

Neural network approaches to Reality Mining: The Reality Mining dataset

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

PRESENTATION MADE BY
ALESSANDRO AND DOMANTAS

Outline

1. Description of the problem
 1. Dataset description
2. Description of our approach
 1. Feature extraction
 2. Classifiers
 3. Validation
3. Implementation
 1. Feature Extraction and Selection
 2. Classification
4. Results and Conclusion

Description of the problem

Main goal - find some correlations between the habits of some students and the type of studies thier belong.

Main idea - recognize the address of studies(or a type job for a worker) considering in which place the subjects are used to pass their time during a day.

Types of the **classification** used:

- Multilayer Perceptron Neural Network (MLP)
- Restricted Boltzmann Machine (RBM)

Dataset description

Reality Mining 1 was an experiment conducted from 2004-2005 at MIT Media Laboratory.

It consisted to detect the follow information of 94 subjects:

- Bluetooth devices (proximity of approximately five meters)
- call logs
- cell tower ID
- application usage
- phone status
- self-report relational data

Data were detect with the help of particular type of cellphone (Nokia 6600)

Categories that have been used in our project: **affiliation type** and **data_mat**.

Dataset description

my_affil - The subject's affiliation:

- 'mlgrad' – Media Lab Graduate Student (not a first year)
- '1st yeargrad' – Media Lab First Year Graduate Student
- 'mlfrosh' – Media Lab First Year Undergraduate Student
- 'mlstaff' – Media Lab Staff
- 'mluop' – Media Lab Undergraduate
- 'professor' – Media Lab Professor
- 'sloan' – Sloan Business School

Dataset description

data_mat - Inferred subject locations at each hour of the day:

- 1 - home
- 2 - work
- 3 - elsewhere
- 0 – no signal
- NaN – phone is off

Features selection

Features are selected from **data_mat** dataset.

Feature selection rules for **Restricted Boltzman Machine(RBM) + Multi Layer Perceptron**:

1. Subject activities data **is not empty**;
2. Activities data should represent **at least 7 days** of subject activity.

Features selection rules for **Multi Layer Perceptron(MLP)**:

1. Affiliation type of subject **is not empty**;
2. Subject activity data **is not empty**.

Classification

Classification selection rules:

1. Affiliation type for subject should not be empty
2. Activity data of subject should not be empty

Classification is used only for **Multi Layer Perceptron**

Validation

Features vector is split into two parts:

1. Training part
2. Testing part

We apply first the train part on the two learning procedure and then we use test set to complete the accuracy.

The two training operation, **RMB** before and **Multilayer-Perceptron** after, can rely on two distinct datasets that remain independent and therefore do not change learning information.

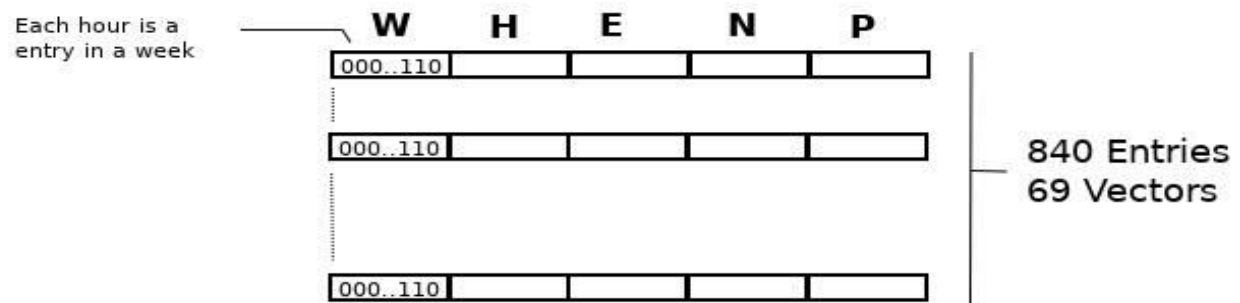
Implementation. Feature selection

Multi Layer Perceptron:

Vector representation is created for each subject activity frequency values. All days activity about subject is gathered.

Restricted Boltzman Machine:

All week activity of each subject hour of each activity type is stored in a 840(5 available activities * 24 hours * 7 days).



Implementation. Classification

Classification are selected from Affiliation type dataset.

Classification types: **Sloan** or **No sloan**.

Sloan – **Business school subjects**.

No sloan :

‘mlgrad’ – Media Lab Graduate Student (not a first year)

‘1 st yeargrad’ – Media Lab First Year Graduate Student

‘mlfrosh’ – Media Lab First Year Undergraduate Student

‘mlstaff’ – Media Lab Staff

‘mluop’ – Media Lab Undergraduate

‘professor’ – Media Lab Professor

Results and Conclusion

While in the Multi Layer Perceptron method has observed an accuracy of 0.6 in the other method the accuracy is floating and can depend from many influences.

On average the accuracy reach from the second method is in a range of $\approx 0.7 - 0.8$. For source of future developments it's seems to be a correlation between the quality of RBM learning (considering maximum pseudo-Likelihood) and final learning through perceptron.

So it can be said that if the RBM network can be useful to identify a reasonable pattern, but that one can't be considered evident by the presence (perhaps) of a few and noisy data.

Questions
