A Multi-Agent Based Implementation of a Delphi Process (Short Paper)

Iván García-Magariño
D. Software Engineering and Artificial Intelligence
Facultad de Informática
Universidad Complutense de Madrid, Spain
ivan_gmg@fdi.ucm.es

Jorge J. Gómez-Sanz
D. Software Engineering and Artificial Intelligence
Facultad de Informática
Universidad Complutense de Madrid, Spain
jjgomez@sip.ucm.es

José R. Pérez-Agüera
D. Software Engineering and Artificial Intelligence
Facultad de Informática
Universidad Complutense de Madrid, Spain
jose.aguera@fdi.ucm.es

ABSTRACT

The Delphi protocol is applied when a community of experts is required to reach a consensus and to deliver an answer. In these cases, consensus stands for reaching an agreement among the experts about what the answer should be. This consensus reaching problem has been already considered in the literature, though its automatisation remains as a challenge. Intuitively, the experts should dialogue, interchange ideas, and change their mind as the discussion progresses. This paper presents the first complete-implementation of the Delphi process. This implementation is achieved with a Multi-agent System(MAS), in which the experts are implemented with agents. The presented case study solves the document relevance evaluation problem where a community of experts decide whether a document is relevant or not. In conclusion, this paper makes an important contribution to people using Delphi processes, because the presented system is the first complete-computerised Delphi process. With respect to multi-agent systems, it has the potential to solve coordination in an original way.

Keywords

agent oriented software engineering, multi-agent systems, development

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Design, Experimentation

1. INTRODUCTION

Towards the intuitive vision of a coordination, this paper addresses the problem with an approach of social sciences, the Delphi protocol. A Delphi survey is a procedure for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem [7]. From the uses this procedure has, this paper focuses on the consensus agreement capabilities it brings. Reaching consensus implies there are experts providing an opinion about a concrete issue and the possibility of a disagreement among those experts. Each expert is supposed to follow different criteria and use different sources of knowledge. In this context, an external client needs to obtain a consensed opinion about an issue. This implies reaching an agreement among experts.

The goal of this paper is to provide a fully computerised Delphi process. Literature tells Delphi has been executed mainly by humans and sometimes with some computer assistance [10].The main obstacle is adapting the Delphi essence, which is very fuzzy, to the context of agents.

The Delphi integration is tested first in a document relevance evaluation domain. The problem consists in deciding if a concrete document is relevant or not in a concrete context. To answer the question, there are several expert agents designed to rate documents according to different criteria. Despite this circumstance, the paper will show how a dialogue among these agents can be established and an agreed answer obtained.

The scenario has been constructed with the INGENIAS [4] methodology. Compared to other alternatives, INGENIAS provides a comprehensive notation as well as a set of tools supporting modelling and implementation of specifications.

The paper is structured as follows. The next section indicates the representation of the Delphi Method with INGENIAS. Section 3 indicates the domain-specific aspects of the presented system. Section 4 evaluates the presented system. Finally, Section 5 mentions the conclusions and the future work.

2. REPRESENTING THE DELPHI METHOD WITH INGENIAS NOTATION

According to the guidelines from [3], there should be rounds of questionnaires and a connection between them. To model them, the delphi specification starts with the definition of these two concerns. There are two main roles: *expert* role, which fills in questionaires, and *monitor* roles, responsible of elaborating questionaires and analysing the answers. There is an additional role, the *client*, which is the one requesting the Delphi. There can be several monitors, at least one, and several experts, at least two, in a Delphi process.

Figure 1 captures the Delphi functionality applied to the document evaluation problem. The *evaluationUC* use case

Cite as: A Multi-Agent Based Implementation of a Delphi Process (Short Paper), I. García-Magariño, J. J. Gómez-Sanz and J. R. Pérez-Agüera, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons(eds.),May,12-16.,2008,Estoril,Portugal,pp. 1543-1546.

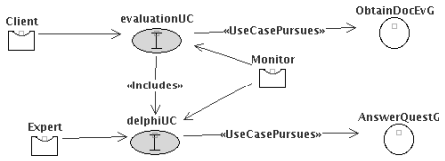

Figure 1: Main use cases considered in the development of the Delphi process

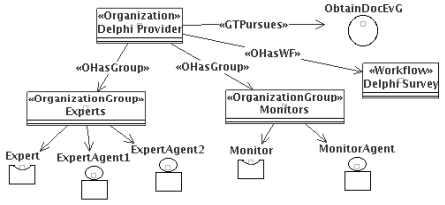

Figure 2: MAS organisation providing the document relevance evaluation

represents a client requesting a service for document evaluation by means of a Delphi survey. The service is provided by an agent playing the *monitor* role. To discover goals, the current version of the IDK permits to associate goals to identified use cases. Therfore, when the *evaluationUC* use case is performed, the *ObtainDocEvG* goal is achieved. This goal represents a future state in the system where a document has been evaluated following a Delphi process. The second use case, *delphiUC*, encapsulates the access to the questionaire filling in service offered by an agent playing the *expert* role. The *monitor* asks an *expert* to fill in a form, following the spirit of a Delphi process. The results are gathered and analysed by the *monitor* who will decide to go again into another round or finishing at the current moment. Like previous use case, this one intends to achieve a concrete goal, the *AnswerQuestG* goal. This goal represents the state of the system reached when an *expert* has filled in the supplied questionaire and a *monitor* has analysed the answer.

Now, according to the methodology, the developer must define ways in which those goals are achieved. Some goals require the involvement of a group of agents, like the *ObtainDocEvG* goal, others the involvement of a few. To satisfy the first goal, *ObtainDocEvG*, an organisation is created, the *Delphi Provider* organisation. This organisation (see Figure 2) is structured into two groups, the *experts* and the *monitors*. In the *experts* group, there will be agents able to play the *expert* role. In this case, agents *ExpertAgent1* and *ExpertAgent2* are responsible of answering the different questionaires delivered by *monitors*. For the sake of initial experiments, two *expert* agents are enough, though it is scalable to many more, provided they can implement the *expert* role.

The organisation is able to provide a service by means of the *monitor* role. The service is implemented as a workflow named *Delphi Survey*. Following again Delphi instructions, the method requires at least two rounds of questionaires. The interaction among individuals in the workflow is controlled by two interactions, *AskingEval* and *DelphiCoop*, whose corresponding protocol appears in Figure 4. The

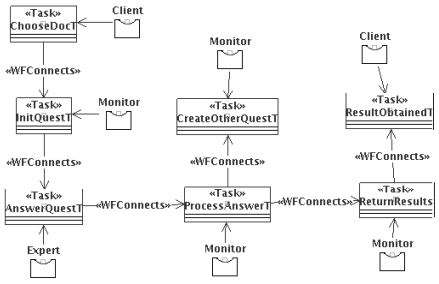

Figure 3: Overview of the workflow used to implement the Delphi process

first one encapsulates the interaction between the *client* and *monitor* roles to request the evaluation service. The second contains the questionaire elaboration, deliver, and answer gathering activities.

The workflow itself gathers the tasks shown in Figure 3. This workflow is relevant since no Delphi formal definition has been made, yet, according to our research. Therefore, this definition is also relevant. The workflow presented in Figure 3 starts with a client requesting the service with the task *chooseDoctT*. This task is supposed to provide the document to be evaluated by a *Delphi provider* organisation. The document is received by the *monitor* and a customised questionaire is elaborated with task *InitQuestT*. The questionaire is answered by experts by means of a task *AnsweQuestT*. The answer is processed by the *monitor* with a task *ProcessAnswerT*. As a result of this task, another round can be derived or not. If a new round occurs, the task *CreateOtherQuestT* should be executed. This would force another elaboration of questionaires and a new answer deliver by experts. If no more rounds occur, then the *monitor* delivers the result to the *client*, which processes the evaluation with task *ResultObtainedT*.

Some of these tasks have the responsibility of launching interactions. This is the case of *ChooseDocT*, *InitQuestT*, and *CreateOtherQuestT*. The first task creates an interaction of type *AskingEval*, while the second and third ones create an interaction of type *DelphiCoop*. As it will be seen later in Figure 4, the interaction complements the workflow definition by telling what information is passed to each agent and what tasks are expected to be triggered as a result of that information transfer.

The specification problem requires incorporating different ways of answering questionaires depending on the experts and still keeps the protocol generic. This is achieved by redefining the content of some tasks.

The protocol for sending questionnaires and receiving answers is presented in Figure 4. The protocol interleaves entities of type *InteractionUnit* with task entities. Each *interaction unit* type entity represents a communication between a *Monitor* and an *Expert* role. It has associated an speech act and the information to be transmitted. For instance, the *DistQuest* interaction unit transmits the questionnaire. When the entity is transferred, the *expert* role is expected to execute several task until the expert creates a reply for the questionnaire.

In this paper, it is assumed this extra processing is provided by tasks associated to external software components,

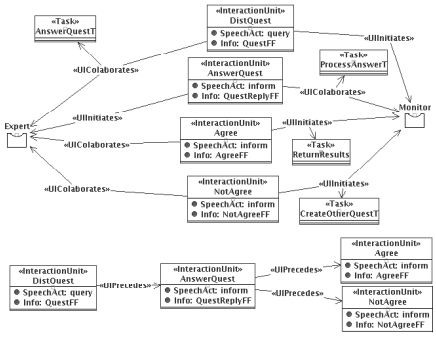

Figure 4: Protocol for passing a questionaire and receiving the answer

which implement the expert criteria. Once received the answer from the expert, the agent playing the *monitor* role either finds a consensus or decides to initiate another round of questionnaires. The first case implies engaging into a *Agree* interaction unit and sending the result of the consensus. In the second case, the *CreateOtherQuestT* task creates another instance of the interaction following the protocol from Figure 4. Also, it informs the expert that there was not an agreement by a *NotAgree* interaction unit.

Therefore, there can be several rounds of queries to the different experts. The dialogue among experts is not a direct one, since it happens as a result of the elaboration of the second round of questionaires. According to the Delphi method, the dialogue happens because each new round of questionaires incorporates results from the answers of all experts in the last round. Therefore, each expert has the opportunity to reconsider its decision according to the new information.

Unfortunately, the elaboration, replies and analysis of questionaires are domain specific. The adaptation of this part to other domains is left for future work. Except this part, this description is generic enough to fit into most applications of Delphi. The problem specific part is considered in more detail into the following section.

3. DOMAIN SPECIFIC DELPHI ASPECTS

In the the Delphi processes, some domain specific aspects are necessary. For this reason, the presented research needs to select, at the beginning, a specific domain for the experimentations. The presented work selects the domain of document relevance. The computerisation of Delphi processes with other specific domains is left for future work

The customisation of the Delphi process requires determining what the questionaires are, how they are constructed, and how they are answered. Due to the document relevance evaluation nature, the customisation requires some insight into information extraction and information retrieval. This section explains how this customisation takes place, explaining how questionaires are built for the first round and subsequent ones.

The questionaires questions considers the most important sentences contained in the document to be evaluated. Experts answers to each question are created with the relevance of the sentence in the opinion of the expert. The opinion of the expert is modelled with a set of documents. This knowledge of each expert is denoted as expert profile

To simulate the human information exchange between experts, i.e. the dialog among individuals, defined in Delphi process, like expert's comments, we propose a pseudo-relevance feedback method where each expert agent tries to append comments, modeled like query expansion terms, extracted from his profile to each entry of the questionaire to show the knowledge of the domain contained is his domain profile.

The delphi process is used before for the document relevance domain. For instance, Green[5] uses the Delphi method to evaluate web sites. However, that work needs human beings for the evaluation of documents. The best contribution of this paper is the following. The presented system is the first complete-computerised delphi process.

To generate the first questionaire, it is necessary to transform the input document in a list of queries corresponding to the questions of the questionaire.

For the elaboration of questionnaires, firstly the sentences are extracted from from the original documents. Then, the TF-ISF (Term Frequency - Inverse Sentence Frequency) [6] algorithm is applied. This algorithm selects the most informative phrases from the document.

To compute the relevance for each question of the questionaire, the presented system uses a similarity function that is able to compare these questions with the documents retrieved from the expert profile.

A relevance value is assigned to each question of the questionaire and a global relevance value is computed for the whole questionaire. This global relevance is computed using the mean average value of our similarity function for the questions contained in the questionaire. The monitor uses this global value in each round to know if the consensus has been reached.

For this task, the presented system uses the default similarity measure implemented in Lucene[1] that is co-related with the cosine in the Vector Space Model [2].

Building the questionaires for the next rounds is necessary to take in account the comments generated by the experts in the first round. For this task, the questions with highest relevance are increased with the words/comments proposed by each expert agent. This method is very similar to query expansion process using pseudo-relevance feedback [2] to extract the terms candidates to become comments.

For each round the system carries out a new expansion on the questions contained in the questionaires.

4. EVALUATION OF THE DELPHI METHOD

The evaluation of the system implementing Delphi follows the guidelines established for the evaluation of an information retrieval system. This evaluation requires, first, determining a test document collection. These collections are usually processed by humans before hand determining, for concrete queries, which documents should be retrieved from the collection.

Once prepared the collection of documents, the system is tested by asking if a document is relevant or not. The relevance is measured with the Delphi method, i.e., asking the system, and without the method, i.e., applying a TF-IDF technique directly to the document. This way, it is checked whether the Delphi method implemented with agents really

[1]http://lucene.apache.org/

improves a stand alone technique.

4.1 Preparing the experiment

Document collections from information retrieval discipline establish, for a given document, which other documents are really related to and which are not. This paper uses the collection provided by CLEF (Cross-Language Evaluation Forum) [8] for the Spanish language. The name of the collection used in this paper is EFE94. It was constructed by the international news agency EFE from all the news received during 1994. The size of the collection is 215.738 documents. The collection includes a set of topics and relevance assessments produced by humans.

Each expert profile is made of 5452 relevant documents extracted from the relevance assessments of the collection. The *train set* is divided between the different experts also whitout overlapping between them. The document test set is made of 104 documents from the relevance assessments of the test collection, 54 relevant and 50 non-relevant. There is no overlapping among the documents of the training set and the documents used for the expert profiles. In our experiments the documents contained in the test set must be judged by Delphi agent system to know their relevance using the consensus among expert agents.

4.2 Evaluation results

Commonly, the evaluation of an information retrieval technique requires measurement of Precision, Recall and F1 [9]. Precision is defined as the ratio of good assessments (relevant and non-relevant) selected to total number of assessments. Recall is defined as the ratio of relevant documents selected to total number of relevant documents available. F1 combines precision and recall into a single number. Increasing both precission and recall is the best result. However, only increasing one of them is the most common. In this evaluation, both precission and recall increase.

The results from our experiments are presented in the following table.

	Only TF-IDF	TF-IDF with DELPHI
Precision	0.86	0.92(+6.5%)
Recall	0.84	0.96(+12.5%)
F1	0.84	0.93(+9.6%)

The improvement is significant in every concern. Nevertheless, alternative measurements were applied to verify the result, concretely with the ROC [1] method. According to this, our experiment results are presented in the following table.

	Without DELPHI	DELPHI
Hit Rate	0.84	0.96(+12.5%)
False Alarm Rate	0.12	0.12(=)

Again, it can be observed the use of Delphi method achieved an improvement of the performance, greater than the one achieved without cooperation among agents. On the other hand, a very good general performance is obtained, because our system is capable to detect on average, 9 out of every 10 relevant documents.

5. CONCLUSIONS AND FUTURE WORK

This paper presents a Multi-agent based Delphi process for the document relevance domain. This system is the first complete-computerised process of the Delphi method.

Furthermore, the Delphi Method is a technique that promises a new way of dealing with the coordination of agents.

A complicate part of this method consists in determining which questions should appear in the questionaire and a proper method of elaborating, as well as analysing, answers. This part is domain specific. For the presented experimentations, the document relevance domain is selected. Nevertheless, there are already some reusable content, like the MAS specification and a part of the MAS implementation. The domain-specific part is encapsulated in certain *Task* entities and certain external components.

The presented system can be adapted to other specific domains. This adaptation is left for future work.

6. ACKNOWLEDGEMENTS

This work has been supported by the project "Methods and tools for agent-based modelling" supported by Spanish Council for Science and Technology with grant TIN2005-08501-C03-01, and by the grant for Research Group 910494 by the Region of Madrid (Comunidad de Madrid) and the Universidad Complutense Madrid.

7. REFERENCES

[1] Introduction to recommender systems: Algorithms and evaluation. *ACM Trans. Inf. Syst.*, 22(1):1–4, 2004.

[2] R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.

[3] K. Cuhls. Delphi method. Technical report, Fraunhofer Institute for Systems and Innovation Research, 2003.

[4] J. J. Gomez-Sanz, R. Fuentes, and J. Pavon. *Agent-oriented Methodologies*, chapter The INGENIAS Methodology and Tools, pages 236–276. Idea Group Publishing, 2005.

[5] J. W. Green. Delphi method in web site selection: Using the experts. *The Reference Librarian*, 33(69-70):299–310, 2001.

[6] J. Larocca Neto, A. D. Santos, C. A. A. Kaestner, and A. A. Freitas. Document clustering and text summarization. In *Proceedings of the 4th International Conference on Practical Applications of Knowledge Discovery and Data Mining*, pages 41–55, London, 2000.

[7] H. Linstone and M. Turoff. *The Delphi Method: Techniques and Applications*. Addison-Wesley Pub. Co., Advanced Book Program, 1975.

[8] C. Peters and M. Braschler. European research letter: Cross-language system evaluation: The clef campaigns. *JASIST*, 52(12):1067–1072, 2001.

[9] J. Rennie. Derivation of the F-Measure. *In other words*, 1:4.

[10] M. Turoff and S. Hiltz. Computer Based Delphi Processes. *Gazing into the Oracle. The Delphi Method and its Application to Social Policy and Public Health, Jessica Kingsley Publishers, London*, pages 56–85, 1996.