Peñas-de Pablo, José Miguel; Portilla-Figueras, José Antonio; Navío-Marco, Julio; Salcedo-Sanz, Sancho

Conference Paper

Identifying Telecommunication strategies and investment Opportunities in Latin American Countries Based on Clustering Analysis

26th European Regional Conference of the International Telecommunications Society (ITS): "What Next for European Telecommunications?", Madrid, Spain, 24th-27th June, 2015

Provided in Cooperation with:
International Telecommunications Society (ITS)

Identifying Telecommunication strategies and investment Opportunities in Latin American Countries Based on Clustering Analysis

José Miguel Peñas-de Pablo[1], José Antonio Portilla-Figueras [1], Julio Navío-Marco [2] & Sancho Salcedo-Sanz [1]

[1] *Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, Spain*
[2] *Department of Business Organization, Universidad Nacional Educación a Distancia, Spain*

ABSTRACT:

In this paper we present a methodology for feature selection and clustering over variables describing countries' economies and ICT indicators to study and identify investment opportunities, based on similarities between European and Latin American countries. We address two different problems. First, the work is based on a feature selection problem carried out with the Coral Reef Optimization algorithm. The CRO is a novel bio-inspired based on the simulation of reef formation and coral reproduction. On the other hand, the K-Means++ method is a high-performance robust tool designed to solve clustering problems. Together, both algorithms are able to successfully identify investment opportunities in Latin America and quantify the potential of the telecommunications industry in both regional areas. The work considers different economical and ICT's variables from different European and Latin America countries datasets (mainly Agenda 21 and other available and global sources) for the period 2002-2012.

Keywords: ICT Market, Coral Reef Optimization algorithm, K-Means++ Clustering, Investment Opportunities, Europe, Latin America

1. INTRODUCTION

Information and communications technology (ICT) industry has grown exponentially in value creation from the mid-90s. This grow has been motivated by several factors such (1) corporate investment in infrastructure deployment, (2) sectorial innovation, (3) evolved network infrastructure and (4) an increasing number of customers with access to a large briefcase of services [1] Statistics recovered and processed by the International Telecommunication Union

(ITU) [2]. show that ICT market is stable, solid and it is raising its relevance in the economic needs of all global actors. Data analysis reflects that a higher percentage of disposable income is dedicated to receive and send information from citizens and institutions.

However, the recent economic recession and its further evolution has caused a deep slowdown in the industrial profit generation, an ICT sector of higher concentration and a supply reduction (there are less corporations but with larger transnational presence) and, above all, an increasing uncertainty on how to invest in infrastructures with a reasonable return of investment (ROI). These decisions substantiated in economic instability are based in industrial fears (high-leverage levels, bad corporate decisions…) and future economic expectations.

European and Latin-American (Latam) ICT markets have evolved differently from the beginning of the crisis. This could be mainly caused by their unequal starting point and the investments previously made in each areas [3]. Telephony and communications market (established as a necessary commodity for benefit creation) has been defined by a growing price war that has reduced margins across the continent, a brand concentration and also an increasing competitiveness. Such competitiveness is not incompatible with a remarkable market opening through the European integration process and the notorious reduction of economic and social barriers in both South and Eastern Europe enabling local markets and new investment opportunities and corporate growth (Balkan countries, former Soviet republics, etc..).

Meanwhile, from the Latam environment, we must avoid falling on a wrong perception assuming that this territory is an untapped market full of business opportunities by the mere fact of considering an emergent economy area outside instabilities observed (at least, less than in Europe) in the capital markets . Nothing further from reality, we can describe Latin America as a heterogeneous conglomerate of countries with large and divergent social, industrial and economic characteristics that make complex global characterization. The incorrect analysis of the industry starting point or the lack of use of predictive methods for the local market devolvement had led many operators have to change their business plans, assuming large losses or selling their subsidiaries in the absence of rigorous classification/ modelling studies. The largest operators in the ICT sector (usually European companies such as Telefonica, Vodafone, T-Mobile and North-Central American companies like AT&T, Verizon, America Mobil Group) [4] [5] are betting and

still rely on this vast territory as a major income source within their balance sheets as other industrial sectors are performing (with mature and consolidated markets) as banking, insurance or automotive. The ICT sector is looking for an horizontal integration as mandatory step to further develop of their global business in a increasingly complex and interdependent corporate environment.

There are many possible economic arguments for the present and future studies [6] [7] to deepen the research of the ICT market and its regional classification such (1) the global growing uncertainty, (2) the emergence of new companies in the ICT ecosystem, which are using the existing infrastructure but have not been responsible for the development or deployment (such VMO, utilities, social networks, etc…), (3) the corporate priority of cash requirements for both investment and financial activities, (4) new technologies within reach of a greater number of users, (5) the failed implementation and the mere copies of business strategy across companies without a prior assessment or obtaining extraordinary corporate profits. All these situations are cause for the development of study-cases and researches [3].

As it is described above we have detected a lack of research studies (at least academic) intended to apply the know-how obtained from the expertise in Europe to the Latam case [9] [5] In this work we propose the development of a soft-computing clustering method to group European and Latin-American countries based on economic, social and ICT indicators obtained from sources as [2] [10] [11]. This work qualifies behaviour patterns under different types of study variables and help decision making for all industry players within the ICT market. The main objectives of this classification are: decreasing the uncertainty associated with the long term investment generated returns and especially the right investment opportunities selection. [12]

The rest of the paper is organized as follows, next section performs the problem description, section 3 describes the basic principles of a Coral Reef Optimization technique applied to feature selection and the K-means clustering algorithm, section 4 shows the model developed for this specific case to show the experiments and results in section 5. Finally we show some conclusions and future worklines.

2. SCENARIO DESCRIPTION

The global objective described above needs to be considered in a process with two phases. Initially we must appropriately identify the starting point of every European and Latin America countries, finding common behaviour patterns and evaluating the evolution of the ICT sector based on a reduced set of key economic and technological indicators. These indicators will be obtained from a large pool obtained from different sources [2] [10] [11], applying a meta heuristic algorithm for feature selection named Coral Reef Optimization (CRO). The CRO performance in this task has been proven in energy conversion and management [13] [14]. On a second stage, and using the indicators obtained in the feature selection phase we will carry on a clustering analysis to define each country in a global environment by developing and technological potential.

Therefore the first step consists of the definition of the pool of relevant indicators and the sources to obtain them. In this project we have used a large database collected from several institutional sources (ICT, World Bank, IMF, other international boards, etc) which is part of Agenda 21 [15]. This Project seeks to promote, through the study of quantitative data, improving homogeneous communications and technological development of OECD and emerging countries. As we need to observe the evolution of the different countries, the indicators have been obtained in the range of years from 2000 to 2012.

The complete set of indicators has been grouped into economic and telecom indicators in order to measure the real impact of the different data sets. An important problem to tackle is the lack of data available for the Latin America countries compared to the information available in the other European or OECD countries. The lack of existing data has traditionally been one of the biggest problems for studies and comparative analysis between countries in different geographical areas. To solve this problem the original database has been completed with other official sources [16] [17] [18] and in case incomplete data series, we have used a linear interpolation to complete them. Table 1 shows the complete set of ICT (technical) social and economic indicators used in this study.

ICT Variables		Economic Variables	
Variable	**Units**	**Variable**	**Units**
Mobile Phones / 100 inhabitants	Ratio	Population	Nominal (units)
Land Line / 100 inhabitants	Ratio	GDP ($ - Nominal)	Nominal ($)
International Telephone Traffic	Nominal (minutes)	% Tertiary Sector	Ratio (%)
Mobile Revenue	Nominal ($)	% Debt s/GDP	Ratio (%)
Land Revenue	Nominal ($)	Net Foreign Investment	Nominal ($)
Total Revenue	Nominal ($)	Gini Index (Classic)	Índix (Base 1)
ICT Productivity	Index (Base:100)	GDP Relative Growth)	Ratio (%)
ICT Expendure / GDP	Ratio (%)	Unemployment	Ratio (%)
% Internet Population	Ratio (%)	Force	Ratio (%)
5 Minutes International Call ($ Purchase Power Parity)	Nominal ($)	Deficit	Ratio (%)
Internet Penetration (%)	Ratio (%)	% Primary Sector	Ratio (%)
Mobile Income	Nominal ($)	% Urban Population	Ratio (%)
Land Income	Nominal ($)	% Immigrant Population	Ratio (%)
Total Income	Nominal ($)	Homes	Nominal (units)
Regular Internet Use (%)	Ratio (%)	Political Confidence Index	Índex (Base:100)
Public Phones	Nominal (unit)	Tax Revenue	Ratio (%)
High Tech Exportation	Nominal ($)	Fiscal Pressure	Ratio (%)
ICT Goods Export	Nominal ($)	GDP / capita (PPA)	Ratio (units)
ICT Goods Import	Nominal ($)	CPI (Consumer Price Index)	Ratio (%)
I+D Expendure s/GDP	Ratio (%)	Education Expenditure (s/GDP)	Ratio (%)
ICT Expendure s/GDP	Ratio (%)	Health Expenditure (s/GDP)	Ratio (%)

Table 1: ICT and Economic Variables

To perform the studies we have defined a dataset with 32 countries where 20 are European and 12 Latin America. We have selected a group of European countries with different economic status, technological progress and geographical representation. In this way, we get a complete view of the European ICT sector in Europe. Concerning Latin America we have chosen those countries that

have regional representation considering also that there must be a comprehensive and coherent set of available data for the study [2] [10] [11]. This constraint (data availability) makes that the number of samples (countries) for Europe in the dataset is higher than Latin America ones. Far from being a drawback, this could be an advantage when classifying countries in Latin America due to the higher quality of the data indicators in European countries. Table 2 shows the complete set of countries separated into regional areas.

European Countries		**Latin America Countries**	
Germany	France	Brazil	Uruguay
Denmark	Finland	Colombia	Ecuador
Sweden	Switzerland	Poland	Venezuela
Belgium	Norway	Argentina	Peru
United Kingdom	Romania	Mexico	Chile
Czech Republic	Spain	*Costa Rica**	*Nicaragua**
Netherlands	Italy	*Honduras**	
Ireland	Austria		
Bulgaria	Greece		
Poland	Portugal		

Table 2: European and Latin American countries considered in the study

At this point we have defined the complete set of indicators and the countries that we are going to apply the clustering process. Next two sections describe, first the algorithms, which perform the determination of the most relevant indicators for our purpose, and second, the proper clustering algorithm and its application to this specific problem.

3. CRO VALIDATION AND FEATURE SELECTION

3.1. CORAL REEF OPTIMIZATION ALGORITHM

The Coral Reefs Optimization Algorithm (CRO) is an evolutionary bio-inspired approach based on the simulation of the processes in coral reefs. ([13] [14]. The CRO can be classified into the family of bio-inspired algorithms which try to artificially simulate the behavior of a specific natural ecosystem to tackle optimization problems, similarly to ant colony optimization , particle swarm optimization algorithm, artificial bee colony approach or the weed colonization algorithm. The CRO has been proven to be effective in several single-objective optimization problems,

obtaining better solutions than alternative optimization algorithms in the literature and also is used in feature selection, classification, clustering and value prediction in social and technological researches

Basically, the CRO algorithm starts from a population of individuals encoding different solutions to a given optimization problem. These solutions are located in an square grid (reef), where there are also empty spaces at the beginning of the algorithm. The algorithm is thought to simulate the process of coral reproduction (sexual and asexual reproduction operators are applied), and the process of coral reef formation, where a fight for space occurs. Thus, in each step of the CRO algorithm a coral larvae formation is carried out, and each larva tries to occupy a place in the reef. It depends on how strong the larva is (how good the solution to the optimization problem is), or if it is lucky enough to find an empty place in the reef. Note that empty places in the reef are scarce after some generations of larvae, though a process of corals depredation ensures the possibility of empty places in the reef even at the final stages of the algorithm.

After the reef initialization described above, a second phase of reef formation is artificially simulated in the CRO algorithm: a simulation of the corals' reproduction in the reef is done by sequentially applying different operators. This sequential set of operators is then applied until a given stop criteria is met. Several operators to imitate corals' reproduction are defined, among them: a modelling of corals' sexual reproduction (broadcast spawning and brooding), a model of asexual reproduction (budding), and also some catastrophic events in the reef, i.e. polyp's depredation. After the sexual and asexual reproduction, the set of larvae formed (new solutions to the problem), try to locate a place to grow in the reef. It could be in a free space, or in an occupied once, by fighting against the coral actually located in that place.

1. Broadcast Spawning (external sexual reproduction): the modeling of coral reproduction by broadcast spawning consists of the following steps:
    a. In a given step k of the reef formation phase, select uniformly at random a fraction of the existing corals pk in the reef to be broadcast spawners. The fraction of broadcast spawners with respect to the overall amount of existing corals in the reef will be denoted as Fb. Corals that are not selected to be broadcast spawners will reproduce by brooding later on, in the algorithm.
    b. Select couples out of the pool of broadcast spawner corals in step k. Each of such couples will form a coral larva by sexual crossover, which is then released out to the water. Note that, once two corals have been selected to be the parents of a larva,

they are not chosen anymore in step k. These couple selection can be done uniformly at random or by resorting to any fitness proportionate selection approach

2. Brooding (internal sexual reproduction): as previously mentioned, at each step k of the reef formation phase in the CRO algorithm, the fraction of corals that will reproduce by brooding is $1 - Fb$. The brooding modelling consists of the formation of a coral larva by means of a random mutation of the brooding-reproductive coral (self-fertilization considering hermaphrodite corals).
3. Larvae setting: once all the larvae are formed at step k either through broadcast spawning (1.) or by brooding (2.), they will try to set and grow in the reef. First, the health function of each coral larva is computed. Second, each larva will randomly try to set in a square (i, j) of the reef. If the square is empty (free space in the reef), the coral grows therein no matter the value of its health function. By contrast, if a coral is already occupying the square at hand, the new larva will set only if its health function is better than that of the existing coral. We define a number κ of attempts for a larva to set in the reef
4. Asexual reproduction: in the modelling of asexual reproduction (budding or fragmentation), the overall set of existing corals in the reef are sorted as a function of their level of healthiness from which a fraction Fa duplicates itself and tries to settle in a different part of the reef. Note that a maximum number of identical corals (μ) are allowed in the reef.
5. Depredation in polyp phase: corals may die during the reef formation phase of the CRO algorithm. At the end of each reproduction step k, a small number of corals in the reef can be depredated, thus liberating space in the reef for next coral generation.

Figure 1 shows the flow diagram of the CRO algorithm, summarizing the steps explained above

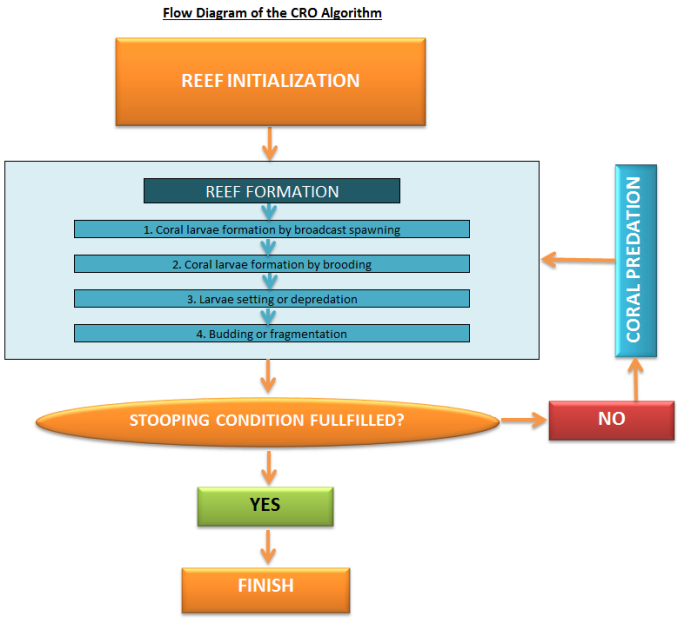

Figure 1: Flow Diagram of the CRO Algorithm

3.2. FEATURE SELECTION

This section explains the application of the CRO algorithm to the feature selection problem of this paper. The objective is to determine, from the complete set of 42 indicators a reduced set of 20, half economic, half technical, which describes best the features of the different countries. Please note that some indicators could be correlated and therefore they can be eliminated from the data set without losing relevant information.

Let us consider a set of countries $C\ (1,2,..,N)$. Each country has a set of macro-economic variables defined as $E_{Ei}(1,2,\dots,E)$ and also a set of ICT's variable defined as $E_{Ti}(1,2,\dots,T)$. Let us also consider the accumulated value of industrial ICT profit in each geographical area (either Latin America, Europe or Global), Y, as the indicator of the sector potential because it describes appropriately the evolution of corporate expectations$Y = \sum_i^N (y_1, y_2, .., y_N)$. The value of Y may be estimated, $\tilde{Y}$, using the exponential model described in Equation [1] that has been previously used in [19] for energy demand production. The feature selection procedure tries to obtain from the original dataset (E_{Ei} and E_{Ti}) a subset of indicators (called E'_{Ei} and E'_{Ti}) and their corresponding weights w_{iN} and w_{jN} which estimates best objective target based on the function described by Equation [1] :

$$\tilde{Y} = \sum_{i=1}^{Max} w_{i1}\, {E'_{Ti}}^{wj1} + \sum_{i=1}^{Max} w_{i2}\, {E'_{Ei}}^{wj2} + w_0 \qquad [1]$$

Where:

- $\tilde{Y} =$ Estimated CRO prediction (Fitness Function)
- E'_{Ti} = Selected economic indicators
- E'_{Ei} = Selected ICT Indicators
- w_{iN}= Product Weight
- w_{jN}= Exponential Weight

We have divided the experiments into three classes. The first one (1) considering only economic indicators, the second one (2) considering only ICT indicators and finally the third set (3) considering the complete set of them (Economic + ICT). Each class is also divided into three subclasses referred to the regional areas, in such a way that the first subclass corresponds to European countries, the second one to Latin America countries and the third one to both regions together. Figure 2 shows a scheme of the experiments performed. Please note that in the first class we try to obtain the 10 out of 21 best economic indicators, in the second one we try to obtain the 10 out of 21 best ICT indicators and in the third run we try to obtain the 20 out of 42 indicators, independently if they are ICT or Economic. This way we get, on the one hand, which indicators best represents the socio-economic position of the countries related to the ICT and, on the other hand, the consistency of the indicator selection, by the comparison of the results obtained in classes (1) and (2) with the obtained in class (3) which will be used in the clustering process.

We have performed 10 runs on each one of the 9 experiments for each one of the 12 years considered, from 2000 to 2012.

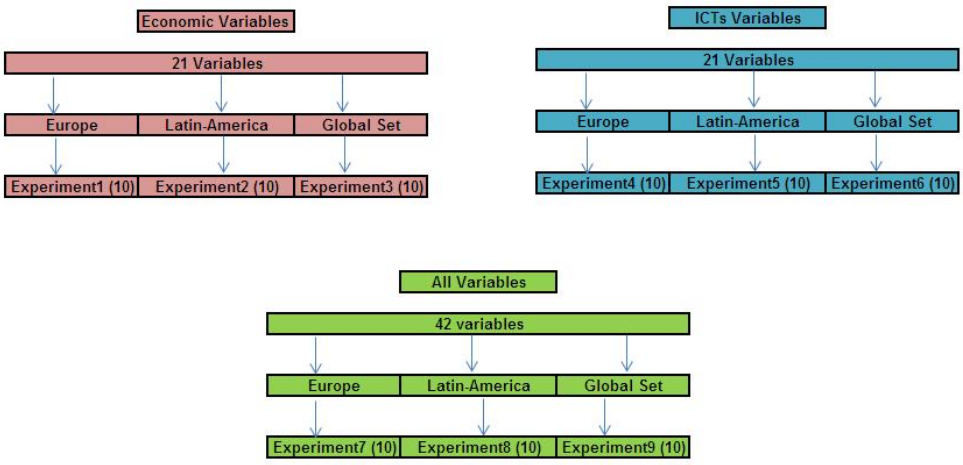

Figure 2: CRO Validation Scheme

Table 3 shows the results of the CRO validations the 9 scenarios for the year 2012. We have calculated the Root-Mean-Square Deviation (Equation [2]) from the sum of industrial ICT profit in each geographical area (Y) and the CRO prediction ($\tilde{Y}$)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y - \tilde{Y})^2} \quad [2]$$

We present the average value, the best value of the ten simulations, the target value and the relative error.

Experiment	Average Value (M€)	Experiment (M€)	Average Value (M€)	Experiment
Experiment 1	16.41 10^3	16.92 10^3	17.62 10^3	6.82 %
Experiment 2	14.21 10^3	13.82 10^3	12.99 10^3	9.39 %
Experiment 3	32.86 10^3	31.51 10^3	30.61 10^3	7.35 %
Experiment 4	18.90 10^3	17.23 10^3	17.62 10^3	7.29 %
Experiment 5	14.74 10^3	14.21 10^3	12.99 10^3	13.46 %
Experiment 6	33.64 10^3	32.94 10^3	30.61 10^3	9.90 %
Experiment 7	17.06 10^3	17.51 10^3	17.62 10^3	3.14 %
Experiment 8	13.85 10^3	13.30 10^3	12.99 10^3	6.58 %
Experiment 9	32.14 10^3	31.84 10^3	30.61 10^3	5.02 %

Table 3: Validation of feature selection with CRO algorithm.

CRO algorithm shows good performance forecasting the global ICT profit with a low relative error in each experiment. Please note that the worse performance in Latin America countries (explained by the lack of data available in this area). The CRO algorithm is programmed to find the best set of features in every area, the absence of data, forces the method to search in the rest of complete features of the data set. We assume a 5% of relative error as adequate prediction result to validate the CRO method.

It should be emphasized that the selection of twenty variables is consistent. From the 42 variables selected in first time and used, CRO algorithm determined that 23 of these are relevant appearing as variables designated in the simulations and 17 are recurrent in each this test.

ICT Variables		Economic Variables	
Variable	**Units**	**Variable**	**Units**
Mobile Phones / 100 inhabitants	Ratio	Population	Nominal (units)
Total Revenue	Nominal ($)	GDP ($ - Nominal)	Nominal ($)
ICT Expenditure s/GDP	Ratio (%)	% Tertiary Sector	Ratio (%)
Total Income	Nominal ($)	% Debt s/GDP	Ratio (%)
ICT Goods Exports	Nominal ($)	Net Foreign Investment	Nominal ($)
ICT Goods Imports	Nominal ($)	Gini Index (Classic)	Índix (Base 1)
I+D Expenditure s/GDP	Ratio (%)	Unemployment	Ratio (%)
ICT Expenditure s/GDP	Ratio (%)	Deficit	Ratio (%)
		% Urban Population	Ratio (%)
		% Immigrant Population	Ratio (%)
		Education Expenditure (s/GDP)	Ratio (%)
		GDP / capita (PPA)	Ratio (units)

Table 4: CRO Validation and Selected Variables in Clustering Method

Therefore, taking into account variables such recurrence and validation of the CRO algorithm, the following variables presented in Table 4, are selected at the time we performed clustering. The CRO algorithm selects 20 variables, 8 of these are ICT variables and 12 are economic variables (it is powerfully striking that many social-demographic variables are selected in the final test and are used in the clustering step)

4. K-MEANS++ CLUSTERING METHOD

Clustering is an unsupervised classification technique that consists in grouping data objects from a general set into disjoint groups. This grouping is done in such a way that elements in the same subgroup are similar in terms of some specified metric and different from members of the rest of groups. Clustering has been applied to a large range of fields as renewable energy [20] pattern recognition [21], or ICT and telecommunications market [22]. There is a large variety of clustering algorithm types, from the well-known K-means [23] [24] [25] to novel approaches as population based metaheuristics as the evolutive algorithms [26]. In this work we are going to use an improved version of the K-means algorithm name K-means++, [27]. K-means algorithm

considers intra-cluster variation (intra-clusters points or centroids) to form the final solution and it works as follows:

Let us consider a set of data $X\,(1,2,..,N)$, in our case the set of countries, each one with its corresponding indicators, E_T, E_E, and a predefined number of clusters k. To get to the final solution the K-means algorithm executes the following steps.

1. Arbitrarily choose an initial k centres C = $(c_1, c_2, \dots, c_k)$.
2. For each i ∈ {1, . . . , k}, set the cluster C_i to be the set of points in X that are closer to c_i than they are to cj for all j ≠ i. In this work for step 2 we are going to consider Euclidean Distance
3. For each i ∈ {1, . . . , k}, set c_1 to be the centre of mass of all points in C_i: $c_i = \frac{1}{|C_i|}\sum x \in C_i$x
4. Repeat Steps 2 and 3 until C no longer changes.

Please note that the basic K-means algorithm the initial centres are randomly chosen from X. K-means algorithm has some limitations referred to the computational time in large sized experiments and that the clustering found could, on some cases, be far away from the optimal one. These shortcomings are solved applying K-means++ algorithm, which determines the starting point of each clusters in this way. Let D(x) denote the shortest distance from a data point to the closest centre already chosen, the following steps are taken

1. Select one center c_1, chosen uniformly at random from X, an compute de distances D(x) from the rest of data
2. Select a new center c_i from the data set X, choosing x ∈ X with probability $\frac{D\,(x^2)}{\sum x \in X\, D\,(x^2)}$
3. Repeat step2 until k centres have been
4. Proceed with the standard k-means algorithm.

5. CLUSTERING RESULTS

With the 20 indicators selected by the CRO algorithm procedure in section 3 and shown in Table 4 we have applied K-Means ++ method to cluster the 32 countries under study. Please note that we applied the CRO feature selection with only 29 out of 32 countries. Honduras, Costa Rica and Guatemala showed bad feature selection with the available data and therefore they were removed

from that part of the experiments. However we have included these three countries to obtain their classification in the experiment. We have performed clustering from K=3 to K=12 to obtain the different country classification.

To analyse the quality of the resulting clusters and determine their optimum number, we have applied two different metrics: Maximal Similarity metric and Cosine distance method.

1. Maximal Similarity Metric: This is a metric specifically designed for this work. The objective of this measure is to minimize the ratio of two distances, see equation [3]. On the numerator it is takes the normalized distance between countries that are in the same cluster (Intra-Cluster) and on the denominator it is the normalized distance between the centroids of the clusters (Inter-Cluster). We define P_{ij} as the Euclidean Intra-Cluster distance and $\sum D_{ij}$ is the Euclidean Inter-Cluster distance.

$$Min\ \frac{min\left[\frac{1}{N_T}\sum P_{ij}\right]}{max\left[\frac{1}{N_C}\sum D_{ij}\right]} \qquad [3]$$

2. Cosine Distance Method: It is one of the most widely used metric to assess clustering and similitude between clusters. By this method, it applies a trigonometric variation of Euclidean distance which is used to obtain as a baseline measurement of the close relationship between the different representatives points of each country. We define x_i and y_i as pair of points of the centroid of each cluster as it is describe in Equation [4]

$$\cos(\vec{x},\vec{y}) = \sum_i^{k=3} \frac{x_i * y_i}{\sqrt{\sum_i x_i^2 * \sum_i y_i^2}} \qquad [4]$$

Therefore, in both cases, the cluster will be more effective as long as the values of the metrics are minimal. The metrics reflect that we obtained a number of clusters that minimizes optimum internal distances and maximize external.

Metrics Validation		
K	**Maximal Similarity**	**Cosine Distance Method**
3	2.302	3.499
4	1.947	2.813
5	1,652	2.216
6	1.356	1.731
7	**1,136**	**1.470**
8	**1,142**	**1.481**
9	1.265	1.713
10	1.395	1.897
11	1.565	2.050
12	1.717	2.319

Table5: Clustering Performance and Metrics Validation (From K=3 to K=12)

Experiments show that K = 7 is the optimal number of clusters for both metrics. However, it is necessary and relevant to draw attention also to the K=8 cluster selection due the small differences that exist between the two metrics methods. Tables 6 and 7 shows the country classification with K=7 and K=8. Latin America countries are marked in bold letter and Tables 8 and 9 show the relative distance between the centroids of the different clusters for the Maximal Similarity metric.

Table6: Euro-Latin America Clustering Evolution (2002-2012) for K=7

CLUSTERS	1	2	3	4	5	6	7

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Germany	1	1	1	1	1	1	1	1	1	1	1
UK	1	1	1	1	1	1	1	1	1	1	1
Sweden	1	1	1	1	1	1	1	1	1	1	1
Finland	1	1	1	1	1	1	1	1	1	1	1
Switzerland	1	1	1	1	1	1	1	1	1	1	1
Denmark	1	1	1	1	1	1	1	1	1	2	2
France	2	2	2	2	2	2	2	2	2	2	2
Norway	2	2	2	2	2	2	2	2	2	2	2
Mexico	2	2	2	2	2	2	2	2	2	2	2
Belgium	2	2	2	2	3	3	3	3	2	2	3
Austria	2	2	2	2	2	2	3	3	3	3	3
Ireland	3	3	3	3	3	3	2	2	2	2	2
Chile	3	3	3	3	3	3	3	2	2	2	2
Netherlands	3	3	3	3	3	3	3	3	3	3	3
Spain	3	3	3	3	3	3	3	3	3	3	3
Italy	3	3	3	3	3	4	3	3	3	3	3
Czech Rep.	3	3	3	3	3	3	3	4	4	4	4
Brazil	4	4	4	4	3	3	3	3	3	3	3
Colombia	5	5	5	4	4	4	4	4	3	3	3
Poland	4	4	4	4	4	4	4	4	4	4	4
Argentina	4	4	4	4	4	4	4	4	4	4	4
Portugal	4	4	4	4	4	4	5	5	5	4	4
Greece	4	4	4	4	4	5	5	5	5	6	6
Romania	5	5	5	5	5	5	5	5	5	4	4
Bulgaria	5	5	5	5	5	5	5	5	5	5	5
Venezuela	5	5	5	5	5	5	5	6	6	6	6
Uruguay	6	6	6	6	6	6	6	6	5	5	5
Perú	6	6	6	6	6	6	6	6	6	6	6
Ecuador	6	6	6	6	6	6	6	6	6	6	6
Costa Rica	7	7	7	7	7	7	7	7	7	7	7
Guatemala	7	7	7	7	7	7	7	7	7	7	7
Honduras	7	7	7	7	7	7	7	7	7	7	7

Table7: Euro-Latin America Clustering Evolution (2002-2012) for K=8

CLUSTERS	1	2	3	4	5	6	7	8

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Germany	1	1	1	1	1	1	1	1	1	1	1
UK	1	1	1	1	1	1	1	1	1	1	1
Sweden	1	1	1	1	1	1	1	1	1	1	1
Finland	1	1	1	1	1	1	1	1	1	1	1
Switzerland	1	1	1	1	1	1	1	1	1	1	1
Denmark	2	2	2	2	1	1	1	2	2	2	2
France	2	2	2	2	2	2	2	2	2	2	2
Norway	2	2	2	2	2	2	2	2	2	2	2
Mexico	2	2	2	2	2	2	2	2	2	2	2
Belgium	2	2	2	3	3	3	3	3	2	2	3
Austria	3	3	3	3	3	3	3	3	3	3	3
Ireland	3	3	3	3	3	3	3	2	2	2	2
Chile	3	3	3	3	3	3	2	2	2	2	2
Netherlands	3	3	3	3	3	3	3	3	3	3	3
Spain	3	3	3	3	3	3	3	3	3	3	3
Italy	3	3	3	3	3	4	4	3	3	3	3
Czech Rep.	4	4	4	4	4	4	4	4	4	4	4
Brazil	4	4	4	4	4	3	3	3	3	3	3
Colombia	5	5	5	4	4	4	4	4	3	3	3
Poland	5	5	5	5	5	5	5	5	5	5	5
Argentina	4	4	4	4	4	4	4	5	5	4	4
Portugal	4	4	4	4	4	4	4	4	4	4	4
Greece	5	5	5	5	5	4	4	4	6	6	6
Romania	6	6	6	6	6	6	6	6	6	4	4
Bulgaria	6	6	6	6	6	6	6	6	6	6	6
Venezuela	6	6	6	6	6	6	7	7	7	7	7
Uruguay	6	6	6	6	6	6	6	5	5	5	5
Perú	7	7	7	7	7	7	7	7	7	7	7
Ecuador	7	7	7	7	7	7	7	7	7	7	7
Costa Rica	8	8	8	8	8	8	8	8	8	8	8
Guatemala	8	8	8	8	8	8	8	8	8	8	8
Honduras	8	8	8	8	8	8	8	8	8	8	8

2012	C1	C2	C3	C4	C5	C6	C7
C1		1.484	4.123	6.231	7.984	11.346	15.016
C2			3.011	5.194	7.253	10.891	13.950
C3				2.318	4.044	7.563	11.652
C4					1.983	5.368	9.047
C5						3.841	7.229
C6							4.648
C7							

Table 8: Euro-Latin America Centroids Distance (2012) for K=7

2012	C1	C2	C3	C4	C5	C6	C7	C8
C1		1.315	3.816	5.713	7.201	9.484	12.126	16.144
C2			2.710	4.881	6.321	8.513	11.413	15.202
C3				3.149	4.895	6.002	9.393	12.845
C4					1.974	3.215	6.542	9.774
C5						1.568	4.841	8.514
C6							3.418	7.212
C7								3.849
C8								

Table 9: Euro-Latin America Centroids Distance (2012) for K=8

Results shown on Table 6 and Table 7 draw a temporal evolution of the position in the clustering of countries in Europe and Latin America over the period 2002-2012. Some countries, as Spain in Europe or Mexico in Latin America, maintain a stable trajectory, which means that industrial profits are similar from year to year. However there is a group of countries that move from one cluster to another along the historical series, as Belgium and Brazil. It is very important to note that economic variables have a larger and earlier impact over time than the ICT's variables (i.e. an increase in income per capita to observe and quantify than the development and deployment of telephony infrastructure), therefore we assume that K-Means++ algorithm is more sensitive to economic variables than to technological variables.

Tables provides some coherent results that prove the robustness of the methodology, for example, there is an large stability in most developed countries of the European Union, as Germany, Sweden and Finland. Furthermore the results show the fall of Greece, from cluster

4, in K=7, where it was with countries as Poland or Portugal, in 2002 to cluster 6 in K=7, grouped with Peru and Venezuela in year 2012. The fall of Venezuela is also represented in the results.

Table 6 and Table 7 also show the existing gap between the two continents. Most European countries located in the first three clusters while Latin America countries are positioned at the last (4 to 7 in K=7 and 4 to 8 in K=8) clusters. There are exceptions such as México and Chile, which have a similar behaviour of a Central European country. The evolution of other countries such as Brazil, which climbs from cluster 4 to cluster 3 about 2006, or Colombia which reflects a notorious economic growth potential in the next years. Colombia is a special case because it is the only country in Latin America subset to advance two positions in the ranking of clusters during the study period from a cluster where it is grouped in Bulgaria and Romanian (Cluster 5, K=7) to a cluster where it is grouped with Spain and Netherlands (Cluster 3, K=7).

We need to bring the attention to Uruguay. Uruguay in 2002 is located in cluster 6 in K=7 and K=8, and it evolves towards cluster 5 in 2009 -2010 (depending on the K). It is the only Latin American country located in the lower part of the table, that is, with low values in ICT and economic indicators, which evolves towards a better situation. In fact in K=8 it is grouped with Poland in cluster 5 but in K=7 Poland is in cluster with 4 while Uruguay remains in cluster 5. This means that Poland and Uruguay has similar structures in terms of ICT and economic development but Poland is a little more evolved. Therefore as Poland is currently growing at 1 % Rate (GDP) and the rate of the ICT in the Poland GDP is forecast to growth from 4.8 % in 2013 to a 9.5 % in 2020, [28] it is reasonable to think that the application of Poland-like ICT strategies in Uruguay may be also a successful story.

6. CONCLUSIONS

In this paper we have presented a methodology to perform clustering over countries based on economic and ICT indicators to study and identify investment opportunities based on similarities between European and Latin American countries. The model is divided into two parts. The first one corresponds to a feature selection procedure where we have applied a metaheuristic algorithm named Coral Reef Optimization, that have shown a good performance in similar previous problems. With this we get, from a large set of ICT and economic indicators, which are the most relevant for our study. The second phase is to apply the well

knowm K-means++ algorithm with the indicators obtained from the previous step to obtain the country clustering. To do that we have made a sweep from K=2 to K=12 clusters, finding that K=7 is the optimum one.

Concerning the ICT / economic conclusions, the clustering results show some relevant facts concerning Latin America countries. We have considered (at least) 4 countries to study: Colombia, Uruguay, Ecuador and Brazil.

- Colombia has evolved in 10 years from cluster 5 to cluster 3. That means that Colombia is seeing sustained economic growth in recent years. This country is sharing cluster with countries like Spain, Italy, the Netherlands and Austria. This development leads to increased demand for ICT services and a strong indicator of future return investment expectations.
- Uruguay has grown up from cluster 6 to cluster 5 in the last years. Due to its hight population density and limited extension, this indicator allows us to state the country's growth potential in the short term.
- Brazil has developed a cluster in the last decade and shares positioning with Colombia. However, its growth has been less noticeable than other neighbouring countries. The reasons observed are high economic inequality and geographic dispersion that does not allow optimizing the infrastructure deployment.
- Ecuador is positioned in one of the last clusters but its structure and economic data reveal future growth and investment opportunities. If we extend the results (K=9), this country belongs to the next cluster separating from Peru, Guatemala or Honduras.

On the other hand, there are many countries in Latin America that are included within the technological underdevelopment and do not have an attractive investment in the short term. In this group we include the Caribbean countries, Peru, Honduras or Guatemala that have a large territory but a low population density.

Economic variables show a large weight in the behaviour of the fitness function. As we have emphasized, it is powerfully striking that many social-demographic variables are selected in the feature selection. The results assume that large countries with low density are less interesting for ICT investment and conversely, high-density areas are the most potential for growth.

It is necessary to highlight the presence of many European companies in Latin America offering services. Argentina, Brazil or Chile have more advanced starting point thanks to initial investments received from this corporations. These countries may emphasize public policies to push international corporations to continue the ICT investment in these countries in the same way they did 20 years ago. Finally Mexico as a country with a highly development market, very similar to the Central Europe countries, has its future growth potential will be given incorporating to the ICT sector many of its rural areas.

REFERENCES:

[1] P.C. Symeou,"Economysize and performance: An efficiency analysis in the telecommunications sector", Telecommunications Policy, vol. 35, 426---440, 2011.

[2] ITU – International Technological Union
http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx, Last review: 15th April 2015

[3] C. Fink, A. Mattoo and R. Rathindran, "An assessment of telecommunications reform in developing countries", Information Economics and Policy, vol. 15, 443---466, 2003.

[4] J. Mariscal and E. Rivera, "New trends in the Latin American telecommunications market: Telefonica & Telmex" Telecommunications Policy, vol. 29, pp. 757---777, 2005.

[5] C. Feijoo, J.L. Gómez-Barroso, S. Ramos, R. Coomonte “The mobile communications role in the next generation netwoks: The case of Spain”. 22nd European Regional Conference of the International Telecommunications Society (ITS2011), Budapest, 18 - 21 September, 2011: Innovative ICT Applications - Emerging Regulatory, Economic and Policy Issues

[6] D. Collico Savio. , “VoD pricing in Latam: A business perspective”, 24th European Regional Conference of the International Telecommunication Society, Florence, Italy, 20-23 October 2013

[7] T. Hsin-yi Sandy “Drivers of fixed and mobile broadband infrastructure adoption and quality”. 20th ITS Biennial Conference, Rio de Janeiro, Brazil, 30 Nov. - 03 Dec. 2014: The Net and the Internet - Emerging Markets and Policies

[8] C. Garbacza and H. G. Thompson Jr, "Demand for telecommunication services in developing countries", Telecommunications Policy, vol. 31, 276---289, 2007.

[9] J.P. Lasso, A. Ashraf Awadelkarim Widaa, J. Markendahl “Mobile network sharing trends in developing and developed mobile marks (regulations and market forces). A comparison between selected Latin American markets ans Sweden”, 24th European Regional Conference of the International Telecommunication Society, Florence, Italy, 20-23 October 2013

[10] WORLD BANK DATABASE

http://data.worldbank.org/indicator, Last review: 22th April 2015

[11] IMF – INTERNATIONAL MONETARY FUND

http://www.imf.org/external/ns/cs.aspx?id=28, Last review: 22th April 2015

[12] E. Baranes, J-C Poudou , “Internet Access and investment incentives for broadband service providers”., 22nd European Regional Conference of the International Telecommunications Society (ITS2011), Budapest, 18 - 21 September, 2011: Innovative ICT Applications - Emerging Regulatory, Economic and Policy Issues

[13] S. Salcedo-Sanz , J. Del Ser , I. Landa-Torres , S. Gil-Lopez2 , and A. Portilla-Figueras “The Coral Reefs Optimization Algorithm: An Efficient Meta-heuristic for Solving Hard Optimization Problems“ Proceedings, 15th Applied Stochastic Models and Data Analysis (ASMDA2013). International Conference, Mataró (Barcelona), Spain 25 - 28 June 2013

[14] S. Salcedo-Sanz, A. Pastor-Sánchez, D. Gallo-Marazuela, A. Portilla-Figueras. “A Novel Coral Reefs Optimization Algorithm for Multi-Objective Problems,” The 14th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'2013), Heifei, China, October, 2013

[15] AGENDA 21

http://web.archive.org/web/20120606053242/http://www.un.org/esa/dsd/agenda21_spanish/index.shtml, Last review: 22th April 2015

[16] CITEL - INTER-AMERICAN TELECOMMUNICATION COMISSION

http://www.citel.oas.org/, Last review: 8th April 2015

[17] ERO - EUROPEAN RADIOCOMMUNICATIONS OFFICE

http://www.cept.org/, Last review: 9th April 2015

[18] CTU - CARIBBEAN TELECOMMUNICATIONS UNION

http://www.ctu.int/, Last review: 10th April 2015

[19] S. Salcedo-Sanz, J. Muñoz-Bulnes, J. A. Portilla-Figueras and J. del Ser, "One-Year-Ahead Energy Demand Estimation from Macroeconomic Variables using Computational Intelligence Algorithms", Energy Conversion and Management, in press, 2015.

[20] Gomez-Muñoz, V. M., & Porta-Gándara, M. A. (2002). Local wind patterns for modeling renewable energy systems by means of cluster analysis techni ques. Renewable Energy, 2, 171–182.

[21] Mitra, S., & Banka, H. (2006). Multi-objective evolutionary biclustering of gene expression data. Pattern Recognition, 39, 2464–2477.

[22] C. Qiuru, L. Ye, X. Haixu, L. Yijun and Z. Guangping, "Telecom customer segmentation based on cluster analysis," 2012 International Conference on Computer Science and Information Processing (CSIP) pp.1179,1182, 24---26 Aug. 2012.

[23] L. Kaufman and P.J. Rousseeuw, “Finding Groups in
ata: An Introduction to Cluster Analysis”. New York: John Wiley & Sons, 1990.

[24] L. Ye, C. Qiuru, X. Haixu, L. Yijun and Y. Zhimin, "Telecom customer segmentation with K---means clustering," 7th International Conference on Computer Science & Education (ICCSE), 2012, pp.648,651, 14---17 July 2012

[25] Q. Li, "An Algorithm of Quantitative Association Rule on Fuzzy Clustering with Application to Cross---Selling in Telecom Industry," International Joint Conference on Computational Sciences and Optimization, 2009, CSO 2009,vol.1, no., pp.759,762, 24---26 April 2009

[26] L. Agustin---Blas, S. Salcedo---Sanz, S. Jimenez---Fernández, L. Carro---Calvo, J. del Ser and J. A. Portilla---Figueras, "A new grouping genetic algorithm for clustering problems," Expert Systems with Applications, vol. 39, no. 10, pp. 9695---9703, 2012.

[27] D. Arthur, Vassilvitskii, “How slow is the k-means method” In SCG ’06: Proceedings of the twenty-second annual symposium on computational geometry. ACM Press, 2006.

[28] B Lublińska-Kasprzak, “ICT sector in Poland The story of dynamic growth”, CeBit 2013, Hannover, March 2013.