Identifying patterns of human behavior:

an analysis on experimental data of the Public Goods Game

Ferran Español Casanovas

Universitat de Barcelona

Master Thesis

12 September 2018

Table of Contents

Table of Contents

Chapter 1

Objectives

- Application of Machine Learning (ML) techniques for the identification of patterns in a Public Goods Game.
- Evaluation of Unsupervised and Supervised learning algorithms using experimental game theoretical data.
- In more general terms, contribute to the debate about Collective Risk Dilemmas.
- Questions we want to answer:
  - Which kind of ML tools are potentially good?
  - Why?
  - And what for?

Game Theoretical approach

- "Theory of Games and Economic Behavior" of John von Newmann & Oskar Morgenstern in 1944.
- Progression: 50 articles published annually at 1982 to 200 at 1998.
- Game Theory structure: agents, actions & payoff function.
- Keystone concepts: Equilibrium & Generalizability.

Public Goods Game

- Reference: John O. Ledyard - Public Goods: A Survey of Experimental Research (1995).
- Game-Theoretical prediction & Sociologic-Psycologic prediction.
- 40% of the endowment invested to public goods.
- Traditional strategies: Free-Riding, Conditional Cooperation or Altruistic.

Collective Risk Dilemma

- A collective has to solve a risk achieving certain goal cooperating among them.
- Individuals do not know a priori which could be better to have the maximum profit.
- This schema fits very well in the analysis of the political relations to paliate climate change.
- Treatment groups: Homogeneity, Endowment Heterogeneity, Loss Heterogeneity.
- Open Question: Which context benefits cooperation and the success rate in these games.

References CRD

- Milinski, Manfed et al. (2008). "The collective-risk social dilemma and the prevention of simulated dangerous climate change". In: *Proceedings of the National Academy of Sciences*.
- Milinski, Manfred, Torsten Röhl, and Jochem Marotzke (2011). "Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets". In: *Climatic Change*.
- Tavoni, Alessandro et al. (2011). "Inequality, communication, and the avoidance of disastrous climate change in a public goods game". In: *Proceedings of the National Academy of Sciences*.
- Burton-Chellew, Maxwell N., Robert M. May, and Stuart A.West (2013). "Combined inequality in wealth and risk leads to disaster in the climate change game". In: *Climatic Change*.
- Waichman, Israel et al. (2018). "Challenging conventional wisdom: Experimental evidence on heterogeneity and coordination in avoiding a collective catastrophic event". In: *Economics Working Paper*

Sample size

Table: Total number of participants in related literature

Publication	# of participants
Milinski et al. 2008	180
Milinski et al. 2011	342
Burton-Chellew et al. 2013	192
Brown & Kroll 2017	378
Waichman et al. 2018	510
Current work	612

Data

- Three experiments: DAU (324 participants), STREET (108 participants) & VIL (180 participants).
- Contributions at each round and socio-demographic information.
- Average age: between 25 and 35 years old.
- Academic background: vocational school (30-50%).
- Two treatment groups: Homogeneous & Endowment Heterogeneity.

Chapter 2

2 Methodology
- Experimental procedure
- Normalization
- Principal Component Analysis
- Optimal K

Experimental design

- "Lab in Field" experimentation (Sagarra et al. 2016)
- General audiences participates in the generation of data.
- The objective of this experimental procedure is to avoid the problems of generality and bias of the sample.

Normalization

Table: Normalized contributions per round according selection

Initial Endowment	0	2	4
20	0	1.00	2.00
30	0	0.67	1.33
40	0	0.50	1.00
50	0	0.40	0.80
60	0	0.33	0.67

Principal Component Analysis

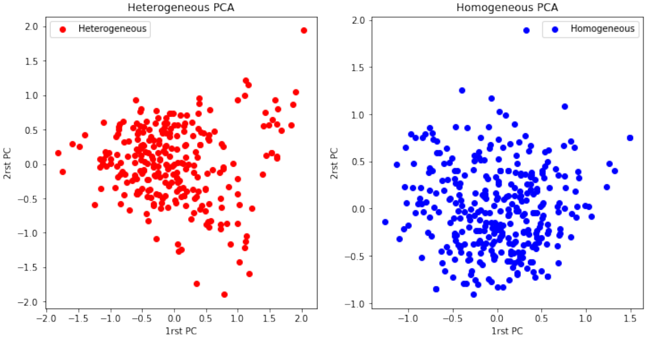


Figure: Scatter plot of the two first principal components for both treatments (the left for heterogeneous games and the right for homogeneous ones

Optimal number of clusters

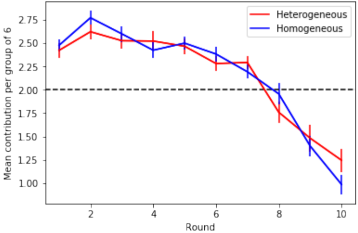
Table: The optimal number of clusters for each criteria

Dataset	NbClust	GAP	Cal & Hara	Krz & Lai	Hartigan	Silhouette
Heterogeneous	3	2 (-0.68)	2 (58.87)	6 (11.86)	3 (19.62)	2 (0.22)
Homogeneous	3	2 (-1.63)	2 (54.55)	4 (7.16)	3 (15.51)	2 (0.13)
Het DAU	3	2 (-0.51)	2 (38.17)	3 (14.71)	3 (8.40)	2 (0.23)
STREET	3	2 (-0.94)	2 (22.34)	8 (4.85)	3 (7.41)	3 (0.18)
Hom DAU	4	2 (-1.47)	2 (30.32)	4 (16.46)	4 (11.04)	2 (0.15)
VIL	3	2 (-1.75)	2 (28.02)	10 (3.90)	3 (14.71)	2 (0.12)

Chapter 3

3. Exploratory Analysis - Game Theoretical Results
   - Average Contribution per Round
   - Accumulated Average Contribution per Round
   - Total Contribution Ratio
   - Contribution according the endowment
   - Composition of groups
   - Inequality - Gini Coefficient

Average Contribution per Round



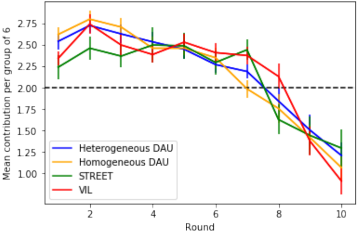
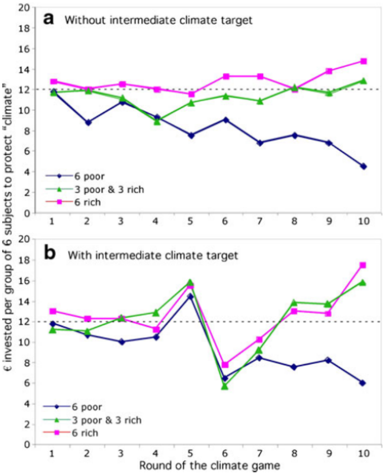


Figure: The normalized average contribution per group of 6 with the standard error. Dotted line represents the fair average contribution.(Left) Treatment level. (Right) Dataset level.

Average Contribution (Milinski et al. 2011)



Accumulated Average Contribution per Round

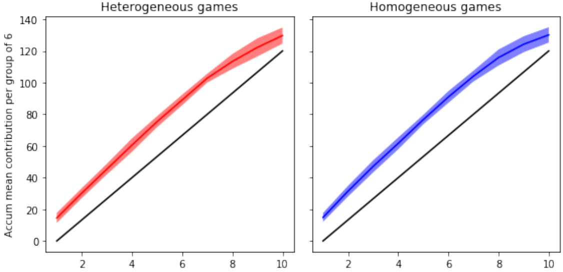


Figure: Average accumulated contribution for each treatment, the straight line shows the fair accumulated contribution and the shadow represents the standard deviation. (Left) Heterogeneous games. (Right) Homogeneous games.

Accumulated Average Contribution (Milinski et al. 2008)

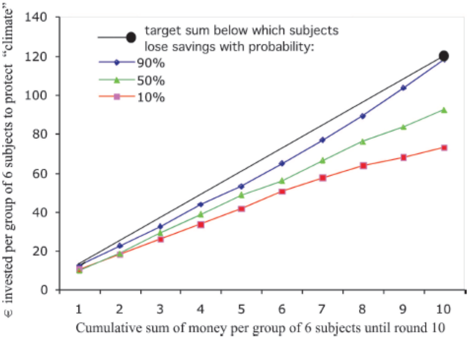


Fig. 2. Cumulative sum of money per group and round provided for the climate account. The target sum to be achieved after 10 rounds was €120; the treatments differed in the probability, i.e., 90%, 50%, and 10%, with which all subjects in a group lost their individual savings when the group did not supply the target sum for the climate account.

Total Contribution Ratio

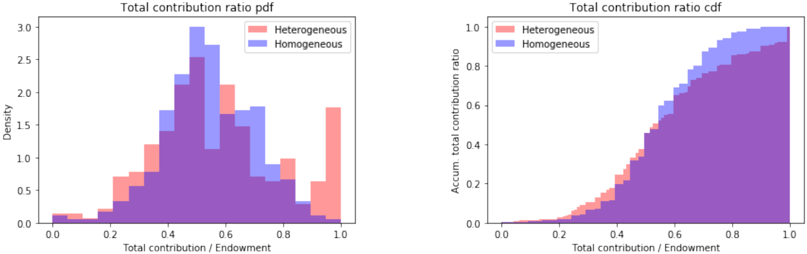
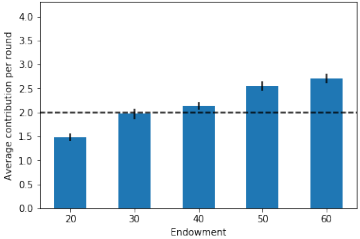


Figure: (Left) PDF and (Right) CDF of the TCR distribution according the treatment group.

Contribution according the endowment



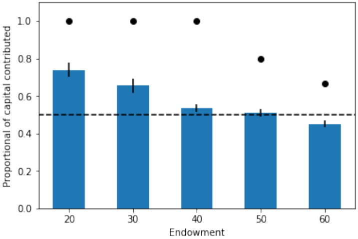


Figure: (Left) Average endowment contributed with standard error. (Right)Average proportional endowment contributed with the standard error. Dots lines represent the fair average selection.

Contribution according the endowment

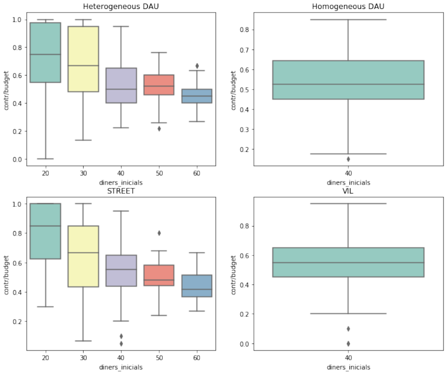


Figure: Boxplot with the average proportion of endowment contributed for each initial endowment per each dataset.

Composition of groups

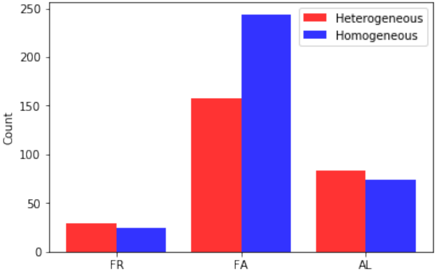


Figure: This bar plot represents the number of users of each category (free-rider, fair, altruist) according the treatment

Gini Coefficient

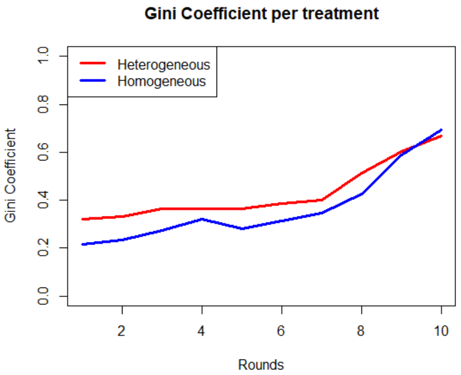


Figure: Evolution of the Gini coefficient for both treatments.

Chapter 4

Hierarchical Clustering

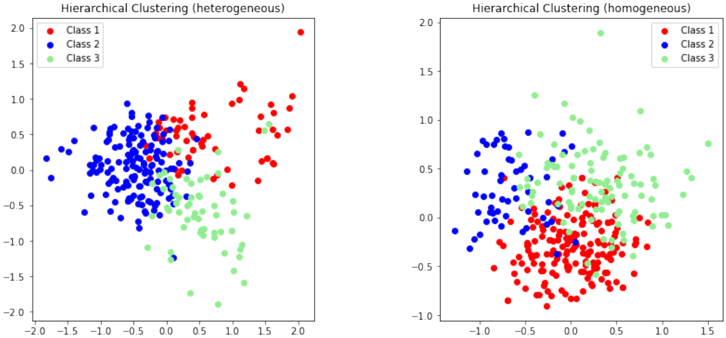


Figure: Clustering results on the PC plane. Hierarchical Clustering results for (Left) Heterogeneous and (Right) Homogeneous games.

Agglomerative Clustering

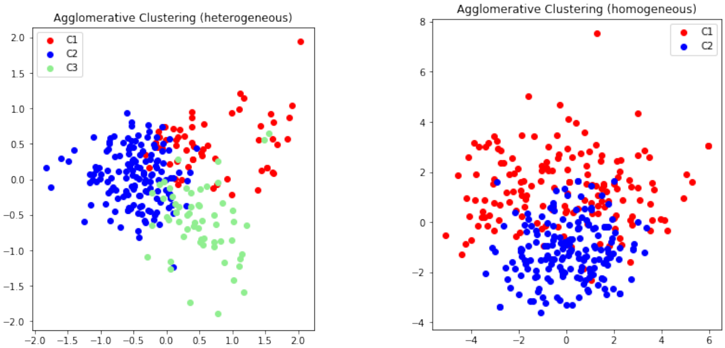


Figure: Clustering results on the PC plane. Agglomerative Clustering results for (Left) Heterogeneous and (Right) Homogeneous games.

K-Means

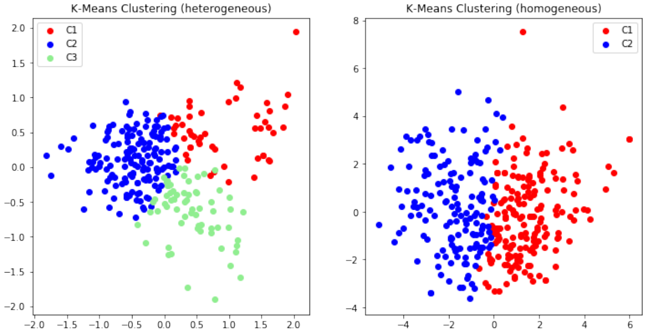


Figure: Clustering results on the PC plane. K-Means Clustering results for (Left) Heterogeneous and (Right) Homogeneous games.

Clustering results - Statistical Information

Heterogeneous:

- Cluster 1: 51 participants (18.89%) with an average contribution of 2.24 $\pm$0.56 MU (0.79 $\pm$0.17 MU).
- Cluster 2: 152 participants (56.30%) with an average contribution of 1.90 $\pm$0.75 MU (0.42 $\pm$0.12 MU).
- Cluster 3: 67 participants (24.81%) with an average contribution of 2.70 $\pm$0.80 MU (0.75 $\pm$0.14 MU).

Homogeneous:

- Cluster 1: 186 participants (54.39%) with an average contribution of 1.76 $\pm$0.43 MU (0.44 $\pm$0.11 MU).
- Cluster 2: 156 participants (45.61%) with an average contribution of 2.66 $\pm$0.40 MU (0.66 $\pm$0.10 MU).

Initial endowment per cluster (Heterogeneous)

Figure: Percentage barplot with the percentage of population from each cluster and initial endowment.

Percentage of population per dataset (Homogeneous)

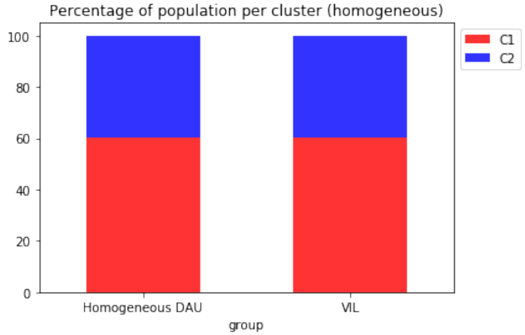
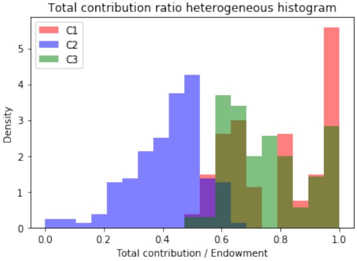


Figure: Proportional population assigned to each cluster for homogeneous DAU and VIL datasets.

Total Contribution Ratio per cluster



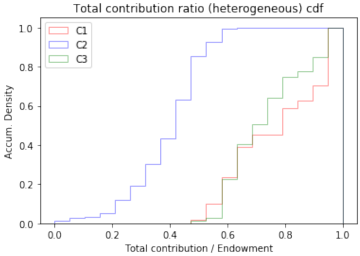


Figure: (Left) PDF and (Right) CDF of the TCR distribution per cluster. Heterogeneous games.

Total Contribution Ratio per cluster

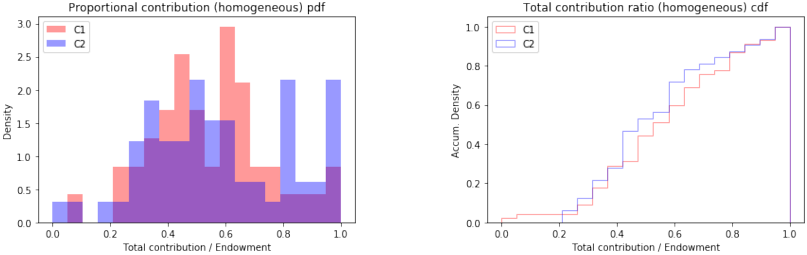


Figure: (Left) PDF and (Right) CDF of the TCR distribution per cluster. Homogeneous games.

Evolution of clusters

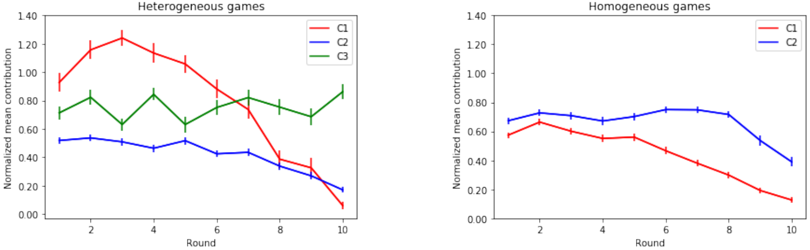


Figure: Evolution of the normalized average contribution with the SE.

Ending Round Effects for heterogeneous games

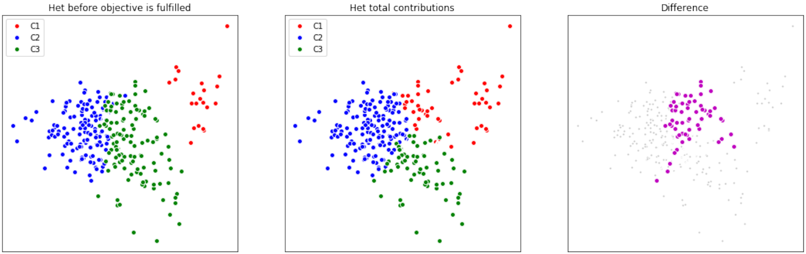


Figure: Individuals in the PC plane colored according their clusters before and after the objective is fulfilled (heterogeneous)

Ending Round Effects for homogeneous games

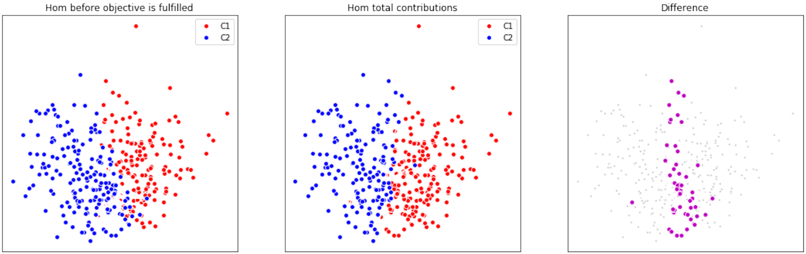


Figure: Individuals in the PC plane colored according their clusters before before and after the objective is fulfilled (homogeneous).

Initial endowment (before the objective is fulfilled)

Figure: Percentage stacked bar plot of members of each cluster according their initial endowment.

Total Contribution Ratio

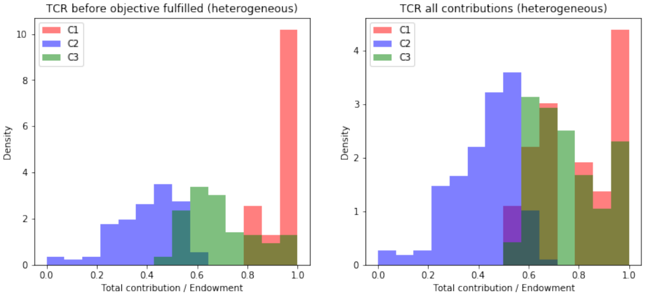


Figure: Histograms with the TCR for each cluster (Left) before and (Right) after the objtective is fulfilled.

Total Contribution Ratio

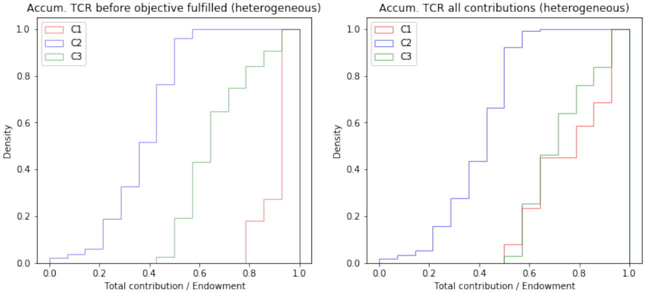


Figure: Cdf of the total contribution ratio for each cluster (Left) before and (Right) after the objtective is fulfilled.

Classification results

Table: Results for classifications algorithms

Dataset	LogReg	DecTree	KNN	LDA	GNB	SVM
Heterogeneous	0.519	0.510	0.551	0.510	0.528	0.579
Homogeneous	0.678	0.572	0.608	0.670	0.652	0.681

Classification report

Table: Results for classifications report

Dataset	Best Classifier	av. Precision	av. Recall	f1-score
Heterogeneous	SVM	0.33	0.57	0.42
Homogeneous	SVM	0.48	0.70	0.57

Confusion matrix

Heterogeneous

$$\begin{array}{c} \\ class_m \\ class_f \end{array} \begin{array}{c} \begin{array}{cc} pred_m & pred_f \end{array} \\ \begin{pmatrix} 31 & 0 \\ 23 & 0 \end{pmatrix} \end{array}$$

Homogeneous

$$\begin{array}{c} \\ class_m \\ class_f \end{array} \begin{array}{c} \begin{array}{cc} pred_m & pred_f \end{array} \\ \begin{pmatrix} 48 & 0 \\ 21 & 0 \end{pmatrix} \end{array}$$

Figure: Confusion matrix for SVM classifier for (Left) Heterogeneous (Right) Homogeneous.

Confusion matrix

LogReg (heterogeneous)

$$\begin{array}{c} \\ class_m \\ class_f \end{array} \begin{array}{cc} pred_m & pred_f \\ \left(\begin{array}{cc} 27 & 4 \\ 17 & 6 \end{array}\right) \end{array}$$

K-NN (homogeneous)

$$\begin{array}{c} \\ class_m \\ class_f \end{array} \begin{array}{cc} pred_m & pred_f \\ \left(\begin{array}{cc} 38 & 10 \\ 17 & 4 \end{array}\right) \end{array}$$

Figure: Confusion matrix for (Left) Logistic Regression classifier (Heterogeneous) and (Right) K-Nearest Neighbors classifier (Homogeneous).

Chapter 5

5 Discussion
- Clustering Discussion
- Game Theoretical Discussion

Clustering Discussion

- We conclude that ML techniques are useful to study experimental data on CRD.
- Unsupervised Learning has identified consistent groups based only on the contributions of the participants.
- Still, we have a little amount of data to work properly with Supervised Learning techniques.

Game Theoretical Discussion

- We have not found significant differences between both treatments in terms of succes rate (100% in our case) or the total contribution ratio.
- We detected imbalances in the heterogeneous groups. Contrary to Milinski et al. 2008 and Waichmann et al. 2018 in our case individuals with high initial endowment contributes proportionally less than the participants with low initial endowment.

Thank You