



HAL
open science

Contribution of Probabilistic Grammar Inference with K-Testable Language for Knowledge Modeling: Application on aging people

Catherine Combes, Jean Azéma

► **To cite this version:**

Catherine Combes, Jean Azéma. Contribution of Probabilistic Grammar Inference with K-Testable Language for Knowledge Modeling: Application on aging people. ICAART, 15-17 feb 2013, Mar 2013, Barcelone, Spain. pp.451-460. hal-00801983

HAL Id: hal-00801983

<https://hal.science/hal-00801983>

Submitted on 18 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Contribution of Probabilistic Grammar Inference with K-Testable Language for Knowledge Modeling: Application on aging people

Catherine Combes¹ and Jean Azéma²

¹ *University of Lyon - Hubert CURIEN Laboratoy – UMR CNRS 5516, University of Jean Monnet -18 rue Benoît Lauras, 42023 Saint-Etienne cedex 2- France*

² *University of Jean Monnet, 23 avenue du Docteur Paul Michelon, 42023 Saint-Etienne cedex 2- France
{combes, azéma}@univ-st-etienne.fr*

Keywords: grammar inference, k-testable language in strict sense, probabilistic deterministic finite automata, time series, evolution of elderly people disability.

Abstract: We investigate the contribution of unsupervised learning and regular grammatical inference to respectively identify profiles of elderly people and their development over time in order to evaluate care needs (human, financial and physical resources). The proposed approach is based on k-Testable Languages in the Strict Sense Inference algorithm in order to infer a probabilistic automaton from which a Markovian model which has a discrete (finite or countable) state-space has been deduced. In simulating the corresponding Markov chain model, it is possible to obtain information on population ageing. We have verified if our observed system conforms to a unique long term state vector, called the stationary distribution and the steady-state.

1 INTRODUCTION

Demographic shifts in the population and the fact that people are living longer have created an awareness that the health care system is and will be increasingly difficult to control, organize and finance especially where the ageing population are concerned. The senior citizen population is increasing along with the diversity of their health backgrounds and medico-social needs which cannot be provided easily because of health aspects, social conventions and lifestyles that are intertwined with the ageing process. Long-term care is a variety of services that includes medical and non-medical care to people who have a chronic illness or disability. This illness or disability could include a problem with memory loss, confusion, or disorientation. This is called cognitive impairment and can result from conditions such as Alzheimer's disease. Care needs often progress as age or as chronic illness or disability progresses. Long-term care helps meet health or personal needs. Most long-term care is to assist people with support services such as activities of daily living like dressing, bathing, and using the toilet. Approximately 70% of individuals over the age of 65 will require at least some type of long-term care services during their lifetime. Over 40%

will need care in a nursing home for some period of time. Nursing homes provide long-term care to people who need more extensive care, particularly those whose needs include nursing care or 24-hour supervision in addition to their personal care needs. We focus our interest on nursing homes. This project is being carried out in close collaboration with a French mutual benefit organization called "Mutualité Française de la Loire" which manages several nursing homes. The steps of the project consist in:

1. The specification of elderly people profiles in using unsupervised learning approach (Combes and Azéma, 2013),
2. The study of the development of these profiles over time in using a probabilistic graph of transitions between the clusters inferred by k-TSSI (k-Testable Languages in the Strict Sense Inference) algorithm. The objective is to deduce Markov process which has a discrete (finite or countable) state-space.
3. Discrete-time Markov chain simulation is used to forecast population ageing. It allows to identify the elderly people care needs and the workload in short-term, medium-term and long-term and to predict the future costs. An

application is presented in (Combes *et al.*, 2008).

This presentation is split up into seven sections. After an introduction describing the scope of the study, we introduce the characteristics of the collected data in section 2. In section 3, we describe the profiles of residents obtained in using cluster analysis. A brief review of previous works is presented in section 4. The section 5 treats the techniques used (regular probabilistic grammar inference) to model the automaton symbolizing the changing profiles and their development over time. Starting from this automaton, a Markov model is deduced. Thereby, it is possible to verify if our system is achieving a steady state. The section 6 presents the obtained results concerning the four medical nursing homes (called Bernadette, Soleil, Les Myosotis, Val Dorlay situated in France) and dementia disease and more particular, Alzheimer's disease. We conclude with some perspectives.

2 DATA COLLECTED

The quantitative data arises from the databases and the corresponding information system deals with the evaluation of autonomy/disability of elderly people. Dependence evaluation in France is carried out using a specific national scale called AGGIR: Autonomy-Gerontology-Group-Iso-Resources. The quantitative data concerns 628 residents and more than 2,200 observations of independence evaluations. The evaluations are made by the resident doctor in collaboration with the medical staff. An item can be evaluated using the four adverbs (see figure 1):

- Spontaneously corresponding to the letter *S*,
- Entirely corresponding to the letter *T*,
- Correctly corresponding to the letter *C*,
- Usually corresponding to the letter *H*.

The codification is the following. If all four adverbs are marked, the code is *C*. If less than four adverbs are checked (three or two or one), the code is *B*. If no adverb is checked, the code is *A*.

The proposed algorithm uses numerical data. So, the corresponding values are:

- 0 for code *A* meaning the person can do it alone,
- 1 for code *B* meaning the person can do partially it,
- 2 for code *C* meaning the person cannot do it alone.

The first step is to analyze the degree of autonomy-disability in order to identify clusters.

Activity	NE FAIT PAS	S	T	C	H	Code
Transferring	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	C
Moving indoors	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	B
Washing: Upper body	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	A
Washing: Lower body	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	B
Toilet: Urinary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	B
Toilet: Faecal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	B
Dressing: Upper body	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	B
Dressing: Middle body	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	B
Dressing: Lower body	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	C
Food: To serve	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	A
Food: To eat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	A
Orientation: In those days	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	B
Orientation: In the space	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	B
Coherence: Communication	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	B
Coherence: behavior	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	A

Figure 1: A.G.G.I.R scale

3 IDENTIFICATION OF RESIDENTS' PROFILES

The aim is to find feature-patterns related to the autonomy-disability level of elderly people living in nursing homes. These levels correspond to profiles based on the people's ability to perform activities of daily living like being able to wash, dress and move. To achieve this aim, an unsupervised learning approach is proposed (Combes and Azéma, 2013). It is based on principal component analysis technique to direct the determination of the clusters with self-organizing partitions. Cluster analysis is made on the 8 variables: Transferring to or from bed or chair, Moving indoors, Washing, Toilet, Dressing, Food, Orientation, Coherence. The cluster analysis identifies two kinds of patterns:

- The decline in executive functions regarding to motor and functional abilities called apraxia disorders,
- The cognitive impairment and neuropsychological deficits.

By combining clustering with a machine learning process, we could be able to predict the development of physical autonomy loss or mental autonomy loss in elderly people over time. To reach this objective, we use machine learning approach based on grammar inference in order to infer a probabilistic automaton. In the article, we only present the patients' profiles evolution regarding to upper function disorders (cognitive impairment).

REFERENCES

- Alur, R., Courcoubetis, C., Dill, D., 1990. Model-checking for real-time systems. In *Proceedings of the Fifth IEEE Symposium on Logic in Computer Science*, 414-425.
- Alur, R., Courcoubetis, C., Dill, D., 1991. Model-checking for probabilistic real-time systems. In *Automata, Languages and programming: Proceedings of the 18th ICALP*, Lecture Note in Computer Science 510.
- Alur, R., Dill, D., 1994. A theory of timed automata. *Theoretical Computer Science*, 126, 183-235.
- Angluin, D., Smith, C.H., 1983. Inductive inference: Theory and methods, *ACM Computing Surveys* 15 (3), 237-269.
- Angluin, D., 1987. Learning regular sets from queries and counterexamples. *Information and Computation*, 75, 87-106.
- Balczar, J.L., Daz, J., Gavald R., 1997. Algorithms for learning finite automata from queries: A unified view. In *Advances in Algorithms, Languages, and Complexity*, 53-72.
- Bugalho, M., Oliveira, A., 2005. Inference of regular languages using state merging algorithms with search. *Pattern Recognition*, 38(9), 1457-1467.
- Combes, C., Azéma, J., Dussauchoy, A., 2008. Coupling Markov model – optimization: an application in medico-social care, *7e International Conference MOSIM'08* - march 31th - April 2th, Paris – France, 1310-1319.
- Combes, C., Azéma, J., 2013. Clustering using principal component analysis applied to autonomy-disability of elderly people. *Decision Support System*, <http://dx.doi.org/10.1016/j.dss.2012.10.016>.
- Dupont, P., Denis, F., Esposito Y. 2005. Links between probabilistic automata and hidden Markov models: probability distributions, learning models and induction algorithms. *Patterns Recognition*, 38(9), 1349-1371.
- Garcia, P., Vidal, E., 1990a. Inference of k-testable languages in the strict sense and applications to syntactic pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (9), 920-925.
- Garcia, P., Vidal, E., Oncina, J., 1990b. Learning Locally Testable Language In Strict Sense. In *Proceedings of the Workshop on Algorithmic Learning Theory*, by Japanese Society for Artificial Intelligence. (<http://users.dsic.upv.es/grupos/tlcc/papers/fullpapers/GVO90.pdf>).
- Grinchtein, O., Jonsson, B., Leucker, M., 2005. Inference of Timed Transition Systems. In *Proceedings of International Workshop on Verification of Infinite State Systems*, *Electronic Notes in Theoretical Computer Science*, 138(3):87-99.
- Khachaturian, Z. S., 2007. A chapter in the development of Alzheimer's disease research: a case study of public policies on the development and funding of research programs. *Alzheimer's & dementia: the journal of the Alzheimer's Association* 3(3), 243-258.
- Parekh, R., Honavar, V., 2001. Learning DFA from simple examples. *Machine Learning*, 44(1/2), 9-35.
- Parekh, R., Nichitiu C.M., Honavar, V., 1998. A polynomial time incremental algorithm for learning DFA. In *Proceeding of International Colloquium on Grammatical Inference: Algorithms and Applications*, 37-49.
- Rico-Juan, J., Calera-Rubio, J., Carrasco, R., 2000. Probabilistic k-testable tree languages, In A. Oliveira (Ed.), *Proceedings of 5th International Colloquium, ICGI*, Lisbon (Portugal), Lecture Notes in Computer Science, vol. 1891, Springer, 221-228.
- Rivest, R.L., Schapire. R.E., 1993. Inference of finite automata using homing sequences. *Information and Computation*, 103, 299-347.
- Verwer, S., de Weerd M., Witteveen, C., 2007. An algorithm for learning real-time automata. In *proceedings of the 18th Benellearn*, P. Adriaans, M. van Someren, S. Katrenko (eds.).
- Verwer, S., de Weerd M., Witteveen, C., 2011. The efficiency of identifying timed automata and the power of clocks. *Information and Computation*, 209 (3), 606-625.
- Vidal, E., Thollard, F., de la Higuera, C., Casacuberta, F., Carrasco, R.C., 2005. Probabilistic Finite-States Machine, *IEEE Transactions on Patterns Analysis and Machine Intelligence*, 27 (7), 1013-1039.