The Brazilian Conference on Intelligent Systems (BRACIS) and Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)

Query Rules Study on Active Semi-Supervised Learning using Particle Competition and Cooperation

Fabricio Breve fabricio@rc.unesp.br

Department of Statistics, Applied Mathematics and Computation (DEMAC), Institute of Geosciences and Exact Sciences (IGCE), São Paulo State University (UNESP), Rio Claro, SP, Brazil

Outline

- Introduction
  - Semi-Supervised Learning
  - Active Learning
- Particles Competition and Cooperation
- Computer Simulations
- Conclusions

Semi-Supervised Learning

- Learns from both labeled and unlabeled data items.
  - Focus on problems where:
    - Unlabeled data is easily acquired
    - The labeling process is expensive, time consuming, and/or requires the intense work of human specialists

[1] X. Zhu, “Semi-supervised learning literature survey,” Computer Sciences, University of Wisconsin-Madison, Tech. Rep. 1530, 2005.
[2] O. Chapelle, B. Schölkopf, and A. Zien, Eds., *Semi-Supervised Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, MA: The MIT Press, 2006.
[3] S. Abney, *Semisupervised Learning for Computational Linguistics*. CRC Press, 2008.

Active Learning

- Learner is able to interactively query a label source, like a human specialist, to get the labels of selected data points
  - **Assumption:** fewer labeled items are needed if the algorithm is allowed to choose which of the data items will be labeled

[4] B. Settles, “Active learning,” *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 6, no. 1, pp. 1–114, 2012.
[5] F. Olsson, “A literature survey of active machine learning in the context of natural language processing,” Swedish Institute of Computer Science, Box 1263, SE-164 29 Kista, Sweden, Tech. Rep. T2009:06, April 2009.

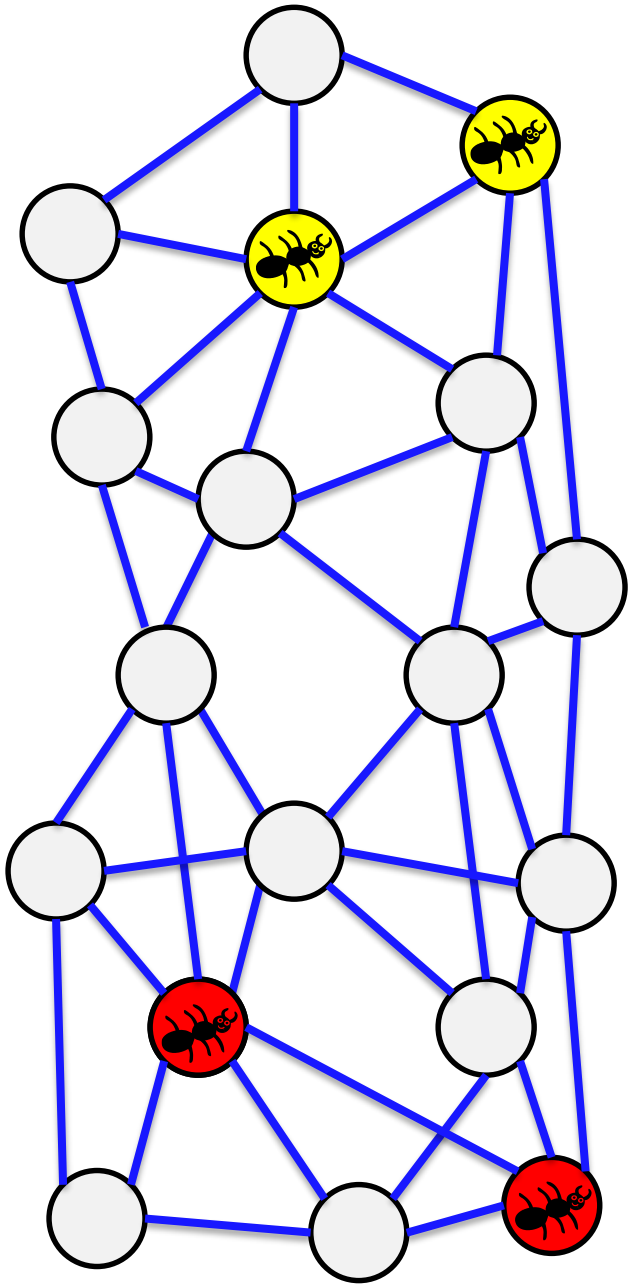
SSL+AL using Particles Competition and Cooperation

- Semi-Supervised Learning and Active Learning combined into a new nature-inspired method
  - Particles competition and cooperation in networks combined into an unique schema
    - Cooperation:
      - Particles from the same class (team) walk in the network cooperatively, propagating their labels.
      - **Goal**: Dominate as many nodes as possible.
    - Competition:
      - Particles from different classes (teams) compete against each other
      - **Goal**: Avoid invasion by other class particles in their territory

[15] F. Breve, “Active semi-supervised learning using particle competition and cooperation in networks,” in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, Aug 2013, pp. 1–6.
[12] F. Breve, L. Zhao, M. Quiles, W. Pedrycz, and J. Liu, “Particle competition and cooperation in networks for semi-supervised learning,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 24, no. 9, pp. 1686 –1698, sept. 2012.

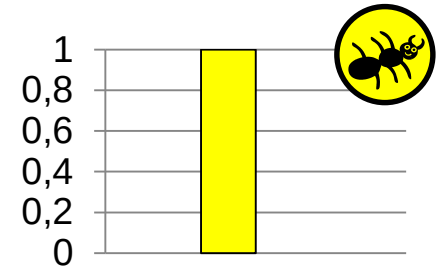
Initial Configuration

- An undirected network is generated from data by connecting each node to its k-nearest neighbors
- A particle is generated for each labeled node of the network
- Particles initial position are set to their corresponding nodes
- Particles with same label play for the same team

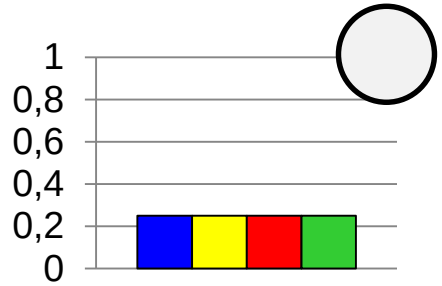


Initial Configuration

- Nodes have a domination vector
  - Labeled nodes have ownership set to their respective teams (classes).
  - Unlabeled nodes have ownership levels set equally for each team



Ex: [0.00 1.00 0.00 0.00]
(4 classes, node labeled as class B)

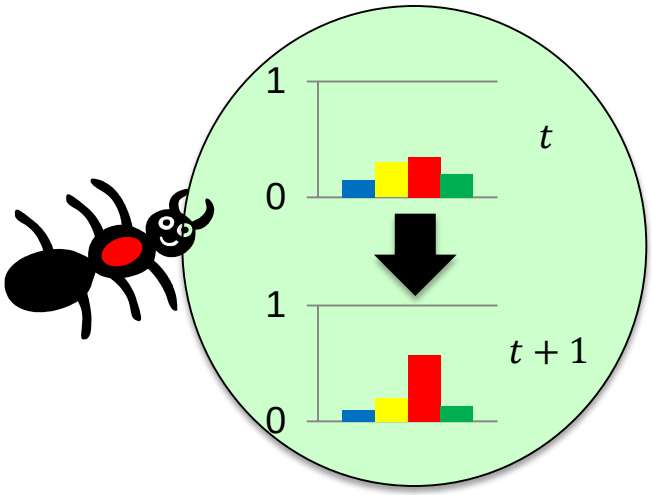


Ex: [0.25 0.25 0.25 0.25]
(4 classes, unlabeled node)

$$v_i^{\omega_\ell} = \begin{cases} 1 & \text{if } y_i = \ell \\ 0 & \text{if } y_i \neq \ell \text{ e } y_i \in L \\ 1/c & \text{if } y_i \notin L \end{cases}$$

Node Dynamics

- When a particle selects a neighbor to visit:
  - It **decreases** the domination level of the **other** teams
  - It **increases** the domination level of its **own** team
  - Exception: **labeled** nodes domination levels are **fixed**

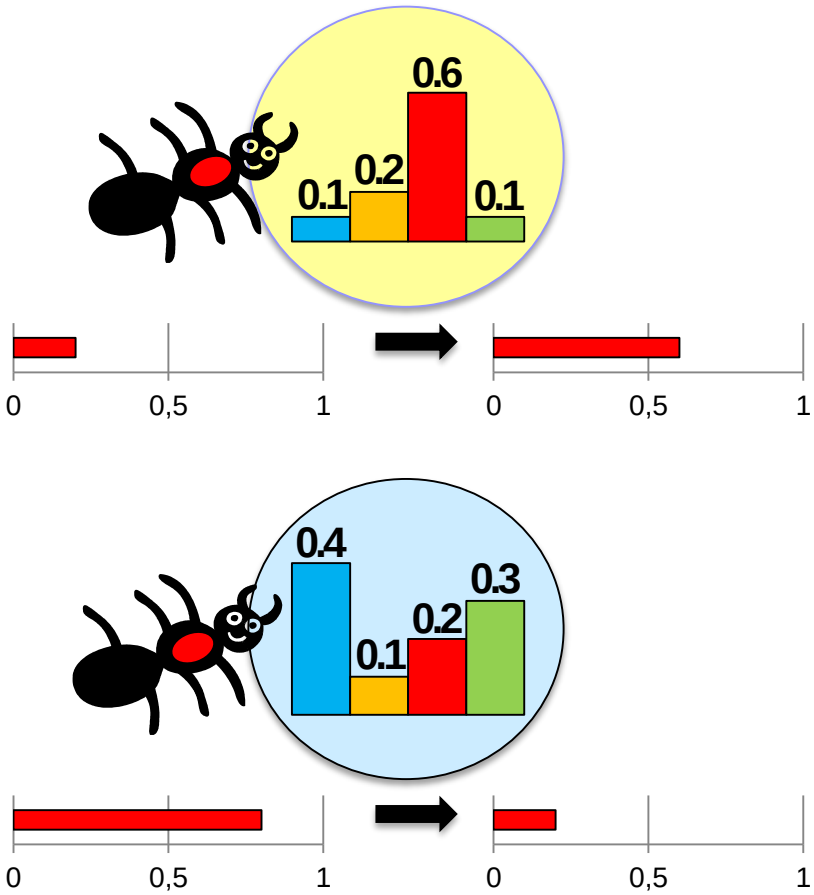


$$v_i^{\omega_\ell}(t+1) = \begin{cases} \max\left\{0, v_i^{\omega_\ell}(t) - \dfrac{0.1\,\rho_j^{\omega}(t)}{c-1}\right\} & \text{if } \ell \neq \rho_j^f \\ v_i^{\omega_\ell}(t) + \sum_{r \neq \ell} v_i^{\omega_r}(t) - v_i^{\omega_r}(t+1) & \text{if } \ell = \rho_j^f \end{cases}$$

Particle Dynamics

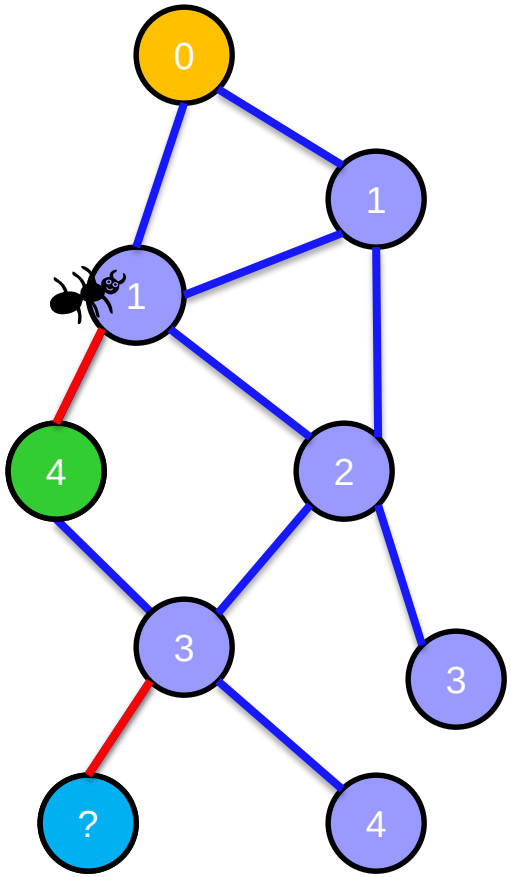
- A particle gets:
  - **Strong** when it selects a node being dominated by **its own** team
  - **Weak** when it selects a node being dominated by **another** team

$$\rho_j^{\omega}(t) = v_i^{\omega_\ell}(t)$$



Distance Table

- Each particle has a distance table.
- Keeps the particle aware of how far it is from the closest labeled node of its team (class).
  - Prevents the particle from losing all its strength when walking into enemies neighborhoods.
  - Keeps the particle around to protect its own neighborhood.
- Updated dynamically with local information.
  - No prior calculation.



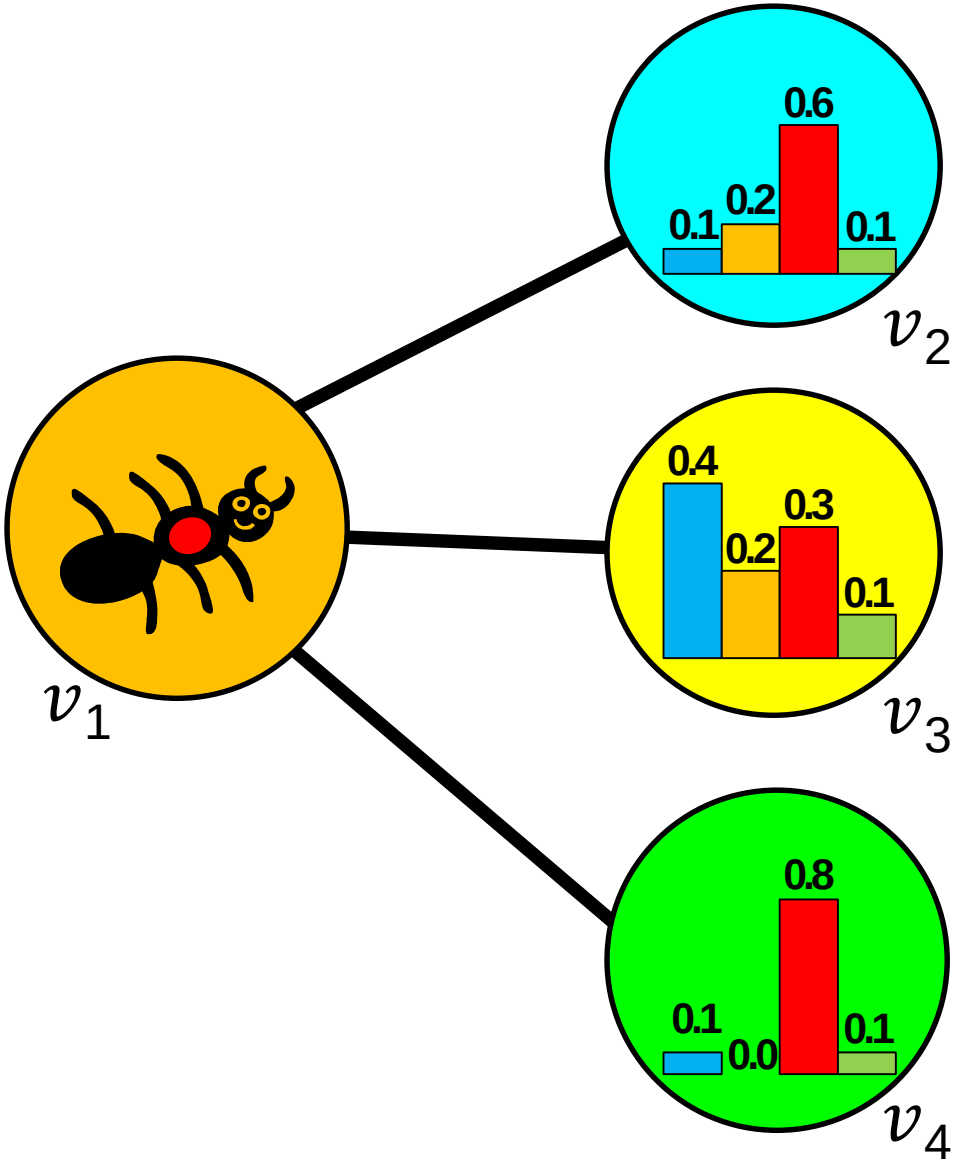
$$\rho_j^{d_k}(t+1) = \begin{cases} \rho_j^{d_i}(t) + 1 & \text{se } \rho_j^{d_i}(t) + 1 < \rho_j^{d_k}(t) \\ \rho_j^{d_k}(t) & \text{otherwise} \end{cases}$$

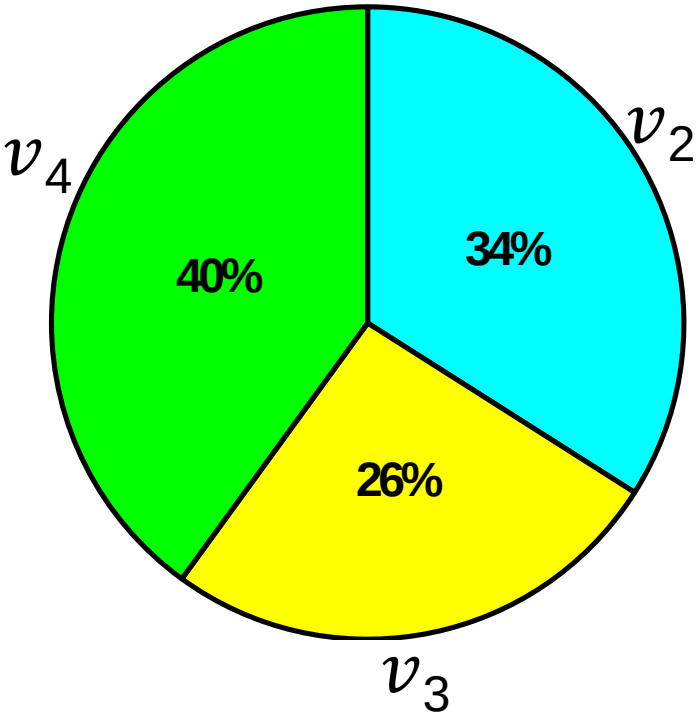
Particles Walk

- Random-greedy walk
  - Each particles randomly chooses a neighbor to visit at each iteration
  - Probabilities of being chosen are higher to neighbors which are:
    - Already dominated by the particle team.
    - Closer to particle initial node.

$$p(v_i|\rho_j) = \frac{W_{qi}}{2\sum_{\mu=1}^{n} W_{q\mu}} + \frac{W_{qi} v_i^{\omega_\ell}\left(1+\rho_j^{d_i}\right)^{-2}}{2\sum_{\mu=1}^{n} W_{q\mu}\, v_\mu^{\omega_\ell}\left(1+\rho_j^{d_\mu}\right)^{-2}}$$

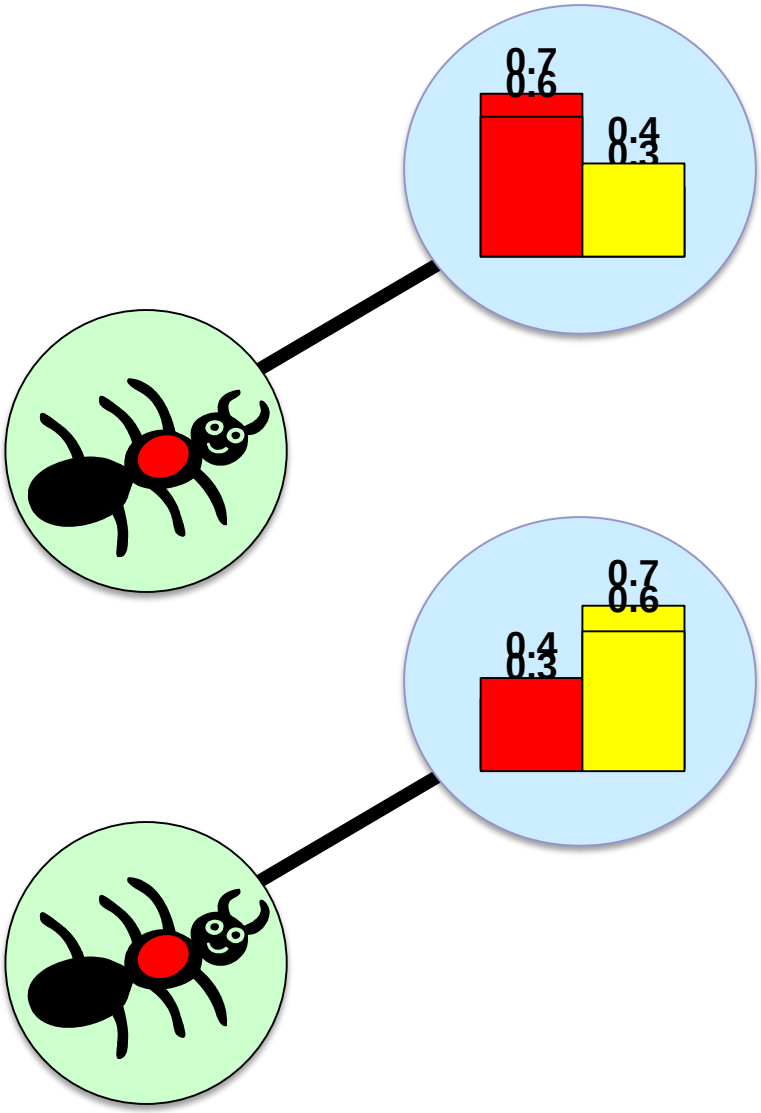
Moving Probabilities





Particles Walk

- Shocks
  - A particle really visits the selected node only if the domination level of its team is higher than others;
  - Otherwise, a shock happens and the particle stays at the current node until next iteration.



Label Query

- When the nodes domination levels reach a fair level of stability, the system chooses a unlabeled node and queries its label.
  - A new particle is created to this new labeled node.
  - The iterations resume until stability is reached again, then a new node will be chosen.
  - The process is repeated until the defined amount of labeled nodes is reached.

Query Rule

- There were two versions of the algorithm:
  - **AL-PCC v1**
  - **AL-PCC v2**
- They use different rules to select which node will be queried.

[15] F. Breve, “Active semi-supervised learning using particle competition and cooperation in networks,” in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, Aug 2013, pp. 1–6.

AL-PCC **v1**

- Selects the unlabeled node that the algorithm is most uncertain about which label it should have.
  - Node the algorithm has least confidence on the label it is currently assigning.
  - Uncertainty is calculated from the domination levels.

$$q(t) = \arg\max_{i, y=\emptyset} u_i(t)$$

$$u_i(t) = \frac{v_i^{\lambda_{\ell**}}(t)}{v_i^{\lambda_{\ell*}}(t)}$$

$$v_i^{\ell*}(t) = \arg\max_{\ell} v_i^{\ell}(t)$$

$$v_i^{\ell**}(t) = \arg\max_{\ell, \ell \neq v_i^{\ell*}(t)} v_i^{\ell}(t)$$

AL-PCC **v2**

- Alternates between:
  - Querying the most uncertain unlabeled network node (like AL-PPC v1)
  - Querying the unlabeled node which is more far away from any labeled node
    - According to the distances in the particles distance tables, dynamically built while they walk.

$$q(t) = \arg\max_i u_i(t)$$

$$u_i(t) = \frac{v_i^{\ell **}(t)}{v_i^{\ell *}(t)}$$

$$v_i^{\ell *}(t) = \arg\max_{\ell} v_i^{\ell}(t)$$

$$v_i^{\ell **}(t) = \arg\max_{\ell, \ell \neq v_i^{\ell *}(t)} v_i^{\ell}(t)$$

$$s_i(t) = \min_j \rho_j^{d_i}(t)$$

$$q(t) = \arg\max_i s_i(t)$$

The new Query Rule

- Combines both rules into a single one

$$q(t) = \arg\max_i \beta u_i'(t) + (1-\beta) s_i'(t)$$

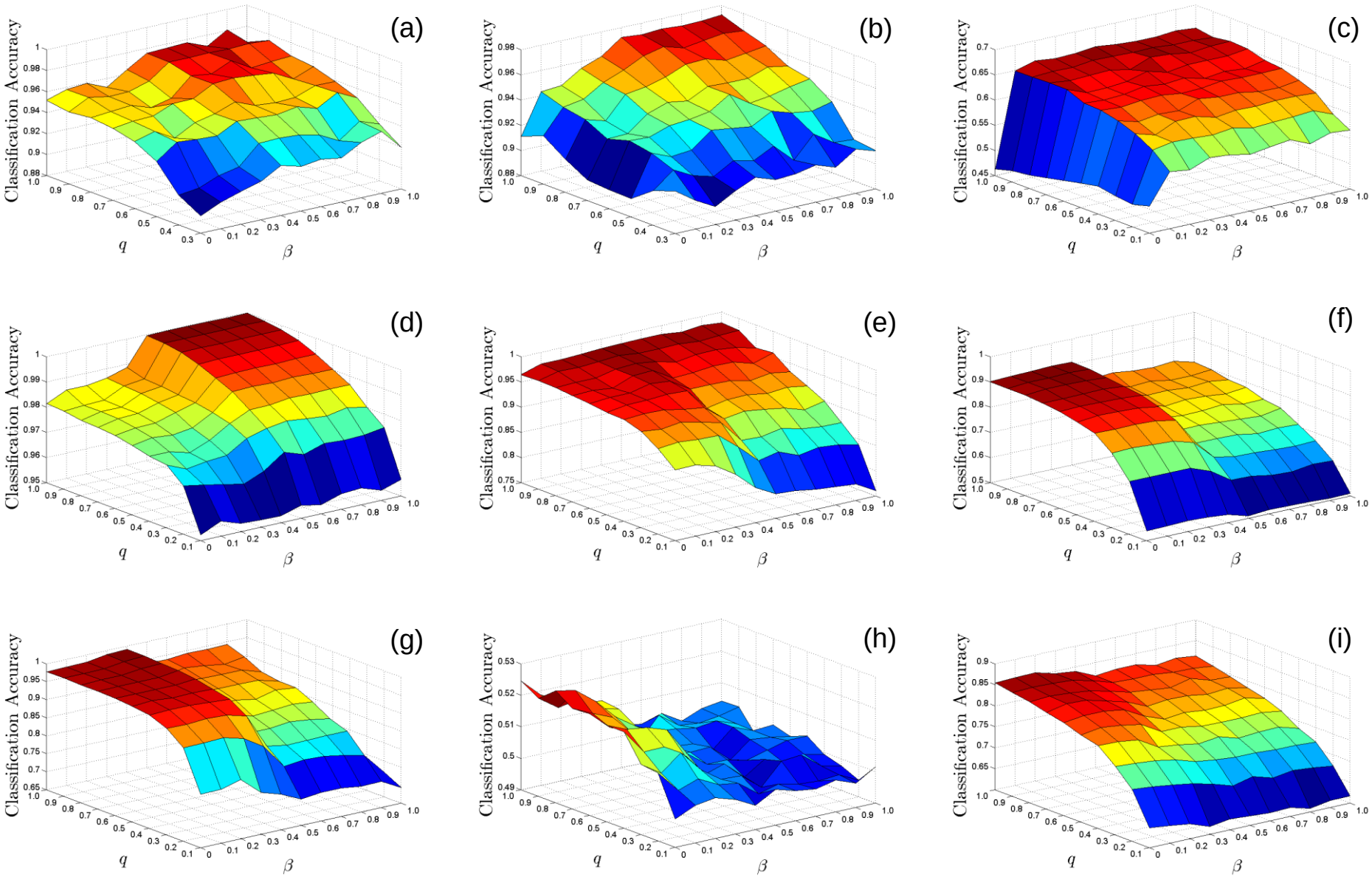
  - β define weights to the *assigned label uncertainty* and to the *distance to labeled nodes* criteria on the choice of the node to be queried.

Computer Simulations

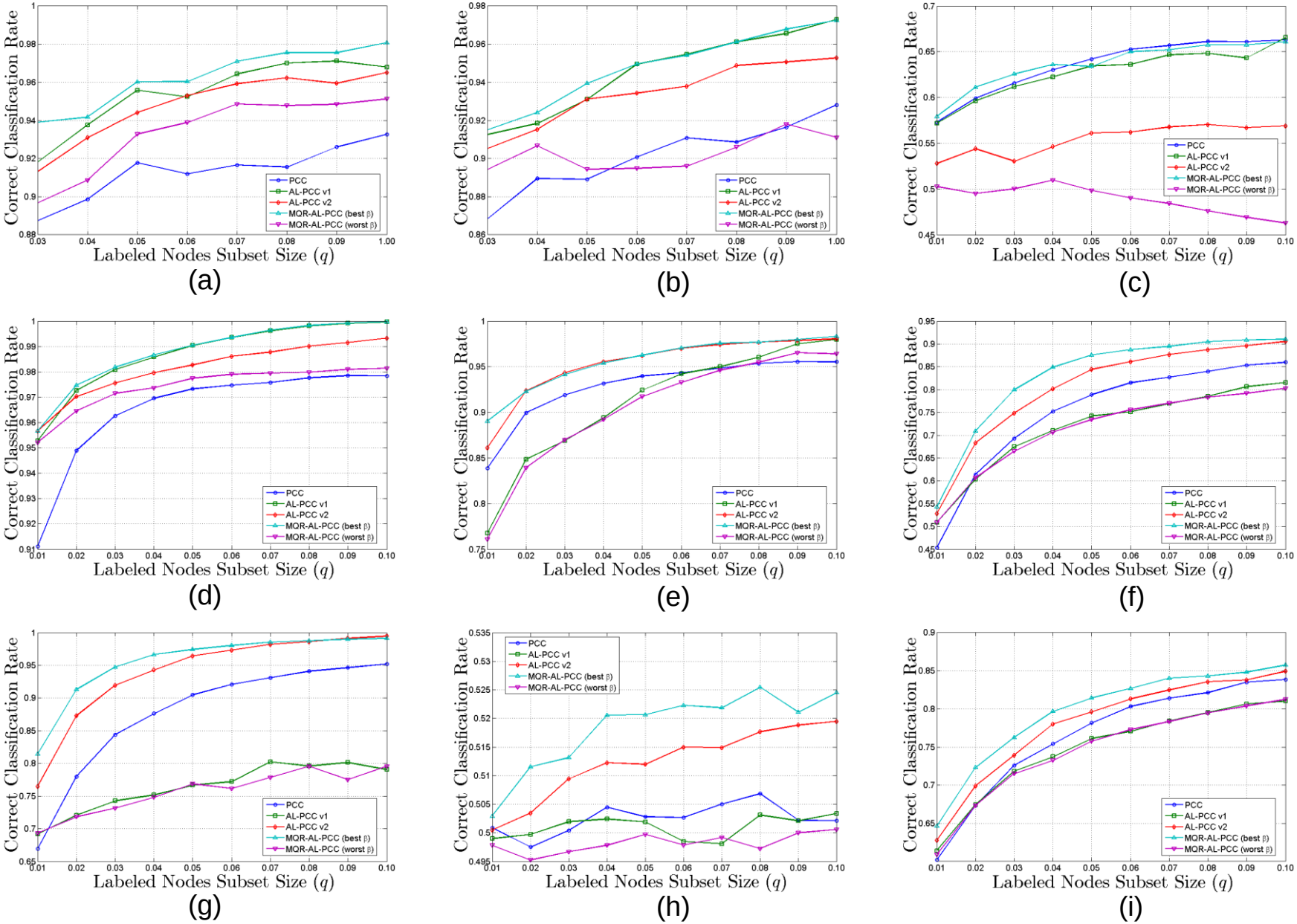
- 9 data different data sets
- $\beta = [0, 0.1, 0.2, \ldots, 1.0]$
- $k = 5$
- 1% to 10% labeled nodes
  - Starts with one labeled node per class, the remaining are queried
- All points are the average of 100 executions

Data Set	Classes	Dimensions	Points	Reference
Iris	3	4	150	[16]
Wine	3	13	178	[16]
g241c	2	241	1500	[2]
Digit1	2	241	1500	[2]
USPS	2	241	1500	[2]
COIL	6	241	1500	[2]
$COIL_2$	2	241	1500	[2]
BCI	2	117	400	[2]
Semeion Handwritten Digit	10	256	1593	[17,18]

[2] O. Chapelle, B. Schölkopf, and A. Zien, Eds., *Semi-Supervised Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, MA: The MIT Press, 2006.
[16] K. Bache and M. Lichman, “UCI machine learning repository,” 2013. [Online]. Available: http://archive.ics.uci.edu/ml
[17] Semeion Research Center of Sciences of Communication, via Sersale 117, 00128 Rome, Italy.
[18] Tattile Via Gaetano Donizetti, 1-3-5,25030 Mairano (Brescia), Italy.



Classification accuracy when the proposed method is applied to different data sets with different β parameter values and labeled data set sizes (q). The data sets are: (a) Iris [16], (b) Wine [16], (c) g241c [2], (d) Digit1 [2], (e) USPS [2], (f) COIL [2], (g) $COIL_2$ [2], (h) BCI [2], and (i) Semeion Handwritten Digit [17], [18]



Comparison of the classification accuracy when all the methods are applied to different data sets with different labeled data set sizes (q). The data sets are: (a) Iris [16], (b) Wine [16], (c) g241c [2], (d) Digit1 [2], (e) USPS [2], (f) COIL [2], (g) $COIL_2$ [2], (h) BCI [2], and (i) Semeion Handwritten Digit [17], [18]

Discussion

- Most data sets have some predilection for the query rule parameter
  - The thresholds, the effective ranges of β and the influence of a bad choice of β vary from one data set to another
    - Distance X Uncertainty criteria
      - May depend on data set properties
        - Data density
        - Classes separation
        - Etc.

Discussion

- **Distance** criterion is useful when...
  - Classes have highly overlapped regions, many outliers, more than one cluster inside a single class, etc.
    - Uncertainty wouldn’t detect large regions of the network completely dominated by the wrong team of particles
      - Due to an outlier or the lack of correctly labeled nodes in that area
- **Uncertainty** criteria is useful when...
  - Classes are fairly well separated and there are not many outliers.
    - Less particles to take care of large regions
    - Thus new particles may help finding the classes boundaries.

Conclusions

- The computer simulations show how the different choices of query rules affect the classification accuracy of the active semi-supervised learning particle competition and cooperation method applied to different real-world data sets.
- The optimal choice of the newly introduced β parameter led to better classification accuracy in most scenarios.
- **Future work**: find possible correlation between information that can be extracted from the network a priori and the optimal β parameter, so it could be selected automatically.

The Brazilian Conference on Intelligent Systems (BRACIS) and Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)

Query Rules Study on Active Semi-Supervised Learning using Particle Competition and Cooperation

Fabricio Breve fabricio@rc.unesp.br

Department of Statistics, Applied Mathematics and Computation (DEMAC), Institute of Geosciences and Exact Sciences (IGCE), São Paulo State University (UNESP), Rio Claro, SP, Brazil