PhD position

Animal welfare : characterizing the diversity between and within livestock farming situations with **data mining methods** used on information from dairy herd sensors.

Alexandre Termier (PhD supervisor) and Véronique Masson, Irisa/Inria RBA (Rennes), Lacodam team

firstname.lastname@irisa.fr

Keywords :

Data Mining, Pattern Mining, Animal Welfare, Dairy cows

Context

The objective of the Lacodam team from Inria (the French national Research Institute for digital sciences) is to considerably facilitate the process of making sense from large quantities of data, either for deriving new knowledge or for taking better actions. The solutions we envision requires to bridge data mining techniques with artificial intelligence approaches, both for taking knowledge into account in a principled way, and to introduce automated reasoning techniques in knowledge discovery workflows.

This PhD will be funded by the #DigitAg Convergence Institute and Inria. #DigitAg is the premier French project on the exploitation of numeric technologies in Agriculture. The PhD will be co-directed by computer scientists of the Inria Lacodam team and zoologists of INRA which will bring domain expertise.

Assignment

General objective

Nowadays, the consumer as well as the citizen are expecting herd management strategies that take into account animal welfare. Animal welfare is satisfied if the animal is in good physical and psychological health, feels good and does not suffer. This led to the design of certification systems. Most of them are based on means obligations that are easy to verify (building type, surface per animal, equipment…). However, these are not sufficient to guarantee **animal welfare**. To guarantee an accountable performance obligation, using monitoring technologies (sensors for reproduction or health,…) can contribute to better evaluate the different dimensions of animal welfare, after a proper processing of the data. To date, monitoring data analysis is mainly focused on the individual animal, and do not allow to understand the diversity and variability of herds. Such an analysis would allow establishing a precise “map” of the behaviors observed in cattle herds and to distinguish behaviors coming from individual animals and those coming from the management strategy. **Data Mining techniques** are especially well suited for this task. “**Discriminative pattern mining**” approaches can discover regularities or irregularities in data for groups of animals or groups of herds. Such information can be used to establish a taxonomy of herd behaviors, allowing a quick positioning of each herd depending on its management strategy. The farmer would then have precise information

on the welfare status of its herd and the possibilities of improvement. Labelling systems could also better answer to consumer expectations.

Technical approach

Our general approach will be to use discriminant pattern extraction algorithms. Such patterns can isolate groups of farms with particular behaviors, and be used to present these behaviors to stakeholders. In particular, we will focus on subgroup discovery algorithms, which are suitable for managing data mixing numeric and symbolic attributes.

A first task will be the management of the heterogeneity of the data. Subgroup discovery algorithms are not designed to handle sets of time series. In the available data, each animal will be able to produce several time series (ex: temperature + activity). A first approach will therefore be to determine what are the simple digital features that can be extracted from the time series in order to simplify the data and to be able to use existing algorithms. If the results are not satisfactory, it will mean that the patterns to be discovered must exhibit some of the time variations of the series (eg: shapelet patterns, i.e. time series subparts). The second approach will be to adapt a subgroup discovery algorithm to directly process time series.

A second task will be to characterize the types of patterns sought. Existing algorithms are designed to discover regularities in any data. It will be necessary to propose a more precise definition of the regularities interesting us, making it possible to discover behaviors related to the breeding (driving, conditions, ...), and having an impact on the animal welfare. The development of this definition will require many back and forth interactions between the two teams (Inria and Inra). The main difficulty will be to distinguish between what is computable in practice and what is interesting from a zootechnical point of view.

The third and final task will be to determine the best return to breeders based on patterns discovered by the algorithm and in accordance with the recommendations of animal welfare experts. First, this will require considering the patterns discovered on all farms, and construct some form of taxonomy of the farms based on their welfare patterns. It will then be easy to place any new farm in that taxonomy, by determining the welfare patterns in activates. This will also allow proposing a simple benchmark to the breeder, and suggestions of improvement. To calibrate this approach, we will work closely with volunteer breeders that will give us feedback on early results. This will be possible thanks to the choices to work with some partners of LIT Ouesterel.

References

M. Atzmueller, Subgroup discovery, Wiley interdisciplinary reviews: data mining and knowledge discovery, vol. 5, iss. 1, pp. 35-49, 2015.

W. Duivesteijn, A. J. Feelders, and A. Knobbe, Exceptional model mining, Data mining and knowledge discovery, vol. 30, iss. 1, pp. 47-98, 2016.

G. Dong, J. Li: Efficient Mining of Emerging Patterns: Discovering Trends and Differences. KDD 1999: 43-52

G. Dong, V. Taslimitehrani: Pattern Aided Classification. SDM 2016: 225-233

K. Fauvel, V. Masson, E. Fromont, P. Faverdin, A. Termier : Towards sustainable dairy management – a machine learning enhanced method for estrus detection. KDD 2019, to appear.

Main activities

The PhD candidate will be asked to:

- Thoroughly review the state of the art of discriminative pattern mining and subgroup discovery, and keep up to date on new papers;
- Analyze real data in order to get insights to what is missing to existing approaches;
- Develop novel algorithms to analyze farm data and discover welfare patterns;
- Interact with Inra partners and possibly farmers in order to understand the precise requirements, and evaluate the first results;
- Implement research prototypes of the proposed approaches.

Skills

Technical skills required:

- Strong interest and ability in developing novel algorithms
- Strong programming skills
- Good understanding of data mining approaches
- Capacity to analyze actual data

Soft skills required:

- Scientific curiosity, interest to work with scientists of other domains (zoology/agriculture)
- Team working
- Ability to communicate scientific reasoning and results either in written or oral form