Interactive Evolution of Parameters for Clustering Textual Data in Social Science with Ant Based Algorithms

N. Monmarché, A. Durand, R. Charier, and M. Slimane

Laboratoire d'Informatique de l'Université de Tours
64 Av. Jean Portalis, 37200 Tours, France
nicolas.monmarche@univ-tours.fr

1 Introduction

In this work, the goal is to provide a tool for sociological analysis of textual data based on interactive evolution of a population of parameters in order to find, or at least to converge towards, a meaningful representation.

The data we have consist in a collection of texts obtained by a team of social science researchers after about one hundred interviews with farmers in the context of the project SOPHY. This project is focusing on how farmers are using chemical products and how they manage these chemicals in their every day work. The texts that are gathered are the direct transcriptions of the recordings made by sociologists. Each interview lasts at least two hours and the texts, obtained through a transcription with the help of Sonal software (see [1]) are then full of onomatopoeia, repetitions (and several transcription mistakes) to keep the original feeling of someone who is speaking naturally. Consequently, these data can be considered as highly noisy in comparison with textual data that are usually used to study textual analysis (such as newspaper articles for instance).

In order to provide another way to explore these texts to social science specialists, we have decided to use metaheuristics to build different representations of one text. We have focused our research on ant based algorithms, mainly because we have experimented these algorithms for clustering data but also because textual applications of artificial ants are not widespread.

The main difficulty for people who are not very familiar with metaheuristics is to understand how they can tune the metaheuristic behavior through the tuning of its parameters. Then, the idea we are providing to by-pass this complex tuning task is to use an interactive evolutionary algorithm which could help the social science researcher during his/her exploratory analysis work.

2 Clustering with ants

In the context of textual data, we could consider several levels for the clustering task. As there are many mistakes in the text, we could use a N-grams technique but the different obtained clusters would be difficult to interpret from the user point of view. Thus, we have basically considered that ants would be used to build clusters of words which could be useful for the social science field. A graph is built from the text and ants are moving from one vertex to another one as ants usually move around their nest: they use pheromons to mark their path and the more they use the same path, the more they deposit pheromons. Thus clusters are obtained when analyzing the pheromone density between vertices.

Several parameters have to be decided, either at the beginning of the ants'work (offline parameters) or while they are working (online parameter turnnig). In this work, we focus our attention to off-line parameter settings. For instance, we have to choose the number of ants, the evaporation rate of pheromones,...

Thus one parameter settings can be considered as one individual's genome in evolutionary computation field. Then, this genome can be more or less randomly mutated or combined with other similar genomes and the evolutionary process can be described.

3 Interactive evolution

The genetic and evolutionary metaphor is well suited when we can not compute a fitness for individual/genome: we just have to ask his opinion to the user. Then, according to his preferences, individuals are used to build the next generation with new individuals (i.e. new parameters).

In order to be as less complicated as possible and intuitive for the user, we have used our previous experiments in interactive evolution to build a framework that can be used for various applications. Initially, we use an interactive evolutionary algorithm proposed for generative arts [2].

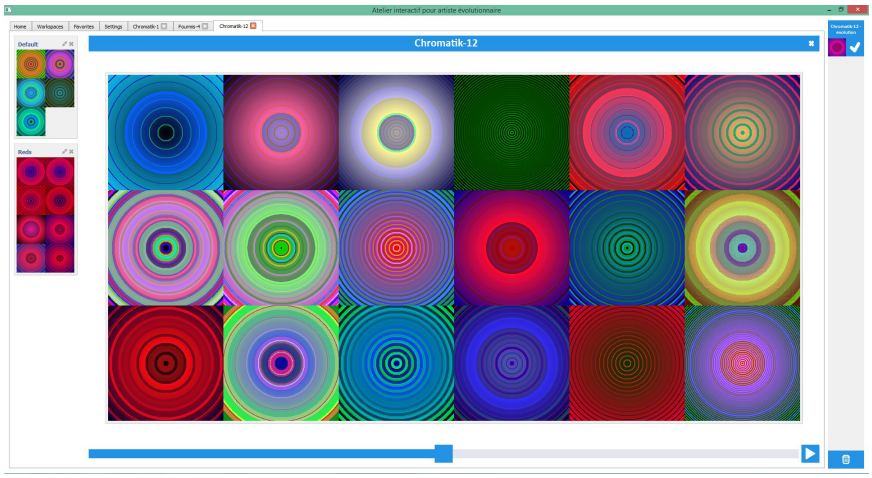

Fig. 1. Interactive Evolutionary Framework (artistic design application here)

In figure 1, each parameter settings (colors,size of circles) leads to a different individual. The user has to select which of the painting he/she prefers and then ask for the next generation. Then, hopefully, the nex generation will contain individuals which better correspond to what the user is looking for.

4 Conclusion

In our context, the social science researcher will have to play with this tool and his data. The evaluation of the framework is not easy: to study the clustering quality we will compare the results obtained on standard databases.

This project is a long term project and the developed framework will be useful for totally different fields. At time of writing, the interactive evolutionary framework is finished, but ant algorithms are still under development and particularly because they have to produce a visual result of the clustering.

We do not know if the obtained clusters will be meaningful for the social science researcher who is analyzing his interviews but we already know that the interactive and evolutionary process can lead to original results.

References

1. Alber A., “ Voir le son : réflexions sur le traitement des entretiens enregistrés dans le logiciel Sonal ”, Socio-logos. Revue de l’association françcaise de sociologie http://socio-logos.revues.org/2482 (2010).
2. Monmarcheé N., I. Mahnich, and M. Slimane: Artificial Art made by Artificial Ants. In Romero, J. and P. Machado (editors): The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music, Natural Computing Series, pages 227?247. Springer Berlin Heidelberg, 2007.