



**HAL**  
open science

## CarrotAge, a software for mining land-use data

Florence Le Ber, Jean-François Mari, Marc Benoît, Catherine Mignolet,  
Céline Schott

► **To cite this version:**

Florence Le Ber, Jean-François Mari, Marc Benoît, Catherine Mignolet, Céline Schott. CarrotAge, a software for mining land-use data. Fourth International Workshop on Environmental Applications of Machine Learning - EAML'2004, 2004, Bled, Slovenia, 2 p. inria-00099884

**HAL Id: inria-00099884**

**<https://inria.hal.science/inria-00099884v1>**

Submitted on 26 Sep 2006

**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.

## CARROTAGE, a software for mining land-use data

F. Le Ber<sup>1,2</sup>, J.-F. Mari<sup>2</sup>, M. Benoît<sup>3</sup>, C. Mignolet<sup>3</sup>, and C. Schott<sup>3</sup>

<sup>1</sup> ENGEES, 1 quai Koch, BP 1039, F-67070 Strasbourg CEDEX  
fleber@engees.u-strasbg.fr, phone: 33 388248230, fax: 33 388248284  
<sup>2</sup> UMR 7503 LORIA, BP 239, F-54506 Vandœuvre-lès-Nancy CEDEX  
jfmari@loria.fr

<sup>3</sup> INRA SAD, Domaine du Joly, 662 avenue Louis Buffet, F-88500 Mirecourt  
{benoit,mignolet,schott}@mirecourt.inra.fr

**Keywords:** Data Mining, Markov Models, Land Use, Cropping Patterns, Crop Rotation.

### Introduction

Mining sequential patterns is an active area of research in artificial intelligence. One basic problem in analyzing a sequence of items is to find frequent episodes, i.e. collections of events occurring frequently together. We rely on new numerical algorithms, based on high-order stochastic models – the second-order hidden Markov models (*HMM2*) – that were initially specified for speech recognition purposes (Mari *et al.*, 1997). We have shown that, with minor changes, they can extract spatial and temporal regularities that can be explained by human experts and may constitute elements of a *knowledge discovery process* (Mari *et al.*, 2002).

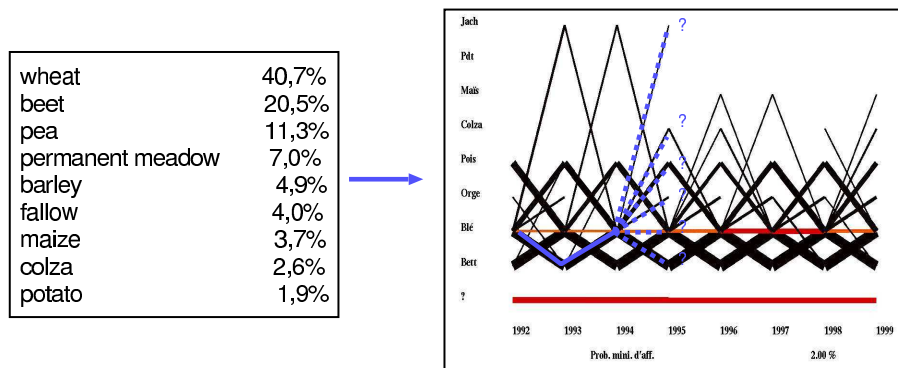
The *HMM2*'s are based on the probabilities and statistics theories. They are implemented with an unsupervised training algorithm, the EM algorithm (Dempster *et al.*, 1977), that allows to estimate a model parameters from a corpus of observations and an initial model. The resulting model is capable to segment each sequence in stationary and transient parts and to build up a classification of the data together with the *a posteriori* probability of this classification. This characteristic makes the *HMM2*'s appropriate to discover temporal and spatial regularities. For all these reasons the *HMM2*'s have been chosen as a basis for the land-use data mining software CARROTAGE (<http://www.loria.fr/~jfmari/App/>).

### Results and Discussion

CARROTAGE has been developed for studying the cropping patterns of an agricultural territory. It uses therefore a french agricultural database, named *Ter-Uti*, which records every year the land-use category of about 500,000 sites regularly spaced on french territory (Ledoux & Thomas, 1992). One *Ter-Uti* site represents roughly 100 hectares. The collected land-use categories (wheat, corn, potato, forest, ...) are logged in a matrix in which the rows are the sites and the columns the time slots. CARROTAGE takes this matrix as an input and builds a partition together with its *a posteriori* probability.

We work within an interdisciplinary research program which aims to develop a tool for forecasting water quality in the Seine river watershed, based on assumptions upon agricultural changes. Thus, we analyse the agricultural activities in the watershed, their dynamics and their spatial organisations, focusing on the crop (temporal) rotations and the associated agricultural practices that are able to explain a part of the risk of nitrate loss (Mignolet *et al.*, 2004). The

data mining software CARROTAGE has been used on *Ter-Uti* data from the Seine watershed. As shown in the figure below for a small agricultural district from the north-east of France, CARROTAGE allows to compute the crop distribution in a given periode (here from 1992 to 1999, left), and to view the annual transitions between crops (right).



These results are analysed, with respect to the main transitions between crops. For example the dashed lines represent the possible transitions between the triple *wheat-beet-wheat* and the other crops: beet, pea, wheat, barley, colza or fallow. The same analysis has been done for all small agricultural districts in the Seine watershed. The districts are then clustered according to their main crop rotations (and their evolutions), for modelling water pollution risks wrt spatial characteristics of the Seine watershed.

## Conclusion

The *HMM2*'s have proven to be appropriate to discover temporal and spatial regularities. Furthermore, the CARROTAGE software, based on *HMM2*'s, has proven to be useful for analysing spatial and temporal cropping patterns. The models and visualisation tools have been designed within a collaboration between agronomists and computer scientists. CARROTAGE has been used successfully for studying the link between nitrate contamination of groundwater and surface water in the Seine watershed and the evolution of agricultural activities. It has also been used for helping satellite images interpretation, with an aim of irrigation needs forecast. Other applications could be developed, regarding ecological or environmental spatio-temporal data.

## References

- Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum-Likelihood From Incomplete Data Via The EM Algorithm. *Journal of Royal Statistic Society, Ser. B*, 39:1 – 38, 1977.
- Ledoux, M., and Thomas, S. De la photographie aérienne à la production de blé. *AGRESTE, la statistique agricole*, (5), 1992.
- Mari, J.-F., Haton, J.-P., and Kriouile, A. Automatic Word Recognition Based on Second-Order Hidden Markov Models. *IEEE Transactions on Speech and Audio Processing*, 5:22–25, 1997.
- Mari, J.-F., Le Ber, F., and Benoît, M. Segmentation temporelle et spatiale de données agricoles. *Revue Internationale de Géomatique*, 12(4):439–460, 2002.
- Mignolet, C., Schott, C., and Benoît, M. Spatial dynamics of agricultural practices on a basin territory: a retrospective study to implement models simulating nitrate flow. The case of the Seine basin. *Agronomie*, 24, 2004. To be published.