Clustering dialogue knowledge with self -organizing maps

Mauri Kaipainen, Kristiina Jokinen, Timo Koskenniemi, Antti Kerminen, Kari Kanto

Media Lab, University of Art and Design Helsinki

kristiina.jokinen@uiah.fi

1. Introduction .

One of the biggest prob lems in building large dialogue systems is the need for defining in advance the number and type of categories, necessary for the operation of the system. Apart from requiring an amount of labor, this approach requires restrictive a priori judgments. Furthermore, this leads to data-specificity of the systems and difficulties to adapt to novel data. In this paper we report progress of an ongoing research of how categorisation could be automated by using elementary data description and data -driven self-organising maps as a tool. The experiments deal with deriving different types of dialogue acts using a small corpus of information-seeking dialogues. We compare clusters of dialogue acts obtained by the SOM to the manually tagged dialogue corpus.

2. Dialogue information

The Interact project aims at developing a generic interaction model that would enable users of online services to interact with various applications in a flexible and natural way. In order to demons trate this model and to explore the applicability of various methods in dialogue processing we are building a dialogue system that deals with public transportation timetable inquiries.

Based on the view that communication is action in context, we describe dialogues with the help of two well-known concepts: dialogue acts and topics. The former describes the act that the speaker performs by particular utterance, and can be regarded as representing application-independent dialogue information, while the latter describe the semantic content of the utterance and provide information related to the application domain itself. Dialogue acts and topics seem to provide a useful first approximation of the utterance meaning by abstracting over possible linguistic realisations. Consequently, the system's internal states can be reduced to a combination of dialogue acts and topics, both of which form an independent source of information for the system to decide on the next move. In this pape r we focus only on dialogue acts , although topics form an analogous problem related to the underlying task and application domain.

The number and type of dialogue acts is not fixed but depend s on the theoretical premises as well as on the type of the dialogues being studied (e.g. information-seeking, negotiation, argum entation). Although there are on -going activities for standardizing the set of dialogue acts and their definitions (e.g. the DRI -initiative, Carletta et al. 1997), the practical dialogue systems exploit their own classifications especially designed for the application in hand. To facilitate the classification of the data into necessary and meaningful categories for the purposes of practical systems , we explore possibilities to learn intrinsic categories of the dialogue data automatically. In particular, we study categorization of utterances into dialogue act classes on the basis of morph o-syntactic features of the words that occur in the utterances, and use the self -organising map algorithm to produce data -driven clusterings. Using minimally analysed information as an input and a specified method for distance comparison, the SOM forms clusters of the input in a manner which can be interpreted as a classification of dialogue acts.

3. Self-organising maps

The self -organizing map (SOM) , originally introduced by Kohonen (1982), is an unsupervised artificial neural network model. The input data for the model is described in terms of vectors, each of which consists of components representing an elementary featu re of the data item, expressed as a numeric value. The output is a similarity -based map of the data items, similarity being defined as proximity of items in the feature space. The clusters of near-similar items are fuzzy sets (Zadeh 1965) by nature. As a hint of the relation of such clusterings to the physically implemented mind, they ar e reminiscent of electro -magnetic response fields observed on cortex (e.g. Wall 1988). In addition, such maps can even represent hierarchical structures in terms of mutually embedded fields , as was shown by the study of Ritter and Kohonen (1989), in which meaningful semantic relations were found from linguistic data.

The self-organizing maps differ from the supervised learning methods in that the SOM, needs no external teacher in the learning phase. We can thus avoid restrictive pre - categorizations of data that may hinder seeing meaningful relations among the data items that do not necessarily follow chosen guidelines. Unlike alternative statistical methods, such as multidimensional s caling and principal component analysis, the SOM approach also offers a platform for continuous upgrading of the data, a requirement native to online services, and important to learning dialogue systems (Jokinen 2000).

4. Description of data

Our dialo gue corpus consists of spoken dialogues recorded at the Helsinki regional transport (HKL) service centre and contains 22 dialogues between customers and the service agent. The corpus was transcribed, manually segmented and tagged with dialogue acts. Transcription takes into account simultaneous speaking so that different types of feedback could be distinguished.

The corpus contains 352 turns and 492 utterances. Each utterance is assigned one dialogue act. The list of dialogue acts with their frequences is given in Table 1.

acknowledgement	*141*
question	*82*
check	*55*
repetition	*42*
statement	*37*
thanking	*26*
call_to_continue	*26*
ending	*23*
opening	*23*
answer	*21*
confirmation	*12*
addition	*2*
wait	*2*

Table 1. Distribution of dialogue acts.

Two different types of feedback are distinguished. *Acknowledgement* is the most frequent act and represents feedback given by the speakers that they have understood and accepted what the partner said. Usually it inc ludes turntaking, i.e. the speaker continues with a statement or a question of her own. The act *call_to_continue* is similar but refers to back - channelling whereby the speaker gives simultaneous acknowledgment of her understanding and encourages the partner to continue without actually taking the turn. The act *addition* is a separate act representing the speaker's completion of the partner's utterance after a short pause.

5. Description of features

In our experi ment, we wanted to test how well some simple features of the utterance distinguish different dialogue acts . We used the following features: the speaker, *wh*-words, the question morpheme -*kO*, the part -of-speech category of the stem of the *kO*-particle, whether the utterance contains a verb, conditional, negation, the particle *entä*, words used in greetings, words used in thankings, words used in acknowledgements and the scaled number of word forms shared with the previous utterance.

P1	speaker
P2	*Wh*-word
P3	-*kO*
P4	verb & -*kO*
P5	conditional
P6	negation
P7	*entä*
P8	greeting
P9	includes a verb
P10	thanking
p11	feedback
p12	Scaled number of shared words betw. current and previous utterance

Table 2. The set of elementary features used to describe utterances.

Preprocessing of the corpus for the SOM dealt with normalizing spoken language and slang expressions and converting inflected word forms to the base form by the Two-level morphological analyser (Koskenniemi 198 3). Ambiguous word forms were solved heuristically by favoring the shortest and simplest analysis. The normalized and morphologically analysed utterances were used as sample instances for the SOM. Each utterance was converted into a vector consisting of the twelve morpho-syntactic features as the components.

6. Results

We computed a map of 12 x 8 units [1] and the resulting map is given in Figure 1 . The light areas in the map depict data points that are close to each other, the dark areas represent distortions of the plane where data points are far apart, i.e. cluster boundaries. The dialogue acts that fall into a particular cluster are listed with the frequency at each point in the map. The labels refer to the dialogue act which has been assigned to the corresponding utterance in the manually tagged corpus.

[1] Using the standard algorithm as defined in SOM-PAK (http://www.cis.hut.fi/research/som_lvq_pak.shtml), 12 dimensions, bubble neighborhood

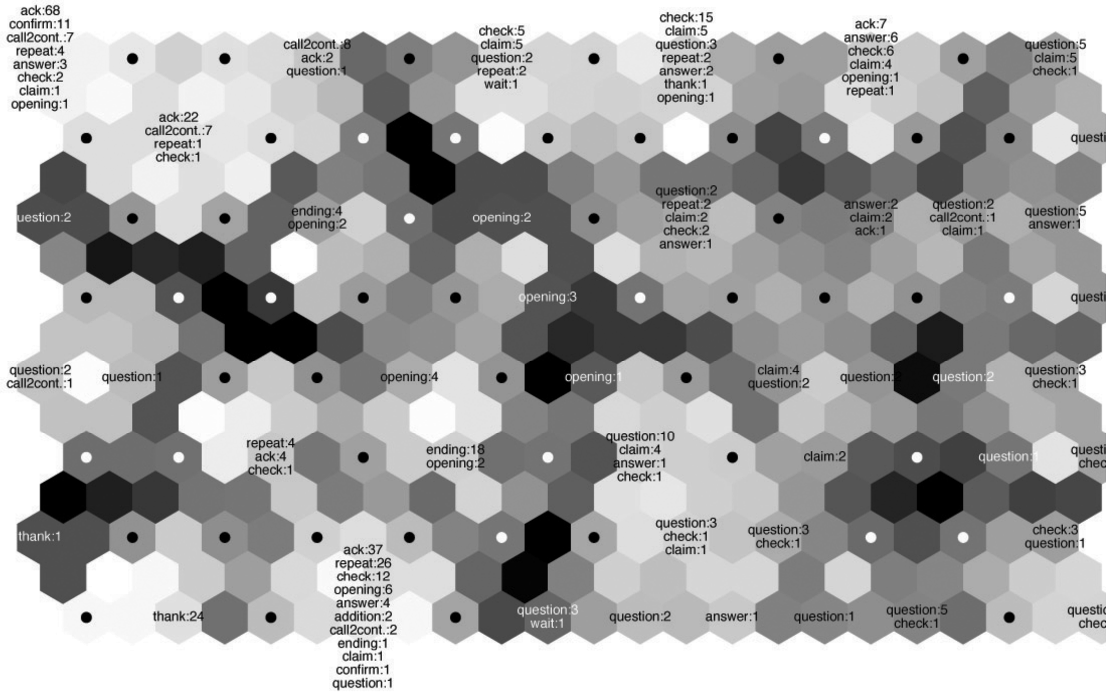

Fig. 1. The distribution of the dialogue acts with respect to each of the 12 features is shown in Figure 2. Labels on the map correspond to names of vectors, given manually in the

As can be seen from Figure 1, the map seems to distinguish rather clear clusters areas of the map for such metacommunication acts as opening and ending (in the middle of the map), and thanking (in the lower left corner of the map), but there is overlap among questions, answers, and claims. Acknowledgements also seem to cluster neatly, although the method found two different clusters, one together with call_to_continue (upper left corner of the map) and the other with repeats and checks (lower leftish of the map). The two separate clusters for acknowledgements suggest that the category under that name is not uniform in the feature space. There is also a cluster of checks (upper middle part of the map), although they are also dispersed around the map, and especially with acknowledgements.

In order to see the distinguishing power of each feature, we also studied their distribution among the data points, as is given in Figure 2. Each of the little maps shows the clusters where the influence of the feature is at its most. (The map in Figure 1 is a map that is achieved by combining the effect of the individual features, Figure 2 pulls the features apart.)

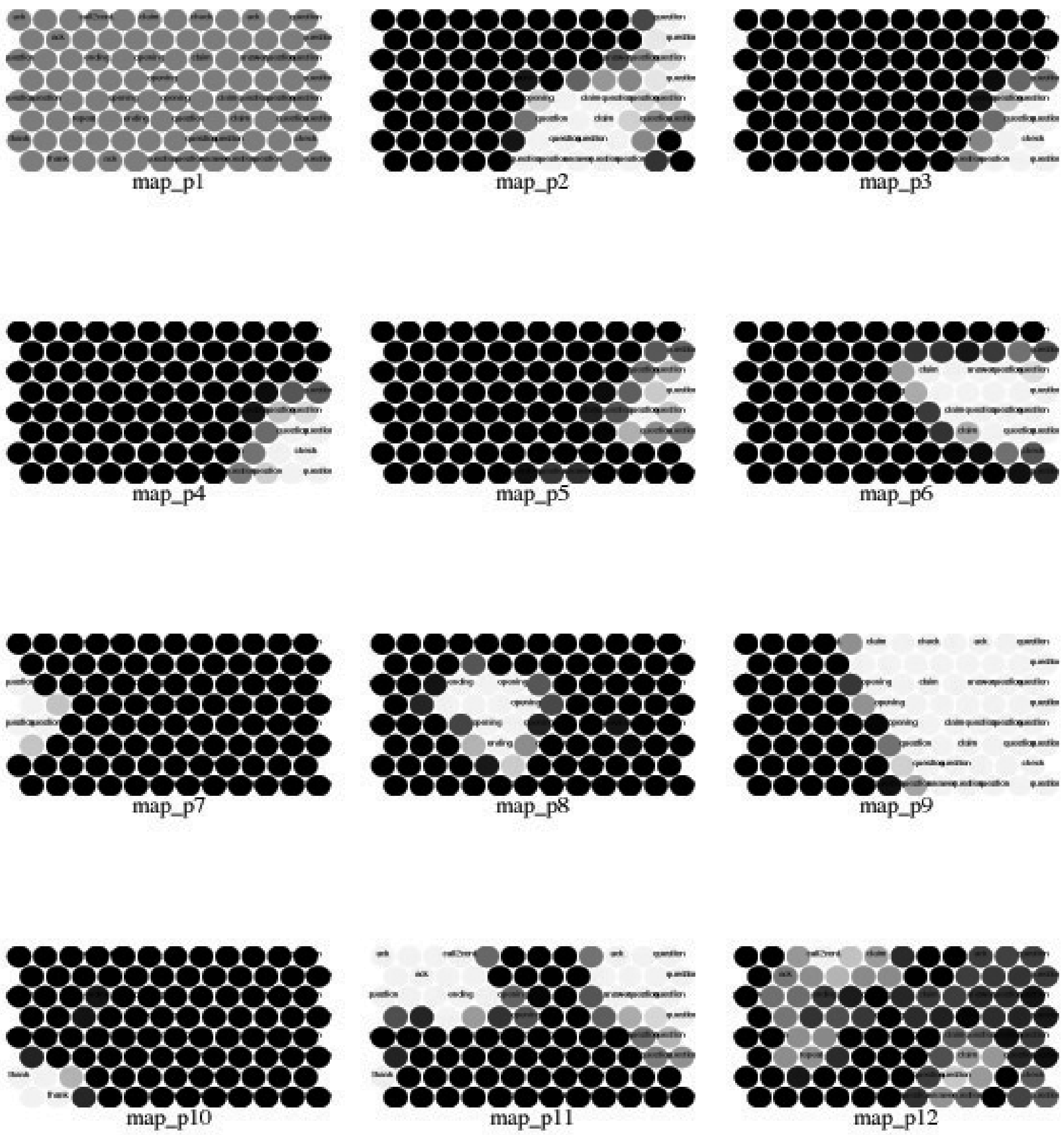

Fig. 2. The response fields corresponding to individual features, numbered as map_p1, map_p2 etc., respective to features listed in Table 1.

As was expected, the main distinction between the various clusters seem to be due to the help of the features 'question word' (map_p2) and 'verb' (map_p9), i.e. checking whether the utterance contains a wh-word and whether it has a verb. The metacommunication acts like openings, endings, thanks, call_to_continue and acknowledgements seem to cluster mainly with the feature ‘no-verb’, and form clear clusters. The characteristic features of the two acknowledgement classes also seem to show a clear tendency for the division: the first cluster deals with the utterance having no verb (map_p9) but containing feedback words (map_p11), whereas the other cluster has no verb and no special feedback words either. The latter cluster also contains several repeats and checks, and the feature distribution seems to explain why this is so: these utterances are the ones that also lack of feedback words. On the other hand, the presence of feedback words seem to be the reason why the first acknowledgement cluster includes the act call_to_continue: both are expressed by the same kind of words. These overlapping clusters point to the fact that with the current feature set we cannot distinguish well enough between the different feedback-types, and obviously more features, possibly derived from the utterance prosody should be taken into account.

An interesting result is that can be seen from the map is that most check acts contain a verb. The checks without a verb get clustered together with acknowledgements and repeats.

7. Discussion and future work

The results show that with the simple feature set used, it is easy to distinguish metacommunication acts but there were no clear clustering separating questions, answers, and claims. This is an area where we expect a longer sequential context to add distinctiveness. For example, questions and answers form adjacency pairs and one might expect that the previous dialogue act strongly suggests the next act type Since there is no general solution to this within the self-organizing paradigm, we are currently developing a recurrent variant of the SOM, based on the ROSOM (Kaipainen & Karhu 2000) which includes the activity history of the whole network in the past in the current input.

The study reported here was carried through with a preliminary data that was not large enough to support any final conclusions but which served as a valuable guide to refine the elementary description as well as the methodology. Significant improvement in the distinctive power is to be expected with larger data. Clearer clusters can also be achieved by refining the elementary feature description.

Acknowledgements

The research is carried out in the context of the USIX-Interact project, a collaboration project between four Finnish universities, and funded by the National Technology Agency (TEKES), Sonera Oyj, ICL Invia Oyj, Lingsoft Oy, Gurusoft Oy, Tampere Technology Center Oy, the Arla Institute, and the Finnish Association of the Deaf.

Bibliography

Carletta, J., N. Dahlbäck, N. Reithinger, and M.A. Walker (eds.) 1997. *Standards for Dialogue Coding in Natural Language Processing*. Dagstuhl-Seminar-Report 167.

Jokinen, K. (2000). Learning Dialogue Systems. Proceedings of the LREC workshop 'From Spoken Dialogue to Full Natural Interactive Dialogue. P.13-17.

Kaipainen, M.; Karhu, P. (2000). *Bringing Knowing-When and Knowing-What Together. Periodically Tuned Categorization and Category-Based Timing Modeled with the Recurrent Oscillatory Self-Organizing Map (ROSOM)*, Minds and Machines 10: 203-229, 2000.

Kohonen, T. (1982). *Self-organized formation of topologically correct feature maps*, Biological Cybernetics 43:59-69.

Koskenniemi, K. (1983). *Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production*. Publications of the Department of General Linguistics, University of Helsinki.

Ritter, H.; Kohonen, T. (1989). *Self-Organizing Semantic Maps*, Berlin: Springer. Biological Cybernetics 61, 241-254.

Wall, J. T. (1988). *Variable organization in cortical maps of the skin as an indication of the lifelong adaptive capabilities of circuits in the mammalian brain*, TINS, Vol. 11, N0. 12.

Zadeh, L. A. (1965). *Fuzzy sets*, Information and Control, 8, 338-353.