An Operator Assisted Call Routing System

Chun-Jen Lee[1,2]
[1] Telecommunication Labs.
Chunghwa Telecom Co., Ltd.
Chung-Li, Taiwan, R.O.C.
cjlee@cht.com.tw

Jason S. Chang[2]
[2] Department of Computer Science
National Tsing Hua University
Hsin-chu, Taiwan, R.O.C.
jschang@cs.nthu.edu.tw

Abstract

A system to assist call routing task for telephone operators at the Directorate General of Telecommunications (DGT) in Taiwan is reported in this paper. The system was developed based on DGT organization profile with description of its six divisions instead of a corpus of recorded and transcribed call-routing dialogs. An acoustic module and an information retrieval module were built specifically for this task. The construction of IR module was based on term extraction and thesaurus discovery processes. By integrating acoustic and IR module, the system achieves satisfactory performance and provides a promising approach to call routing. Simulation results indicated that the proposed algorithm outperforms standard classification methods. A working system based on the proposed approach has been implemented and experimental results are presented.

1 Introduction

We proposed an end-to-end system for operator assisted call routing. The system was designed based on the idea that a telephone operator is capable to capture the intent of caller's inquiry but the operators may not be familiar or kept up to date with all the business of the organization. Therefore, he or she can then operate the call router via a speech interface. The system transforms the natural speech, matches the information related to directory, and finally determines the routing destination. The output is then displayed on the screen for the operator to carry out the actual routing action. The goal is to assist the operator in selecting the desired destination that matches caller's intension. This will expedite operator's response since there is no need to search a printed directory or enter the query in Chinese text.

Numerous previous works on call routing has been reported (Riccardi et. al., 1997; Lee et al., 1998; Chu-Carroll and Carpenter, 1999). Most approaches in the literature require a corpus of routed calls to train a routing matrix or language model. Such corpora of dialog are sometimes difficult to obtain due to the concerns over invasion of privacy right. In addition to the acoustic module, a call router also needs to classify caller's request according to routing destination. This task of classification is similar to text categorization (Lewis et al., 1996) in IR or topic identification (McDonough et al., 1994) in speech research. When a corpus of calls is available, researchers tend to adopt a vector space model, under which terms in the corpus is weighted statistically and treated as independent. Many systems adopted a form-based approach that is commonly used in spoken understanding system (Seneff, Glass, and Goddeau, 1991; Lamel, 1998; Chu-Carroll, 1999; Papineni, Roukos, and Ward, 1999). A form consists of a set of keywords all (or most) of which need to be present in the caller's request in order for the destination to be activated. This idea amounts to formulating an AND query of many independent keywords.

There are problems with adopting a form-based approach. Firstly, the acoustic module produces incorrect output, which may cause the IR module to fail. If we set the precision of the acoustic output too high, there could be insufficient information for the IR module to produce a destination. On the other hand, if the precision is set low for better recall, there could be too much noise causing the IR module to produce more than one destination. The second problem is manual effect needed to construct forms for each destination each time a new call router is built. Furthermore, one needs to provide synonyms to these independent keywords in the form. Otherwise, the system will be very fragile when faced with alternative ways of expressing the same routing request.

To cope with the problems mentioned above, we design the proposed call router with the following considerations:

- Tighter integration of acoustic and IR module.
- Automatic construction of forms.
- Automatic discovery of similar words.

By integrating acoustic and IR module, the system achieves satisfactory performance and provides a promising approach to call routing. To easy the burden of developing forms for destination and ensure consistency and coverage, we develop an automatic algorithm to extract keywords based on the description of destinations as well as a corpus of news. A collection of forms can be generated semi-automatically for a specific routing task. An application developer is allowed to fine tune them in this framework.

2 System Overview

A block diagram of the system is shown in Figure 1. The speech query goes through speech recognizer and transforms into a keyword lattice. Then, call-routing classification is applied to generate a set of destination hypotheses with corresponding scores to be described latter. The score are then compared with a predefined threshold. If more than one candidate destination pass the threshold, a disambiguation module should be invoked. On the contrary, if none of candidates passes the threshold, the request is rejected.

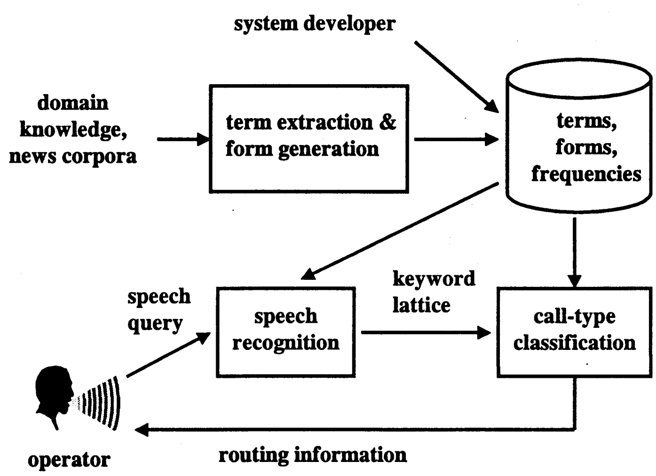

Figure 1: System Block Diagram.

To get a better understanding of the DGT call routing task, collection of domain knowledge about how a operator transfer a call to the desired destination is very important. To avoid imposing on privacy of callers, no actual dialogue between a caller and a human operator has been collected. The system was developed based on DGT organization profile with description of its six divisions. The system consists of an acoustic module and an information retrieval module built specifically for this task. A generic ASR engine was used as the kernel of the acoustic module without adapting to the application domain. The construction of IR module was based on term extraction and thesaurus discovery processes. By integrating acoustic and IR module, the system achieved satisfactory performance and provides a promising approach to call routing. After field study of operators' routing task, we roughly classified the calls into three classes, destination name, activity, and indirect request in a way similar to Chu-Carroll and Carpenter (1999). We also know that most calls can be classified according to destination name or activity. Hence, we focus on the classes of destination name and activity.

In the DGT call routing task, there are six routing destinations, including *General Planning Department, Public Telecommunications Department, Dedicated Telecommunications Department, Radio and TV Broadcast Technology Department, Radio Wave Regulatory Department,* and *Legal Office*. The organization profile of DGT describes the main responsibilities of its six departments. The mission of *GPD* and *PTD* is showed in Tables 1 and 2 respectively.

Table 1: Mission of the General Planning Department (GPD).

1	Drafting and supervising over the implementation of telecommunications policies.
2	Planning, promotion and supervision over the administration and cooperation of international telecommunications.
3	Supervision over radio communication policies and technical R&D.
4	Planning for the telecommunication related businesses and telecommunication tariffs policies.
5	Planning, operation and maintenance of the computerized information system in DGT.

Table 2: Mission of Public Telecommunications Department (PTD).

1	Issuance and change of telecom corporate licenses.
2	Management, supervision and guidance on the operations of telecom enterprises.
3	Inspection and management on the establishment of radio stations and telecom devices for the Type I telecom enterprises.
4	Approval on the operation of Type II telecom enterprises.
5	Inspection and management on telecom premise equipment.

For each of the main responsibilities of DGT departments, a form is formulated. For instance, the destination *GPD* can be represented by a number of forms representing its missions as follows:

FORM: GPD-1

1	Telecom		
2	Strategy		
3	Drafting	Supervising	Implementation

FORM: GPD-2

1	International		
2	Telecom		
3	planning	Promotion	Supervision

FORM: GPD-3

1	Radio
2	Communication
3	Supervision

FORM: GPD-4

1	Telecom		
2	Tariff		
3	Planning	Promotion	Supervision

FORM: GPD-5

1	Information		
2	System		
3	Planning	Operation	Maintenance

Each form can be viewed as representing a Boolean requirement for activation of the destination. The logical relation between slots in a form is represented by the AND operator and the relation between various admissible values in a slot is represented by the OR operator. For instance, to activate GPD via FORM GPD-1, all three slots have to be mentioned by the caller. For each slot, at least one of the values has to be mentioned.

In total, there are 95 forms constructed and 157 terms extracted from DGT profile. The wording in the profile is very formal. On the contrary, the caller tended to use informal, colloquial expressions and jargons when it comes to communication related products and services. For instance, the profile uniformly uses "cellular phones," while the caller tends to use "mobile phone" or "da-ge-da" for the same telecommunication device.

3 Automatic Identification of Keywords and Similar Words

To easy the burden of developing forms for destinations and ensure consistency and coverage, we develop an automatic algorithm to extract keywords from the description text of destinations and a corpus of newspaper articles. A collection of forms can be generated semi-automatically for a specific routing task. An application developer is allowed to fine tune them in this framework.

The first step to the problem of topic classification is word and concept extraction. Although a domain expert can complete the step manually, it is labor-intensive and can be done easily only for small application. The manual approach does not scale up to a large task. For this reason, we proposed an automatic key-phrase detection and synonym extraction method. This step will produce a number of forms with keywords representing each destination. In order to broaden the coverage of these forms, each keyword is augmented with a set of similar words. Taking advantage of work in finding similar word for word sense disambiguation, we proposed a procedure for discovery of similar words to the original keywords.

3.1 Keyword Extraction

The important keywords exhibit distinctive linguistic properties at the discourse level. Justeson and Katz (1995) pointed out domain-specific terms in a text could be characterized as noun phrases that appear more than once. This recurrence property is one of discourse phenomena called lexical cohesion. Reiteration includes identity of reference via repetition of the same words or the use of hypernyms, hyponyms, and synonyms. Due to the lack of a suitable thesaurus, we consider identical recurring words only. With this in mind, we gather the statistics of destination descriptions as well as a newspaper corpus for identification of keywords. For each term *t* in the destination description ***r*** we compute the following

$tf_{r,t}$ = Term frequency of ***t*** in routing destination description ***r,***
$tf_{d,t}$ = Term frequency of ***t*** in document ***d*** in the newspaper corpus,
atf_t = Average value of $\mathbf{tf}_{d,t}$ across documents containing ***t*** in the newspaper corpus,
df = Document frequency (number of documents containing ***t*** in the newspaper corpus),
$w_{d,t} = \mathbf{tf}_{d,t} \log (\mathbf{N/df})$,
N = Number of documents in in the newspaper corpus,
$aw_{d,t}$ = Average value of $w_{d,t}$ across documents containing *t* in the newspaper corpus.

We use the following formula to rank and extract terms from each destination description:

$$tf_{r,t} \times atf_t \times aw_{d,t}$$

The quality of extracted keywords rivals that of keyword picked by expert. Extracted keywords tend to cover the topic of destination better by providing more important terms.

3.2 Discovery of Similar Words

As mentioned above, we used the organization profile to develop the call router. One of the problems with this approach is the difference between user's colloquial vocabulary and written formal style of the material to develop the routing algorithm. In order to bridge the gap, we used an English thesaurus and bilingual dictionary to find synonyms related to the domain of telecommunication regulation. The words appearing in the organization profile written in English were extracted. Sets of English synonyms of those words were obtained in the thesaurus, Longman Lexicon of Contemporary English. Based on class-based probabilistic translation model (Brown et al. 1988), we experimented with a new approach to find compatible and appropriate translations for a group of synonymous words. Those sets (Table 3) of translations were subsequently used to expand keywords for more robust call routing.

Under the statistical machine translation model (Brown et al. 1988), translation probability $P(C|E)$ is the probability an English word E is translated into a Chinese word C. We extended the basic definition of $P(C|E)$ to include the consideration of translation probability for a class of words. For this, we tie together the probability $P(C|E)$ for a set of word E in a class K and form the class-based translation probability $P(C|K)$.

$P(C|K)$ = the probability of translating any word E in class K into C

To overcome the problem of data sparseness, we used a linear combination of probabilities in terms of n Chinese characters U_i and $n-1$ character bigram B_i in C to approximate $P(C|K)$.

$$P(C \mid K) = \frac{1}{2n} \sum_{U_i inC} P(U_i \mid K) + \frac{1}{2n-2} \sum_{B_i inC} P(B_i \mid K)$$

where n is the number of characters in C. After a couple of iterations through Expectation and Maximization Algorithm, the translations stabilized and resulted in translations which were quite suitable to use in our call router. The average accuracy of translations was around 90%. See Table 3 for examples of synonyms generated for words related to the domain of telecommunication regulation.

Table 3: Thesaurus groups related to telecommunication and their translation generated based on a class-based translation model.

Thesaurus Group	English words	Translations
C194: *nouns* Licences and permits	authority authorization charter licence pass passport permission permit sanction visa warrant	權威 職權 憲章 許可證 護照 護照 允許 正式許可文件 認可 簽證 搜查令
G215: *nouns* Telephone and telegraph	phone telecommunications telegraph telephone telephonist teleprinter	電話 電信 電報機 電話 電話接線員 電傳打字機

4 Call-Type Classification

Under the form-based representation scheme, operators provide caller's request and the system performs speech recognition and call-type classification and finally produces routing result. The difficulties of call-type classification are due to insufficient or ambiguous information provided by the operators. Moreover, even when users give an exact and unambiguous speech utterance, the correct routing destination may still be difficult to obtain due to imperfect speech recognition. Therefore, it is necessary to include a ranking function to measure how close user's request is to one of the destination.

The similarity $P(D \mid Q)$ between the query Q and the form, i.e. destination, D is formulated by combining the acoustic likelihood scores of all matching terms within-form provided by the HMM keyword spotter and information retrieval relevance score of form D to query Q provided by document retrieval formula (Chen, Jiang, and Gey, 2001). In our experiments, two scoring functions *SF1* and *SF2* as well as their corresponding weighting factors λ_1 and λ_2 were incorporated into Eq. (1) to tune the likelihood function of $P(D \mid Q)$. The ranking function can be formulated as following:

$$P(D \mid Q) = \lambda 1 \times SF1(AcousticScore) + \lambda 2 \times SF2(IRScore) \quad (1)$$

where *SF1* and *SF2* are the scoring functions of *AcousticScore* and *IRScore* respectively. Although, a discriminating optimization framework (Chiang, Lin, and Su, 1996) has been proposed in finding the suitable weights to integrate the two different scores. An ad hoc approach is employed in our preliminary experiments.

The acoustic score is formulated as follows:

$$AcoustiScore = \frac{1}{dl}\sum_{i=1}^{N}\log(p_{HMM}(kw_i)) \quad (2)$$

where

N is the number of matching terms, i.e. slots, between a form and a query,
dl is the form length (number of terms in a form),
$p_{HMM}(kw_i)$ is acoustic likelihood score of matching term kw_i.

We adopt the probabilistic approach (Chen, Jiang, and Gey, 2001) to formulate information retrieval score:

$$\begin{aligned} IRScore = &-3.51 + 37.4 * X1 + 0.330 * X2 - \\ &0.1937 * X3 + 0.929 * X4 \end{aligned} \quad (3)$$

with

$$X1 = \frac{1}{\sqrt{N}+1}\sum_{i=1}^{N}\frac{dtf_i}{dl + c_1}$$

$$X2 = \frac{1}{\sqrt{N}+1}\sum_{i=1}^{N}\log\frac{qtf_i}{ql + c_2}$$

$$X3 = \frac{1}{\sqrt{N}+1}\sum_{i=1}^{N}\log\frac{ctf_i}{cl}$$

$$X4 = N$$

where

N is the number of matching terms between a form and a query,
qtf_i is the within-query frequency of the ith matching term,
dtf_i is the within-form frequency of the ith matching term,
ctf_i is the occurrence frequency in a collection of the ith matching term,
ql is the query length (number of terms in a query),
dl is the form length,
cl is collection length, i.e. the number of occurrences of all terms in our collection,
c_1 and c_2 are two constants.

5 Experimental Results

The evaluation of the system so far has been limited to a laboratory setting. The system has been tailored to recognize not only keywords and their synonyms, but also fluent query speech. It has been greatly facilitated by the concerted operators, and really put on field trial currently. The understanding module has been subjected to speech and text input. We tested the system in two ways: short query and long query. In the short query experiment, the operator uttered only a sequence of telegraphic key phrases. In the long query experiment, the operator was allowed to utter spontaneous inquiries, which might at times be ungrammatical.

As we mentioned above, we ignored the call type of indirect request in our experiments. We recorded 127 utterances with 91 short queries and 36 long queries. Some query examples are given in the following:

- 證照作業。
  (Licensing operation.)
- 第一類特許證。
  (First class licence.)
- 電信器材審驗。
  (Telecom equipment inspection.)
- 幫我轉接至綜合規劃處。
  (Please transfer me to *General Planning Department.*)
- 負責電信編碼規劃之部門。
  (Department in charge of telecom coding and planning.)
- 要如何申請電信器材的審驗?
  (How to apply for inspection of telecom equipment?)

In order to see whether combining the scores from acoustic and IR module produce the intended effect, we conducted three experiments and compared the results. In the first experiment, we set λ_2 to zero, which means ignoring IR score in Eq. (1). In the second experiment, we set λ_1 to zero ignoring acoustic score in Eq. (1). The third experiment, both acoustic and IR scores were used. No matter for short or long queries, which has an effective error reduction, as shown in Table 4. Evidently, the framework of integrating acoustic and IR score help to achieve a high correct routing rate.

Table 4: Routing error rates in various experimental setting.

Experiments	Query types	Accuracy rate	
		Top 1	Top 2
Experiment 1	Short query	78%	96%
	Long query	64%	81%
Experiment 2	Short query	58%	85%
	Long query	61%	75%
Experiment 3	Short query	81%	96%
	Long query	69%	81%

Table 5: Top-1 errors of transcription input.

Query types	Text transcription accuracy rate
Short query	100%
Long query	92%

We noticed when considering only the top 1 routing result, speech recognition errors led to about 20% of routing errors. However, top 2 does increase the accuracy rate. That means a disambiguation dialogue module definitely will enhance the performance of a call router.

In order to know the upper bound performance of the destination routing, we also experimented with text input. The results are shown in Table 5. Clearly, as shown in Table 5, there is still some room for improvement with acoustic module, since for top 1 20-30% errors are due to imperfect speech recognition. It can be explained by that we employ a generic ASR engine without any fine-tune, so there are a lot of non-relevant keywords included in the candidate set of speech recognition. For example, the words "電視(television)" and "電腦(computer)" are easily to accompany the word "電信(telecom)" which is embedded in the query example "department in charge of telecom coding and planning" shown above. Thus, With the objective of detecting possibly erroneous keywords due to the results of ASR, we undertake a research on developing a speech recognizer with not just only a keyword spotting capability to recognize keywords embedded in a utterance, but also with a keyword verification technique (Rahim et al., 1995; Kawahara et al., 1998; Lee et al., 1998) to reject keywords that have low confidence scores.

6 Conclusion

In conclusion, we have described the operator assisted call router for DGT in Taiwan. We have exploited the statistical methods for extraction of keywords and similar words in order to build a broad coverage and robust form for linking callers' request and destination. An effective way of integrating acoustic score and information retrieval score to classify call routing destination is also described. The preliminary evaluation showed that the approach was quite effective in routing calls in face of speech noise. The approach needs only minimal information and does not require a large call routing corpus. The system proved to be quite useful in providing one-shot question answering service to operators. We are currently developing a keyword verification technique to reduce false alarm rate and an additional dialogue module to facilitate the communication between the operator and the system for better performance.

Acknowledgements

The authors would like to thank Dr. J.-H. Liang, Director of CHT-TL, Dr. B.-S. Jeng and Dr. F.-Y. Hung, Deputy Director of CHT-TL, for their fruitful support. The authors also would like to thank Dr. S.-W. Sun, Dr. J.-K. Chen and Dr. E.-F. Huang for their invaluable advice and timely encouragement. We also thank the help from the team at the NLP Lab. in NTHU for their carrying out a certain part of the work described in this paper.

References

Brown, P. F., Cocke J., Della Pietra S. A., Della Pietra V. J., Jelinek F., Mercer R. L., and Roosin P. S. 1988. A statistical approach to language translation. In *Proc. Int. Conf. Computational Linguistics*, pages 71-76.

Chen, Aitao, Hailing Jiang and Fredric C. Gey. 2001. Berkeley at NTCIR-2: Chinese, Japanese, and English IR Experiments. In *Proc. of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization*, pages 5-32 - 5-40.

Chiang, Tung-Hui, Yi-Chung Lin, and Keh-Yih Su. 1996. On jointly learning the parameters in a character-synchronous integrated speech and language model. *IEEE Trans. on Speech and Audio Processing*, 4(3):167-189.

Chu-Carroll, Jennifier. 1999. Form-based reasoning for mixed-initiative dialogue management in formation-query systems. In *Proc. EUROSPEECH*, pages 1519-1522.

Chu-Carroll, Jennifier and Bob Carpenter. 1999. Vector-based natural language call routing. *Computational Linguistics*, 25(3):361-388.

Justeson, John S. and Slava M. Katz. 1995. Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text. *Natural Language Engineering*, 1(1):9-27.

Kawahara, T., C.-H. Lee, and B.-H. Juang. 1998. Key-phrase detection and verification for flexible speech understanding. *IEEE Trans. on Audio and Speech Processing*, 6(6):558-568.

Lamel, Lori. 1998. Spoken language dialog system development and evaluation at LIMSI. In *Proc. Inter. Symposium on Spoken Dialogue*, pages 9-17.

Lee, C.-H., R. Carpenter, W. Chou., J. Chu-Carroll, W. Reichl, A. Saad, and Q. Zhou. 1998. A study on natural language call routing. In *Proc. Interactive Voice Technology for Telecommunications Applications (IVTTA)*, pages 37-42.

Lee, C.-J., E.-F. Huang, and J.-K. Chen. 1998. Using keyword spotting and utterance verification to a prank call rejection system. In *Conf. Research on Computational Linguistics*, pages 151-162.

Lewis, D. D., R. E Schapire, J. Callan and R. Papka. 1996. Training algorithm for linear text classifiers, In *Proc. SIGIR*, pages 298-306.

McDonough, J., K. Ng, P. Jeanrenaud, H. Gish, and J. R. Rohlicek. 1994 . Approaches to topic identification on the switchboard corpus. In *Proc. IEEE Int. Conf. Acoustic., Speech, Signal Processing,* pages 385-388.

Papineni, K. A., S. Roukos, and R. T. Ward. 1999. Free-flow dialog management using forms. In *Proc. EUROSPEECH*, pages 1411-1414.

Rahim, M. G, C.-H. Lee and B.-H. Juang. 1995. Robust utterance verification for connected digits recognition. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* , pages 285-288.

Riccardi, G., A. L. Gorin, A. Ljolje, and M. Riley. 1997. A spoken language system for automated call routing. In *Proc. IEEE Int. Conf. Acoustic., Speech, Signal Processing,* pages 1143-1146.

Seneff, S., J. Glass, and D. Goddeau. 1991. Development and preliminary evaluation of the MIT ATIS system. In *Proc. DARPA Speech and Natural Language Workshop*, pages 88-93.