Cluster-Based Graphs for Conceiving Dialog Systems

Jean-Leon Bouraoui[1] and Vincent Lemaire[1]

Orange Labs, 2, avenue Pierre Marzin 22300 Lannion, France
jeanleon.bouraoui@orange.com, vincent.lemaire@orange.com
http://vincentlemaire-labs.fr

Abstract. We describe an unsupervised modeling of the structure of task-oriented dialogs. The dialogs are related to a given domain (for instance, train booking). This modeling is used as a basis for conceiving a dialog system architecture. The modeling is represented by a graph. It displays the main stages of the dialogs and the transitions between them. Our approach consists in applying coclustering to a representative dialog corpus. Thus we obtain main topics that appear in the corpus. We then compute the topics transitions within each dialog. The resulting graph describes the main topics in the corpus, and their overall sequential organization.

Keywords: Co-clustering, Natural Language Processing, Dialog Systems, Graphs

1 Introduction

In artificial intelligence field, dialog systems are gaining popularity with the general public; especially as they benefit from advances in understanding of the contents and contexts. Mobile applications such as Siri (Apple), Google Now (Google), Cortana (Microsoft) or Alexa (Amazon) are the most popular. To quantify this growing interest in the technology of dialog interfaces, and dialog systems in particular, let us cite the recent study by the analyst firm Gartner [12]. It places dialog systems among the 10 strategic technologies for 2017.

One of the current trends is to propose software devices to assist the design of dialog systems. These devices are customizable according to the conceiver needs, and the field of application (for example, reservation of trips, ordering of products or services, etc.). One of the matters of these devices is that they can hardly be set up quickly. Indeed, there is currently no generic system, and the adaptation of a dialog system to a given application field takes time.

In this context, we present a methodology to set up a semi-automatic assistance solution for the creation or adaptation of a dialog system for any application field. We describe the details of the process that we set up, and our position in regard to the related works.

In: P. Cellier, T. Charnois, A. Hotho, S. Matwin, M.-F. Moens, Y. Toussaint (Eds.): Proceedings of DMNLP, Workshop at ECML/PKDD, Skopje, Macedonia, 2017.

2 Description of the Problem and Related Works

In this article, we will denominate a dialog as an exchange of information between two interlocutors (knowing that a dialog can involve more than two interlocutors). An interlocutor can be a human or a machine (in a broad sense: an artificial system, software or hardware). We are interested in the finalized dialogs, which aim to achieve a goal: the interlocutors will collaborate to achieve this goal.

We call "dialog corpus" a set of n dialogs relating to a particular domain. Such a set can be composed, for instance, of transcripts of train reservation dialogs, or of interaction chats between a phone provider advisor and a client. Each dialog is composed of t speech turns; a speech turn corresponds to what is said by one of the interlocutors without any interruption (usually one or more sentences).

In a first step, we try to associate each speech turn with a given "class" denoted L_c. A class corresponds to the intention which the interlocutor expresses with his speech turn; let us take as an example a dialog between a customer and the customer service of a phone provider: the speech turn where the agent asks the customer to identify himself belongs to a specific class (named for example *AskIdent*); the one where the customer identifies himself is relative to another class (named for example *AnswerIdentCustomer*).

Next, the topics T_t, which group a set of classes relative to a common subject, are determined. Let us use again our example of the customer service: the topics can be the identification of the client (*IdentCustomer*, the corresponding classes being the request by the adviser, and the response of the client), the discussion of the problem (*ExpProb*, the corresponding classes being the presentation by the customer, and the request for precision by the advisor), etc.

The association of the topics to the speech turns of the different dialogs can be represented as in Fig. 1.

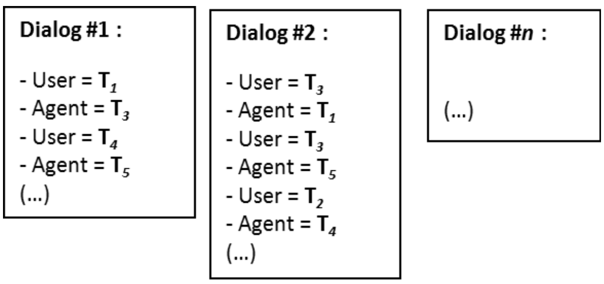

Fig. 1. Association of topics and speech turns

The speech turns are grouped in classes, classes in topics, and topics in dialog as represented, in a simplified way, in Fig. 2.

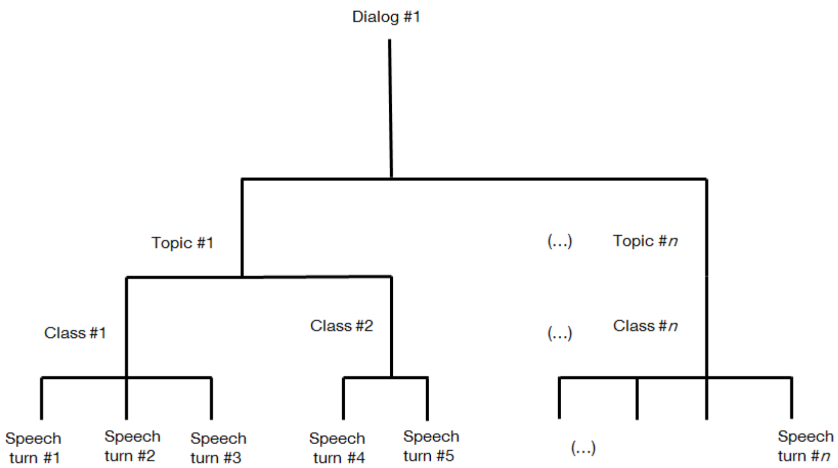

Fig. 2. Hierarchical structure of dialogs, from speech turns to topics

Our first goal is to automatically determine: (i) classes; (ii) topics; (iii) the transitions between topics within each dialog. Thus we could obtain a representation of the typical patterns of the dialogs from the corpus. The desired representation is a directed graph showing the main transitions between topics among all dialogs. Such a (simplified) representation is illustrated in Fig. 3.

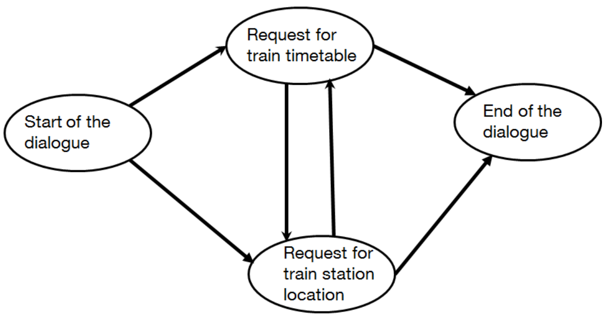

Fig. 3. Prototypical dialog architecture as a graph

This representation presents multiple interests. The main one is the initialization of the design of the dialog system. It can be used as a basis for modeling the architecture of a specialized dialog agent on the target domain. Its execution is thus facilitated and accelerated. Usually, this task is mostly done manually. Either from an *a priori* representation of the designer of the possible dialogs related to a task and a given domain. Or *a posteriori*, from the consultation of existing corpus; in both cases, the process is costly in time.

Moreover, the graph, as well as the steps taken to obtain it, corresponds to the most relevant information about the dialogs. They will enable the designer, without prior knowledge of the field of application, to have a first understanding of the thematic content of the dialogs, their structuring, for the realization of the dialog system.

Related Works It is possible to group in three categories the different approaches used in the literature to tackle this problem. The grouping of the works is based on whether they identify the topics and their sequentiality, or they use Deep Learning, or they adopt an *ad hoc* approach.

- Identification of the topics and sequences of topics: the works belonging to this category differ mainly according to the method used to identify the topics of the dialogs. For example, [3] and [9] use clusters for this purpose, whereas [21] and [26] use the Latent Dirichlet Allocation (LDA) to identify the topics in a document or set of textual documents. In all these studies, the authors use the Hidden Markov Models (HMM) to model the transitions between topics;
- Deep Learning: to our knowledge, the first article that proposed the use of techniques belonging to Deep Learning for modeling dialog systems is [25]. The authors use the *seq2seq* model to model in a recurrent neural network the sequence of speech turns between interlocutors. Two fields of application are described, one of which pertaining to finalized dialogs (computer troubleshooting);
- Other approaches: we group in this category the works that use *ad hoc* methods instead of, or in addition to, data modeling. Thus, [17] apply software heuristics to constitute a model of dialogs from an application database. Heuristics are also used in [8] and [20] to move from a cluster representation to a dialog model. We also cite [19] which integrates Information Retrieval techniques into the architecture of a dialog system.

3 The Methodology Employed

To summarize, our methodology consists, firstly to identify the main topics that appear in the set of dialogs. In a second time we model the relations between these topics among the different dialogs. Thus, we obtain a graph which represent the prototypical dialog continuity in the application field to which the dialog corpus belongs. This section describes the details of this methodology. But first, we address the matter of the relevance of modeling a dialog as a graph.

3.1 Preamble: Hypothesis of a Correspondence between Dialog and Graph

Our approach relates to the first category of the state of the art mentioned above: the identification of the topics and sequences of topics. In our perspective, each speech turn from the dialogs corresponds to a text document, being composed of words. They can be represented by a bipartite graph; in this type of graphs, one set of nodes corresponds to the source of the edges, and the other set to the target of the edges ([16], p. 37-41). In the context of dialog, one set of the bipartite graph would correspond to the set of speech turns, and the other to the words set. This bipartite graph can reciprocally be represented as a texts/words adjacency matrix, where the lines correspond to the speech turn set from the dialog corpus, and the columns to the word set from the same corpus. Figure 4 illustrates this twofold perspective.

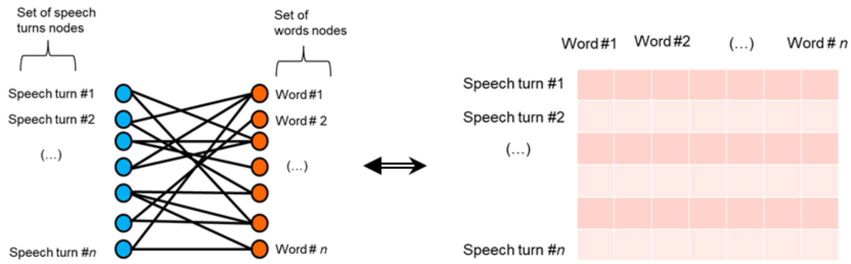

Fig. 4. Dialog as a bipartite graph and an adjacency matrix

It is also well-known that the co-clustering performs well on bipartite graph problems. It has for instance been demonstrated on the Nova, Gina and Hiva datasets ([5], p. 31-32).

Consequently, our assumption is that a co-clustering method would be interesting to tackle the problem described above. We expect the turn of speech to be distributed in different classes which differ in the vocabulary they use. We also postulate that a classic clustering approach would only work to a single dimension of the graph (either the axis of the words or of the speech turns), and consequently, it does not take into account the underlying structure of the texts/words graph. This hypothesis is supported by the superior predictive performance of a co-clustering approach on the Nova text datasets ([5]). The works of [23] and [11] also confirm the interest of applying co-clustering to Natural Language Processing tasks. Besides, the uses of co-clustering as an unsupervised classifier, or for data interpretation in bio-informatics, have also been validated (respectively [1] and [18]).

3.2 Determining Classes and Topics

We use here a technique of coclustering to obtain a “copartition” of the matrix words x speech turns. The aim of co-clustering is to discover the “best reordering and grouping” of lines and columns: grouping turns of speech and words in clusters. The notion of "best reordering and grouping" is a notion of "contrast": (i) the "best reordering and grouping" has a maximum contrast with respect to what would occur if counts were distributed at random keeping the marginals of each cluster fixed; (ii) "best reordering and grouping" maximizes the mutual information between the two clusterings. In the dialog context, one of the partitions correspond to the group of speech turns, and the other to the group of words.

Given these two categorical variables, their simultaneous partitioning is achieved: the values of categorical variables are grouped into clusters - which amounts to a coclustering problem. The product of partitions forms a multivariate partition of the representation space, i.e., a grid or a matrix of cells, as represented in Fig.4. It also represents a joint density estimator of the variables. In order to choose the "best" grid (knowing the data) of the model space, a Bayesian approach called Maximum A Posteriori (MAP) is used. The method used is based on the MODL approach described in [4] and in [5]. The analysis of the observations linking both entities brings naturally mutually consistent similarity notions on both entities.

The MODL approach allows to obtain the best quality co-partition of the speech turns/words matrix. The method finds by itself the right number (K^*) of clusters: the number of coclusters is therefore an output; the method does not require any parameter, which is useful for non-expert people. We discuss in section 4.1 the notion of the quality of a co-cluster. As with any automatic learning method, this approach requires a quantity of data: a minimum volume is required for clustering to be relevant. We observed some relevant results on a relatively small corpus of 73 dialogs (5.010 speech turns)[1].

A tool for the visualization of the generated coclusters is then used [14]. It allows a fine analysis and a "profiling" of the clusters obtained. We do not describe here all the details, but the user may navigate using a hierarchical ascending clustering from the best fine grid (using K^* clusters) to a granularity of the grid needed for his analysis (using $K^{'}$ coclusters, $K^{'} < K^*$) while controlling either the number of parts or the rate of information. The tool also allows to tag the obtained clusters (for example using keywords instead of numerical IDs).

3.3 Determining Transitions between Topics

In our approach of the problem, a topic corresponds to a cluster of "classes", itself constituted of speech turns (cf. section 2). A topic defined in this way can be linked to one or more other topics, depending on the observed frequency of their successions in the dialogs of the corpus.

From there, each dialog can be browsed as a sequence of clusters/topics. Therefore, it is possible to calculate, on the whole set of dialogs, the frequencies of transition between each cluster. The beginning and the end of each dialog are taken into account so as to avoid erroneous transitions.

[1] The "Air France corpus", available at the url: `http://www.loria.fr/projets/asila/corpus_en_ligne.html`

We assume that, depending on its position (its moment of occurrence) in the dialog, a given turn of speech is more likely to belong to a given topic (i.e. a cluster) than another; the position information is therefore taken into account during the process. One of the objectives of our work is to verify the validity of this assumption.

The representation obtained is a directed graph, whose nodes are the clusters, and the arcs are the successions between clusters. The co-clustering, applied to the dialog corpus, removes any information about the sequential order of topics. Therefore, the generation of the graph from the clusters requires several stages to retrieve these information. They are described below:

1. The first phase consists in obtaining a standardized corpus. This is done using a pre-processing tool, internal to Orange Labs and not described here. This tool is parameterized with the list of stopwords used for the French stemming by the NLTK library [2]. This list of stopwords makes it possible to delete words *a priori* useless for learning (e.g. articles, prepositions, etc.);
2. The representation obtained is then used as an input for the aforementioned coclustering approach. It generates the clusters to which each speech turn of the corpus belongs.

   Depending on the maximization of the contrast between the distribution of the co-parts (here: the word coclusters), the utterances are associated to the classes (here: clusters of speech turns) and the distribution expected under the independence hypothesis (knowing the marginals). The partitions induced by a coclustering on the two entities are clusterings. The notion of associated similarity is related to how individuals in a cluster of one entity distribute themselves on clusters of the other entity.

   Thus, a set of clusters of speech turns is obtained. Each one corresponds to a given topic, and is labeled by a unique identifier (ID). The ID is automatically attributed during the coclustering;
3. The cluster IDs are then projected onto the initial corpus, in order to retrieve the sequentiality of the clusters. In other words, every speech turn in the initial corpus is thus associated with the identifier of the cluster to which the turn belongs.

The transition frequencies between clusters are stored in a matrix. It is used to generate the corresponding graph. It is possible to choose a minimum threshold of transition frequencies, from which the transitions are displayed. This makes it possible to generate graphs more or less complex to read: only the clusters linked to others with a transition frequency equal to or above the threshold are displayed. On the other hand, the possibility of displaying only transitions greater than a threshold makes it possible to limit the possibility to take into account some irrelevant transitions. Figure 5 displays the whole process chain of our system.

[2] `http://www.nltk.org/nltk_data`

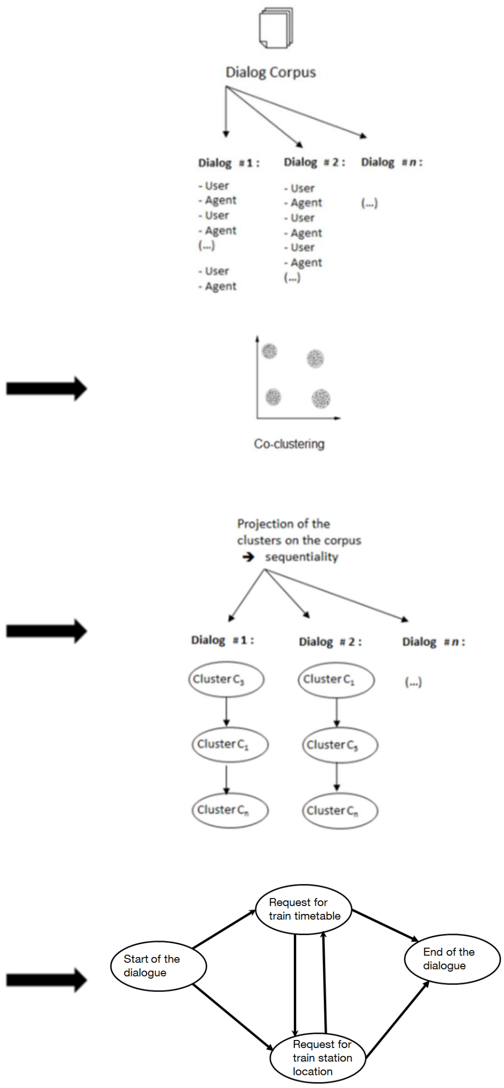

Fig. 5. Processing phases

4 Experimental Results

4.1 Overview of Existing Evaluation Protocols

Our approach, described above, relies heavily on the quality of the co-clusters obtained. It is consequently mandatory to evaluate the quality of the clusters to assess the quality of the whole generated dialog graph. There is hardly some precise method to directly evaluate the quality of a text clustering, contrary to supervised methods. Indeed, supervised methods rely on labeled data and the corresponding set of labels. By definition, it is not the case of unsupervised approaches such as co-clustering (cf. notably [7], p. 574, section 1 and [13], p. 244, section 6.3).

The simplest method would be to use an already annotated corpus as a baseline (such as the Dialog State Tracking Challenge (DSTC) corpus[3]). We would perform a co-clustering on this corpus. Then, we would use the Adjusted Rand Index (ARI) to assess the validity of the clusters thus obtained with regard to the handcrafted annotations.

Another possible but indirect way would be to carry out a clustering of the same data by the k-means method. In that case, the Davies-Bouldin Index ([10]) would be used to select the number of clusters. The result would then be used as a baseline: a domain expert would then determine if the baseline clusters obtained by k-means are more or less interpretable than the co-clusters obtained with the MODL approach.

Another way to assess the quality of the coclusters is defined with respect to the partition of the clusters, in regards to the independence hypothesis. The most useful indicator in this context is the Mutual Information ([15] and [5]). If the documents (here, the dialogs) and the words were distributed randomly, the mutual information would be equal to 0% (independence between all the words). On the contrary, the Mutual Information would be equal to 100% if the clustering would result in one cluster for each word. But then the results would not allow any interpretation. The MODL approach remedies to this issue by the regulation of the Mutual Information. It consists in partitioning the clusters in such a way that the Mutual Information is maximized, whatever the number of clusters is. This allows to select the best trade-off between the value of Mutual Information and a human interpretable number of clusters. This number is "regularized" since it is controlled (for more details, see [15]).

This matter also relates to the issue of the error analysis of the model. There can be some errors of construction of the cluster (average distance in regard to a k-means). For the clustering by k-means of data such as text, the approach Silhouette ([22]) is often used. It consists in determining if a value in a given cluster is adapted to this cluster, by swapping this value among all others clusters. For co-clustering, this corresponds to the typicity of a value; more precisely, there is one typicity for each variable of the co-clustering (for instance, one for

[3] `https://www.microsoft.com/en-us/research/event/dialog-state-tracking-challenge`

the speech turns and one for the words); usually only one of these typicity is used (in our example, the typicity of the speech turns).

4.2 Use Case

The data are a corpus of chat dialogs of online assistance (phone company customers speaking with human advisors). This corpus contains 3.012 dialogs. The average length of the dialogs is about 15 speech turns, for a total of 49.004 different speech turns and 21.700 different words. We applied to this corpus the pre-processing and the coclustering mentioned in the previous section. We obtained 24 different clusters regrouping the speech turns of the corpus.

After the co-clustering phase evaluated in section 3.2, we carried out the generation of the graphs (phase 3 described in section 3.3).

By applying different transitions threshold values to generate the graphs, we obtained globally relevant graphical visualizations. We analyzed the contents of the clusters obtained with a high threshold. We observed that most of the selected clusters correspond to very specific phases of this category of dialogs: phases of greetings, identification of the client, description of the problem encountered, elaboration of a solution by the agent, thanks and end of the dialog.

The scheduling of these clusters/topics in terms of dialog architecture also displays a typical sequential continuity of these phases, notwithstanding some clustering errors. These errors belongs to two main categories: cluster heterogeneity (one cluster mixes several topics instead of a single one) or cluster redundancy (several clusters correspond to the same topic).

As an illustration, Fig.6 presents one of the graphs obtained. For the purposes of readability, this simplified graph displays only 17 clusters: those which are connected with more than 10% of the total number of observed transitions between clusters in the initial graph. The numbers appearing next to each arc correspond to the transition percentage. Each cluster is labeled by a set of 5 of the words which participate the most to the cluster (in terms of Mutual Information).

We used a specific tool (internal to Orange Labs) for a better handling of these graphs. This tool allowed us to rename the clusters, to merge some of them, and to make the presentation even more readable. More generally, this tool allowed us to carry out a fine-grained analysis of the contents of the clusters. An instance of the most user-friendly visualization is displayed in Fig. 7; it is generated from the same clusters and data described above.

From a qualitative perspective, we evaluate the results obtained according to two criteria. The first is the homogeneity of the clusters obtained. The second is the quality and regularity of transitions between clusters; this second criterion strongly depends on the first. Evaluation is underway by the authors.

The results thus obtained were used by an ergonomist to bootstrap a first architecture of a dialog system corresponding to this application field, without any prior knowledge of the domain.

Some of the clusters were also used to represent the intents of the users. In dialog analysis, an intent corresponds to one of the tasks that the user wants to

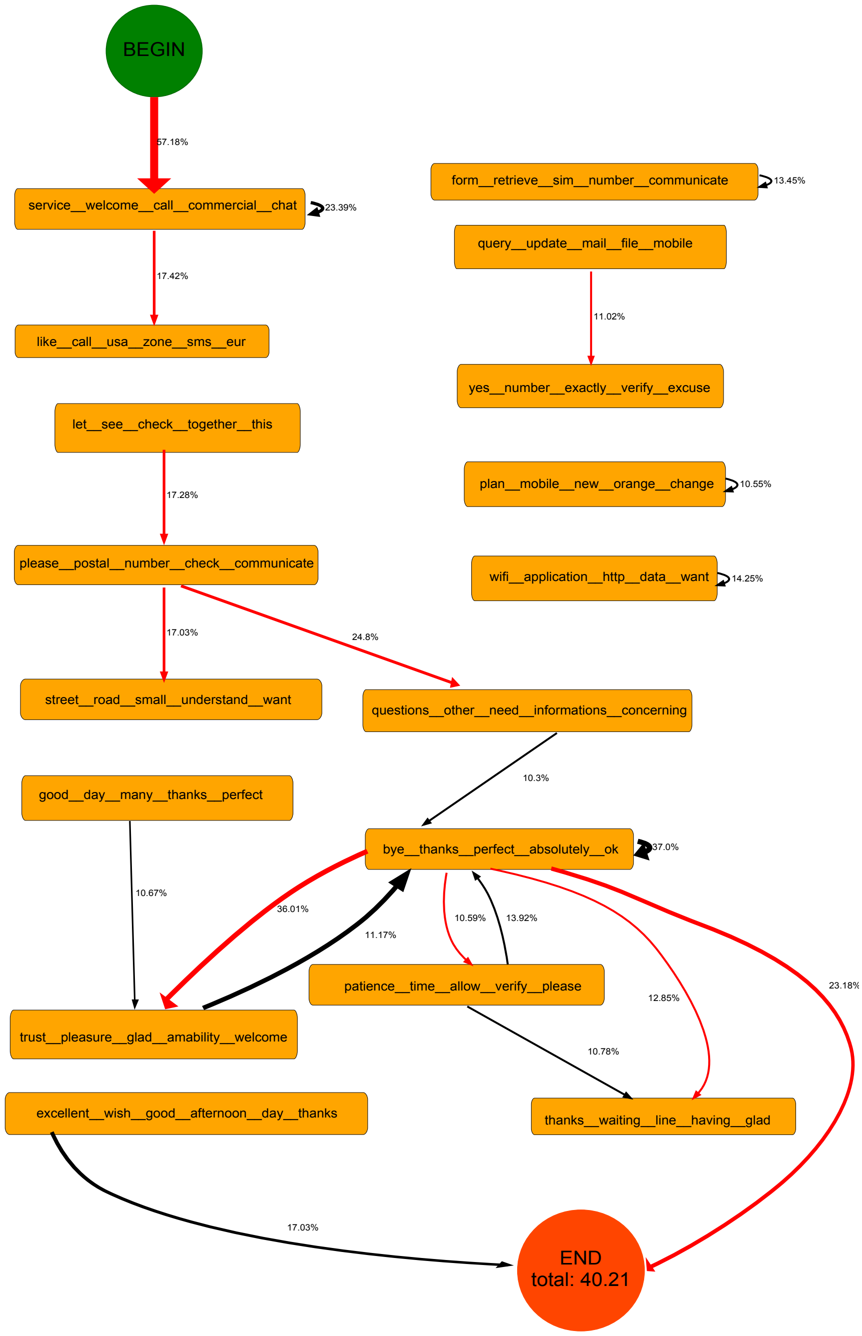

Fig. 6. Instance of an initial graph

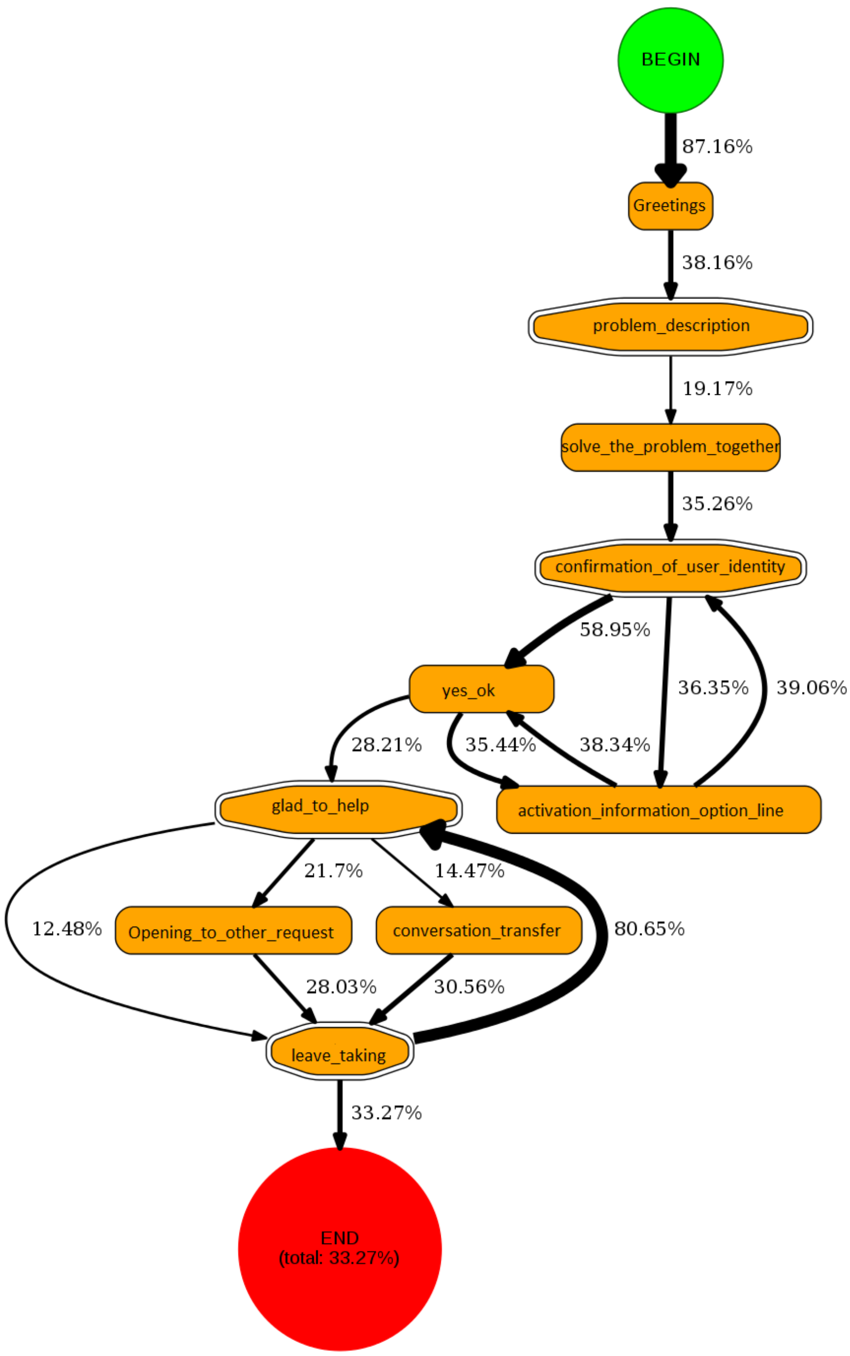

Fig. 7. Graph used for direct dialog system architecture conception

perform (for instance, knowing the price of a phone call)[4]. Here, the correspondence between an intent and a given cluster allows to consider the speech turns of this cluster as some different formulations of the intent. From this point of view also, our approach allows a fast bootstrap of the dialog system.

As a side note, we can also mention the use of our approach for another application field. This field is related to a personal home assistant conceived in Orange. A first prototype has been used by several families. The collected data notably include the vocal interactions between the human users and the assistant. These interactions aim to perform tasks such as setting up alarms, giving the weather, etc. We applied our approach on the interactions (3.322 speech turns). We showed that most of the speech turns clusters are specific to the different possible commands, and to the vocabulary associated. Further investigations are still in progress.

Note on the results reproducibility: the actual implementation of the MODL approach is the *Khiops suite*. It is accessible through Internet ([5]). Thus, any researcher who would want to reproduce our results could access to the same implementation that we used.

4.3 Note on Clustering

We do not have enough room to detail the results but we also tried to use a k-means as a baseline to benchmark the quality of the co-clusters obtained with MODL. We used an usual k-means clustering approach on the same data; center-reduction of features, Kmeans++ for initialization and 20 replicates. We analyzed the results in regards to the Davies-Bouldin Index; the value of this Index has to be the lowest possible to be optimal (cf. Fig. 8).

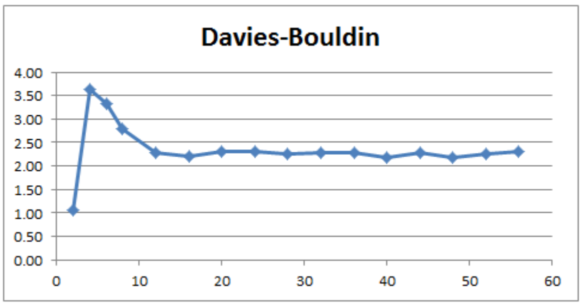

Fig. 8. Davies-Bouldin Index evolution in regard to the number of clusters (k-means)

[4] For more details on this topic, cf. [24], chap. 4.
[5] `https://khiops.predicsis.com`

The k-means does not clearly indicates a clear number of clusters since many values of k, for $k \in \{3, 60\}$ produce a close value for this index. Besides, the k-means seems to indicate that this corpus is only constituted of 2 "classes"[6]. That is clearly not true for our dialog corpus: our analysis shows that it correspond to a larger number of topics. Consequently, the k-means seems not be be the right algorithm for this kind of purpose.

The complexity of the algorithm must also be taken into account, notably in terms of memory and time complexities. [6] demonstrated that the MODL co-clustering algorithm displays high performances in regard to the main other discretization methods.

Besides, k-means clustering needs, as an input parameter, the *a priori* desired number of clusters. The user has to determine empirically this number; this induces the necessity to possibly carry out several times the clustering, until observation of relevant results. On the contrary, the MODL approach computes itself the optimal number of cluster, and produces it as an output.

5 Conclusion

We presented an unsupervised method to obtain a first version of an architecture of a dialog system, irrespective of its application field. The goal is to avoid a "cold start" conception of the system. The method consists of two main steps: firstly, the application of a coclustering algorithm on a corpus of dialogs belonging to the concerned application field; on the other hand, taking into account the sequentiality of the clusters obtained to represent the prototypical process of a dialog as a graphical representation; finally, many graphs with more or less fine grained level can be generated and used directly for the conception of the system.

We showed that our co-clustering approach of the issue shows better results than a classical clustering method. We also observed that the sequential property of the dialogue architecture is well captured by our approach.

There are still many scientific problems to tackle. Several are related to the clustering. We refer notably to the problematic of cluster homogeneity: how to determine the most representative words or rounds of words of the cluster? The matter of the selection of the most relevant clusters, (the most homogeneous ones with regard to a given topic), also arises. It is correlated with the optimal granularity/size of the clusters. It would also be interesting to study the efficiency of linguistic optimizations of the corpus: for example, the lemmatization of words or the neutralization of Named Entities. We have not used them in our work, but they could be of interest to reduce the number of parameters used for clustering. In a next version, we would also like to evaluate more thoroughly the quality of the clusters, notably with the use of the DSTC corpus as a baseline, such as mentioned in section 4.1.

[6] Since it is with this number of classes that the Davies-Bouldin Index has its lowest value (1).

Concerning the modeling of the succession of clusters, many problems also arise. As seen in the literature, it is often an HMM that is used for this purpose.

As seen in section 4.2, the graphs obtained are always very complex initially. To overcome this difficulty, we set up some transitions threshold to limit the complexity of the displayed graphs. But there are more systematic and robust techniques to reduce the initial complexity of the graphs. For instance, we consider using the spectral clustering technique.

Finally, the question arises of the reusability of the information obtained (in particular the classes and topics) from one domain to another. A deployment of this information to domains close to the initial domain is feasible and provided by the Khiops suite. For more remote domains, we envision a transfer learning approach.

Acknowledgements We thanks Nicolas Voisine, who provided us the pre proccessing tool for Khiops, Fabien Dupont and Laurent Roussarie for their essential participation to the clustering and graph visualization process chain, and Marc Boulle for his advices. We also thanks Nathalie Legay for her participation to the graph analysis. Thanks also to Romain Laroche, who contributed to the introduction of this paper.

References

1. Ahmed, M., Abdun, N., Maher, M. J.: Heart Disease Diagnosis Using Co-clustering. Scalable Information Systems: 5th International Conference, INFOSCALE 2014, Seoul, South Korea, September 25-26, 2014, Revised Selected Papers, 61-70 (2015)
2. Alexandersson, J., Reithinger, N. : Learning Dialogue Structures From A Corpus. Eurospeech 1997, 8-15 (1997)
3. Bangalore, S., DiFabbrizio, G., Stent, A. : Learning the Structure of Task-driven Human-human Dialogs. Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL, 201–208 (2008)
4. Boulle, M., Guigoures, R., Rossi, F. : Analyse exploratoire par k-Coclustering avec Khiops CoViz. Advances in Knowledge Discovery and Management, volume 527, 15-35, (2014)
5. Boullé, M.: Data grid models for preparation and modeling in supervised learning. Hands-On Pattern Recognition: Challenges in Machine Learning, volume 1,Guyon, I. and Cawley, G. and Dror, G. and Saffari, A., 99-130, Microtome Publishing, (2011)
6. Boullé, M. : MODL: a Bayes optimal discretization method for continuous attributes. Machine Learning, 65(1):131–165, (2006)
7. Candillier, L., Tellier, I., Torre, F., Bousquet, O.: Cascade evaluation of clustering algorithms. 17th European Conference on Machine Learning (ECML 2006), Berlin (Germany), LNCS 4212, 574-581 (2006)
8. Chalamalla, A., Negi S., Joshi S., Subramaniam, L. V.: Identification of Class Specific Discourse Patterns. CIKM '08 (2008)
9. Chotimongkol, A.: Learning the Structure of Task-Oriented Conversations from the Corpus of In-Domain Dialogs. Carnegie Melon University, PhD Thesis (2008)
10. Davies, D. L., Bouldin, D. W.: A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-1 (2): 224–227, (1979)

11. Dhillon I. S.: Co-clustering documents and words using bipartite spectral graph partitioning. KDD '01 Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 269-274 (2001)
12. Forni, A.: Gartner Identifies the Top 10 Strategic Technology Trends for 2017. Article on Gartner Website `http://www.gartner.com/newsroom/id/3482617` (last consulted in June 2017)
13. Gabor K., Zargayouna, H., Tellier, I., Buscaldi, D., Charnois, T.: Unsupervised Relation Extraction in Specialized Corpora Using Sequence Mining. XVIth Symposium on Intelligent Data Analysis, Oct 2016, Stockholm, Sweden. 237-248 (2016)
14. Guerraz, B., Boulle, M., Gay, D., Lemaire, V., Clerot, F. : Analyse exploratoire par k-Coclustering avec Khiops CoViz Atelier CluCo, Extraction et Gestion des Connaissances (EGC),(2015)
15. Guigourès, R. and M. Boullé and F. Rossi: Discovering patterns in time-varying graphs: a triclustering approach. Advances in Data Analysis and Classification, 1-28,(2015)
16. Guigourès, R.: Utilisation des modeles de co-clustering pour l'analyse exploratoire des données. Universite Pantheon-Sorbonne - Paris I, PhD Thesis (2013)
17. D'Haro, L. F., Cordoba, R., Lucas, J. M., Barra-Chicote, R., San-Segundo, R. : Speeding Up the Design of Dialogue Applications by Using Database Contents and Structure Information. SIGDIAL 2009, 160–169 (2009)
18. Klema, J., Malinka, F., Zelezny, F. : Semantic Biclustering: a New Way to Analyze and Interpret Gene Expression Data. Proceedings of The 12th International Symposium on Bioinformatics Research and Applications (ISBRA), Springer, LNBI 9683, 332-3, (2016)
19. Laroche, R.: Speeding Up the Design of Dialogue Applications by Using Database Contents and Structure Information. 18th International Conference on Intelligence in Next Generation Networks, 231-238,(2015)
20. Negi, S., Joshi, S., Chalamallay, A., Subramaniam, L. V.: Automatically Extracting Dialog Models from Conversation Transcripts. 2009 Ninth IEEE International Conference on Data Mining, (2009)
21. Paul, M.: Mixed membership Markov models for unsupervised conversation modeling. EMNLP-CoNLL '12, 231-238 (2012)
22. Rousseeuw, P. J.: Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. Computational and Applied Mathematics. 20: 53–65, (1987)
23. Takeuchi, K., Takahashi, H. : Co-clustering with Recursive Elimination for Verb Synonym Extraction from Large Text Corpus. IEICE Transactions on Information and Systems Vol. E92.D, 2334-2340 (2009)
24. Tur G., Mori, R. D., Eds.: Spoken Language Understanding: Systems for Extracting Semantic Information from Speech. New York, NY: JohnWiley and Sons (2011)
25. Vinyals, O., Le Quoc V.: A neural conversational model. International Conference on Machine Learning, 231-238 (2015)
26. Zhai, K., Williams, J. : Discovering Latent Structure in Task-Oriented Dialogues. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 36-46, (2015)