Constructing Empirical Models for Automatic Dialog Parameterization

Mikhail Alexandrov[1], Xavier Blanco[1], Natalia Ponomareva[2], and Paolo Rosso[2]

[1] Universidad Autonoma de Barcelona, Spain
[2] Universidad Politecnica de Valencia, Spain

Abstract. Automatic classification of dialogues between clients and a service center needs a preliminary dialogue parameterization. Such a parameterization is usually faced with essential difficulties when we deal with politeness, competence, satisfaction, and other similar characteristics of clients. In the paper, we show how to avoid these difficulties using empirical formulae based on lexical-grammatical properties of a text. Such formulae are trained on given set of examples, which are evaluated manually by an expert(s) and the best formula is selected by the Ivakhnenko Method of Model Self-Organization. We test the suggested methodology on the real set of dialogues from Barcelona railway directory inquiries for estimation of passenger's politeness.

1 Introduction

1.1 Problem Setting

Nowadays, dialogue processing is widely used for constructing automatic dialogue systems and for improving service quality. By the word "dialogue" we mean a conversation between a client and a service center, and by the word "processing" we mean a classification of clients. Politeness, competence, satisfaction, etc. are very important characteristics for client classification but its formal estimation is quite difficult due to the high level of subjectivity. Thus, these characteristics usually are not taken into account or they are estimated manually [1].

In this work, we aim to construct an empirical formula to evaluate the mentioned characteristics, which is based on:

(i) objective lexical-grammatical indicators related to a given characteristic;
(ii) subjective expert opinion about dialogues.

The selection of lexical-grammatical indicators depends on expert experience. However, some simple indicators are often obvious, e.g. polite words for estimation of politeness, "if-then" expressions for the estimation of competence, or objections for estimation of a level of satisfaction. The technical problem is to find an appropriate tool for revealing such indicators and include this linguistic tool into the automatic process of dialogue parameterization.

V. Matousek and P. Mautner (Eds.): TSD 2007, LNAI 4629, pp. 455–463, 2007.

Subjective expert opinion(s) may be obtained by means of manual evaluation of a set of dialogues. For this, a fixed scale is taken and each dialogue is evaluated in the framework of this scale. Usually symmetric normalized scale [-1,1] or positive normalized scale [0,1] is considered.

In order to construct an empirical formula we use the Inductive Method of Model Self-Organization (IMMSO) proposed by Ivakhnenko [7]. This method allows to select the best formula from a given class using the training and the control sets of examples.

For definiteness, in this paper we consider only client's politeness. And it should be emphasized that we have no aim to find the best way for numerical estimation of politeness. Our goal is only to demonstrate how one may transform the lexical-grammatical properties of a text and the subjective expert opinion to these numerical estimations.

The paper is organized as follows. Section 2 describes the linguistic factors that should be taken into account in the formula to be constructed. Section 3 shortly describes the Ivakhnenko method. Section 4 contains the results of experiments. Conclusions and future work are drawn in Section 5.

In order to construct an empirical formula we use the Inductive Method of Model Self-Organization (IMMSO) proposed by Ivakhnenko [7]. This method allows to select the best formula from a given class using the training and the control sets of examples.

For definiteness, in this paper we consider only client's politeness. And it should be emphasized that we have no aim to find the best way for numerical estimation of politeness. Our goal is only to demonstrate how one may transform the lexical-grammatical properties of a text and the subjective expert opinion to these numerical estimations.

The paper is organized as follows. Section 2 describes the linguistic factors that should be taken into account in the formula to be constructed. Section 3 shortly describes the Ivakhnenko method. Section 4 contains the results of experiments. Conclusions and future work are drawn in Section 5.

1.2 Related Works

The existing automatic tools related with the estimation of politeness only detect polite (impolite) expressions in dialogues but do not give any numerical estimation of the level of politeness [2,3]**.** And it can be easy explained: such estimations are too subjective. In the work [11], some formal factors of politeness are proposed and the empirical formula based on these factors is constructed. Nevertheless this formula was not properly justified: it was given in advance and fitted to data.

The Ivakhnenko method (it is better to say 'approach'), mentioned above has many applications in Natural Sciences and Techniques [7]. It has been applied in Computational Linguistics for constricting empirical formulae for testing word similarity [4,9].

2 Models for Parameter Estimation

2.1 Numerical Indicators

The model to be constructed represents a numerical expression, which depends on various indicators of politeness of a given text and determines a certain level of politeness. This level is measured by a value between 0 and 1, where 0 corresponds to a regular politeness, and 1 corresponds to the highest level of politeness. We do not consider any indicators of impoliteness, although in some cases it should be done.

In this paper we take into account the following 3 factors of politeness: the first greeting (*G*), polite words (*W*) and polite grammar forms (*V*). As examples of polite words such well-known expressions as "please", "thank you", "excuse me", etc. can be mentioned. We considered verbs in a subjunctive mood as the only polite grammar forms, e.g. "could you", "I would", etc.

We take into account the following two circumstances:

(i) The level of politeness does not depend on the length of dialogue. It leads to the necessity to normalize a number of polite expressions and polite grammar forms on the length of dialogue. The dialogue's length here is the number of client's phrases.

(ii) The level of politeness depends on the number of polite words and polite grammar forms non-linearly: the greater number of polite words and grammar forms occur in a text the less contribution new polite words and grammar forms give. It leads to the necessity to use any suppressed functions as the logarithm or the square root, etc.

Therefore, we consider the following numerical *indicators* of politeness:

$$\boldsymbol{G} = \{0,1\}, \quad \boldsymbol{W} = Ln(1 + N_w/L), \quad \boldsymbol{V} = Ln(1 + N_V/L), \tag{1}$$

where N_w, N_v are a number of polite words and polite grammar forms respectively and L is a number of phrases.

It is evident that: a) $W = V = 0$, if polite words and polite grammar forms do not occur; b) $W = V = Ln(2)$, if polite words and polite grammar forms occur in every phrase. All these relations are natural and easy to understand.

2.2 Example

Here we demonstrate how the mentioned indicators are manifested and evaluated. Table 1 shows the example of dialogue (the records are translated from Spanish into English). Here *US* stands for a user and *DI* for a directory inquire service. This example concerns the train departure from Barcelona to other destinations both near the Barcelona and in other provinces of Spain.

Table 2 shows the results of parameterization of this dialogue and its manual estimation by a user.

Table 1. Example of a real dialogue between passengers and directory inquires

US: ***Good evening***. ***Could you*** tell me the schedule of trains to Zaragoza for tomorrow? *DI*: For tomorrow morning? *US*: Yes *DI*: There is one train at 7-30 and another at 8-30 *US*: And later? *DI*: At 10-30 *US*: And till the noon? *DI*: At 12 *US*: ***Could you*** tell me the schedule till 4 p.m. more or less? *DI*: At 1-00 and at 3-30 *US*: 1-00 and 3-30 *DI*: hmm, hmm <SIMULTANEOUSLY> *US*: And the next one? *DI*: I will see, one moment. The next train leaves at 5-30	*US*: 5-30 *DI*: hmm, hmm < SIMULTANEOUSLY > *US*: Well, and how much time does it take to arrive? *DI*: 3 hours and a half *US*: For all of them? *DI*: Yes *US*: Well, ***could you*** tell me the price? *DI*: 3800 pesetas for a seat in the second class *US*: Well, and what about a return ticket? *DI*: The return ticket has a 20% of discount *US*: Well, so, it is a little bit more than 6 thousands, no? *DI*: Yes *US*: Well, ***thank you very much*** *DI*: Don't mention it, good bye

Table 2. Parameterized dialogue

First greeting ***g***	Number of polite words ***Nw***	Number of polite grammar forms ***Nv***	Indicator ***G***	Indicator ***W***	Indicator ***V***	Estimation
Yes	1	3	1	0.07	0.21	0.75

In our work all factors *G, w, v* are detected by means of the *NooJ* resource [10]. Early for the same goal we used morphological analyzers described in [6]. The *NooJ* is a linguistic tool to locate morphological, lexical and syntactic patterns used for raw texts processing. The results of *NooJ* work were fixed in a file for further processing by Ivakhnenko method.

2.3 Numerical Models

Having in view the three factors described above the following series of polynomial models can be suggested for automatic evaluation of the level of politeness:

$$
\begin{aligned}
&\text{Model 1: } F(\boldsymbol{G},\boldsymbol{W},\boldsymbol{V}) = A_0\\
&\text{Model 2: } F(\boldsymbol{G},\boldsymbol{W},\boldsymbol{V}) = C_0\boldsymbol{G}\\
&\text{Model 3: } F(\boldsymbol{G},\boldsymbol{W},\boldsymbol{V}) = A_0 + C_0\boldsymbol{G}\\
&\text{Model 4: } F(\boldsymbol{G},\boldsymbol{W},\boldsymbol{V}) = A_0 + C_0\boldsymbol{G} + B_{10}\boldsymbol{W} + B_{01}\boldsymbol{V}\\
&\text{Model 5: } F(\boldsymbol{G},\boldsymbol{W},\boldsymbol{V}) = A_0 + C_0\boldsymbol{G} + B_{10}\boldsymbol{W} + B_{01}\boldsymbol{V} + B_{11}\boldsymbol{VW} \qquad (2)\\
&\text{Model 6: } F(\boldsymbol{G},\boldsymbol{W},\boldsymbol{V}) = A_0 + C_0\boldsymbol{G} + B_{10}\boldsymbol{W}^2 + B_{01}\boldsymbol{V}^2\\
&\text{Model 7: } F(\boldsymbol{G},\boldsymbol{W},\boldsymbol{V}) = A_0 + C_0\boldsymbol{G} + B_{11}\boldsymbol{VW} + B_{20}\boldsymbol{W}^2 + B_{02}\boldsymbol{V}^2\\
&\text{Model 8: } F(\boldsymbol{G},\boldsymbol{W},\boldsymbol{V}) = A_0 + C_0\boldsymbol{G} + B_{10}\boldsymbol{W} + B_{01}\boldsymbol{V} + B_{11}\boldsymbol{VW} + B_{20}\boldsymbol{W}^2 + B_{02}\boldsymbol{V}^2
\end{aligned}
$$

etc.

Here: $A_{0,}$ $C_{0,}$ B_{ij} are undefined coefficients. It is easy to see that all these models are the polynomials with respect to the factors *W* and *V*. Such a representation is enough general for various function $\psi(W,V)$ and this a reason of its application. Of course, one can suggest the other type of models.

3 The Ivakhnenko Method

3.1 The Contents of the Method

Inductive method of model self-organization (IMMSO) was suggested and developed by Ivakhnenko and his colleagues at 80s. This method allows to determine the model of optimal complexity, which well describe a given experimental data. Speaking 'model' we mean a formula, equation, algorithm, etc.

This method does not require any a priori information concerning distribution of parameters of objects under consideration. Just for this reason the Ivakhnenko method proves to be very effective in the problems of Data and Text Mining. Nevertheless it should be said that if such a priori information exists then the methods of Pattern Recognition will give better results.

This method has one restriction: it cannot find the optimal model in any continuous class of models because its work is based on the competition of the models. So this method is titled as an *inductive* one. The main principle of model selection is the principle of stability: the models describing different subsets of a given data set must be similar.

Here are the steps of the Ivakhnenko method

(1) An expert defines a sequence of models, from the simplest to more complex ones.
(2) Experimental data are divided into two data sets: training data and control data, either manually or using an automatic procedure.
(3) For a given kind of model, the best parameters are determined using, first, the training data and, then, the control one. For that any internal criteria of concordance between the model and the data may be used (e.g., the least squares criterion).
(4) Both models are compared on the basis of any external criteria, such as the criterion of regularity, or criterion of unbiasdness, etc. If this external criterion achieves a stable optimum, the process is finished; otherwise, more complex model is considered and the process is repeated from the step (3).

Why we expect the external criterion to reach any optimum? The fact is the experimental data are supposed to contain: (a) a regular component defined by the model structure and (b) a random component-noise. A simplified model does not react to noise, but simultaneously it does not reflect the nature of objects. Otherwise, a sophisticated model can reflect very well real object behavior but simultaneously such a model will reflect a noise. In both cases the values of the penalty function (external criterion) are large. The principle of model *self-organization* consists in that an external criterion passes its minimum when the complexity of the model is gradually increased.

3.2 Application of Method

There are two variants of the Ivakhnenko method:

I Combinatorial Method
II Method of Groupped Arguments

In the first case all variants of model are considered step-by-step. And in the second one the models are filtered [8]. In this work we use only the first method and consequently consider all 8 models presented in the section 2.3.

Parameters of the concrete model are determined by means of the least square method. For this we fix one of the models (2) and construct the system of lineal equations for a given set of dialogs:

$$F(g_i,\ w_i,\ v_i) = P_i\ , \quad i=1,..,N \tag{3}$$

Here: g,w,v are the factors, P_i are the manual estimations of dialog, N is the number of dialogs. For example, the dialog described above forms the following equation for the 4[th] model: $A_0 + C_0 + 0.07B_{10} + 0.021\ B_{01} = 0.75$.

The system (3) is a system of lineal equations with respect to undefined coefficients. It can be solved by the least square method. It should take into account that the number of equations must be several times more then the number of parameters to be determined. It allows to filter a noise in the data. Speaking 'noise' we mean first of all fuzzy estimations of politeness.

According to IMMSO methodology for the series of models starting with the first model from (2) any external criterion is calculated and checked whether it achieved an optimal point. Depending on the problem different forms of this criterion can be proposed [8]. In our case we use the criterion of regularity. It consists in the following:

– model parameters (coefficients A_0, C_0, etc.) are determined on the training data set
– this model is applied to control data set and 'model' politeness is calculated
– the relative difference between the model politeness and the manual politeness of an expert is estimated

All these actions can be reflected by the following formula

$$K_r = \frac{\sqrt{\sum_N \left(P_i\ (T) - P_i\ \right)^2}}{\sqrt{\sum_N \left(P_i\ \right)^2}} \tag{4}$$

where $P_i(T)$ are the 'model' estimations of politeness on the control data set, that is the left part of the equations (3), P_i are the manual estimations of dialogs from the control data set, N is the number of dialogs in control data set. It should emphasize that the model parameters are determined on the training data set.

4 Experiments

The data we used in our experiments represent a corpus of 100 person-to-person dialogues of Spanish railway information service. The short characteristics of the corpus (length of talking, volume of lexis) are described in [5]**.** From the mentioned corpus of dialogues we took randomly N = 15 dialogues for training data set and N=15 dialogues for control data set. The level of politeness was estimated manually in the framework of scale [0, 1] with the step 0.25. Table 3 represents the part of data used for the experiments.

Table 3. Example of data used in the experiments

G	W	V	W^2	WV	V^2	**Manual estimation**
1	0.134	0.194	0.0178	0.0259	0.0377	1
0	0.111	0.057	0.0124	0.0064	0.0033	0.75
1	0.000	0.074	0.0000	0.0000	0.0055	0.25
1	0.000	0.031	0.0000	0.0000	0.0009	0
1	0.000	0.118	0.0000	0.0000	0.0139	0.75
1	0.043	0.043	0.0018	0.0018	0.0018	0.5
1	0.000	0.000	0.0000	0.0000	0.0000	0.25
1	0.043	0.083	0.0018	0.0035	0.0070	0.5
0	0.000	0.074	0.0000	0.0000	0.0055	0
1	0.134	0.069	0.0178	0.0092	0.0048	1

We tested all 8 models (2) and calculated the criterion of regularity (4). The results are presented in the Table 4.

Table 4. Criterion of regularity

Model-	Model-	Model-	Model-	Model-	Model-	Model-	Model-
0.505	0.567	0.507	***0.253***	0.272	0.881	1.875	0.881

It is easy to see, that the lineal model is a winner. The fact that the model reflects only trend could be explained by imperfectness of a given class of models and/or a high level of noise. Joining together all 30 examples we determined the final formula as:

$$F(g, w, v) = -0.14 + 0.28\boldsymbol{G} + 3.59\boldsymbol{W} + 3.67\boldsymbol{V} \quad (5)$$

This formula provides 24% of relative mean square root error.

In order to evaluate the sensibility of results to the volume of data the same calculations were completed on the basis 10 dialogs taken for training and 10 data taken for control. We considered only first 4 models: more complex models needs more data. The results presented in the Table 5 show that the dependence on the volume is insignificant with respect to the behavior of external criterion.

Table 5. Criterion of regularity for shorten data set

Model-	Model-	Model-	Model-
0.497	0.503	0.502	0.319

5 Conclusions and Future Work

In this paper, we suggested the simple methodology for automatic estimation of various 'fuzzy' dialogue characteristics, which have a large level of subjectivity. We applied this methodology for the estimation of politeness. The constructed formula correctly reflects the contribution of selected factors of politeness: all factors have positive coefficients. The obtained error is comparative with the step of the manual dialogue estimation.

In the future, we intend to consider more complex empirical models for estimation of politeness, culture and competence, satisfaction.

Acknowledgements

The authors thank Dr. Alexander Gelbukh from the National Polytechnic Institute of Mexico for his valuable notes and help in this investigation.

This work has been partially supported by MCyT TIN2006-15265-C06-04 research project.

References

1. Alexandrov, M., Sanchis, E., Rosso, P.: Cluster analysis of railway directory inquire dialogs. In: Matoušek, V., Mautner, P., Pavelka, T. (eds.) TSD 2005. LNCS (LNAI), vol. 3658, pp. 385–392. Springer, Heidelberg (2005)
2. Alexandris, C., Fotinea, S.E.: Discourse particles: Indicators of positive and non-positive politeness in the discourse structure of dialog systems for modern greek. Intern. J. for Language Data Processing Sprache Datenverarbeitung, 1-2, 19–29 (2004)
3. Ardissono, L., Boella, C., Lesmo, L.: Indirect speech acts and politeness: A computational approach. In: Proceedings of the 17th Cognitive Science Conference, pp. 113–117 (1995)
4. Blanco, X., Alexandrov, M., Gelbukh, A.: Modified Makagonov's Method for Testing Word Similarity and its Application to Constructing Word Frequency List. In: Advances in Natural Language Processing, Mexico, pp. 27–36 (2006)
5. Bonafonte, A.: Desarrollo de un sistema de dialogo oral en dominios restringidos (in Spanish). In: I Jornadas en Tecnologia de Habla (2000)
6. Gelbukh, A., Sidorov, G.: Approach to construction of automatic morphological analysis systems for inflective languages with little effort. In: Gelbukh, A. (ed.) CICLing 2003. LNCS, vol. 2588, pp. 215–220. Springer, Heidelberg (2003)

7. Ivakhnenko, A.: Inductive method of model self-organization of complex systems (in Russian), Tehnika Publ. (1982)
8. Ivakhnenko, A.: Manual on typical algorithms of modeling (in Russian). Tehnika Publ. (1980)
9. Makagonov, P., Alexandrov, M.: Constructing empirical formulas for testing word similarity by the inductive method of model self-organization. In: Advances in Natural Language Processing. Mexico, pp. 239–247 (2004)
10. NooJ description: http://www.nooj4nlp.net
11. Ponomareva, N., Blanco, X.: Example-based empirical formula for politeness estimation in dialog processing. In: Proceedings of 10th Intern. Workshop on NooJ applications (to be published, 2007)