Fuzzy clustering of WTO members position in the Doha Agricultural Negotiation

dario bruzzese
dario.bruzzese@unina.it

University of Naples
Department of Preventive Medical Sciences

Rome - 18 Settembre 2007

Outline

Outline

Outline

Outline

The work of Bjørnskov and Lind

In the paper *Where Do Developing Countries Go After Doha?*, Bjørnskov and Lind attempted to identify potential coalitions among WTO members by rating their official positions on a wide range of issues considered as crucial for the forthcoming round of trade negotiations concerning agricultural goods.

The work of Bjørnskov and Lind

In the paper *Where Do Developing Countries Go After Doha?*, Bjørnskov and Lind attempted to identify potential coalitions among WTO members by rating their official positions on a wide range of issues considered as crucial for the forthcoming round of trade negotiations concerning agricultural goods.

The 13 rated issues according to competence macro-area

Market Access	Tariffs; Tariffs Peak and Escalation; Tariffs Rate Quotas; Special Safeguard Clause
Export Support	Export Subsidies; Export Credit
Domestic Support	Green Box; Blue Box; Development Box; AMS; De Minimis Level
Non-trade concern	Broad vs. Narrow Round; Labour Standards and Environments

The work of Bjørnskov and Lind

The rating system

The rating system is based on a 5 point ordinal scale (ranging from 0 to 4) were higher values indicate an higher commitment for reducing real and potential barriers in international trade.

The work of Bjørnskov and Lind

The rating system

The rating system is based on a 5 point ordinal scale (ranging from 0 to 4) were higher values indicate an higher commitment for reducing real and potential barriers in international trade.

The Statistical Analysis

- First Step: *Hierarchical Classification*
- Second Step: *k-means*

The work of Bjørnskov and Lind

The Results

Gradual liberalisation	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Congo, Cote d'Ivoire, Djibouti, Gabon, the Gambia, Ghana, Guinea, Guinea-Bissau, Lesotho, Madagascar, Malawi, Mali, Mauritania, Mauritius, Morocco, Mozambique, Niger, Rwanda, Senegal, Sierra Leone, Swaziland, Tanzania, Togo, and Zambia
Market Access	Cuba, Haiti, Nicaragua, and Peru
Average	Albania, Antigua, Brunei, Bulgaria, Burma, Canada, Croatia, the Czech Republic, Dominica, Ecuador, Estonia, Grenada, Hungary, Jamaica, Japan, Jordan, the Kyrgyz Republic, Latvia, Lithuania, Mongolia, Poland, St. Kitts, St. Lucia, St. Vincent, Singapore, the Slovak Republic, Slovenia, South Korea, Suriname, Switzerland, Trinidad, Tunisia, Turkey and Venezuela
Narrow Round	Argentina, Australia, Bolivia, Brazil, Chile, Colombia, Costa Rica, Fiji, Guatemala, Honduras, Indonesia, Malaysia, Namibia, New Zealand, Paraguay, South Africa, Thailand, Uruguay and the US
Small changes	India, Mexico and the Philippines
-	The EU and Israel
Free Trade	The Dominican Republic, Egypt, El Salvador, Kenya, Nigeria, Pakistan, Sri Lanka, Uganda and Zimbabwe
-	Norway
Developing multifunctionality	Barbados and the Democratic Republic Congo

The work of Bjørnskov and Lind

Con’s

The work of Bjørnskov and Lind

Con's

- Because of the data collection method, the input data set was charachterized by a huge number of missing ratings. The adopted strategy to face the sparsity of the data array was to replace each missing by the column mean.

The work of Bjørnskov and Lind

Con's

- Because of the data collection method, the input data set was charachterized by a huge number of missing ratings. The adopted strategy to face the sparsity of the data array was to replace each missing by the column mean.
- A *crisp* classification strategy does not allow to capture the inter-class similarities among objects as each country can belong to only one cluster.

Fuzzy Sets

Definition

Given a collection of objects $\Omega \subseteq R^n$ a Fuzzy Set A is defined as an ordered couple $\{\Omega, \mu_A(\omega)\}$ where

$$\mu_A(\omega) : \omega \longrightarrow [0, 1]$$

is a *membership function* which measures the coherence of each element $\omega \in \Omega$ with the properties that characterize the fuzzy set A.

Fuzzy Sets

Definition

Given a collection of objects $\Omega \subseteq R^n$ a Fuzzy Set A is defined as an ordered couple $\{\Omega, \mu_A(\omega)\}$ where

$$\mu_A(\omega) : \omega \to [0, 1]$$

is a *membership function* which measures the coherence of each element $\omega \in \Omega$ with the properties that characterize the fuzzy set A.

Fuzzy Set Theory has been successfully used in several applicative and methodological contexts:

- Credit scoring
- Medicine
- Expert System
- Pattern Recongition
- Regression
- Classification

Fuzzy Sets

Definition

Given a collection of objects $\Omega \subseteq R^n$ a Fuzzy Set A is defined as an ordered couple $\{\Omega, \mu_A(\omega)\}$ where

$$\mu_A(\omega) : \omega \rightarrow [0, 1]$$

is a *membership function* which measures the coherence of each element $\omega \in \Omega$ with the properties that characterize the fuzzy set A.

Fuzzy Set Theory has been successfully used in several applicative and methodological contexts:

- Credit scoring
- Medicine
- Expert System
- Pattern Recongition
- Regression
- Classification

Fuzzy Clustering

Definition

A Fuzzy Partition of n objects in C fuzzy sets is given by a $\mathbf{M_{nxc}}$ membership matrix such that:

- $\mu_{i,k} \in [0, 1]\ \forall i, k$ where $\mu_{i,k}$ measures the degree of membership of object i to fuzzy set A_k for $i = 1, .., n$; $k = 1, .., C$
- $\sum_{k=1}^{C} \mu_{i,k} = 1\ \forall i$ *i.e. the membership function is normalized*
- $\sum_{i=1}^{n} \mu_{i,k} > 0\ \forall k$ *i.e. no empty sets are allowed*

Fuzzy Clustering

Crisp Partition

Objects	*Cluster*			
	1	**2**	**...**	**c**
1	1	0	...	0
2	0	1	...	0
3	0	1	...	0
...	...	...	...	..
n	0	0	...	1

Fuzzy Clustering

Crisp Partition

Objects	*Cluster*			
	1	**2**	**...**	**c**
1	1	0	...	0
2	0	1	...	0
3	0	1	...	0
...	...	...	...	..
n	0	0	...	1

Fuzzy Partition

Objects	*Cluster*			
	1	**2**	**...**	**c**
1	0.81	0.02	...	0.1
2	0.31	0.54	...	0.12
3	0.01	0.99	...	0
...	...	...	...	..
n	0.5	0	...	0.5

Fuzzy Clustering

Different algorithms have been developed to produce fuzzy partition. In our research we adopted the Fuzzy c-means (also known as Fuzzy Isodata) [Dunn, 1973; Bezdeck 1981] where the final partition is obtained through the iterative minimization of the following objective criterion:

$$J_m(M, v) = \sum_{k=1}^{C} \sum_{i=1}^{n} \mu_{i,k}^{m} \|x_i - v_k\|^2$$

where $m > 0$ is a fuzzification parameter, x_i is the i-th p-dimensional object to classify, v_k is the prototype of the k^{th} cluster and $\|x\|^2$ is any p-norm, usually the Euclidean norm: $\|x\|^2 = \sum_{j=1}^{p} x_{ij}^2$

Fuzzy Clustering

The optimization of the objective function, after an inizialization phase, is obtained by iteratively zeroing the gradient of $J_m(M, v)$ with respect to $\mu_{i,k}$ and to v_k.

Fuzzy C-means (FCM)

1. Fix n, C and ϵ. Inizialize the prototypes $v^{(0)} \subset R^p$. Then at step r, $r = 0, 1, 2, ...$
2. Compute $M^{r+1} = \arg\min_M \left\{ J_m \left(M, v^{(r)} \right) \right\}$
3. Compute $v^{r+1} = \arg\min_v \left\{ J_m \left(M^{(r+1)}, v \right) \right\}$
4. Compare v^{r+1} to v^r using $\left\| v^{r+1} - v^r \right\| < \epsilon$. If True then stop. Otherwise set $r = r + 1$ and return to step 2.

Fuzzy Clustering

The Fuzzy clustering procedure ends with a defuzzification phase where each point is assigned to the cluster with the highest membership values. Ties are resolved by random selection.

Object l belongs to cluster k iff

$$\mu_{l,k} = \max_{k=1,..,C} \mu_{l,k}$$

Fuzzy Clustering

The Fuzzy clustering procedure ends with a defuzzification phase where each point is assigned to the cluster with the highest membership values. Ties are resolved by random selection.

Fuzzy Partition

Objects	*Cluster*			
	1	**2**	**...**	**c**
1	0.81	0.02	...	0.1
2	0.31	0.54	...	0.12
3	0.01	0.99	...	0
...	...	...	...	..
n	0.5	0	...	0.5

Defuzzification

Objects	*Cluster*			
	1	**2**	**...**	**c**
1	1	0	...	0
2	0	1	...	0
3	0	1	...	0
...	...	...	...	..
n	0	0	...	1

Fuzzy Clustering

Pro's

- FCM, and more generally all the Fuzzy Clustering algorithms, allow to preserve the natural *fuzziness* of the real problems along the whole classification procedure.

Con's

- The FCM algorithm references each component of each datum and thus it cannot be directly applied to data sets containing missing features values
- It is mandatory to define the number of clusters in the inizialization phase.

Fuzzy Clustering

Missing Features estimates

In order to face with incomplete data set, Hathaway and Bezdek (2001) proposed a modified version of the FCM were the missing values are iteratively estimated during the optimization process. In particular, the optimization of $J_m(M, v)$ is pursued by adding one more step where missing values estimates are obtained by zeroing the gradient of $J_m(U, v)$ with respect to the missing features.

$$X_{Miss}^{r+1} = \arg\min_{X_{Miss}} \left\{ J_m \left(M^{(r+1)}, v^{(r+1)} \right) \right\}$$

where

$$X_{Miss} = \{x_{k,j} = ? \text{ for} 1 \leq i \leq n, 1 \leq j \leq p\}$$

Fuzzy Clustering

Finding the sub-optimal number of clusters

The *suboptimal* number of clusters can be achieved by repeating the FCM clustering procedure for C=2,... and computing for each classification a validity index which measures the quality of the resulting classification. Among the several cluster validity index proposed in the literature we have adopted the Partition Entropy Coefficient:

$$PE = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{c}\mu_{i,k}\log(\mu_{i,k})$$

The results of FCM

1 Ecuador (0.95), Estonia (0.95), Mongolia (0.95), Slovenia (0.95), Jordan (0.24), Singapore (0.2), Slovak Republic (0.2), Albania (0.13), Canada (0.13), Croatia (0.12), Brunei (0.12), Myanmar (0.12), Morocco (0.1), El Salvador (0.1), India (0.1) and Poland (0.1)

2 Uruguay (0.1), Namibia (0.1), Thailand (0.1), Jamaica (0.1), Ghana (0.1), Indonesia (0.1), St Vincent (0.1), Antigua (0.1), South Africa (0.1), Honduras (0.1), Argentina (0.1), Egypt (0.1), Zimbabwe (0.1), Barbados (0.09), Turkey (0.09) and Congo DR (0.08).

3 Australia (0.78), Brazil (0.78), Guatemala (0.4), Fiji (0.35), Paraguay (0.23), CostaRica (0.17), Chile (0.17), Colombia (0.15), USA (0.14), New Zealand (0.13), Bolivia (0.13), Malaysia (0.12), Philipines (0.1) and Mexico (0.1).

4 Grenada (0.98), St Kitts (0.98), St Lucia (0.98), Surinam (0.98), Trinidad (0.98), Tunisia (0.98), Venezuela (0.98), Dominica (0.2), South Korea (0.16), Switzerland (0.15), Swaziland (0.14), Cote d'Ivoire (0.1), Lesotho (0.11) and Mauritius (0.1).

5 Haiti (0.98), Nicaragua (0.98), Peru (0.98), Uganda (0.56), Kenya (0.23), Cuba (0.22), Dominican Republic (0.2), Pakistan (0.14), Nigeria (0.13) and Sri Lanka (0.12).

6 Bulgaria (0.97), Kyrgyz Republic (0.97), Latvia (0.97), Hungary (0.45), Lithuania (0.4) and Czech Republic (0.19).

7 Angola (1), Botswana (1), Burundi (1), Central African Republic (1), Chad (1), Congo Republic (1), Djibouti (1), Gambia (1), Guinea-Bissau (1), Madagascar (1), Malawi (1), Mauritania (1), Niger (1), Rwanda (1), Tanzania (1), Togo (1), Zambia (1), Gabon (0.34), Mozambique (0.28), Senegal (0.22), Benin (0.21), Sierra Leone (0.17), Burkina Faso (0.15), Cameroon (0.13), Guinea (0.1) and Mali (0.1).

8 Norway (1)

9 Austria (1), Belgium (1), Denmark (1), Finland (1), France (1), Germany (1), Greece (1), Ireland (1), Italy (1), Luxemburg (1), Netherlands (1), Portugal (1), Spain (1), Sweden (1), United Kingdom (1), Israel (0.1) and Japan (0.1).

Fuzzy Clustering vs Hard Clustering

- Unlike Crisp classification, in each cluster stemmed from FCM it is quite direct to identify leader members that are those with the highest values of the membership function.

Cluster n. 9

Austria (1), Belgium (1), Denmark (1), Finland (1), France (1), Germany (1), Greece (1), Ireland (1), Italy (1), Luxemburg (1), Netherlands (1), Portugal (1), Spain (1), Sweden (1), United Kingdom (1), Israel (0.1), Japan (0.1)

Fuzzy Clustering vs Hard Clustering

- Unlike Crisp classification, in each cluster stemmed from FCM it is quite direct to identify leader members that are those with the highest values of the membership function.

Cluster n. 9

Austria (1), Belgium (1), Denmark (1), Finland (1), France (1), Germany (1), Greece (1), Ireland (1), Italy (1), Luxemburg (1), Netherlands (1), Portugal (1), Spain (1), Sweden (1), United Kingdom (1), Israel (0.1), Japan (0.1)

Fuzzy Clustering vs Hard Clustering

- Cluster number 2 does not own this property.

Cluster n. 2

Uruguay (0.1), Namibia (0.1), Thailand (0.1), Indonesia (0.1), South Africa (0.1), Honduras (0.1), Argentina (0.1), Ghana (0.1), Egypt (0.1), Zimbabwe (0.1),Jamaica (0.1), St Vincent (0.1), Antigua (0.1), Turkey (0.09), Barbados (0.09), Congo DR (0.08)

Fuzzy Clustering vs Hard Clustering

- Cluster number 2 does not own this property.

Cluster n. 2

Uruguay (0.1), Namibia (0.1), Thailand (0.1), Indonesia (0.1), South Africa (0.1), Honduras (0.1), Argentina (0.1), Ghana (0.1), Egypt (0.1), Zimbabwe (0.1),Jamaica (0.1), St Vincent (0.1), Antigua (0.1), Turkey (0.09), Barbados (0.09), Congo DR (0.08)

- Developing multifunctionality

Fuzzy Clustering vs Hard Clustering

- Cluster number 2 does not own this property.

Cluster n. 2

Uruguay (0.1), Namibia (0.1), Thailand (0.1), Indonesia (0.1), South Africa (0.1), Honduras (0.1), Argentina (0.1), Ghana (0.1), Egypt (0.1), Zimbabwe (0.1),Jamaica (0.1), St Vincent (0.1), Antigua (0.1), Turkey (0.09), Barbados (0.09), Congo DR (0.08)

Developing multifunctionality

Narrow Round

Fuzzy Clustering vs Hard Clustering

- Cluster number 2 does not own this property.

Cluster n. 2

Uruguay (0.1), Namibia (0.1), Thailand (0.1), Indonesia (0.1), South Africa (0.1), Honduras (0.1), Argentina (0.1), Ghana (0.1), Egypt (0.1), Zimbabwe (0.1),Jamaica (0.1), St Vincent (0.1), Antigua (0.1), Turkey (0.09), Barbados (0.09), Congo DR (0.08)

- Developing multifunctionality
- Narrow Round
- Average

Fuzzy Clustering vs Hard Clustering

- The counterpart of the fusion of the *Developing multifunctionality* cluster is represented by the splitting of the *Average* group mainly in to the Cluster 1 and Cluster 6.

Cluster 1

Morocco (0.1), El Salvador (0.1), India (0.1), Ecuador (0.95), Estonia (0.95), Mongolia (0.95), Slovenia (0.95), Jordan (0.24), Singapore (0.2), Slovak Republic (0.2), Albania (0.13), Canada (0.13), Croatia (0.12), Brunei (0.12), Poland (0.1)

Cluster 6

Bulgaria (0.97), Kyrgyz Republic (0.97), Latvia (0.97), Hungary (0.45), Lithuania (0.4) and Czech Republic (0.19).

Fuzzy Clustering vs Hard Clustering

- The counterpart of the fusion of the *Developing multifunctionality* cluster is represented by the splitting of the *Average* group mainly in to the Cluster 1 and Cluster 6.

Cluster 1

Morocco (0.1), El Salvador (0.1), India (0.1), Ecuador (0.95), Estonia (0.95), Mongolia (0.95), Slovenia (0.95), Jordan (0.24), Singapore (0.2), Slovak Republic (0.2), Albania (0.13), Canada (0.13), Croatia (0.12), Brunei (0.12), Poland (0.1)

Cluster 6

Bulgaria (0.97), Kyrgyz Republic (0.97), Latvia (0.97), Hungary (0.45), Lithuania (0.4) and Czech Republic (0.19).

Concluding remarks

- Fuzzy clustering guarantees richer and deeper information than a crisp classification as it allows to differenziate the role of the different countries in each cluster
- By using three way data analysis methods on the membership matrices (obtained for instance in different temporal occasions), it could be possible to highlight even small changes in the trade negotiation strategies
- The FCM algorithm does not necessarily converge to a global minimum (but even k-means does not); countries characterized by very fuzzy membership could change their belonging due to the inizialization phase. This drawback could be faced by estimating fuzzy confidence intervals on the punctual memebership values