



Data management and user profiling based on neural networks

W. Funika, F. Szura, J. Kitowski

Department of Computer Science AGH-UST
ACC CYFRONET AGH

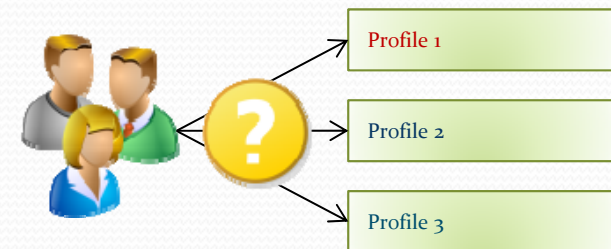


Motivation

- Continuously monitor local resources and users' activities
- Observe user behavior and needs in terms of data access and provide them with an ability to increase the storage quality for particular types of users' tasks

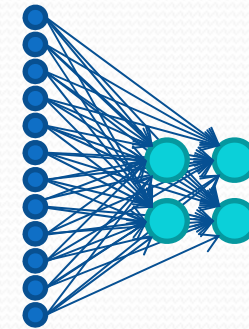
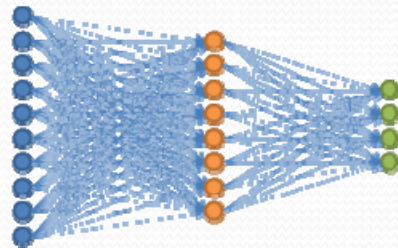
Key functions

- User profiling based on historical data accesses
- Management of data sets based on user profile and identified patterns of data usage
- Self-learning enabling to detect possible user profiles and to classify user actions to an appropriate profile



Key features

- Use of neural network for clustering and classification



- Storage usage and user monitoring
- Simple and fast user profiling allowing to adapt to changes in the behavior of the user
- Actions to be performed are associated with an individual profile, not exactly with the user

Our poster

- Poster number: 15

We invite you to our poster.

Data management and user profiling based on neural networks

Włodzimirz Funiak(1,2), Filip Szura(1,2) and Jacek Kitowski(1,2)
 (1) ACC CYFRONET AGH, ul. Nawojki 11, 30-950 Kraków
 (2) Institute of Computer Science AGH, al. Mickiewicza 30, 30-059 Krakow, Poland

MOTIVATION

The presented system is designed to reduce the time which is needed to access data during user's computation. The user is profiled based on his / her history and current operation to an appropriate profile. Each profile defines the manner how the data should be accessed or how it should be stored. The system assigns this profile to the user for every his / her task. E.g., it can determine that the system should provide access to faster devices in read operations when user's tasks are mainly aimed at reading large amounts of data but do not produce much data.

- Profile 1
- Profile 2
- Profile 3

KODMAS MANAGEMENT TOOL

KODMAS (KnoWledge based Data Management Support) exploits neural network as a knowledge engine which is responsible for user clustering and classification. This is done to define the profile for the individual user. The engine in its current development uses a Kohonen network for the clustering and afterwards, once enough data is collected it is using an MLP (multilayered perceptron) network for the online classification. The flow of the information in our system contains four major types which are as follows:

1. Statistics about user data access – it is used to discover available user profiles
2. Learning data – this data is used to create and train an MLP network. Afterwards this network is able to receive information no. 3 (see in Architecture diagram) and assign the user to a profile online
3. Current data which describes the user's activity – this information contains the data which defines how many tasks the user perform in the system and what are their requirements
4. User profile which is assigned based on information from points 2 and 3. This profile is assigned to the user for their task and represents their type of activity in the infrastructure.

Data manager

Data placement manager

This part of our system being under development is connected with data management. It uses the information collected from devices with regard to their performance and information which describes the user profile. The system is able to manage data most suitably to user requirements. The data of users which only read data and do not write a lot of data are stored on the devices which are the fastest ones in terms of read operations speed but not as fast as other ones in write operations. The data placement is strongly connected with the user and their profile and the type of computations which they performed.

Classification of users realised in KODMAS

The classification of users is performed by MLP network. This network is created and learned with the use of information which were stored at the user clustering stage. This information comprises the number of neurons in the MLP network. This data is also used to learn the new network how to recognize a user and classify him/her to an appropriate profile.

DEVELOPMENT SUMMARY AND FUTURE WORK

Current tests were aimed to check the first component of the developed system, i.e. the user profiler. The user were assigned correctly to the appropriate profile. The profiles were discovered by the use of Kohonen network which was operating on the historical data. Each profile defines a different access manner due to that the location of data for users was done in a different way being dependent on the profile assigned.

- The use of neural network gives us an ability to create a solution which is able to adapt to the observed situation and find the user profiles which defines their activity with regard to the data access.
- In the future we are going to extend our system with a component which will be able to manage user data (shown above as the data manager).

REFERENCES

[1] W. Funiak, F. Szura, J. Kitowski: Automation of System Monitoring Based on Fuzzy Logic or Rules: Comparison of Two Designed Approaches with Regard to Computational Infrastructures. UNCS vol. 7136, Springer, 2012, pp. 142-156

[2] W. Funiak, F. Szura, J. Kitowski: Data management with AI methods in distributed environment. Abstracts of CGW'12, ACC Cyfronet AGH, Krakow, 2012, pp.57-58

ACKNOWLEDGMENTS

The research presented in this paper has been partially supported by the European Union within the European Regional Development Fund program no. POIG.02.03.00-00-096/10 as part of the PL-Grid PLUS Project (<http://www.plgrid.pl/projekty/plus>) and ACC Cyfronet AGH grant 500-08.