Hume-Nash Machines

Context-Aware Models of Learning and Recognition

Rocco Tripodi, Marcello Pelillo {rocco.tripodi,pelillo}@unive.it
ECLT – European Centre for Living Technology - Ca' Foscari University of Venice

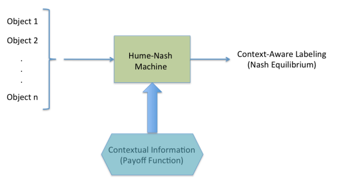
Introduction

With the spectacular growth and the ubiquity of network data, classical (feature-based) approaches to machine learning and pattern recognition are no longer viable.
More sophisticated **"context-aware" approaches**, which exploit local environmental information, are needed in order for next-generation computers to cope with the increasing complexity of real-world applications.
In particular, this is of the utmost importance in real-world scenarios involving several agents interacting in a complex environment using multiple cues and modalities.

Methodology

We propose **Hume-Nash machines**, a context-aware classification model based on:

- ✓ the use of similarity principles which go back to the work of British philosopher David Hume ("assign *similar* labels to *similar* objects")
- ✓ game-theoretic equilibrium analysis introduced by Nobel laureate John Nash ("no incentive to unilaterally deviate from equilibrium")



The intuition is to view learning problems as non-cooperative games, whereby the competition between the hypotheses of class membership is driven by contextual and similarity information encoded as payoff functions [1].

According to this perspective, the focus will shift from optima of objective functions to equilibria of (non-cooperative) games.
The learning algorithms is based on **evolutionary game dynamics** [2], a large, thoroughly-studied class of dynamical systems. **This may open up an exciting era where the theory of learning and classification is combined with (nonlinear) dynamical systems theory.**

Example Setup

Consider:

- ✓ A set of objects $B = \{b_1, \dots, b_n\}$
- ✓ A set of labels $\Lambda = \{1, \dots, m\}$

The goal is to label each object of B with a label of Λ.
To this end, two sources of information are exploited:

- ✓ Local measurements which capture the salient features of each object viewed in isolation
- ✓ Contextual and similarity information, expressed in terms of a payoff matrix of $R = \{r_{ij}(\lambda,\mu)\}$.

Context-Aware classification as a Non-Cooperative Game

We shall formulate the context-aware classification problem as a non-cooperative game, where:

- ✓ Objects to be labeled = players
- ✓ Class labels = pure strategies
- ✓ Weighted labeling assignments = mixed strategies
- ✓ Contextual constraints = payoff function R

The coefficient $r_{ij}(\lambda,\mu)$ measures the strength of compatibility between the two hypotheses: "b_i is labeled λ" and "b_j is labeled μ", or in other words the payoff that players get when player i plays strategy λ and player j plays strategy μ.
According to this formulation, context-aware classification = **Nash equilibrium**
Payoff function:

$$q_i(\lambda, p_i) = \sum_{j=1}^{n}\sum_{\mu=1}^{m} r_{ij}(\lambda,\mu)\, p_j(\mu)$$

Where p is a probability distribution that denotes, for each player i, the probability that it chooses a determined strategy.
System dynamics:

$$p_i(\lambda)^{t+1} = \frac{p_i(\lambda)^t q_i(\lambda)}{\sum_{\mu=1}^{m} p_i(\mu)^t q_i(\mu)}$$

In this way better than average strategies grow at each iteration.

Word Sense Disambiguation and Entity Linking [3]

Word Sense Disambiguation is the task of identifying the intended meaning of a word based on the **context** in which it appears. It has been studied since the beginnings of Natural Language Processing and today it is still a central topic of this discipline. This because it is important for many NLP tasks such as **text understanding**, **text entailment**, **machine translation**, **opinion mining**, **sentiment analysis** and **information extraction**. All these applications can benefit from the disambiguation of ambiguous words, as a preliminary process; otherwise they remain on the surface of the word, compromising the **coherence** of the data to be analyzed.
We modeled:

- ✓ The words to be disambiguated as the **players** of the games
- ✓ The senses of the words as the **strategies** that the players can choose
- ✓ The **interactions** among the players as a word similarity graph
- ✓ The **payoff** function as a sense similarity function
- ✓ We used the **replicator dynamics equation** to compute the Nash equilibria of the games.

In this way, it is possible to maintain the textual coherence associating each word to the most appropriate sense according to the senses that other words in the text are choosing.
The system was validate against unsupervised, semi-supervised and supervised algorithms. We used four datasets for the WSD task and on each of them we obtained higher or comparable results compared to state-of-the-art systems. For the **Entity Linking** task we validated our system on two datasets and also on this task we obtained higher or comparable results.

		S7CG	S7CG (N)	S7	S3	S2
unsup.	*Nav10*	–	–	43.1	52.9	–
	PPR_{w2w}	80.1	83.6	41.7	57.9	59.7
	WSD_{games}	**80.4***	**85.5**	**43.3**	**59.1**	**61.2**
semi sup.	*IRST-DDD-00*	–	–	–	58.3	–
	MFS	76.3	77.4	54.7	62.8	65.6*
	MRF-LP	–	–	50.6*	58.6	60.5
	Nav05	**83.2**	84.1	–	60.4	–
	PPR_{w2w}	81.4	82.1	48.6	63.0	62.6
	WSD_{games}	82.8	**85.4**	**56.5**	**64.7***	**66.0**
sup.	*Best*	82.5	82.3*	**59.1**	65.2	**68.6**
	Zhong10	**82.6**	–	58.3	**67.6**	68.2

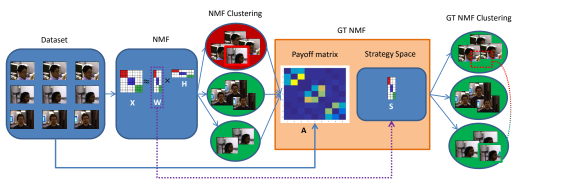
WSD

	S13	KORE50
WSD_{games}	**70.8**	**75.7**
Babelfy	69.2	71.5
SUDOKU	66.3	–
MFS	66.5*	–
PPR_{w2w}	60.8	–
KORE	–	63.9*
GETALP	58.3	–

Entity Linking

Context Aware Nonnegative Matrix Factorization Clustering [4]

We propose a method to **refine** the clustering results obtained with the nonnegative matrix factorization technique, imposing **consistency constraints** on the final labeling of the data.



Normalized Mutual Information

Dataset	SymNMF	SymNMF+GT	NMF	NMF+GT	NMFS	NMFS+GT	NNDSVD	NNDSVD+GT
NIPS ★	0.385 (±0.011)	**0.405** (±0.016)	0.375 (±0.022)	**0.386** (±0.011)	0.388 (±0.006)	**0.403** (±0.007)	0.388	**0.399**
NIPS	0.387 (±0.007)	**0.418** (±0.017)	0.401 (±0.016)	**0.406** (±0.016)	0.393 (±0.008)	**0.412** (±0.018)	0.388	**0.421**
Reuters ★	0.502 (±0.014)	**0.51** (±0.016)	0.451 (±0.026)	**0.49** (±0.02)	0.505 (±0.014)	**0.511** (±0.015)	**0.427**	0.425
Reuters	0.517 (±0.007)	**0.525** (±0.006)	0.442 (±0.006)	**0.497** (±0.003)	0.518 (±0.006)	**0.527** (±0.005)	0.488	**0.493**
RCV1 ★	0.406 (±0.007)	**0.422** (±0.007)	0.51 (±0.007)	**0.516** (±0.005)	0.404 (±0.009)	**0.42** (±0.01)	**0.403**	0.402
RCV1	0.411 (±0.006)	**0.422** (±0.006)	0.462 (±0.009)	**0.483** (±0.007)	0.413 (±0.006)	**0.424** (±0.006)	0.398	**0.407**
PIE-Expr	0.95 (±0.004)	**0.968** (±0.004)	0.939 (±0.008)	**0.959** (±0.006)	0.89 (±0.005)	**0.931** (±0.006)	0.86	**0.889**
ORL	0.888 (±0.006)	**0.921** (±0.006)	0.691 (±0.015)	**0.844** (±0.014)	0.889 (±0.006)	**0.918** (±0.004)	0.808	**0.892**
COIL-20	0.871 (±0.009)	**0.875** (±0.012)	0.619 (±0.017)	**0.669** (±0.016)	0.877 (±0.013)	**0.883** (±0.013)	0.824	**0.836**
ExtYaleB	0.308 (±0.005)	**0.313** (±0.005)	**0.356** (±0.006)	0.355(±0.007)	0.309 (±0.007)	**0.314** (±0.005)	0.288	**0.315**

Accuracy

Dataset	SymNMF	SymNMF+GT	NMF	NMF+GT	NMFS	NMFS+GT	NNDSVD	NNDSVD+GT
NIPS ★	0.462 (±0.013)	**0.483** (±0.016)	**0.426** (±0.02)	0.425(±0.014)	0.465 (±0.011)	**0.485** (±0.017)	0.474	**0.509**
NIPS	0.379 (±0.01)	**0.415** (±0.037)	**0.396** (±0.022)	0.39(±0.018)	0.384 (±0.014)	**0.407** (±0.031)	0.466	**0.503**
Reuters ★	0.517 (±0.044)	**0.528** (±0.043)	0.322 (±0.024)	**0.401** (±0.026)	0.516 (±0.037)	**0.525** (±0.037)	0.403	**0.427**
Reuters	0.324 (±0.029)	**0.363** (±0.032)	0.222 (±0.011)	**0.282** (±0.02)	0.339 (±0.023)	**0.378** (±0.024)	0.277	**0.339**
RCV1 ★	**0.292** (±0.015)	0.289(±0.014)	0.383 (±0.009)	**0.387** (±0.01)	**0.298** (±0.017)	0.297(±0.017)	**0.285**	0.276
RCV1	0.243 (±0.008)	**0.247** (±0.008)	0.279 (±0.01)	**0.295** (±0.011)	0.242 (±0.01)	**0.245** (±0.011)	0.239	**0.24**
PIE-Expr	0.81 (±0.021)	**0.85** (±0.019)	0.783 (±0.023)	**0.809** (±0.024)	0.617 (±0.019)	**0.7** (±0.02)	**0.536**	0.513
ORL	0.776 (±0.017)	**0.811** (±0.018)	0.465 (±0.019)	**0.608** (±0.026)	0.77 (±0.013)	**0.804** (±0.015)	0.653	**0.71**
COIL-20	0.727 (±0.036)	**0.729** (±0.037)	0.478 (±0.023)	**0.507** (±0.025)	0.739 (±0.046)	**0.741** (±0.046)	**0.674**	0.672
ExtYaleB	**0.235** (±0.008)	0.228(±0.007)	0.194 (±0.007)	**0.197** (±0.009)	**0.237** (±0.012)	0.23(±0.01)	0.229	**0.242**

Document Clustering Games in Dynamic Scenarios [5]

Document clustering is a particular kind of clustering that involves textual data. The objects to be clustered can have different characteristics, varying in length and content. Popular applications of document clustering aims at organizing **tweets**, **news**, **novels** and **medical documents**. It is a fundamental task in text mining, with different applications that span from document organization to language modeling. Clustering algorithms tailored for this task are based on **generative models**, **graph models** and **matrix factorization** techniques. A general problem, common to all these approaches, involves the **temporal dimension**. In fact, for these approaches is difficult to deal with **datasets that evolve over time** and in many real world applications documents are **streamed continuously**. Furthermore, this problem can be more severe in case of **huge datasets**, because of scalability issues.
With our approach we overcome these problems, in fact, we can classify new objects according to the information on previous clusterings. The problem of clustering new objects is defined as a game, in which we have **labeled players** (clustered objects), which always play the strategy associated to their cluster and **unlabeled players** which try to learn their strategy according to the strategy that their co-players are choosing. The geometry of the data is modeled as a similarity graph, whose nodes are players (documents), and the games are played only among similar players. In this way new data points can be clustered at different times and their classification adapt to the structure of the whole dataset. The validation of the algorithm has been conducted on 12-folds datasets and as we can see on the table below, the accuracy of the system is constant on each fold.

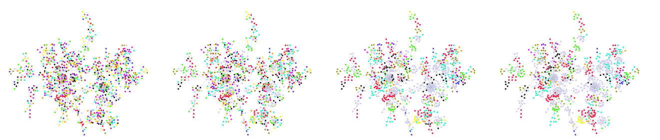
Data	2	3	4	5	6	7	8	9	10	11	12
ng3sim	0.85 ± 0.03	0.84 ± 0.03	0.84 ± 0.02	0.84 ± 0.01	0.84 ± 0.02	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01
classic	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.00	0.94 ± 0.00	0.94 ± 0.00	0.94 ± 0.00	0.94 ± 0.00	0.94 ± 0.00	0.94 ± 0.00
k1b	0.94 ± 0.02	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01
hitech	0.61 ± 0.04	0.61 ± 0.03	0.61 ± 0.03	0.61 ± 0.03	0.60 ± 0.02	0.60 ± 0.02	0.60 ± 0.02	0.60 ± 0.02	0.60 ± 0.02	0.60 ± 0.02	0.60 ± 0.02
reviews	0.92 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01
sports	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01
la1	0.82 ± 0.03	0.82 ± 0.03	0.82 ± 0.02	0.82 ± 0.02	0.82 ± 0.02	0.82 ± 0.02	0.82 ± 0.02	0.82 ± 0.02	0.82 ± 0.01	0.82 ± 0.01	0.82 ± 0.01
la12	0.85 ± 0.02	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.00	0.84 ± 0.01	0.84 ± 0.00	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.00
la2	0.83 ± 0.02	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01
tr11	0.72 ± 0.07	0.72 ± 0.08	0.71 ± 0.08	0.70 ± 0.07	0.69 ± 0.07	0.69 ± 0.06	0.69 ± 0.06	0.69 ± 0.06	0.69 ± 0.06	0.69 ± 0.06	0.69 ± 0.06
tr23	0.73 ± 0.08	0.71 ± 0.08	0.69 ± 0.08	0.69 ± 0.07	0.69 ± 0.07	0.68 ± 0.07	0.68 ± 0.07	0.68 ± 0.07	0.68 ± 0.07	0.68 ± 0.07	0.68 ± 0.07
tr41	0.90 ± 0.04	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.03	0.90 ± 0.02	0.90 ± 0.02	0.90 ± 0.02	0.90 ± 0.02	0.90 ± 0.02	0.90 ± 0.02	0.90 ± 0.02
tr45	0.80 ± 0.04	0.81 ± 0.04	0.82 ± 0.04	0.82 ± 0.04	0.82 ± 0.04	0.82 ± 0.03	0.82 ± 0.04	0.82 ± 0.03	0.82 ± 0.03	0.82 ± 0.04	0.82 ± 0.04

The Dynamics of Opinion Diffusion in Social Networks [6]

*"That **social influences** shape every person's practices, judgments and beliefs is a truism to which anyone will readily assent. A child master his native dialect down to the finest nuances; a member of a tribe accepts cannibalism as altogether fitting and proper. All the social sciences take their departure from the observation of the profound effects that groups exert on their members."* (Solomon E. Asch, Opinions and Social Pressure, 1955).

The exponential growth in the popularity of the online social networks such as **Facebook**, **Twitter** has led to a lot of renewed research in understanding basic sociological phenomena such as **opinion and consensus formation**. Psychologists discovered that people tend to adapt their opinions to their social environment trying to minimize the divergence among their opinions and the one of their friends.
How the opinion on a particular topic can influence the opinions of other people in a social network?
We modeled:

- ✓ The players of the games as the users on the network.
- ✓ The strategies of the games as opinions
- ✓ The payoff of the games as opinion similarity
- ✓ The interactions among the players as a weighted similarity graph



Iteration 1 Iteration 5 Iteration 20 Iteration 40

This model will be tested on real world datasets to test its prediction power. The datasets will be constructed using the **reviews** posted by users on **Amazon**, **Reddit** and **Tripadvisor**. In this way it is possible to analyze the reviews and to measure the influence that **positive** and **negative** reviews have on the diffusion of the opinions, about a particular **product**, on the underling social network.

References

1. Pelillo, M. "The dynamics of nonlinear relaxation labeling processes." Journal of Mathematical Imaging and Vision 7.4 (1997): 309-323.
2. Weibull, J. W. 1997. *Evolutionary game theory*. MIT press.
3. Tripodi R; Pelillo M, A Game-Theoretic Approach to Word Sense Disambiguation, Computational Linguistics, MIT Press. (in press)
4. Tripodi R.; Vascon S; Pelillo M, Context Aware Nonnegative Matrix Factorization Clustering, ICPR 2016. (under review)
5. Tripodi R; Pelillo M, Document Clustering Games, In Proceedings of ICPRAM 2016, Rome, 2016. Tripodi.
6. Tripodi R; Pelillo M, Evolutionary Aspects of Network and Opinions, presented at D2NetLang, Lyon, 2016.

Acknowledgment

This work was partially supported by Samsung Global Research Outreach Program.