Exploiting semantic relations for a Spoken Language Understanding application

Catherine Kobus(1)(2), *Géraldine Damnati*(1), *Lionel Delphin-Poulat*(1)
Renato De Mori(2)

(1) France Télécom R& D - 2, avenue Pierre Marzin 22307 Lannion - France
(2) LIA - University of Avignon, BP 1228 84911 Avignon Cedex - France

Abstract

This article proposes a new confidence measure estimated for concept hypotheses provided by a semantic language model used in the context of a dialog application. This confidence measure is based upon the ontology and more precisely, upon the semantic relations between concepts. It aims at measuring how high a concept hypothesis is related to the other hypotheses of an utterance. The semantic relation confidence measure is evaluated alone, and in combination with a classical acoustic confidence measure. The two measures are also used as parameters of a decision tree. It is shown that the two confidence measures are complementary and yield good performance in terms of cross entropy relative reduction.

Index Terms: confidences measures, Spoken Language Understanding, ontology, semantic relations, decision tree

1. Introduction

Confidence measures are often used in speech recognition systems in order to measure reliability of answers (or hypotheses) provided by the system. Applications of confidence measures are manyfold; in the context of a dialog application, a rejection strategy or an adaptation of the dialog (confirmation request, etc.) can be achieved. They can be evaluated at different levels : word, concept or utterance level. In a context of a Spoken Language Understanding (SLU) application, it is crucial to know the reliability degree of recognized concepts. In this paper, the confidence measures are estimated for concept hypotheses provided by a semantic language model, applied to a word graph in a second recognition pass. Semantic confidence measures have already been proposed. [2] and [3] defined semantic confidence measure based upon the *a posteriori* probability of the concept. [4] validates a concept hypothesis by analyzing the consensus of different semantic classifier. In this paper, we introduce a new semantic confidence measure, which uses the ontology of the application and precisely the different semantic relations between concepts. The resulting measure gives, for each concept hypothesis of an utterance, a degree of semantic consistency with the other hypotheses of the utterance. The context is a banking application, which allows users to perform transactions and ask questions about the stock market, portfolios and accounts.

The paper is organized as follows : section 2 defines the semantic confidence measure and the acoustic confidence measure; section 3 presents methods to combine confidence measures. Logistic regression is used to directly combine the two confidence measures. A decision tree can take into account of the information on the concept identity. Experimental setup and evaluations of the confidence measures, used individually or in combination, are detailed in section 4.

2. Confidence measures

2.1. Semantic relation confidence measure

The first confidence measure uses the knowledge source provided by the semantic relations between the elementary concepts of a dialog application.

2.1.1. Ontology of a dialog application

An ontology is a structured set of concepts; it describes how the different concepts are organized and related to each other by semantic relations. We used and adapted the KL-ONE description of the ontology; KL-ONE [5] is intended to represent general conceptual information and is typically used in the construction of the knowledge base of a single reasoning entity. It distinguishes two types of concepts : generic and individual concepts. The first one are sets, while the second one are instances of the generic concepts. A generic concept is completely defined by the roles (or relations) with other concepts. An individual concept inherits the roles from the generic concept, whose it is an instance. Figure 1 shows an example of a semantic relation, linking the generic concepts *claIndex* and *claLevel*; the respective individual concepts inherit this relation. *level* is the only one concept instance of the generic concept $claLevel$ while *Nikkei*, *Dow-Jones*, *etc.* are instances of the generic concept *claIndex*, which represents the stock indexes used in different countries.

In order to take into account concepts that occur alone in an utterance, we introduce a special relation type, called "null relation", which is associated to every generic concept. One concept is said

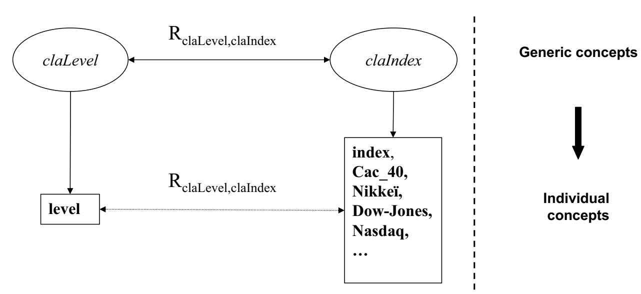

Figure 1: Example of a semantic relation.

to be in "null relation" if it is the only concept hypothesis of an utterance.

2.1.2. Probabilistic model

The different semantic relations of the ontology are used in order to define a probability distribution P^{Rel}, such as, for two individual concepts c_i and c_j, which appear in the same utterance :

- $P^{Rel}(c_i, c_j) = 0$ if the concepts are not related in the ontology, and
- $P^{Rel}(c_i, c_j)$ is a function of the probability of the relation between associated generic concepts otherwise.

As the semantic relations are initially defined for two generic concepts, we introduce a function γ that gives, for each concept c_i, the only associated generic concept. The relation probability $P^{Rel}(c_i, c_j)$ is given by :

$$\begin{aligned} P^{Rel}(c_i, c_j) &= P(c_i, c_j|\gamma(c_i), \gamma(c_j)) \cdot P(\gamma(c_i), \gamma(c_j)) \\ &\simeq P(c_i|\gamma(c_i)) \cdot P(c_j|\gamma(c_j)) \\ &\quad \cdot P(\gamma(c_i), \gamma(c_j)) \end{aligned} \tag{1}$$

The probability $P(c_i|\gamma(c_i))$ represents the probability that an instance of the generic concept $\gamma(c_i)$ is the individual concept c_i. The semantic relation, linking the two generic concepts $\gamma(c_i)$ and $\gamma(c_j)$, is denoted $R_{\gamma(c_i),\gamma(c_j)}$; as the semantic relation does not consider the appearance order in the utterance :

$$P(\gamma(c_i), \gamma(c_j)) = P(\gamma(c_j), \gamma(c_i)) = \frac{1}{2} P(R_{\gamma(c_i),\gamma(c_j)}) \tag{2}$$

The semantic relation probability $P^{Rel}(c_i, c_j)$ becomes :

$$P^{Rel}(c_i, c_j) \approx \frac{1}{2} P(c_i|\gamma(c_i)) P(c_j|\gamma(c_j)) P(R_{\gamma(c_i),\gamma(c_j)}) \tag{3}$$

A relation probability $P^{Rel}(c_i)$ is defined for each concept c_i :

$$P^{Rel}(c_i) = \sum_{c} P^{Rel}(c_i, c) \tag{4}$$

2.1.3. Mutual information

The semantic relation measure is based upon the estimation of the "mutual information quantity", denoted $i^{Rel}(c_i; c_j)$, brought by the co-occurrence of concepts c_i and c_j related in the ontology with the probability distribution $P^{Rel}(c_i, c_j)$:

$$i^{Rel}(c_i; c_j) = \log \frac{P^{Rel}(c_i, c_j)}{P^{Rel}(c_i) \cdot P^{Rel}(c_j)} \tag{5}$$

The mutual information quantity $i^{Rel}(c_i; c_j)$ allows measuring the influence of an occurrence of the concept c_i on the occurrence of the another concept c_j (and *vice versa*) in the same utterance.

2.1.4. Confidence measure for a non-isolated concept hypothesis

Let W be a sentence hypothesis, provided by the semantic language model [1], which contains n concepts ($n > 2$). The semantic relation confidence measure of a concept c_i denoted $CM^{Rel}(c_i)$, is the mean of the set of mutual information quantity with the other concepts of the utterance :

$$CM^{Rel}(c_i) = \frac{1}{n-1} \sum_{j \neq i / Rel(c_i, c_j) = 1}^{n} i^{Rel}(c_i; c_j) \tag{6}$$

The notation $Rel(c_i, c_j) = 1$ means that the concepts c_i and c_j are semantically related. The confidence measure is normalized by the total number of potential relations (i.e. $n - 1$) where as the summation is only on the co-ocurrences that are really in relation. It allows having the measure decreasing when a concept co-occurs with concepts that are not in relation with it. A concept hypothesis is considered to be more reliable if it is semantically related to the other concept hypotheses of the utterance and if these semantic relations are highly probable.

2.1.5. Confidence measure for an isolated concept hypothesis

The introduction of a "null relation" for each generic concept enables to extend the definition of the semantic relation confidence measure to the special case of an isolated concept in an utterance. Some concepts are more frequently observed as single hypotheses that other concepts which explains our choice of estimating a semantic relation confidence measure for such concept hypotheses. The confidence measure $CM^{Rel}(c)$ for an isolated concept c is given by :

$$CM^{Rel}(c) = i^{Rel}(c; c_\epsilon) = \log \frac{P^{Rel}(c, c_\epsilon)}{P^{Rel}(c) \cdot P^{Rel}(c_\epsilon)} \tag{7}$$

c_ϵ represents the "null" concept.

2.2. Acoustic confidence measure

An acoustic confidence measure is also used; initially, the measure is estimated for a word hypothesis. It is equal to the difference between the acoustic likelihood provided by the speech recognition model for a given hypothesis and the one that would be provided by a totally unconstrained phoneme loop model (the difference is normalized by the number of frames of the word). This definition is extended to a concept hypothesis, given by the semantic language model. Let c be a concept hypothesis, given by the semantic language model, constituted with D words, $w_1 ... w_D$. The acoustic confidence measure, $CM^{Acous}(c)$, is estimated by averaging the acoustic confidence measure $CM^{Acous}(w_i)$ of each word w_i, weighted by the associated number of frames $T(w_i)$:

$$CM^{Acous}(c) = \frac{\sum_{i=1}^{D} T(w_i) \cdot CM^{Acous}(w_i)}{\sum_{i=1}^{D} T(w_i)} \tag{8}$$

2.3. Confidence measure calibration

The previously defined confidence measures can be individually used in a first time; a confidence measure CM estimated for a concept hypothesis allows providing the *a posteriori* probability, denoted $P(Cor.|CM)$, probability that the hypothesis is correct given the value of the associated confidence measure. A calibration step is required in order to compute the *a posteriori* probability distribution. By using the logistic regression method, the posterior probability is approximated by :

$$P(Cor.|CM) = \frac{1}{1 + \exp^{-(a_0 + a_1 \cdot CM)}} \tag{9}$$

The parameters a_0 and a_1 are estimated so that the cross entropy for the development set is minimized.

3. Confidence measures combination

As the previously defined confidence measures used different knowledge sources, they can be combined in order to get the best of the two measures and improve the prediction power on the correctness of concept hypotheses.

3.1. Combination by logistic regression

Logistic regression is also a formalism that enables to fuse predictors. The posterior probability $P(Cor.|CM^{Rel}, CM^{Acous})$ that the concept is correct given its associates confidence measures CM^{Rel} and CM^{Acous} is given by :

$$P(Cor|CM^{Rel}, CM^{Acous}) = \frac{1}{1 + e^{-(a_0 + a_1 \cdot CM^{Rel} + a_2 \cdot CM^{Acous})}} \quad (10)$$

The parameters a_0, a_1 and a_2 are estimated so that they minimize the cross entropy for the development set.

3.2. Combination by using a decision tree

The information on the concept identity is also crucial and we wanted to combine this information with the confidence measures previously defined. To this aim, we used a decision tree which can automatically decide, after a training process, on the correctness of a concept hypothesis. The decision tree type used is a Semantic Classification Tree, denoted *SCT*; these decision trees were initially used for natural language processing [6]. The decision tree is trained on a corpus of concept hypotheses; each hypothesis is tagged with "Correct" if it is correct, with "Incorrect" otherwise. A set of parameters or criteria, including the concept identity, describe each sample to classify. The main advantage of a decision tree strategy is that we do not need to have *a priori* knowledge about the relevance of the chosen criteria; it is the decision tree itself that selects the relevant ones. The decision tree is trained in order to minimize the impurity of the distribution of "Correct" and "Incorrect" samples; the training process stops when no further impurity gain can be achieved or when the size of samples attached to a node is below a given threshold. Questions associated to each node to discriminate samples are regular expressions of criteria attached to samples. The score associated to each leaf is the ratio between the number of "Correct" samples and the total number of samples attached to the leaf. The score represents the confidence given by the classifier that a sample in this leaf is correct. Hence, each concept hypothesis is classified in a particular leaf of the decision tree and the probability that the hypothesis is correct is given by the probability associated to the leaf.

4. Experiments

4.1. Experimental context

Experiments have been carried out on a corpus collected at France Telecom R&D for a banking application, which allows users performing transactions and asking questions about the stock market, portfolios and accounts. Different instantiation and relation probabilities have been estimated, by counting and smoothing on a training corpus of 24604 sentences, which contains 34623 concepts. The application ontology contains 67 generic concepts and 141 individual concepts. The semantic language model (described in [1]) is applied, in a second recognition pass, to word graphs and provides the best joint words/concepts sequence hypothesis. The first set denoted Dev is the development set, constituted of 4472 concept hypotheses; the second set denoted $Test$ is the test set, constituted of 4177 concept hypotheses. The development and test set have been split in two. Dev_1 (resp. $Test\_1$) contains isolated concept hypotheses from the corpus Dev (resp. $Test$), while Dev_2+ (resp. $Test\_2+$) contains non-isolated concept hypotheses from the corpus Dev (resp. $Test$). Table 1 gives the number of concept hypotheses (*# Conc.*), and the precision (*Prec.*) of the semantic language model on the (concept/value), used to estimate the initial cross entropy H_{init}.

	# Conc.	Prec. (%)
Dev_1	1516	87.9
Dev_2+	2956	92.1
$Test\_1$	1445	89.3
$Test\_2+$	2732	92.3

Table 1: Statistics of each sub-corpus

4.2. Evaluation criterion

Different methods can be used to evaluate a confidence measure; we have chosen the relative reduction of the cross entropy for a given test set as the criterion to evaluate the confidence measure. The cross entropy, for a test set of N concepts c_i, is defined as :

$$H = -\frac{1}{N}\sum_{i=1}^{N} \delta(c_i)\log p_i + (1-\delta(c_i))\log(1-p_i) \quad (11)$$

$\delta(c_i)$ is an indicator, equal to 1 if the hypothesis c_i is correct, equal to 0 otherwise. p_i is the *a posteriori* probability that the concept c_i is correct. With no confidence measure, p_i is the same for all hypotheses and is equal to the precision on the set. Hence, the initial cross entropy is :

$$H_{init} = -Prec. \log Prec. - (1 - Prec.)\log(1 - Prec.) \quad (12)$$

When a confidence mesure CM is introduced, the *a posteriori* probability becomes the probability $P(Cor.|CM(c_i))$, which is the posterior probability that the hypothesized concept c_i is correct, given its associated confidence measure $CM(c_i)$. The resulting cross entropy H_{CM} is :

$$\begin{aligned} H_{CM} = \quad & -\tfrac{1}{N}(\textstyle\sum_{i=1}^{N} \delta(c_i)\log(P(Cor.|CM(c_i)) + \\ & (1-\delta(c_i))\log(1 - P(Cor.|CM(c_i)))) \end{aligned} \quad (13)$$

The evaluation criterion for a confidence measure is then the relative diminution of cross entropy induced by the introduction of the confidence measure : $\Delta H = \frac{H_{init} - H_{CM}}{H_{init}}$. For a completely random confidence measure CM, $H_{CM} = H_{init}$; for a perfect one, $H_{CM} = 0$. ΔH varies between 0 and 100%. The bigger ΔH is, the more predictive the confidence measure is.

4.3. Results with confidence measures used individually

The semantic relation and acoustic confidence measures are calibrated on the development set Dev. As the semantic relation confidence measure is differently defined for isolated and non-isolated concept hypotheses, a first experiment was conducted to determine wether the development set needs to be split, that is to say if the confidence measure for isolated concepts (resp. non-isolated concepts) needs to be calibrated on a development set of only isolated concepts (resp. non-isolated concepts). The results showed

	Dev_1	Dev_2+	$Test\_1$	$Test\_2+$
H_{init}	0.36965	0.32760	0.36492	0.34104
CM^{Rel}	4.06	4.86	3.70	3.89
CM^{Acous}	13.11	8.97	10.67	8.21

Table 2: Confidence measures evaluation in terms of $\Delta H(\%)$

that the choice of the development set influenced semantic relation confidence measure performances and it was better to split the development set. Table 2 gives the initial cross entropy of each sub-corpus and details the results obtained in terms of relative cross entropy diminution by using the confidence measures individually. The acoustic confidence measure gives better results that the semantic relation one. Some differences appear; the semantic relation confidence measure is as effective for the isolated concepts as for the non-isolated ones, which comforts our choice to introduce the "null" relation in our model. The acoustic confidence measure is less effective on the sub-corpora of non-isolated concepts (Dev_2+ and $Test\_2+$). A detailed analysis of these corpora shows that these corpora contain much more short concepts in terms of number of frames (on average, 4 times more). A large part of the concepts are numbers, for which many ambiguities may subsist (for example, in French, *"dix"*, *"six"*, etc.). The acoustic modeling is more delicate on short words.

4.4. Results with the combination of confidence measures

This section details the results obtained by combining confidence via logistic regression or via a decision tree process. As the decision tree handles only discrete values, the confidence measures, which are numerical values, have to be quantified. We introduced two levels for each confidence measure : H for a high confidence measure and L for a low confidence measure. A confidence measure is set to level H (resp. L) if its value gives a posterior probability $P(Cor.|CM)$ greater (resp. less) than 0.9. We trained 4 different classifiers, which used different describing parameters :

- SCT_I uses the concept identity and the semantic relation confidence measure level,
- SCT_II uses the concept identity and the acoustic confidence measure level,
- SCT_III uses the concept identity and the two confidence measures levels,
- SCT_IV uses the concept identity and the level of a measure result of the combination by logistic regression of the two confidence measures.

Table 3 details the results obtained in terms of $\Delta H(\%)$ by combining the confidence measures via logistic regression ($Log.Reg$) or via the previously defined decision trees. The results obtained show the efficiency of the logistic regression method to combine measures using different knowledge sources. Improvements brought by each confidence measure separately are almost added, which proves the complementarity of the two confidence measures. The semantic relation confidence measure gives better results when it is integrated in a decision tree; the relative cross entropy diminutions are until 4 times bigger. The acoustic one also gives better results when it is integrated in a decision tree but the performance differences between the two measures are less obvious. The integration into a decision tree seems to be more advantageous for the semantic relation confidence measure; the information on the identity of the concept hypothesis is more complementary to this measure than to the acoustic one. We can still notice the performances differences of the acoustic confidence measure on the isolated and non-isolated concepts corpora $Test\_1$ and $Test\_2+$. The combination of the two via the decision tree (SCT_III) seems to be less relevant than the combination via the logistic regression (*cf.* Table 2). Indeed, the resulting confidence mesure gives better results on the isolated concepts corpus $Test\_1$ but is less powerful on the non-isolated concepts corpus $Test\_2+$, on which the acoustic confidence measure is less relevant. The decision tree SCT_IV, which uses the level of the measure resulting from the combination via logistic regression of the two confidence measures, gives the best results in terms of cross entropy reduction for the two sub-corpora $Test\_1$ and $Test\_2+$.

	Dev_1	Dev_2+	Test_1	Test_2+
$Log.Reg.$	17.00	12.25	13.00	11.83
SCT_I	13.10	17.43	9.89	11.83
SCT_II	21.88	20.07	14.62	11.18
SCT_III	20.13	24.01	14.13	13.10
SCT_IV	21.30	25.27	15.18	14.27

Table 3: Results obtained in terms of $\Delta H(\%)$ with the combination of confidence measures

5. Conclusion

This paper proposes a new semantic relation confidence measure estimated for concept hypotheses of a semantic language model. This measure exploits the different semantic relations linking concepts of an ontology together. It gives, for a concept hypothesis, a consistency degree with the other hypotheses of the sentence. This semantic relation confidence measure gives promising results when it is combined with an acoustic one via logistic regression or via an integration into a decision tree.

6. References

[1] Kobus, C. and Damnati, G. and Delphin-Poulat, L., "Conceptual Language Model Design for Spoken Language Understanding", Eurospeech, 2005

[2] Hacioglu, K. and Ward, W., "A Concept Graph Based Confidence Measure", ICASSP, 2002

[3] Sarikaya, R. and Gao, Y. and Picheny, M. and Erdogan, H., "Semantic Confidence Measurement for Spoken Dialog Systems", IEEE Transactions on Speech and Audio Processing, 2005

[4] Raymond, C. and Béchet, F. and Camelin, N. and De Mori, R. and Damnati, G., "Semantic Interpretation with Error Correction", ICASSP, 2005

[5] Brachman, T., "What's in a concept: Structural foundations for semantic networks", International Journal of Man-Machine Studies, 1977

[6] Kuhn, R. and De Mori, R., "The application of semantic classification trees to natural language understanding", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995