Directory Assistance:
Learning User Formulations for Business Listings

C. Popovici, M. Andorno, P. Laface, L. Fissore, M. Nigra, C. Vair ^*

* Politecnico di Torino, Italy
{andorno,laface}@polito.it
^ Loquendo, Torino, Italy
(Cosmin.Popovici,Luciano.Fissore,Mario.Nigra,Claudio.Vair}@loquendo.com

Abstract

One of the main problems in automatic Directory Assistance (DA) for business listings is that customers formulate their requests for the same listing with a great variability.
In this paper we show that an automatic approach allows to detect, from field data, user formulations that were not foreseen by the designers, and that can be added, as variants, to the denominations already included in the system to reduce its failures.

1. Introduction

Telecom Italia has deployed since the beginning of this year a nationwide automatic DA system, jointly developed with Loquendo (formerly CSELT), that routinely serves customers asking for either residential or business listings. Whenever the automatic system is unable to terminate the transaction with the customer, the call is routed to a human operator. A description of the system, related to the management of the residential calls, has been presented in [1].
Since about 80% of the DA customer accesses are related to business listings, it is important to improve the percentage of success of the system for this class of calls.
Among the problems that must be solved to design of an automatic DA system for business listings, one of the hardest is that customers formulate their requests for the same business listing with great variability. Several approaches can be considered to face this problem: statistical language modelling and continuous speech recognition technology, or word spotting are possible candidates. However, since the perplexity of the task is very large, the risk of obtaining poor performance with a weakly constrained speech recogniser is relevant. Furthermore, it is more difficult to develop reliable confidence measures and rejection strategies for continuous speech. It was decided, thus, to design the DA service using a large vocabulary isolated word recognition technology, where the sequence of words of a business listing is concatenated and transcribed as a single word, with possible silences in between. Since the content of the original records in the database does not, typically, match the linguistic expressions used by the callers, a complex processing step is needed for deriving a set of possible formulation variants (FVs) from each original records in the listing book. Using this approach, a large percentage of expressions will not yet be perfectly covered by the FV database, and also the complexity of the search increases because the size of the system vocabularies increases.
This paper presents the results of a study aiming at automatic learning, from field data, of expressions typically used by customers to formulate their requests for the most frequent business listings. The methodology used to tackle the problem is based on partitioning the field data into phonetically similar clusters from which new user formulations can be derived.
The paper is organized as follows: Section 2 gives a short overview of the Loquendo DA system. Section 3 details the generation of FVs from the records in the book listings and describes the databases that have been used. Sections 4 and 5 present our approach for learning new formulations from the field data. Our conclusions are given Section 6.

2. Loquendo DA system overview

As stated before, the basic technology for this DA application is isolated word recognition, carried out in two steps.
The first step decodes the user utterance by means of a Hybrid HMM-NN model, where the emission probabilities of the HMM states are estimated by a Multi Layer Perceptron. This step generates also the best phonetic string of a looped model, and its score. The second step, based on Continuous Density HMMs, decodes the same utterance with the vocabulary of the N-best hypotheses produced by the first step. The added value of the second step is twofold since the combination of the hypothesis scores of the two steps, not only increases the recognition accuracy, but allows also the production of a reliability score for the best hypothesis. The dialog manager module uses the reliability and phonetic scores to reject unreliable hypotheses, or to reduce unnecessary request turns.

3. Generation of formulation variants

The Italian telephone book listings include more than 25.000.000 records, about 3.500.000 of which are business listings.

Pref: 011 *Tel*: 5175296
City: TORINO *Prov*: TORINO
Address: 63, C. VITTORIO EMANUELE II
***Den*: BAR RISTORANTE LA FORCHETTA D'ORO DI MARIO ROSSI**
***Description*: PIZZA**
***Category*: RISTORANTI**

RISTORANTE LA FORCHETTA D'ORO
BAR LA FORCHETTA D'ORO
BAR RISTORANTE LA FORCHETTA D'ORO
PIZZERIA LA FORCHETTA D'ORO
LA FORCHETTA D'ORO

Figure 1*:* Example of some fields in a record, and its formulation variants

Each record includes several fields. Of particular interest for this work are two of them: denomination and category. These fields were set up according to the indication of the subscriber, or added afterward by operators for easing their search, without any standard, and thus with a very large variability in the included data for different categories of business. The category field was not filled for several entries and has been available only in the oldest releases of the database. This information has been used for the generation of the list of expressions for business listings (FVs). Each record is classified according to 30 macro-categories, obtained from an a priori mapping of the original 1600 categories. For records that do not include the category field - about 20% of the business listings - the macro-category is obtained by means of a fuzzy classifier. Then, a lexical, grammatical, and semantic analysis of the denomination field is performed using a grammar specifically designed for each macro-category. These analyses produce a table that summarizes the semantic content of the record, reducing the variability of the information inserted by the operators in the record fields. Using the semantic table, it is possible to generate the formulation variants of a given denomination with an associated score. An example of a record, and its generated formulation variants, ordered by score, is shown in Fig. 1. The best scoring formulation is also played back to the customer for confirmation.

Several turns were performed for evaluating the coverage of the user formulations by the FVs. In a first phase, real user data were collected from the interactions with human DA operators located in Turin, then from calls to a prototype DA serving the Catania telephone district. From these preliminary tests it has been verified that the coverage of the original FVs was about 40%. It was mandatory, thus, to generate more accurate formulations for frequently requested listings, in particular for those presenting high failure rates.

Another large database (DB20000) was collected from a month and a half of customer calls to an automatic system operating in Rome. In particular, 8848 business calls, routed to the human operator by the system, because it was unable to deliver the desired information, were selected and transcribed. Another set was selected, and transcribed, from the traffic managed by 13 call centers distributed in several regions of Italy. All these calls correspond to the most frequently asked listings. The database includes 20216 transcribed calls associated to the phone number provided by the human operators.

To generate new, more accurate, FVs, the transcribed denominations were analyzed, and generation rules derived, depending on the business category, according to a priori knowledge and data evidence. The FVs that received most attention were those related to hospitals, social services, public utilities, communication and transportation agencies, and the like, because they account for the majority of the calls.

By using the FVs rules derived from this new field data, the coverage of the FVs increased from 40% to more than 60%, using an average of 5 FVs per denomination. This also means that many users are rather collaborative and that the system prompts elicit concise linguistic expressions.

4. Automatic learning of formulation variants

An analysis of the DA system failures has been done to discover the main causes of its errors. The errors can be grouped in three main classes:

1. User formulations are slightly different (due to articles, prepositions, etc.) with respect to the stored set of FVs.
2. User formulations differ because extra words or sentences are inserted, or due to the deletion of words, even though part of the information is still present.
3. User formulations are completely different with respect to the stored set.

We focus, in this work, on the errors of the first and third class. In particular, frequent errors of the third class for a specific entry can give an indication that the insertion of new formulations in the DA database is required for that entry.

4.1. Phonetic transcription

From the calls routed to the operators, the list of the most frequently requested phone numbers (provided by the operator) was selected. The log and the recording of the call have been associated to each phone number.

As introduced in Section 2, the recognition module of the system produces, together with the lexical constrained word hypotheses, the phonetic transcription of each utterance as the best sequence of phones obtained using a looped phone model. The phonetic strings associated to a given phone number are, thus, automatic transcriptions of the user formulations for the corresponding business listing.

Table 1 shows, as an example, a small set of unconstrained phonetic transcriptions associated to the most requested phone number in the DB20000 database, corresponding to *FSInforma*, a widely used automatic call center for train timetable information, developed by CSELT, and managed by the Italian railways service provider *Ferrovie dello Stato*.

These phonetic strings are widely different, and some of them can hardly be decoded. Recall, please, that these utterances were not completed by the automatic DA system for several reasons such as endpoint detection failures, extra-linguistic

Table 1: Samples of transcriptions of user requests for the railway information service *FSInforma*

ufiCoinfoRmaZionilstaZiuneditaeni	
enomaladelataZaneditoRenopoRtanovelomRaveRda	
feRoviodelostato	
ifuoRmaZionifeRoiedelostato	
esaZionetiboRtina	**fveRuilstato**
skaZione	**oRaRiodetReni**
tReno	**fRovionalostato**
nomaRomeRbefese	**saZinoCentRale**

phenomena, low confidence scores, recognition errors due to the lack of a suitable transcription in the current database, etc. Another cause of system failures is that the user request was ambiguous, incomplete or embedded in a sentence, so that only several turns of dialog with the user allowed the operator to deliver the information.

On the other hand, it is possible to detect in Table 1 phonetic sequences that are easily interpreted since they are correct or nearly correct transcriptions of a denomination such as <feRoviodelostato> and <skaZione> for “Ferrovie Dello Stato” and “Stazione” respectively, and several variants with relatively few phonetic distortions.

It is also worth noting that, given a huge number of requests for the same phone number, there is a high probability of obtaining clusters of phonetically similar strings. The distance between two strings of phones is obtained by Viterbi alignment of the two strings using the log-probability of insertion, deletion and confusion among phones. These probabilities were trained using another set of field data, aligning each phonetic sequence with its corresponding correct transcription.

These data were also exploited for training field adapted acoustic models.

4.2. Clustering and selection of new formulations

For the most frequently requested phone numbers, each set of phonetic strings was clustered into similar subsets by using a furthest neighbor hierarchical cluster algorithm based on the mutual distance between each phonetic string. The set of phonetically similar utterances is detected by setting a threshold on the within-cluster distance. The clusters with few elements and large within cluster variance are discarded.
As can be argued, the number of available samples for the Catania database is too small for deriving reliable phonetic transcriptions for new formulations. However, if a large enough database is available, it is possible to select significant clusters, characterized by high cardinality and small dispersion

Table 2 – Central elements of the three significant clusters related to the denomination FSInforma

Central element	System nearest variant	No of elements	Cluster variance	Distance
feRoviedelostato	**feRoviedelostato**	**156**	**2.13**	**0.00**
staZioneCentRale	**staZionefeRoviaRia**	**198**	**3.27**	**3.22**
staZione	**staZionefeRoviaRia**	**25**	**1.9**	**4.43**

of the included phonetic strings. For example, using the 458 formulations that were available for the phone number of service *FSInforma* in the DB20000 database, the procedure generated several phonetically similar clusters, but only three of them were significant according to a selection criterion related to the number of elements in the cluster (> 20 in this case) and to a low (< 4.0) dispersion of the elements within the cluster. The central element of the three clusters, defined as the string that has the minimum sum of the distance from all the other elements of the cluster, is shown in Table 2.
It is worth noting that, when the number of elements of a cluster is large enough, the central element of the cluster gives a very good transcription of the required denomination.
For the central elements in Table 2, good formulation variant candidates are the phonetic strings <staZioneCentRale> and <staZione> that are quite distant from the already present formulation <staZionefeRoviaRia>, while <feRoviedelostato> exactly matches a formulation already in the system.

5. Assessment of the procedure

As stated in Section 4.1, the Catania and DB20000 databases were processed to include the phone number information, provided by the human operators. However, to assess the capabilities of our approach we need a huge amount of phonetic strings and, for the first experiment, their associated phone numbers.

5.1. Automated Calls

It is, currently, cumbersome to associate the phonetic strings with the phone number that will be eventually delivered by the operator, rather than relying on the calls that the automatic DA failed to serve. Thus, we cluster the phonetic strings of calls successfully processed by the automatic system. Using these data, available at our will, we can assess the quality of our clustering procedure, and its ability to produce, as a

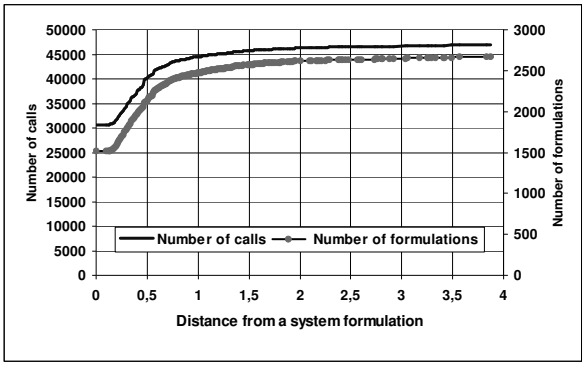

Figure 2 - Number of formulations (calls) matching a system formulation variant within a phonetic distance

central element of a cluster, a formulation variant that has been included in the system. During February 2001 it was very easy to collect, from a platform located in Turin, serving part of the North-West region of Italy, 340K phonetic strings related to business listing calls, and the related phone number provided by the automatic DA service. There are about 156K different listing references in this database (DBFEB01), 111K of which have been requested only once. We processed the denominations that were requested more than 5 times, for a total of 127K calls, obtaining 2761 clusters. Again, the most frequently requested calls refer to hospitals, police, cinema, public offices, television, and transportation.
Fig. 2 shows the number of formulations (calls) that match a system FV, corresponding to that phone number, within a phonetic distance. It can be observed that 1515 automatically detected FVs (54.9%) perfectly match a FV that was included in the system. These formulations cover 30626 (63.6%) requests. 2472 formulations (89.5%), covering 44534 calls (92.5%) are within a distance of 1.0 from the nearest system FVs. This distance corresponds to slight formulation variations like the ones shown in the first row of Table 3.

Table 3 - Examples of detection of new formulations

Cluster central element	Cluster variance	Distance	System formulation variant
kameRadikomeRCo	0.73	0.94	kameRekomeRCo
ospedalebesta	0.83	2.28	ospedale
atenaseRviZi	1.09	2.06	atenaesepia

These results demonstrate that the cluster algorithm is able to detect very well most of the old formulations. In particular, the example in the first row of the Table 3 shows that users correctly pronounce <kameRadikomeRCo>, including the Italian preposition “di” (of) that does nor appear in the current FVs.
Moreover, the automatic DA service may succeed in completing a user request, even though his formulation is distant from a system FV, because the dialog manager may ask the address information to the user. Thus, the DBFEB01 database includes associations of the correct phone number with phonetic strings that are far from the nearest system FV, but are correct transcriptions of new formulations. Example of these formulations are shown in the last two rows of Table 3: in the third row the name of the hospital “*Besta*” is normally formulated by the users, while the acronym S.P.A. is substituted with word “*Servizi*” (services).

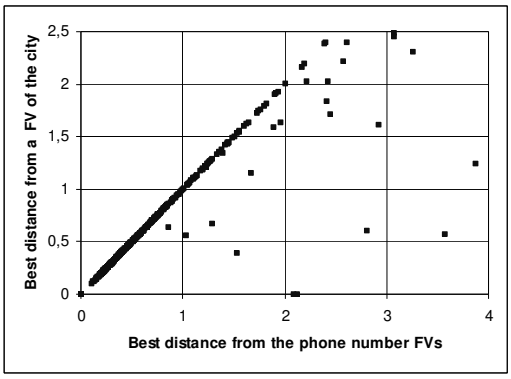

Figure 3 - Scatter plot of best distances of a central element of a cluster from the FVs of its phone number (x-axis), and from all the FVs of the city (y-axis)

As a final assessment we computed, for each of the 2761 central elements produced by our clustering procedure, its distance from *all* the system FVs of the city corresponding to the area code of the phone number associated with the cluster.
Fig. 3 is a scatter plot where the *x* and *y* coordinates of a point represent the distance of a phonetic string from the nearest phone number FV, and from the nearest FVs of the city listings respectively. Thus, the 2618 points laying on the 45° right line correspond to automatic phonetic transcriptions that are close to one of the FVs associated to the correct denomination. As better shown in Fig. 2, 1515 of them perfectly match a system FV. The 143 points laying below the 45° right line indicate that a central element is closer to an incorrect denomination. It is worth noting, however, that the points located on the right side of the figure may correspond to new formulations if the cardinality of the corresponding cluster is high and the within cluster variance is low. We obtained, for example, the phonetic string <CentRokomeRCalemetRopoli> as the central element of a cluster including 48 user formulations, having a cluster variance of 0.94, and a distance 3.54 from the current best system FV <CentRokomeRCale>. Since it satisfies the criterion of cluster robustness and "purity", and since it is far from the nearest system FV, it can be added to the current FVs as a new formulation.
The distribution of the points in Fig. 3, concentrated on the (lower part of) the 45° right line, assesses the viability of the approach for incremental learning of new (phonetically accurate) formulations if a large enough database is available for the requests routed to the operators.

5.2. Calls routed to the operator

In the DBFEB01 database there are about 700K calls routed to the operators, and related to phone numbers of 7368 cities. Since for these calls we don't have the phone number information associated with the phonetic string, we cluster all the strings associated with a given city name. The number of these strings can be as large as 164K for a big city like Milan.
For these experiments it was necessary to adapt the clustering procedure to deal with a logically huge, but sparse distance matrix. To compute the distance among all the phonetic strings, we use a recursive tree to tree matching procedure, where a tree branch is a phonetic string, as sketched in Fig. 4. Since the distance is defined to be symmetric, a phonetic string is matched only against the tree branches at its right using Viterbi beam search. The trellis computed during the

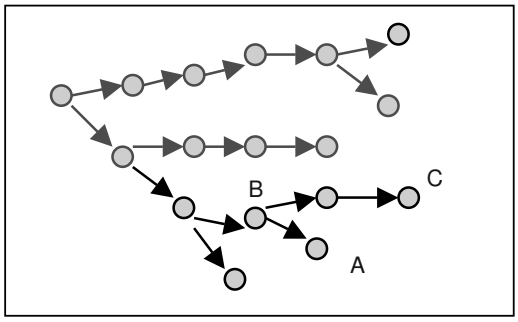

Fig. 4 - Tree of the phonetic strings

dynamic programming alignment of branch A against the branches at its right is kept, up to the column corresponding to node B, to minimize the computation load when matching branch C.
Using a rather conservative approach for clustering we obtained 1427 clusters (and the phonetic string for their central elements) referring to 363 cities. These clusters cover 37k calls, about 7% of the total calls for these cities (542K).
431 of the detected clusters have a central element that is already covered by a system formulation variant (its distance from the system FV is lower than a threshold).
The remaining 996 central elements, covering 24K calls, can be inspected to eliminate few inappropriate formulations, such as "doctor", or "hospital" for a big city, or "operator".
From this relatively small database we obtained 976 new formulations, covering 23K calls. Some formulations derived from routed calls are shown in Table 4.

Table 4: Some formulations derived from routed calls

Nr	Phonetic string	Distance	System formulation	Phone Nr
17	istitutodetumoRe	1.68	-> istitutodemaRki	0257991
26	longonispoRta	1.84	-> longonibaR	024694816
69	atenaseRviZe	1.93	-> GemeseRviCe	0248004792
50	aeRopoRtolinate	0.83	-> aeRopoRtodilinate	0274851
31	kameRadikomeRCo	1.17	-> kameRekomeRCo	0303770248
38	infostRada	2.00	->polstRada	02326781
21	infostRadaesepia	1.34	-> autostRadaesepia	0235201
50	sekondamano	2.33	-> skuolamaino	025510171
16	meRkatoneuno	1.21	-> meRkatone	0143609611

6. Conclusions

We have shown that an unsupervised approach is able to detect user formulations that were not foreseen by the designers of a DA system. These formulations can be added to the system to reduce its failures. Conversely, the system formulations generated from the book listings for a given denomination, that never appeared in a huge amount of real data can be eliminated, to reduce the search costs.
This approach has been used to automatically detect new formulations, discover why existing formulations exhibit high rejection rates, and to compute the percentage of routed versus automated calls per phone number.

7. References

[1] R. Billi, F. Canavesio, C. Rullent, "Automation of Telecom Italia Directory Assistance Service: Field Trials results", Proc. IVTTA 1998, Turin, pp. 11-16, 1998.