly identified by our approach only	43
Correctly identified by the benchmark approach only	10
Total sample size	53

Table 5.15: Analysis on the number of sentences that are correctly identified by either of our approach or the benchmark approach.

In the significance test, the null hypothesis(H_0) and alternative hypothesis(H_1) are:

$$H_0\colon p = 0.5$$

$$H_1\colon p > 0.5$$

where α (the significance level) is 0.001, hence $Z_{0.001}$ is -3.09.

p_{H_0} is hypothesized proportion of the population that using our approach is better than using the benchmark approach.

q_{H_0} is hypothesized proportion of the population that using the benchmark approach is better than using our approach.

$\bar{p}$ is proportion of successes in the sample.

$\bar{q}$ is proportion of failures in the sample.

p_{H_0} is 0.5.

q_{H_0} is 0.5.

n is 53.

$\bar{p}$ is $43/53 = 0.8113$.

$\bar{q}$ is $10/53 = 0.1887$.

The test statistic is:

$$Z_0 = \frac{\bar{p} - p_{H_0}}{\sigma_{\bar{p}}} = 4.5313 \tag{5.2}$$

where

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} \qquad = \sqrt{\frac{(0.5)(0.5)}{53}} \qquad = 0.0687 \tag{5.3}$$

We reject H_0 if $Z_0 <$ -Z_α.

Since Z_0=4.5313 $>$ 3.09, we conclude that our approach can give us better goal identification accuracy than the benchmark approach and the performance difference is statistically significant.

5.3.5.3 Error Analysis

We conclude that our extended framework benefits the language understanding task on the spoken Chinese ISIS stocks domain. We found that out of the 495 sentences in our collected testing corpus, 442 queries obtained the same goal identification results under the both of the approaches using integration score and BN probability. Among the 442 queries, 382 were correctly identified by both approaches. They usually refer to the sentences that can be correctly recognized by the speech recognizer, especially as the first-best hypotheses (Table 5.16, Example 1). Therefore, training and testing on such hypotheses are similar to the cases that are done on the reference sentences. However, we have 60 sentences in which their goals cannot be correctly identified by both approaches. Most probably, they are referring to the cases that the recognition errors are overwhelming and rare (Table 5.16, Example 2).

On the other hand, we have 53 queries that have different goal identification performance by the two approaches. Among these, 43 queries are correctly identified by our approach but not the benchmark approach (Table 5.16, Example 3) while 10 queries are vice versa (Table 5.16, Example 4). Our approach outperforms the benchmark approach mainly due to the fact that we can fully utilize the speech information into natural language understanding.

Example 1: Sentence goal can be correctly identified by both approaches
Original Query : 我想知中電控股最新既報價
Original Goal : REQUEST_REALTIME_QUOTES
Result from both integration score an BN probability approaches
Hypothesis 1 : 我想知中電控股最新既報價 (Correct, sentence is perfectly recognized)
Inferred Goal : REQUEST_REALTIME_QUOTES (Correct)
Example 2: Sentence goal can be identified by neither approaches
Original Query : 幫我沽三百股香港電訊丫
Original Goal : SELL
Result from both integration score an BN probability approaches
Hypothesis 1 : 放沽左萬股零毛年上 (Incorrect, recognition errors are overwhelming in the hypothesis)
Inferred Goal : REQUEST_ACCOUNT_INFO (Incorrect)
Example 3: Sentence goal can be correctly identified by the integration score approach but not the BN probability approach
Original Query : 長實多入一萬股
Original Goal : AMENDMENT
Result from the BN probability approach
Hypothesis 1 : 長實多入一萬股
Inferred Goal : BUY (Incorrect)
Result from the integration score approach
Hypothesis 1 : 長實多入一萬股
Inferred Goal : AMENDMENT (Correct)
Example 4: Sentence goal can be correctly identified by the the BN probability approach but not the integration score approach
Original Query : 幫我係十二個半買晒所有長實既貨
Original Goal : BUY
To be continued ...

continued
Result from the BN probability approach
Hypothesis 1 : 幫我係十二個半買晒所有長實既貨
Inferred Goal : BUY (Correct)
Result from the integration score approach
Hypothesis 1 : 幫我係十二個半買晒所有長實既貨
Inferred Goal : SELL (Incorrect)

Table 5.16: Example queries and their corresponding inferred goals generated by the BN framework implementing both of our approach and the benchmark approach.

5.3.6 Justification of Using Max-of-max Classifier in Our Single Goal Identification Scheme

So far, we have proposed an extended NLU framework implementing the max-of-max goal identification scheme as depicted in Figure 5.1. One may argue to replace the max-of-max classifiers in our approach with an alternate scheme, namely the max-of-sum classifiers. In the max-of-sum scheme, the goal of a sentence is determined by the BN that obtains the maximum summed scores across all the hypotheses. Table 5.17 tabulates the results under the max-of-sum scheme. The performance we obtained from the max-of-sum scheme is not as good as the max-of-max scheme. As observed, the performance has been degraded by 3.6%. The major drawback of the max-of-

	Training	**Testing**
Performance (%)	85.5	82.2

Table 5.17: Goal identification accuracies under the max-of-sum scheme.

sum scheme is the cancellation effect of the recognition scores incorporated in the integration scores. It in turn hinders the utilization of the recognition information into natural language understanding. Table 5.18 shows a general example explaining how the max-of-sum scheme ignores the contribution of recognition scores. In the example, we have shown a sentence with 2 hypotheses. To proceed, we need to check if the summed integration scores of

For GOAL 1:			
Hyp.	**Recognition Scores**	**Belief Network Scores**	**Integration Scores**
1	R_1	BN_{11}	$I_{11} = w \times R_1 + (1-w) \times BN_{11}$
2	R_2	BN_{21}	$I_{21} = w \times R_2+(1-w) \times BN_{21}$
For GOAL 2:			
Hyp.	**Recognition Scores**	**Belief Network Scores**	**Integration Scores**
1	R_1	BN_{12}	$I_{12} = w \times R_1+(1-w) \times BN_{12}$
2	R_2	BN_{22}	$I_{22} = w \times R_2+(1-w) \times BN_{22}$

Table 5.18: An example showing the use of the max-of-sum scheme.

GOAL 1 is greater than that of the GOAL 2 as follows:

If

$$(I_{11} + I_{21}) - (I_{12} + I_{22}) > 0 \tag{5.4}$$

where I_{ij} *is the integration score of the hypothesis* i *for the goal* G_j.

Canceling the common factors, what have left are just the BN scores:

If

$$BN_{11} + BN_{21} > BN_{12} + BN_{22} \tag{5.5}$$

where BN_{ij} *is the natural log value of BN probability of the hypothesis* i *for the goal* G_j.

The expression is now independent of the recognition scores. It therefore explains why the results obtained under the max-of-sum scheme is regardless the presence of recognition confidence scores.

5.4 Chapter Summary

This chapter describes our initial attempt in the integration of speech recognition with BNs. We utilize the speech recognition information provided by the speech recognizer. Our approach focuses on (i) learning the recognition errors of recognition transcriptions that are behavior of the recognizer coupled with our NLU framework, (ii) processing the N-best hypothesis list for each input utterance and (iii) integrating speech recognition confidence scores with the goal identification process performed by the BNs. Our experiments suggest that all factors are favorable to the goal identification performance for spoken language understanding. Testing BNs implemented with the three features on our collected corpus gives us satisfactory goal identification accuracy of 85.9%. We have also compared the goal identification accuracies of our approach with that of the benchmark approach and have obtained 6.7% improvement. In the benchmark approach, none of the three features are implemented. We have performed a significant test and have showed that the performance improvement is statistically significant.

Chapter 6

Conclusions and Future Work

This chapter summarizes this thesis and discusses its main contributions. Future research directions for our work are also suggested.

6.1 Conclusions

In this thesis, we have described a natural language understanding (NLU) framework using the Belief Networks (BNs). This framework consists of a semantic tagging procedure followed by informational goal inference using BNs. The BNs model the causal relationships between the semantic concepts in the user's query and its underlying informational goal. Each BN is developed for a particular informational goal under the prescribed domain. By formulating the N-binary classifications, each performed by a BN, our NLU framework has the capability to identify queries with multiple informational goals, and reject out-of-domain (OOD) sentences.

The NLU framework has been previously used for English in the ATIS (air travel) domain. In order to demonstrate the portability of this NLU framework across both the application domains and languages, the framework is extended to handle both English and Chinese for the ISIS (stocks) domain. However, the stocks domain presents new complexities for NLU, which were not found in the ATIS domain. The prominent ones are the disambiguation of various kinds of numeric expressions and new words (OOV). We applied a transformation-based parsing (TBP) technique, which utilizes both the left and right contexts of the entities, to disambiguate the numeric expressions and infer the possible category of the OOV words. Our ported frameworks gave goal identification accuracies of 92.0% for English queries and 93.7% for Chinese queries on the ISIS stocks domain. These suggest that our NLU framework is applicable to the ISIS stocks domain and the Chinese language.

After justifying the portability of our NLU framework, we have to face another problem encountered by the NLU system development – the lack of training data. Portability of a NLU framework to new domains or languages often implies the lack of domain-specific training data. In our framework, both the semantic concept selection and BN probabilities estimation require large amounts of training data. To enhance the data collection process, a NLU system with limited capabilities must first be developed. In view of this, we have devised a set of principles that requires only handful amount of training data for the NLU system development. Our principles cover the guidelines for the semantic concept selection as well as the hand-assignment of BN probabilities. The NLU systems developed with these principles are found to have satisfactory performance with only slight degradation from the trained BNs.

With the increasing use of conversational systems in recent years, NLU systems should be able to understand the user's requests as interpreted by the speech recognizers. Nonetheless, speech recognition technologies are still far from perfect, speech recognition errors are often unavoidable. Speech recognition errors may cause the NLU system fails and generate inappropriate response. To alleviate the NLU problem originated from mis-recognition, our NLU framework is extended to utilize the information provided by the speech recognizer. In our enhanced NLU framework, we focus (i) the learning of speech recognition patterns of the speech recognizer integrated with our NLU system, (ii) the processing of the N-best hypotheses for each input utterance, and (iii) the integration of speech recognition confidence scores into natural language understanding. Significant improvements are achieved when we adopt this extended NLU framework to process the spoken input from the users than the baseline approach. In the baseline approach, the NLU system is trained with perfect reference sentences. For each input utterance, only the first-best hypothesis will be processed. Moreover, the understanding process considers only the confidence scores output from the BNs – speech recognition scores will not be considered.

6.2 Contributions

Our work has a number of innovative features that are contributive to the field of natural language understanding. The contributions are listed as follows:

1. Our BN-based NLU framework does not only work on the ATIS domain and the English language, but is also applicable to the ISIS stocks domain and the Chinese language. Our successful examples prove the portability of our BN-based NLU framework across both application

domains and languages.

2. Transformation-based parsing technique is applied in our work to disambiguate the various kinds of numeric expressions and classify the out-of-vocabulary words using the contextual information. To the best of our knowledge, it is one of the first attempts to apply the TBP technique towards language understanding.

3. Migration of NLU framework to new domains or new languages usually requires large amounts of training data, which usually forms the bottleneck in NLU system development. We have devised a set of principles that covers the schemes for semantic concept selection as well as the statistical probabilities estimation in the BNs. Developing a NLU system with our devised principles can ameliorate the reliance on the availability of large amounts of training data.

4. We extend our BN-based NLU framework to receive N-best recognition hypotheses as input. We have proved that the utilization of speech recognition information, such as N-best hypotheses, in NLU tasks can significantly improve the spoken language understanding performance.

6.3 Future Work

Possible extensions to this work include:

1. Our BN-based NLU framework couples a semantic tagging procedure with an informational goal inference performed by the BN. The semantic tagging procedure is currently conducted with the hand-designed

grammar rules. Had there been more data collected, we believe we could have applied a semi-automatic grammar induction process for acquiring such structures from unannotated corpora [55]. Moreover, for the informational goal inference, the pre-defined BN topology incorporates the simplifying assumption that all the concepts are dependent only on the goal, but are independent of one another. The topology can be enhanced by learning the inter-concept dependencies from the training data [36].

2. We have successfully extended our NLU framework to handle Chinese speech recognition hypotheses. Therefore, another possible extension of our work is to test the application of our integrated framework on English.

Bibliography

[1] J. Y. Antoine. Parsing Spoken Language without Syntax: A Microsemantic Approach. In *Proceedings of COLING, Copenhagen, Danemark, vol. 1*, 1996.

[2] Kazuhiro Arai, Jeremy H. Wright, Giuseppe Riccardi, and Allen L Gorin. Grammar Fragment Acquisition using Syntactic and Semantic Clustering. In *Proceedings of the Fifth International Conference on Spoken Language Processing (ICSLP 1998)*, 1998.

[3] M. Araki, K. Ueda, T. Nishimoto, and Y. Nimi. A Semantic Tagging Tool for Spoken Dialogue Corpus. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 2000.

[4] L.R. Bahl, F. Jelinek, and R. L. Mercer. A maximum likelihood approach to continuous speech recognition. *IEEE PAMI*, 5(2):179–190, 1983.

[5] E. Brill. Automatic Grammar Induction and Parsing Free Text: A Transformation-based Approach. In *Proceedings of 31st Annual Meeting for Computational Linguistics (ACL 1993)*, pages 259–265, 1993.

[6] E. Brill, J. C. Henderson, and G. Ngai. Automatic Grammar Induction: Combining, Reducing and Doing Nothing. In *International Workshop on Parsing Technologies*, 2000.

[7] B. Carpenter and J. Chu-Carroll. Natural Language Call Routing: A Robust, Self-Organizing Approach. In *Proceedings of the Fifth International Conference on Spoken Language Processing (ICSLP 1998)*, 1998.

[8] M. Cettolo, A. Corazza, and R. D. Mori. A Mixed Approach To Speech Understanding. In *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP 96)*, volume 1, pages 845–848, 1996.

[9] E. Charniak. Statistical Techniques for Natural Language Parsing. In *AI Magazine 18:4, AAAI, Menlo Park,California*, pages 33–44, 1997.

[10] W. N. Choi, Y. W. Wong, T. Lee, and P. C. Ching. Lexical Tree Decoding with a Class-based Language Model for Chinese Speech Recognition. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 2000.

[11] Y. Chow and R. Schwartz. The N-best Algorithm: An Efficient Procedure for Finding Top N Sentence Hypotheses. In *Proceedings of ARPA Workshop on Speech and Natural Language*, pages 190–202, 1989.

[12] M. J. Collins. A New Statistical Parser Based on Bigram Lexical Dependencies. In *Proceedings of the 34th Annual Meeting for Computational Lingistics (ACL 1996)*, pages 184–191, 1996.

[13] Y. Deng, B. Xu, and T. Huang. Chinese Spoken Language Understanding Across Domain. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 2000.

[14] P. Fetter, F. Dandurand, and P. R. Brietzmann. Word Graph Rescoring Using Confidence Measures. In *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP 96)*, volume 1, pages 10–13, 1996.

[15] V. Fischer and S. J. Kunzmann. Acoustic Language Model Classes for a Large Vocabulary Continuous Speech Recognizer. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 2000.

[16] E. Giguet and J. Vergne. From Part of Speech Tagging to Memory-based Deep Syntactic Analysis. In *Proceedings of the Fifth International Workshop on Parsing Technologies, Boston*, 1997.

[17] J. Glass. Challenges for spoken dialogue systems. In *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop '99, Keystone, CO, December*, 1999.

[18] J. Glass, G. Flammia, D. Goodine, M. Phillips, J. Polifroni, S. Sakai, S. Seneff, and V. Zue. Multingual spoken-language understanding in the MIT VOYAGER system. In *Proceedings of Speech Communications, Vol. 17, No. 1, March*, pages 1–18, 1995.

[19] J. Glass, J. Polifroni, S. Seneff, and V. Zue. Data Collection and Performance Evaluation of Spoken Dialogue Systems: The MIT Experience. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 2000.

[20] D. Goddeau, E. Brill, J. Glass, C. Pao, M. Phillips, J. Polifroni, S. Seneff, and V. Zue. GALAXY: A Human-language Interface to On-line Travel Information. In *Proceedings of the Third International Conference on Spoken Language Processing (ICSLP 94)*, pages 707–710, 1994.

[21] A. L. Gorin, G. Riccardi, and J. H. Wright. How May I Help you? In *Proceedings of Speech Communications, Vol. 23, Nos. 1–2*, pages 113–127, 1997.

[22] R. Grishman. A Very Brief Introduction To Computational Linguistics. In *Proceedings of DARPA Speech and Natural Language Workshop*, pages 37–52, 1989.

[23] T. J. Hazen, S. Seneff, and J. Polifroni. Recognition confidence scoring and its use in speech understanding systems. *Computer Speech and Language*, 16(1):49–67, January 2002.

[24] D. Heckerman and E. Horvitz. Inferring Informational Goals from Free-Text Queries: A Bayesian Approach. In *Proceedings of the Fourteenth*

Conference on Uncerntainty in Artificial Intelligence, 1998, pages 230–238, 1998.

[25] L. Hirschman, F. M. Lang, J. Dowding, and C. Weir. Porting PUNDIT to the Resource Management Domain. In *Proceedings of DARPA Speech and Natural Language Workshop, Feb*, pages 277–282, 1989.

[26] L. Hirschman, S. Seneff, D. Goodine, and M. Philips. Integrating Syntax and Semantics into Spoken Language Understanding. In *Proceedings of DARPA Speech and Natural Language Workshop, Feb*, pages 366–371, 1991.

[27] E. Jackson, D. Appelt, J. Bear, R. Moore, and A. Podlozny. A Template Matcher for Robust NL Interpretation. In *Proceedings of DARPA Speech and Natural Language Workshop*, pages 190–194, 1991.

[28] D. Jurafsky, C. Wooters, G. Tajchman, J. Segal, A. Stolcke, E. Foster, and N. Morgan. The Berkeley Restaurant Project. In *Proceedings of the third International Conference on Spoken Language Processing (ICSLP 1994)*, pages 2139–2142, 1994.

[29] L. Kathryn, A. Baker, M. Franz, and P. W. Jordan. Coping with Ambiguity in Knowledge-based Natural Language Analysis. In *Proceedings of FLAIRS*, 1994.

[30] M. Z. Kurdi. Semantic Tree Unification Grammar: a new formalism for spoken language processing. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 2000.

[31] R. I. Levin and D. S. Rubin. *Statistics for Management.* Prentice Hall, Upper Saddle River, New Jersey, 1998.

[32] Y. C. Lin and H. M. Wang. Error-tolerant Language Understanding for Spoken Dialogue Systems. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 2000.

[33] D. M. Magerman. Statistical Decision-Tree Models for Parsing. In *Proceedings of the 33rd Annual Meeting of the Association for Computation Linguistics*, pages 276–283, 1995.

[34] H. Meng, S. F. Chan, Y. F. Wong, C. C. Chan, Y. W. Wong, T. Y. Fung, W. C. Tsui, K. Chen, L. Wang, T. Y. Wu, X. L. Li, T. Lee, W. N. Choi, P.C. Ching, and H. S. Chi. ISIS: A Learning System with Combined Interaction and Delegation Dialogs. In *Proceedings of the Seventh European Conference on Speech Communication and Technology (EUROSPEECH'2001)*, 2001.

[35] H. Meng, S. F. Chan, Y. F. Wong, T. Y. Fung, W. C. Tsui, T. H. Lo, C. C. Chan, K. Chen, L. Wang, T. Y. Wu, T. Lee, W. N. Choi, Y. W. Wong, P. C. Ching, and H. S. Chi. ISIS: a multilingual spoken dialog system developed with corba and kqml agents. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 2000.

[36] H. Meng, W. Lam, and K. F. Low. Learning Belief Networks for Natural Language Understanding. In *Proceedings of ASRU*, 1999.

[37] H. Meng, W. Lam, and C. Wai. To Believe is to Understand. In *Proceedings of the Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99)*, volume 5, pages 2015–2018, 1999.

[38] H. Meng and W. C. Tsui. Comprehension Across Application Domains and Languages. In *Proceedings of the ISCSLP*, 2000.

[39] S. Miller, D. Stallard, R. Bobrow, and R. Schwartz. A Fully Statistical Approach to Natural Language Interfaces. In *Proceedings of 34th Annual Meeting for Computational Linguistics (ACL 1996)*, pages 55–61, 1996.

[40] W. Minker. An English Version of the LIMSI L' ATIS System. In *Proceedings of Technical Report 9512, LIMSI-CNRS, April. Notes et Documents LIMSI*, 1995.

[41] W. Minker. Stochastically-Based Natural Language Understanding Across Tasks and languages. In *Proceedings of the Fifth European Conference on Speech Communication and Technology (EUROSPEECH'97)*, pages 1423–1426, 1997.

[42] W. Minker, S. Bennacef, and J. L. Gauvain. A Stochastic Case Frame Approach for Natural Language Understanding. In *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP 1996)*, 1996.

[43] E. Noth, M. Boros, J. Haas, V. Warnke, and F. Gallwitz. A Hybrid Approach to Spoken Dialogue Understanding: Prosody, statistics and partial parsing. In *Proceedings of the Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99)*, volume 5, pages 2019–2022, 1999.

[44] S. Ortmanns, H. Ney, and X. Aubert. A Word Graph Algorithm for Large Vocabulary Continuous Speech Recognition. *Computer, Speech and Language*, 11(1):43–72, January 1997.

[45] D. Pallett. Benchmark Tests for DARPA Resource Management Database Performance Evaluations. In *Proceedings of ICASSP*, pages 536–539, 1989.

[46] C. Pao, P. Schmid, and J. Glass. Confidence Scoring For Speech Understanding System. In *Proceedings of the Fifth International Conference on Spoken Language Processing (ICSLP 98)*, volume 3, pages 815–818, 1998.

[47] A. N. Pargellis and A. Potamianos. Cross-domain Classification Using Generalized Domain Acts. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*, 2000.

[48] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc., San Mateo, California, 1988.

[49] J. Pearl and S. Russel. Bayesian Networks. *M. Arbib (Ed.), Handbook of Brain Theory and Neural Networks, MIT Press*, 2001.

[50] P. Price. Evaluation of Spoken Language Systems: The ATIS Domain. In *Proceedings of ARPA Human Language Technology Workshop*, pages 91–95, 1990.

[51] R. Schwartz, S. Miller, D. Stallard, and J. Makhoul. Language Understanding Using Hidden Understanding Models. In *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP 96)*, volume 2, pages 997–1000, 1996.

[52] S. Seneff. Robust Parsing for Spoken Language System. In *Proceedings of ICASSP*, pages 189–192, 1992.

[53] S. Seneff. TINA: A Natural Language System for Spoken Language Applications. In *Proceedings of Computational Linguistics, 18(1)*, pages 61–86, 1992.

[54] S. Seneff, R. Lau, and J. Polifroni. Organization, Communication and Control in the Galaxy-II Conversational System. In *Proceedings of the Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99)*, volume 3, pages 1271–1274, 1999.

[55] K. C. Siu and H. Meng. A Semi-Automatic Approach for Acquiring Domain-Specific Semantic Structures. In *Proceedings of the Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99)*, volume 5, pages 2039–2042, 1999.

[56] A. Stolcke, Y. Konig, and M. Weintraub. Explict Word Error Minimization in N-best List Rescoring. In *Proceedings of the Fifth European Conference on Speech Communication and Technology (EUROSPEECH'97)*, pages 163–166, 1997.

[57] C. J. van Rijsbergen. *Information Retrieval*. London Butterworth, 1979.

[58] C. Wai, R. Pieraccini, and H. Meng. A Dynamic Semantic Model for Rescoring Recognition Hypotheses. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2001.

[59] C. Wang. Porting the Galaxy System to Mandarin Chinese. In *S.M. thesis, MIT Department of Electrical Engineering and Computer Science, May*, 1997.

[60] C. Wang, J. Glass, H. Meng, J. Polifroni, S. Seneff, and V. Zue. YINHE: A Mandarin Chinese Version of the GALAXY System. In *Proceedings of the Fifth European Conference on Speech Communication and Technology (EUROSPEECH'97)*, pages 351–354, 1997.

[61] Y. Y. Wang. A Robust Parser for Spoken Language Understanding. In *Proceedings of Wang*, 1999.

[62] T. Ward, S. Roukos, C. Neti, J. Gros, M. Epstein, and S. Dharanipragada. Towards Speech Understanding Across Multiple Languages. In *Proceedings of the Fifth International Conference on Spoken Language Processing (ICSLP 98)*, volume 5, pages 2243–2246, 1998.

[63] W. Ward. The CMS Air Travel Information Service: Understanding Spontaneous Speech. In *Proceedings of DARPA Speech and Natural Language Workshop*, pages 123–129, 1990.

[64] V. Zue. Conversational Interfaces: Advances and Challenges. In *Proceedings of the Fifth European Conference on Speech Communication and Technology (EUROSPEECH'97)*, 1997.

[0] References without page numbers are those from CD Rom or World Wide Web.

Appendix A

Semantic Frames for Chinese

GOAL:REQUEST_CHART

{
    stock := <StockRicHK>|<StockName>|<All>|
             <HSI>|<MarketPosition>
    type := <TimeFrequency>|<RelativeDate>
}

Figure A.1: The semantic frame of the goal REQUEST_CHART in Chinese.

GOAL:REQUEST_NEWS

{
    stock := <StockRicHK>|<StockName>|<All>|
             <HSI>|<Market>
    type := <RelativeDate>
}

Figure A.2: The semantic frame of the goal REQUEST_NEWS in Chinese.

GOAL:ASK_TRENDS

{

stock	:=	<StockRicHK>|<StockName>|<All>| <HSI>|<MarketPosition>
time	:=	<RelativeDate>
Market	:=	<PriceValue>
+Market	:=	<+PriceValue>
-Market	:=	<-PriceValue>
direction	:=	<MarketMovement>|<AskTrend>

}

Figure A.3: The semantic frame of the goal ASK_TRENDS in Chinese.

GOAL:AMENDMENT

{

stock	:=	<StockRicHK>|<StockName>|<All>
price	:=	<PriceValue>|<MarketPrice>
+price	:=	<+PriceValue>
-price	:=	<-PriceValue>
lot number	:=	<LotNumber>
+lot number	:=	<+LotNumber>
-lot number	:=	<-LotNumber>
share number	:=	<ShareNumber>|<AllOfStock>
+share number	:=	<+ShareNumber>
-share number	:=	<-ShareNumber>
new action	:=	<BuyAction>|<SellAction>|<AmendAction>
prev action	:=	<CancelledAction>|<CompanyStat>
prev lot number	:=	<CancelledLotNumber>
prev share number	:=	<CancelledShareNumber>
prev price	:=	<CancelledPriceValue>
prev stock	:=	<CancelledStockRicHK>|<CancelledStockName>

}

Figure A.4: The semantic frame of the goal AMENDMENT in Chinese.

GOAL:CANCELLATION

```
{
    stock           :=  <StockRicHK>|<StockName>|
                        <CancelledStockRicHK>|
                        <CancelledStockName>|<All>
    action          :=  <BuyAction>|<SellAction>|
                        <CancelledAction>
    price           :=  <PriceValue>|<MarketPrice>|<CancelledPriceValue>
    lot number      :=  <LotNumber>|<CancelledLotNumber>
    share number    :=  <ShareNumber>|<CancelledShareNumber>|
                        <AllOfStock>
}
```

Figure A.5: The semantic frame of the goal CANCELLATION in Chinese.

GOAL: REQUEST_ACCOUNT_INFO

```
{
    stock           :=  <StockRicHK>|<StockName>|
                        <CancelledStockRicHK>|
                        <CancelledStockName>|<All>
    action          :=  <BuyAction>|<SellAction>|<CancelledAction>|
                        <AmendAction>|<AskAcc>|<Processed>
}
```

Figure A.6: The semantic frame of the goal REQUEST_ACCOUNT_INFO in Chinese.

GOAL:REQUEST_REALTIME_QUOTES

```
{
    stock           :=  <StockRicHK>|<StockName>|<All>|
                        <HSI>|<MarketPosition>
}
```

Figure A.7: The semantic frame of the goal REQUEST_REALTIME_QUOTES in Chinese.

GOAL:BUY

```
{
    stock            :=  <StockRicHK>|<StockName>|<All>
    price            :=  <PriceValue>|<MarketPrice>
    +price           :=  <+PriceValue>
    -price           :=  <-PriceValue>
    lot number       :=  <LotNumber>
    share number     :=  <ShareNumber>
}
```

Figure A.8: The semantic frame of the goal BUY in Chinese.

GOAL:SELL

```
{
    stock            :=  <StockRicHK>|<StockName>|<All>
    price            :=  <PriceValue>|<MarketPrice>
    +price           :=  <+PriceValue>
    -price           :=  <-PriceValue>
    lot number       :=  <LotNumber>
    share number     :=  <ShareNumber>|<AllOfStock>
}
```

Figure A.9: The semantic frame of the goal SELL in Chinese.

GOAL:SYSTEM_META_COMMANDS

```
{
    command          :=  <ResponseHelp>|<ResponseBye>|
                         <ResponseConfirm>|<ResponseReject>|
                         <ResponseRefresh>|<RefreshUndo>
}
```

Figure A.10: The semantic frame of the goal SYSTEM_META_COMMANDS in Chinese.

Appendix B

Semantic Frames for English

GOAL:REQUEST_CHART

{

stock := <StockRicHK>|<StockName>|<All>|
<HSI>|<MarketPosition>

type := <TimeFrequency>|<RelativeDate>

}

Figure B.1: The semantic frame of the goal REQUEST_CHART in English.

GOAL:REQUEST_NEWS

{

stock := <StockRicHK>|<StockName>|<All>|
<HSI>|<Market>

type := <RelativeDate>

}

Figure B.2: The semantic frame of the goal REQUEST_NEWS in English.

```
GOAL:ASK_TRENDS
{
    stock              :=  <StockRicHK>|<StockName>|<All>|
                           <HSI>|<MarketPosition>
    time               :=  <RelativeDate>
    Market             :=  <PriceValue>
    +Market            :=  <+PriceValue>
    -Market            :=  <-PriceValue>
    direction          :=  <MarketMovement>|<AskTrend>
}
```

Figure B.3: The semantic frame of the goal ASK_TRENDS in English.

```
GOAL:AMENDMENT
{
    stock              :=  <StockRicHK>|<StockName>|<All>
    price              :=  <PriceValue>|<MarketPrice>
    +price             :=  <+PriceValue>
    -price             :=  <-PriceValue>
    lot number         :=  <LotNumber>
    +lot number        :=  <+LotNumber>
    -lot number        :=  <-LotNumber>
    share number       :=  <ShareNumber>|<AllOfStock>
    +share number      :=  <+ShareNumber>
    -share number      :=  <-ShareNumber>
    new action         :=  <BuyAction>|<SellAction>|<AmendAction>
    prev action        :=  <CancelledAction>|<CompanyStat>
    prev lot number    :=  <CancelledLotNumber>
    prev share number  :=  <CancelledShareNumber>
    prev price         :=  <CancelledPriceValue>
    prev stock         :=  <CancelledStockRicHK>|<CancelledStockName>
}
```

Figure B.4: The semantic frame of the goal AMENDMENT in English.

GOAL:CANCELLATION

{

stock	:=	<StockRicHK>|<StockName>| <CancelledStockRicHK>|<CancelledStockName>| <All>
action	:=	<BuyAction>|<SellAction>|<CancelledAction>
price	:=	<PriceValue>|<MarketPrice>| <CancelledPriceValue>
lot number	:=	<LotNumber>|<CancelledLotNumber>
share number	:=	<ShareNumber>|<CancelledShareNumber>| <AllOfStock>

}

Figure B.5: The semantic frame of the goal CANCELLATION in English.

GOAL: REQUEST_ACCOUNT_INFO

{

stock	:=	<StockRicHK>|<StockName>| <CancelledStockRicHK>|<CancelledStockName>| <All>
action	:=	<BuyAction>|<SellAction>|<CancelledAction>| <AmendAction>|<AskAcc>|<Processed>

}

Figure B.6: The semantic frame of the goal REQUEST_ACCOUNT_INFO in English.

GOAL:REQUEST_REALTIME_QUOTES

{

stock	:=	<StockRicHK>|<StockName>|<All>| <HSI>|<MarketPosition>

}

Figure B.7: The semantic frame of the goal REQUEST_REALTIME_QUOTES in English.

```
GOAL:BUY
{
    stock           :=  <StockRicHK>|<StockName>|<All>
    price           :=  <PriceValue>|<MarketPrice>
    +price          :=  <+PriceValue>
    -price          :=  <-PriceValue>
    lot number      :=  <LotNumber>
    share number    :=  <ShareNumber>
}
```

Figure B.8: The semantic frame of the goal BUY in English.

```
GOAL:SYSTEM_META_COMMANDS
{
    command         :=  <ResponseHelp>|<ResponseBye>|
                        <ResponseConfirm>|<ResponseReject>|
                        <ResponseRefresh>|<RefreshUndo>
}
```

Figure B.9: The semantic frame of the goal SYSTEM_META_COMMANDS in English.

Appendix C

The Concept Set of Positive Evidence for the Nine Goals in English

GOAL:REQUEST_CHART		
StockRicHK	StockName	HSI
All	TimeFrequency	RelativeDate
MarketPosition	Chart	ShowWord
GOAL:REQUEST_NEWS		
StockRicHK	StockName	HSI
All	Market	RelativeDate
News	NewsSubject	ShowWord
Quest		
GOAL:ASK_TRENDS		
HSI	StockRicHK	StockName
All	RelativeDate	PriceValue
+PriceValue	-PriceValue	MarketMovement
AskTrend	MarketPosition	NewsSubject
Quest	Market	BuyAction
SellAction		
GOAL:AMENDMENT		
StockRicHK	StockName	All
PriceValue	MarketPrice	+PriceValue
-PriceValue	LotNumber	+LotNumber
-LotNumber	ShareNumber	+ShareNumber
-ShareNumber	BuyAction	SellAction
OrderName	CancelledAction	AmendAction
CancelledShareNumber	CancelledLotNumber	CancelledPriceValue
CancelledStockRicHK	CancelledStockName	Instead
CompanyStat	AllOfStock	Order
To be continued ...		

continued		
GOAL:CANCELLATION		
StockRicHK	StockName	All
CancelledStockRicHK	CancelledStockName	BuyAction
SellAction	CancelledAction	OrderName
ShareNumber	CancelledShareNumber	LotNumber
CancelledLotNumber	PriceValue	MarketPrice
CancelledPriceValue	CancelAction	Not
AllOfStock	AmendAction	Order
GOAL:REQUEST_ACCOUNT_INFO		
StockRicHK	StockName	All
CancelledStockRicHK	CancelledStockName	Portfolio
BuyAction	SellAction	CancelledAction
OrderName	AmendAction	PortfoInfo
Processed	AskAcc	ShowWord
NotYet	Quest	Order
AllOfStock	CompanyStat	RelativeDate
GOAL:REQUEST_REALTIME_QUOTES		
StockRicHK	StockName	All
HSI	MarketPosition	CompanyStat
ShowWord	AskQuote	OpenClose
MarketMovement	HowMuch	MarketPrice
Quest	RelativeDate	
GOAL:BUY		
StockRicHK	StockName	All
PriceValue	MarketPrice	+PriceValue
-PriceValue	LotNumber	ShareNumber
BuyAction	MarketMovement	CompanyStat
GOAL:SYSTEM_META_COMMANDS		
ResponseHelp	ResponseBye	ResponseConfirm
ResponseReject	ResponseRefresh	ResponseUndo

Table C.1: The concept set of positive evidence selected for the rest nine goals in English.

Appendix D

The Concept Set of Positive Evidence for the Ten Goals in Chinese

GOAL:REQUEST_CHART		
StockRicHK	StockName	HSI
All	TimeFrequency	RelativeDate
MarketPosition	Chart	ShowWord
GOAL:REQUEST_NEWS		
StockRicHK	StockName	HSI
All	Market	RelativeDate
News	NewsSubject	ShowWord
Quest		
GOAL:ASK_TRENDS		
HSI	StockRicHK	StockName
All	RelativeDate	PriceValue
+PriceValue	-PriceValue	MarketMovement
AskTrend	MarketPosition	NewsSubject
Quest	Market	BuyAction
SellAction		
GOAL:AMENDMENT		
StockRicHK	StockName	All
PriceValue	MarketPrice	+PriceValue
-PriceValue	LotNumber	+LotNumber
-LotNumber	ShareNumber	+ShareNumber
-ShareNumber	BuyAction	SellAction
CancelledAction	CancelledShareNumber	CancelledLotNumber
CancelledPriceValue	CancelledStockRicHK	CancelledStockName
AmendAction	CompanyStat	AllOfStock
Order		
To be continued ...		

continued		
GOAL:CANCELLATION		
StockRicHK	StockName	All
CancelledStockRicHK	CancelledStockName	BuyAction
SellAction	CancelledAction	ShareNumber
CancelledShareNumber	LotNumber	CancelledLotNumber
PriceValue	MarketPrice	CancelledPriceValue
CancelAction	Not	AllOfStock
AmendAction	Order	
GOAL:REQUEST_ACCOUNT_INFO		
StockRicHK	StockName	All
CancelledStockRicHK	CancelledStockName	Portfolio
BuyAction	SellAction	CancelledAction
AmendAction	PortfoInfo	Processed
AskAcc	ShowWord	NotYet
Quest	Order	AllOfStock
CompanyStat	RelativeDate	
GOAL:REQUEST_REALTIME_QUOTES		
StockRicHK	StockName	All
HSI	MarketPosition	CompanyStat
ShowWord	AskQuote	OpenClose
MarketMovement	HowMuch	MarketPrice
Quest	RelativeDate	
GOAL:BUY		
StockRicHK	StockName	All
PriceValue	MarketPrice	+PriceValue
-PriceValue	LotNumber	ShareNumber
BuyAction	MarketMovement	CompanyStat
GOAL:SELL		
StockRicHK	StockName	All
Portfolio	PriceValue	MarketPrice
+PriceValue	-PriceValue	LotNumber
ShareNumber	AllOfStock	SellAction
MarketMovement	CompanyStat	
GOAL:SYSTEM_META_COMMANDS		
ResponseHelp	ResponseBye	ResponseConfirm
ResponseReject	ResponseRefresh	ResponseUndo

Table D.1: The concept sets of positive evidence selected for the ten goals in Chinese.

Appendix E

The Complete Concept Set including Both the Positive and Negative Evidence for the Ten Goals in English

GOAL:REQUEST_CHART		
StockRicHK	StockName	HSI
All	TimeFrequency	RelativeDate
MarketPosition	Chart	ShowWord
GOAL:REQUEST_NEWS		
StockRicHK	StockName	HSI
All	Market	RelativeDate
News	NewsSubject	ShowWord
Quest		
GOAL:ASK_TRENDS		
HSI	StockRicHK	StockName
All	RelativeDate	PriceValue
+PriceValue	-PriceValue	MarketMovement
AskTrend	MarketPosition	NewsSubject
Quest	Market	BuyAction
SellAction		
GOAL:AMENDMENT		
StockRicHK	StockName	All
PriceValue	MarketPrice	+PriceValue
-PriceValue	LotNumber	+LotNumber
-LotNumber	ShareNumber	+ShareNumber
-ShareNumber	BuyAction	SellAction
OrderName	CancelledAction	AmendAction
CancelledShareNumber	CancelledLotNumber	CancelledPriceValue
To be continued ...		

continued		
CancelledStockRicHK	CancelledStockName	Instead
CompanyStat	AllOfStock	Order
Processed	NotYet	AskAcc
GOAL:CANCELLATION		
StockRicHK	StockName	All
CancelledStockRicHK	CancelledStockName	BuyAction
SellAction	CancelledAction	OrderName
ShareNumber	CancelledShareNumber	LotNumber
CancelledLotNumber	PriceValue	MarketPrice
CancelledPriceValue	CancelAction	Not
AllOfStock	AmendAction	Order
Processed	NotYet	AskAcc
Instead		
GOAL:REQUEST_ACCOUNT_INFO		
StockRicHK	StockName	All
CancelledStockRicHK	CancelledStockName	Portfolio
BuyAction	SellAction	CancelledAction
OrderName	AmendAction	PortfoInfo
Processed	AskAcc	ShowWord
NotYet	Quest	Order
AllOfStock	CompanyStat	RelativeDate
GOAL:REQUEST_REALTIME_QUOTES		
StockRicHK	StockName	All
HSI	MarketPosition	CompanyStat
ShowWord	AskQuote	OpenClose
MarketMovement	HowMuch	MarketPrice
Quest	RelativeDate	Chart
AskTrend		
GOAL:BUY		
StockRicHK	StockName	All
PriceValue	MarketPrice	+PriceValue
-PriceValue	LotNumber	ShareNumber
BuyAction	MarketMovement	CompanyStat
CancelAction	Instead	AmendAction
Processed	NotYet	AskAcc
Quest		
GOAL:SELL		
StockRicHK	StockName	All
Portfolio	PriceValue	MarketPrice
+PriceValue	-PriceValue	LotNumber
ShareNumber	AllOfStock	SellAction
MarketMovement	CompanyStat	CancelAction
Instead	AmendAction	Processed
NotYet	AskAcc	Quest
GOAL:SYSTEM_META_COMMANDS		
To be continued ...		

continued		
ResponseHelp	ResponseBye	ResponseConfirm
ResponseReject	ResponseRefresh	ResponseUndo

Table E.1: The complete concept sets including both the positive and negative evidence selected for the ten goals in English.

Appendix F

The Complete Concept Set including Both the Positive and Negative Evidence for the Ten Goals in Chinese

GOAL:REQUEST_CHART		
StockRicHK	StockName	HSI
All	TimeFrequency	RelativeDate
MarketPosition	Chart	ShowWord
GOAL:REQUEST_NEWS		
StockRicHK	StockName	HSI
All	Market	RelativeDate
News	NewsSubject	ShowWord
Quest		
GOAL:ASK_TRENDS		
HSI	StockRicHK	StockName
All	RelativeDate	PriceValue
+PriceValue	-PriceValue	MarketMovement
AskTrend	MarketPosition	NewsSubject
Quest	Market	BuyAction
SellAction		
GOAL:AMENDMENT		
StockRicHK	StockName	All
PriceValue	MarketPrice	+PriceValue
-PriceValue	LotNumber	+LotNumber
-LotNumber	ShareNumber	+ShareNumber
-ShareNumber	BuyAction	SellAction
CancelledAction	CancelledShareNumber	CancelledLotNumber
CancelledPriceValue	CancelledStockRicHK	CancelledStockName
To be continued ...		

continued		
AmendAction	CompanyStat	AllOfStock
Order	Processed	NotYet
AskAcc		
GOAL:CANCELLATION		
StockRicHK	StockName	All
CancelledStockRicHK	CancelledStockName	BuyAction
SellAction	CancelledAction	ShareNumber
CancelledShareNumber	LotNumber	CancelledLotNumber
PriceValue	MarketPrice	CancelledPriceValue
CancelAction	Not	AllOfStock
AmendAction	Order	Processed
NotYet	AskAcc	
GOAL:REQUEST_ACCOUNT_INFO		
StockRicHK	StockName	All
CancelledStockRicHK	CancelledStockName	Portfolio
BuyAction	SellAction	CancelledAction
AmendAction	PortfoInfo	Processed
AskAcc	ShowWord	NotYet
Quest	Order	AllOfStock
CompanyStat	RelativeDate	
GOAL:REQUEST_REALTIME_QUOTES		
StockRicHK	StockName	All
HSI	MarketPosition	CompanyStat
ShowWord	AskQuote	OpenClose
MarketMovement	HowMuch	MarketPrice
Quest	RelativeDate	Chart
AskTrend		
GOAL:BUY		
StockRicHK	StockName	All
PriceValue	MarketPrice	+PriceValue
-PriceValue	LotNumber	ShareNumber
BuyAction	MarketMovement	CompanyStat
CancelAction	AmendAction	Processed
NotYet	AskAcc	Quest
GOAL:SELL		
StockRicHK	StockName	All
Portfolio	PriceValue	MarketPrice
+PriceValue	-PriceValue	LotNumber
ShareNumber	AllOfStock	SellAction
MarketMovement	CompanyStat	CancelAction
AmendAction	Processed	NotYet
AskAcc	Quest	
GOAL:SYSTEM_META_COMMANDS		
ResponseHelp	ResponseBye	ResponseConfirm
ResponseReject	ResponseRefresh	ResponseUndo
To be continued ...		

continued	

Table F.1: The complete concept sets including both the positive and negative evidence selected for the ten goals in Chinese.

Appendix G

The Assignment of Statistical Probabilities for Each Selected Concept under the Corresponding Goals in Chinese

GOAL:REQUEST_CHART				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
HSI	0.670000	0.330000	0.250000	0.750000
All	0.330000	0.670000	0.750000	0.250000
TimeFrequency	1.000000	0.000000	0.250000	0.750000
RelativeDate	0.220000	0.780000	0.250000	0.750000
MarketPosition	0.780000	0.220000	0.750000	0.250000
Chart	1.000000	0.000000	0.000000	1.000000
ShowWord	0.670000	0.330000	0.500000	0.500000
GOAL:REQUEST_NEWS				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
HSI	0.670000	0.330000	0.250000	0.750000
All	0.440000	0.560000	0.250000	0.750000
Market	0.890000	0.110000	0.250000	0.750000
RelativeDate	0.220000	0.780000	0.250000	0.750000
News	1.0000000	0.000000	0.050000	0.950000
To be continued ...				

continued				
NewsSubject	0.890000	0.110000	0.050000	0.950000
ShowWord	0.670000	0.330000	0.500000	0.500000
Quest	0.560000	0.440000	0.500000	0.500000
GOAL:ASK_TRENDS				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
HSI	0.670000	0.330000	0.250000	0.750000
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
All	0.330000	0.670000	0.750000	0.250000
RelativeDate	0.220000	0.780000	0.500000	0.500000
PriceValue	0.440000	0.560000	0.250000	0.750000
+PriceValue	0.890000	0.110000	0.250000	0.750000
-PriceValue	0.890000	0.110000	0.250000	0.750000
MarketMovement	0.670000	0.330000	0.250000	0.750000
AskTrend	1.000000	0.000000	0.250000	0.750000
MarketPosition	0.780000	0.220000	0.250000	0.750000
NewsSubject	0.890000	0.110000	0.250000	0.750000
Quest	0.560000	0.440000	0.500000	0.500000
Market	0.890000	0.110000	0.250000	0.750000
BuyAction	0.670000	0.330000	0.250000	0.750000
SellAction	0.890000	0.110000	0.250000	0.750000
GOAL:AMENDMENT				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.500000	0.500000
StockName	0.220000	0.780000	0.500000	0.500000
All	0.440000	0.560000	0.750000	0.250000
PriceValue	0.440000	0.560000	0.500000	0.500000
MarketPrice	0.440000	0.560000	0.500000	0.500000
+PriceValue	0.780000	0.220000	0.500000	0.500000
-PriceValue	0.780000	0.220000	0.500000	0.500000
LotNumber	0.780000	0.220000	0.500000	0.500000
+LotNumber	1.000000	0.000000	0.500000	0.500000
-LotNumber	1.000000	0.000000	0.500000	0.500000
ShareNumber	0.780000	0.220000	0.500000	0.500000
+ShareNumber	1.000000	0.000000	0.500000	0.500000
-ShareNumber	1.000000	0.000000	0.500000	0.500000
BuyAction	0.670000	0.330000	0.500000	0.500000
SellAction	0.670000	0.330000	0.500000	0.500000
CancelledAction	0.780000	0.220000	0.500000	0.500000
CancelledShareNumber	0.780000	0.220000	0.500000	0.500000
CancelledLotNumber	0.780000	0.220000	0.500000	0.500000
CancelledPriceValue	0.780000	0.220000	0.500000	0.500000
CancelledStockRicHK	0.780000	0.220000	0.500000	0.500000
To be continued ...				

continued				
CancelledStockName	0.780000	0.220000	0.500000	0.500000
AmendAction	0.890000	0.110000	0.000000	1.000000
CompanyStat	0.670000	0.330000	0.500000	0.500000
AllOfStock	0.330000	0.670000	0.500000	0.500000
Order	0.780000	0.220000	0.500000	0.500000
Processed	0.890000	0.110000	1.000000	0.000000
NotYet	0.780000	0.220000	1.000000	0.000000
AskAcc	0.890000	0.110000	1.000000	0.000000
GOAL:CANCELLATION				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.110000	0.890000	0.750000	0.250000
StockName	0.110000	0.890000	0.750000	0.250000
All	0.440000	0.560000	0.750000	0.250000
CancelledStockRicHK	0.780000	0.220000	0.250000	0.750000
CancelledStockName	0.780000	0.220000	0.250000	0.750000
BuyAction	0.560000	0.440000	0.750000	0.250000
SellAction	0.560000	0.440000	0.750000	0.250000
CancelledAction	0.780000	0.220000	0.250000	0.750000
ShareNumber	0.670000	0.330000	0.750000	0.250000
CancelledShareNumber	0.780000	0.220000	0.250000	0.750000
LotNumber	0.670000	0.330000	0.750000	0.250000
CancelledLotNumber	0.780000	0.220000	0.250000	0.750000
PriceValue	0.330000	0.670000	0.750000	0.250000
MarketPrice	0.330000	0.670000	0.250000	0.750000
CancelledPriceValue	0.780000	0.220000	0.250000	0.750000
CancelAction	0.890000	0.110000	0.000000	1.000000
Not	1.000000	0.000000	0.500000	0.500000
AllOfStock	0.330000	0.670000	0.250000	0.750000
AmendAction	0.890000	0.110000	0.750000	0.250000
Order	0.780000	0.220000	0.250000	0.750000
Processed	0.890000	0.110000	1.000000	0.000000
NotYet	0.780000	0.220000	1.000000	0.000000
AskAcc	0.890000	0.110000	1.000000	0.000000
GOAL:REQUEST_ACCOUNT_INFO				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
All	0.440000	0.560000	0.250000	0.750000
CancelledStockRicHK	0.780000	0.220000	0.750000	0.250000
CancelledStockName	0.780000	0.220000	0.750000	0.250000
Portfolio	0.890000	0.110000	0.250000	0.750000
BuyAction	0.670000	0.330000	0.250000	0.750000
SellAction	0.670000	0.330000	0.250000	0.750000
To be continued ...				

continued				
CancelledAction	0.780000	0.220000	0.750000	0.250000
AmendAction	0.890000	0.110000	0.250000	0.750000
PortfoInfo	1.000000	0.000000	0.250000	0.750000
Processed	1.000000	0.000000	0.250000	0.750000
AskAcc	1.000000	0.000000	0.250000	0.750000
ShowWord	0.670000	0.330000	0.500000	0.500000
NotYet	0.890000	0.110000	0.250000	0.750000
Quest	0.560000	0.440000	0.500000	0.500000
Order	0.780000	0.220000	0.250000	0.750000
AllOfStock	0.330000	0.670000	0.500000	0.500000
CompanyStat	0.670000	0.330000	0.500000	0.500000
RelativeDate	0.220000	0.780000	0.500000	0.500000
GOAL:REQUEST_REALTIME_QUOTES				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.050000	0.950000
StockName	0.220000	0.780000	0.050000	0.950000
All	0.440000	0.560000	0.750000	0.250000
HSI	0.670000	0.330000	0.050000	0.950000
MarketPosition	0.780000	0.220000	0.050000	0.950000
CompanyStat	0.670000	0.330000	0.250000	0.750000
ShowWord	0.670000	0.330000	0.500000	0.500000
AskQuote	0.780000	0.220000	0.500000	0.500000
OpenClose	0.890000	0.110000	0.500000	0.500000
MarketMovement	0.670000	0.330000	0.250000	0.750000
HowMuch	0.890000	0.110000	0.500000	0.500000
MarketPrice	0.440000	0.560000	0.250000	0.750000
Quest	0.560000	0.440000	0.500000	0.500000
RelativeDate	0.220000	0.780000	0.250000	0.750000
Chart	0.890000	0.110000	1.000000	0.000000
AskTrend	0.890000	0.110000	1.000000	0.000000
GOAL:BUY				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
All	0.330000	0.670000	0.750000	0.250000
PriceValue	0.440000	0.560000	0.250000	0.750000
MarketPrice	0.440000	0.560000	0.250000	0.750000
+PriceValue	0.780000	0.220000	0.750000	0.250000
-PriceValue	0.890000	0.110000	0.250000	0.750000
LotNumber	0.780000	0.220000	0.250000	0.750000
ShareNumber	0.780000	0.220000	0.250000	0.750000
BuyAction	0.670000	0.330000	0.000000	1.000000
MarketMovement	0.670000	0.330000	0.500000	0.500000
To be continued ...				

continued				
CompanyStat	0.670000	0.330000	0.500000	0.500000
CancelAction	0.780000	0.220000	1.000000	0.000000
AmendAction	0.780000	0.220000	1.000000	0.000000
Processed	0.890000	0.110000	1.000000	0.000000
NotYet	0.780000	0.220000	1.000000	0.000000
AskAcc	0.890000	0.110000	1.000000	0.000000
Quest	0.450000	0.550000	1.000000	0.000000
GOAL:SELL				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
All	0.440000	0.560000	0.500000	0.500000
Portfolio	0.890000	0.110000	0.500000	0.500000
PriceValue	0.440000	0.560000	0.250000	0.750000
MarketPrice	0.440000	0.560000	0.250000	0.750000
+PriceValue	0.890000	0.110000	0.250000	0.750000
-PriceValue	0.780000	0.220000	0.750000	0.250000
LotNumber	0.780000	0.220000	0.250000	0.750000
ShareNumber	0.780000	0.220000	0.250000	0.750000
AllOfStock	0.330000	0.670000	0.250000	0.750000
SellAction	0.670000	0.330000	0.000000	1.000000
MarketMovement	0.670000	0.330000	0.500000	0.500000
CompanyStat	0.670000	0.330000	0.500000	0.500000
CancelAction	0.780000	0.220000	1.000000	0.000000
AmendAction	0.780000	0.220000	1.000000	0.000000
Processed	0.890000	0.110000	1.000000	0.000000
NotYet	0.780000	0.220000	1.000000	0.000000
AskAcc	0.890000	0.110000	1.000000	0.000000
Quest	0.450000	0.550000	1.000000	0.000000
GOAL:SYSTEM_META_COMMANDS				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
ResponseHelp	1.000000	0.000000	0.250000	0.750000
ResponseBye	1.000000	0.000000	0.250000	0.750000
ResponseConfirm	1.000000	0.000000	0.250000	0.750000
ResponseReject	1.000000	0.000000	0.250000	0.750000
ResponseRefresh	1.000000	0.000000	0.250000	0.750000
ResponseUndo	1.000000	0.000000	0.250000	0.750000

Table G.1: The assignment of statistical probabilities for each selected concept under the corresponding goals in Chinese.

Appendix H

The Assignment of Statistical Probabilities for Each Selected Concept under the Corresponding Goals in English

GOAL:REQUEST_CHART				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
HSI	0.670000	0.330000	0.250000	0.750000
All	0.330000	0.670000	0.750000	0.250000
TimeFrequency	1.000000	0.000000	0.250000	0.750000
RelativeDate	0.220000	0.780000	0.250000	0.750000
MarketPosition	0.780000	0.220000	0.750000	0.250000
Chart	1.000000	0.000000	0.000000	1.000000
ShowWord	0.670000	0.330000	0.500000	0.500000
GOAL:REQUEST_NEWS				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
HSI	0.670000	0.330000	0.250000	0.750000
All	0.440000	0.560000	0.250000	0.750000
Market	0.890000	0.110000	0.250000	0.750000
RelativeDate	0.220000	0.780000	0.250000	0.750000
News	1.0000000	0.000000	0.0500000	0.950000
NewsSubject	0.890000	0.110000	0.050000	0.950000
ShowWord	0.670000	0.330000	0.500000	0.500000
To be continued ...				

continued				
Quest	0.560000	0.440000	0.500000	0.500000
GOAL:ASK_TRENDS				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
HSI	0.670000	0.330000	0.250000	0.750000
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
All	0.330000	0.670000	0.750000	0.250000
RelativeDate	0.220000	0.780000	0.500000	0.500000
PriceValue	0.440000	0.560000	0.250000	0.750000
+PriceValue	0.890000	0.110000	0.250000	0.750000
-PriceValue	0.890000	0.110000	0.250000	0.750000
MarketMovement	0.670000	0.330000	0.250000	0.750000
AskTrend	1.000000	0.000000	0.250000	0.750000
MarketPosition	0.780000	0.220000	0.250000	0.750000
NewsSubject	0.890000	0.110000	0.250000	0.750000
Quest	0.560000	0.440000	0.500000	0.500000
Market	0.890000	0.110000	0.250000	0.750000
BuyAction	0.670000	0.330000	0.250000	0.750000
SellAction	0.670000	0.330000	0.250000	0.750000
GOAL:AMENDMENT				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.500000	0.500000
StockName	0.220000	0.780000	0.500000	0.500000
All	0.440000	0.560000	0.750000	0.250000
PriceValue	0.440000	0.560000	0.500000	0.500000
MarketPrice	0.440000	0.560000	0.500000	0.500000
+PriceValue	0.780000	0.220000	0.500000	0.500000
-PriceValue	0.780000	0.220000	0.500000	0.500000
LotNumber	0.780000	0.220000	0.500000	0.500000
+LotNumber	1.000000	0.000000	0.500000	0.500000
-LotNumber	1.000000	0.000000	0.500000	0.500000
ShareNumber	0.780000	0.220000	0.500000	0.500000
+ShareNumber	1.000000	0.000000	0.500000	0.500000
-ShareNumber	1.000000	0.000000	0.500000	0.500000
BuyAction	0.670000	0.330000	0.500000	0.500000
SellAction	0.670000	0.330000	0.500000	0.500000
OrderName	0.890000	0.110000	0.500000	0.500000
CancelledAction	0.780000	0.220000	0.500000	0.500000
AmendAction	0.890000	0.110000	0.050000	0.950000
CancelledShareNumber	0.780000	0.220000	0.500000	0.500000
CancelledLotNumber	0.780000	0.220000	0.500000	0.500000
CancelledPriceValue	0.780000	0.220000	0.500000	0.500000
CancelledStockRicHK	0.780000	0.220000	0.500000	0.500000
To be continued ...				

continued				
CancelledStockName	0.780000	0.220000	0.500000	0.500000
Instead	1.000000	0.000000	0.050000	0.950000
CompanyStat	0.670000	0.330000	0.500000	0.500000
AllOfStock	0.330000	0.670000	0.500000	0.500000
Order	0.780000	0.220000	0.500000	0.500000
Processed	0.890000	0.110000	1.000000	0.000000
NotYet	0.780000	0.220000	1.000000	0.000000
AskAcc	0.890000	0.110000	1.000000	0.000000
GOAL:CANCELLATION				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.110000	0.890000	0.750000	0.250000
StockName	0.110000	0.890000	0.750000	0.250000
All	0.440000	0.560000	0.750000	0.250000
CancelledStockRicHK	0.780000	0.220000	0.250000	0.750000
CancelledStockName	0.780000	0.220000	0.250000	0.750000
BuyAction	0.560000	0.440000	0.750000	0.250000
SellAction	0.560000	0.440000	0.750000	0.250000
CancelledAction	0.780000	0.220000	0.250000	0.750000
OrderName	0.780000	0.220000	0.250000	0.750000
ShareNumber	0.670000	0.330000	0.750000	0.250000
CancelledShareNumber	0.780000	0.220000	0.250000	0.750000
LotNumber	0.670000	0.330000	0.750000	0.250000
CancelledLotNumber	0.780000	0.220000	0.250000	0.750000
PriceValue	0.330000	0.670000	0.750000	0.250000
MarketPrice	0.330000	0.670000	0.250000	0.750000
CancelledPriceValue	0.780000	0.220000	0.250000	0.750000
CancelAction	0.890000	0.110000	0.000000	1.000000
Not	1.000000	0.000000	0.500000	0.500000
AllOfStock	0.330000	0.670000	0.250000	0.750000
AmendAction	0.890000	0.110000	0.750000	0.250000
Order	0.780000	0.220000	0.250000	0.750000
Processed	0.890000	0.110000	1.000000	0.000000
NotYet	0.780000	0.220000	1.000000	0.000000
AskAcc	0.890000	0.110000	1.000000	0.000000
Instead	0.890000	0.110000	1.000000	0.000000
GOAL:REQUEST_ACCOUNT_INFO				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
All	0.440000	0.560000	0.250000	0.750000
CancelledStockRicHK	0.780000	0.220000	0.750000	0.250000
CancelledStockName	0.780000	0.220000	0.750000	0.250000
Portfolio	0.890000	0.110000	0.250000	0.750000
To be continued ...				

continued				
BuyAction	0.670000	0.330000	0.250000	0.750000
SellAction	0.670000	0.330000	0.250000	0.750000
CancelledAction	0.780000	0.220000	0.750000	0.250000
OrderName	0.890000	0.110000	0.250000	0.750000
AmendAction	0.890000	0.110000	0.250000	0.750000
PortfoInfo	1.000000	0.000000	0.250000	0.750000
Processed	1.000000	0.000000	0.250000	0.750000
AskAcc	1.000000	0.000000	0.250000	0.750000
ShowWord	0.670000	0.330000	0.500000	0.500000
NotYet	0.890000	0.110000	0.250000	0.750000
Quest	0.560000	0.440000	0.500000	0.500000
Order	0.780000	0.220000	0.250000	0.750000
AllOfStock	0.330000	0.670000	0.500000	0.500000
CompanyStat	0.670000	0.330000	0.500000	0.500000
RelativeDate	0.220000	0.780000	0.500000	0.500000
GOAL:REQUEST_REALTIME_QUOTES				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.050000	0.950000
StockName	0.220000	0.780000	0.050000	0.950000
All	0.440000	0.560000	0.750000	0.250000
HSI	0.670000	0.330000	0.050000	0.950000
MarketPosition	0.780000	0.220000	0.050000	0.950000
CompanyStat	0.670000	0.330000	0.250000	0.750000
ShowWord	0.670000	0.330000	0.500000	0.500000
AskQuote	0.780000	0.220000	0.500000	0.500000
OpenClose	0.890000	0.110000	0.500000	0.500000
MarketMovement	0.670000	0.330000	0.250000	0.750000
HowMuch	0.890000	0.110000	0.500000	0.500000
MarketPrice	0.440000	0.560000	0.250000	0.750000
Quest	0.560000	0.440000	0.500000	0.500000
RelativeDate	0.220000	0.780000	0.250000	0.750000
Chart	0.890000	0.110000	1.000000	0.000000
AskTrend	0.890000	0.110000	1.000000	0.000000
GOAL:BUY				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
All	0.330000	0.670000	0.750000	0.250000
PriceValue	0.440000	0.560000	0.250000	0.750000
MarketPrice	0.440000	0.560000	0.250000	0.750000
+PriceValue	0.780000	0.220000	0.750000	0.250000
-PriceValue	0.890000	0.110000	0.250000	0.750000
LotNumber	0.780000	0.220000	0.250000	0.750000
To be continued ...				

continued				
ShareNumber	0.780000	0.220000	0.250000	0.750000
BuyAction	0.670000	0.330000	0.000000	1.000000
MarketMovement	0.670000	0.330000	0.500000	0.500000
CompanyStat	0.670000	0.330000	0.500000	0.500000
CancelAction	0.780000	0.220000	1.000000	0.000000
Instead	0.890000	0.110000	1.000000	0.000000
AmendAction	0.780000	0.220000	1.000000	0.000000
Processed	0.890000	0.110000	1.000000	0.000000
NotYet	0.780000	0.220000	1.000000	0.000000
AskAcc	0.890000	0.110000	1.000000	0.000000
Quest	0.450000	0.550000	1.000000	0.000000
GOAL:SELL				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
StockRicHK	0.220000	0.780000	0.250000	0.750000
StockName	0.220000	0.780000	0.250000	0.750000
All	0.440000	0.560000	0.500000	0.500000
Portfolio	0.890000	0.110000	0.500000	0.500000
PriceValue	0.440000	0.560000	0.250000	0.750000
MarketPrice	0.440000	0.560000	0.250000	0.750000
+PriceValue	0.890000	0.110000	0.250000	0.750000
-PriceValue	0.780000	0.220000	0.750000	0.250000
LotNumber	0.780000	0.220000	0.250000	0.750000
ShareNumber	0.780000	0.220000	0.250000	0.750000
AllOfStock	0.330000	0.670000	0.250000	0.750000
SellAction	0.670000	0.330000	0.000000	1.000000
MarketMovement	0.670000	0.330000	0.500000	0.500000
CompanyStat	0.670000	0.330000	0.500000	0.500000
CancelAction	0.780000	0.220000	1.000000	0.000000
Instead	0.890000	0.110000	1.000000	0.000000
AmendAction	0.780000	0.220000	1.000000	0.000000
Processed	0.890000	0.110000	1.000000	0.000000
NotYet	0.780000	0.220000	1.000000	0.000000
AskAcc	0.890000	0.110000	1.000000	0.000000
Quest	0.450000	0.550000	1.000000	0.000000
GOAL:SYSTEM_META_COMMANDS				
P(G=0)=0.900000	P(G=1)=0.100000			
CONCEPT	**P(C=0|G=0)**	**P(C=1|G=0)**	**P(C=0|G=1)**	**P(C=1|G=1)**
ResponseHelp	1.000000	0.000000	0.250000	0.750000
ResponseBye	1.000000	0.000000	0.250000	0.750000
ResponseConfirm	1.000000	0.000000	0.250000	0.750000
ResponseReject	1.000000	0.000000	0.250000	0.750000
ResponseRefresh	1.000000	0.000000	0.250000	0.750000
ResponseUndo	1.000000	0.000000	0.250000	0.750000
To be continued ...				

continued	

Table H.1: The assignment of statistical probabilities for each selected concept under the corresponding goals in English.