Learning of User Formulations for Business Listings in Automatic Directory Assistance

Original

Availability:
This version is available at: 11583/1413117 since:

Publisher:

Published
DOI:

Terms of use:
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)
Learning of User Formulations for Business Listings in Automatic Directory Assistance

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

Abstract

Automatic Directory Assistance (DA) for business listings poses many application specific problems. One of the main problems is that customers formulate their requests for the same listing with a great variability.

This paper presents the results of a study aiming at the evaluation of an approach towards automatic learning, from field data, of expressions typically used by customers to formulate their requests for the most frequent business listings.

We use a clustering procedure that exploits the association of the phonetic string produced by a lexical unconstrained search for a given denomination pronounced by the user and the phone number provided by the system or by the human operator, in case of failure of the automatic DA service.

We show that an unsupervised approach allows to detect 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 an automatic DA system. The automated DA service, developed by Telecom Italia and Loquendo (formerly CSELT), routinely serves customers asking for both residential and business listings. Whenever the automatic system is unable to terminate the transaction with the customer, the call is routed to a human operator.

The Italian telephone book listings includes more than 25,000,000 records, about 3,500,000 of which are business listings. Since about 80% of the DA accesses is related to business listings, it is important to improve the percentage of success of the automatic DA system for this class of calls.

The design of an automatic DA system for business listings poses several problems, one of the hardest to be solved is that customers formulate their requests for the same business listing with great variability. Several approaches have been considered to face this problem. By using statistical language modelling and continuous speech recognition technology a quite flexible system with respect to the user formulations, providing high linguistic coverage, could be developed. 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. A word spotting approach presents the same problems enhanced by the size of the vocabulary.

It was decided, thus, to design the DA service using a large vocabulary isolated word recognition technology, where the sequence of words in business listings 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 formulations variants (FVs) from each original record in the book listings.

The advantage of the FV approach is to supply the speech recognizer with some knowledge about the variability of the user formulations, and to allow the use of isolated word recognition technology with the capability of dealing with out of vocabulary words. On the other hand, it is clear that a large percentage of expressions will not be perfectly covered by the FV database, and that the complexity of the search increases because the size of the system vocabularies increases.

This paper presents the results of a study aiming at the evaluation of an approach towards automatic learning, from field data, of expressions typically used by customers to formulate their requests for the most frequent business listings.

We partition the field data referring to a given denomination into phonetically similar clusters from which new user formulations can be derived. Our clustering procedure exploits the association of the phonetic string produced by a lexical unconstrained search for a denomination pronounced by the user and the phone number provided by the system or by the human operator, in case of failure of the automatic DA service.

The paper is organized as follows: Section 2 gives a short overview of the Loquendo DA system. Section 3 recalls the steps for obtaining FVs from the records in the book listings and from field data. Section 4 details our approach for learning new formulations from field data, and discusses the obtained results. Finally, our conclusions are given in Section 5.

2. Loquendo DA system overview

As introduced in the previous section, large vocabulary isolated word recognition is the basic technology embedded in the Loquendo DA application.
includes a total of 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. Since the automatic DA system is currently fully operational, new FVs, and possibly rules, are also derived whenever the service provider signals consistent and important anomalies. 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 are different with respect to the stored set due to the insertion of extra words or sentences or due to the deletion of words, even though part of the information is still there.
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. To each phone number the trace and the recording of the corresponding call has been associated. Setting a threshold of 20 requests per phone number, the most requested listings for the 3434 calls in the Catania database are 16 only. A much higher spreading has been experienced, as expected, for the nationwide calls, where 53 listings only were requested more than 20 times. As said 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, the automatic transcriptions of the different way in which users formulate their request 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: 848888088 corresponding to FSInforma, a widely used automatic train timetable information system, developed by CSEL, 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 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, in Table 1 it is possible to detect phonetic sequences that are easily interpreted since they are correct or nearly correct transcriptions of a denomination such as <ferroviedelostato> and <stazione> 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 measure of the distance between two strings of phones can be obtained by Viterbi alignment of the two strings using the inverse of the log-probability of insertion, deletion and confusion among phones. These probabilities were trained using another set of field data, aligning each phonetic string 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 bottom-up clustering when the number of elements collapsed into a cluster is large enough, the central element of the cluster gives a very good transcription of the required denomination.

**Table 2: Two clusters for different formulations related to main Catania Hospital**

<table>
<thead>
<tr>
<th>French pronunciation</th>
<th>Italian pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>o&amp;ana</td>
<td>aun&amp;an</td>
</tr>
<tr>
<td>vo&amp;ana</td>
<td>aun&amp;an</td>
</tr>
<tr>
<td>fo&amp;ana</td>
<td>auv&amp;a</td>
</tr>
<tr>
<td>fo&amp;an</td>
<td>aun&amp;ian</td>
</tr>
<tr>
<td>ok&amp;an</td>
<td>aun&amp;an</td>
</tr>
</tbody>
</table>

**Table 3: Samples of two clusters of similar pronunciations for the denomination Auchan**

<table>
<thead>
<tr>
<th>Central element</th>
<th>Within - system nearest variant</th>
<th>No of elements</th>
<th>Cluster variance</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ferroviedelostato</td>
<td>feRoviedelostato</td>
<td>156</td>
<td>2.13</td>
<td>0.00</td>
</tr>
<tr>
<td>staZioneCentrale</td>
<td>staZioneFeRoviaRia</td>
<td>198</td>
<td>3.27</td>
<td>3.22</td>
</tr>
<tr>
<td>staZione</td>
<td>staZioneFeRoviaRia</td>
<td>25</td>
<td>1.9</td>
<td>4.43</td>
</tr>
</tbody>
</table>

**Table 4 – Central elements of the three significant clusters related to the denomination FSInforma**

characterized by high cardinality and small dispersion of the included phonetic strings. For example, using the 458 formulations that were available for the phone number of the 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.

An interesting example of output of the clustering procedure is related to the main Catania Hospital, whose phone number is requested using several formulations referring to different clinics within the same hospital. Six denominations were added manually as formulation variants after the preliminary analysis of the errors of the prototype system in operation in Catania. The clustering algorithm detected the same formulations: a few elements of two clusters are shown in Table 2.

Another example of automatic clustering, detecting a pronunciation variant of a foreign denomination, is reported in Table 3, where some phonetic strings of two clusters related to requests for the French word Auchan are shown. 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.
of Italy, 336113 phonetic strings related to business listing calls, and the related phone number provided by the automatic DA service. There are a total of 153043 different listing references in this database (DBFEB01), 108026 of which have been requested only once. To obtain reliable statistics we processed the 358 denominations that were requested more than 50 times, for a total of 46516 calls. Again, the most frequently requested one is the P3informa service, and the majority of the other calls refer to hospitals, police, cinema, public offices, television, and transportation.

Figure 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 448 automatically detected FVs (43.8%) perfectly match a FV that was included in the system. These formulations cover 26200 (56.3%) requests for 270 different denominations (75.4%). 803 formulations (77.9%), covering 40079 calls (86.1%) for 335 denominations (75.4%). 803 formulations are shown in the last two rows of Table 5: 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).

As a final assessment we computed, for each of the 1031 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. Figure 4 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 934 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, 448 of them perfectly match a system FV. The 97 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 upper 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 <CentRokomeRCalmetRopoli> 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 <CentRokomeRCalc>. 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 upper 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 <CentRokomeRCalmetRopoli> 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 <CentRokomeRCalc>. 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 upper 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 <CentRokomeRCalmetRopoli> 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 <CentRokomeRCalc>. 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.

5. 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 by the rule-based system from the book listings for a given denomination, that never appeared in a huge amount of real data can be purged.

6. References