An evolutive approach for the delineation of local labour markets

Francisco Flórez-Revuelta[1], José Manuel Casado-Díaz[2], and Lucas Martínez-Bernabeu[1]

[1] Research Unit on Industrial Computing and Computer Networks, University of Alicante, P.O. Box 99, E-03080, Alicante, Spain
florez@dtic.ua.es, lucas.martinez@ua.es
http://www.ua.es/i2rc
[2] Institute of International Economics, University of Alicante, P.O. Box 99, E-03080, Alicante, Spain
jmcasado@ua.es
http://iei.ua.es

Abstract. This paper presents a new approach to the delineation of local labour markets based on evolutionary computation. The main objective is the regionalisation of a given territory into functional regions based on commuting flows. According to the relevant literature, such regions are defined so that (a) their boundaries are rarely crossed in daily journeys to work, and (b) a high degree of intra-area movement exists. This proposal merges municipalities into functional regions by maximizing a fitness function that measures aggregate intra-region interaction under constraints of inter-region separation and minimum size. Real results are presented based on the latest database from the Census of Population in the Region of Valencia. Comparison between the results obtained through the official method which currently is most widely used (that of British Travel-to-Work Areas) and those from our approach is also presented, showing important improvements in terms of both the number of different market areas identified that meet the statistical criteria and the degree of aggregate intra-market interaction.

1 Introduction

Delineating local labour markets (LLMs) is an exercise that has become very common in the last decades across developed countries [1]. These sets of functional areas are seen as an alternative to the use of local and regional administrative areas as the relevant geography for statistical purposes and for the design, implementation and monitoring of labour market and other public policies in related fields such education and housing markets. The reason for this is that administrative areas are defined by boundaries that very frequently derive from historical reasons, and so it is not assured that they provide a meaningful insight of the territorial functional reality. Most countries have opted for defining markets through the aggregation of units which are intimately related in terms of exchange of flows. Thus accordingly to their nature in most developed countries travel-to-work commuting flows have been used to identify LLMs instead of defining markets characterized by the homogeneity of the constituting geographical basic units in certain attributes (a review of procedures which concentrate on this last option can be found in [20]).

A LLM represents an area where the majority of the interaction between workers seeking jobs and employers recruiting labour occurs. This refers to what Goodman [2] called external perfection (the boundary of the area is rarely crossed in daily journeys to work) and is joined by high degree of intra-market movement (so that the defined market is internally active and so as unified as possible) to form the basis of the ideal LLM. More than a decade ago Eurostat [3] established a code of good practices to guide the selection of a specific procedure: (1) the ideal map of LLMs should be based on statistical criteria, thus defined in a consistent way to allow comparison for statistical and policy purposes, (2) the procedure should allow the delineation of boundaries between areas within which most people both live and work, (3) each basic spatial unit should be in one, and only one LLM, (4) contiguity should be respected, (5) a certain degree of self-containment should be reached, so that most of the LLM's workers live in that area and most of the area's employed residents should work locally, (6) the map should consist on homogeneous units whose size should overpass a minimum threshold, (7) the areas defined should not be unnecessarily complex from a topographic point of view, (8) the map of LLMs should respect where possible the standard administrative top tier boundaries, this being considered advantageous from both statistical and policy points of view and finally (9) the procedure should be flexible enough to allow evaluation and adjustment, although the possibility of varying the statistical criteria between regions must be excluded. The preference for detail (delineating as many criteria-meeting LLMs as possible) is also frequently included as one additional criterion.

Despite sharing a common basic view about the ideal features of such an area, current official methods have a very diverse nature and are mostly based in sets of rules whose sophistication substantially varies nationally and, to a certain degree, temporally. In [4] several classifications of these official procedures are presented. One of the procedures that has been more successfully applied is that of Coombes et al. [5] which has been used in the United Kingdom for the delineation of LLMs (so-called *Travel-to-Work Areas*, TTWAs) since the 80s. This procedure has also been used, with minor changes, to define LLMs in Italy [6], [7], [8], Spain [9], New Zealand [10] and Australia [11], among other countries. This is the procedure that serves as inspiration for the one proposed in the article. In our proposal the regionalisation problem is presented as the maximization of markets' internal cohesion in terms of travel-to-work subject to a number of restrictions among which stands meeting certain self-containment and minimum size (in terms of occupied population) thresholds, with the aim of identifying as many independent markets as possible, and without making use of contiguities constrictions or distance measures. Unlike most current procedures, the method proposed here meet the criteria listed above and means a significant improvement in measurable indicators such as the number of LLMs identified which meet the stated criteria compared with alternative methods.

Given the size of the problem, which can be characterized as NP-complete, an exhaustive search of the solution is not possible, this is the reason why an evolutive approach is proposed where specific operators and strategies have been designed and implemented in experimentation using the latest figures available for Spain [12].

2 Problem description

Let $A=\{A_1,A_2,\ldots,A_n\}$ be a set of areas (territory). The objective is to obtain the set of regions $R=\{R_1,R_2,\ldots,R_m\}$ so as $\bigcup_{i=1}^{m} R_i = A$ and $R_i \cap R_j = \emptyset, \forall i,j \in [1,m], i \neq j$, that maximizes the fitness function

$$\operatorname{card}(R)\cdot\sum_{\forall i\in A} f(i) \tag{1}$$

where

$$f(i)=\frac{W^2_{\{i\},R(i)-\{i\}}}{W_{\{i\},A}\cdot W_{A,R(i)-\{i\}}}+\frac{W^2_{R(i)-\{i\},\{i\}}}{W_{R(i)-\{i\},A}\cdot W_{A,\{i\}}} \tag{2}$$

being $R(i)$ the region containing area i, and

$$W_{R_s,R_t}=\sum_{\forall i\in R_s}\sum_{\forall j\in R_t} W_{ij} \tag{3}$$

where W_{ij} is the number of commuters from area i to area j, that is the number of employed residents in area i that work in area j.

$f(\cdot)$ represents the interaction index between an area and the rest of the region to which it belongs, while the introduction of the number of regions tries to maximize the division of the territory.

Besides, each one of the regions $R_i \in R$ must fulfil two constraints of self-containment (β_1, β_2), $\beta_2 \geq \beta_1$ and minimal size (β_3, β_4), $\beta_3 \geq \beta_4$:

$$\min\left(\frac{W_{R_i,R_i}}{W_{R_i,A}};\frac{W_{R_i,R_i}}{W_{A,R_i}}\right)\geq\beta_1 \tag{4}$$

$$W_{R_i,A}\geq\beta_4 \tag{5}$$

A trade-off between both constraints has been introduced similarly to [5], but in the formulation proposed by Casado [9]. According to this trade off, the self-containment absolute requisite is relaxed for regions which are sufficiently large following a linear relationship. This trade-off establishes a new constraint:

$$\begin{gathered}\min\left(\frac{W_{R_i,R_i}}{W_{R_i,A}};\frac{W_{R_i,R_i}}{W_{A,R_i}}\right)\geq a+b\cdot W_{R_i,A}\\ a=\beta_2-b\beta_4\\ b=\frac{\beta_2-\beta_1}{\beta_4-\beta_3}\end{gathered} \tag{6}$$

We have also included a requisite to guarantee some degree of contiguity by employing only commuting data: an area can only belong to a region if some of the γ areas to/from it has more output/input commuting flows is also part of that region.

3 Evolutive proposal

The structure of the evolutive algorithm for the regionalisation of the territory is:

```
Produce a random initial population of size n
Repeat
      Evaluate fitness of all individuals
      Generate new individuals by recombination
      Generate new individuals by mutation
      Evaluate fitness of all new individuals
      Order all individuals (old and new) by fitness
      Generate a new population choosing the n best indi-
      viduals
Until there were no change in the best individual for a
number of iterations
```

3.1 Individual representation

The individuals of the population represent feasible solutions, that is, the aggregation of all the geographical basic areas composing territory A into no over-lapping local labour markets (regions). Each individual is represented by a vector of n components, each of which corresponds to an area of A, and takes the value of the identifier of the region the area belongs to.

1	2	1	3	2	1	3	2	3	4	Individual

$R_1 = \{a_1, a_3, a_6\}$ $R_2 = \{a_2, a_5, a_8\}$ $R_3 = \{a_4, a_7, a_9\}$ $R_4 = \{a_{10}\}$

3.2 Selection

Selection of the individuals to be affected by recombination and mutation operations is performed following a ranking method [13], according to which those individuals scoring higher in the fitness function have a larger probability of being selected.

3.3 Recombination operators

Due to the large number of constraints that the individuals must fulfil, and very notably to the fact that in a regionalisation exercise it is important to guarantee the exhaustive coverage of the territory and the avoidance of overlapping between regions, the usual operator of recombination does not in many cases lead to feasible solutions.

This is the reason why we have designed a wide group of specific operators which allow a more rapid evolution of the population to acceptable solutions:

– *Rec1*: a crossover point is randomly selected. Offspring is generated by taking the initial part of one of the parents and the final part of the other one. This is the usual operator employed in genetic algorithms. However, unacceptable offspring is a frequent result of this operator in this specific case, since frequently there is not a compatible correspondence between the region identifiers of both parents.

Parent #1	1	2	1	**3**	2	1	**3**	2	3	4	
Offspring	1	2	1	**3**	2	3	**4**	2	4	5	Crossover point= 4
Parent #2	1	2	3	**4**	2	3	**4**	2	4	5	

To avoid such discrepancies in the codification of the regions of both parents two new operators of recombination have been introduced:

– *Rec2*: A region identifier belonging to parent #1 is randomly chosen. The areas with identifiers lower or equal to the chosen one are inherited by the offspring. The rest of the areas are then assigned the identifiers of parent #2, except for the cases when this involves a region for which one or more of its constituting areas were already in the offspring. In such cases, the areas take the identifiers from parent #1.

Parent #1	1	2	1	3	2	1	3	2	3	4	
Offspring	1	2	1	4	2	**1**	4	2	4	5	Crossover region = 2
Parent #2	1	2	3	4	2	3	4	2	4	5	

– *Rec3*: a crossover point is randomly selected. For the areas previous to that point, the offspring takes the values of parent #1. From that crossover point, values from parent #2 are inherited, unless this involves a region with an area already set in the offspring, when the identifier of parent #1 is used

Parent #1	1	2	1	3	2	1	3	2	3	4	
Offspring	1	2	1	3	2	**1**	**3**	2	**3**	5	Crossover point = 4
Parent #2	1	2	3	4	2	3	4	2	4	5	

Since the areas characterised by lower identifiers are also assigned to regions with lower identifiers, their probability of being taken from parent #1 is greater than that of areas with high identifiers. To cope with this we have added two recombination operators (*Rec4* and *Rec5*), as variations of Rec2 and Rec3 respectively. In them a random recoding of the regions in the representation of both parents is performed previously to the recombination.

3.4 Mutation operators

We have designed an extensive set of mutation operators, some of them specifically intended for the delineation of local labour market areas, with the aim to accelerate the obtaining of individuals with adequate fitness:

- *Mut1*: This is the mutation operator usually employed in evolutionary computation. The only difference is that instead of muting just one gene (area), we mute a randomly selected number of genes, changing the region they belong to.
- *Mut2*: This operator is analogous to Mut1. In this case, however, instead of choosing the region assigned to the muted area on a random basis, such area is merged with its optimal region, that is, the region with a higher interaction index with it:

$$R'(i) = \arg\max_{\forall R_j \in R - R(i)} \left(\frac{W^2_{\{i\},R_j}}{W_{\{i\},A} \cdot W_{A,R_j}} + \frac{W^2_{R_j,\{i\}}}{W_{R_j,A} \cdot W_{A,\{i\}}} \right) \quad (7)$$

- *Mut3*: In this case, two randomly selected regions are merged.
- *Mut4*: A region is randomly chosen. Each of its constituting areas is then merged with its optimal region (see Mut2). So, as results of this operator, the number of regions in the offspring is one less compared to its parent.
- *Mut5*: This operator divides a region into two. The splitting process is as follows:
  1. A region R_i is randomly selected. This region must fulfil two constraints: (a) $W_{R_i,A} > 2\beta_4$ and (b) $W_{R_i,A} - W_{focus(R_i),A} > \beta_4$ where $focus(R_i) = \arg\max_{\forall a \in R_i} \left(W_{\{a\},A} + W_{A,\{a\}} \right)$ (that is, the region is large enough).
  2. An area belonging to R_i is randomly chosen. It is then assigned to the new region $R_i^{'}$.
  3. Another area belonging to R_i is randomly chosen. It is then assigned to the new region $R_i^{''}$.
  4. The rest of the areas belonging to R_i are taken at random, being assigned to region $R_i^{'}$ or $R_i^{''}$ which they have a greater interaction index with.
- *Mut6*: This operator creates a new region from another one by removing from the latter a number of areas sufficiently large so as to form a valid market:
  1. Similar to Mut5.
  2. An area belonging to R_i is chosen at random, being assigned to the new region $R_i^{'}$.
  3. If region $R_i^{'}$ does not fulfil the size constraint (equation 5), it takes the area belonging to R_i with which it has a higher interaction index. This process is repeated until $R_i^{'}$ is large enough.
- *Mut7*: This operation removes from a region the areas that score lower in the interaction index when measured with regards the rest of the region. Such areas are then assigned to their optimal regions:
  1. Similar to Mut5.
  2. The area to remove is selected as:

$$s = \arg\min_{\forall j \in R_i} \left(\frac{W^2_{\{j\},R_i-\{j\}}}{W_{\{j\},A} \cdot W_{A,R_i-\{j\}}} + \frac{W^2_{R_i-\{j\},\{j\}}}{W_{R_i-\{j\},A} \cdot W_{A,\{j\}}} \right) \quad \textbf{(8)}$$

3. If $W_{R_i-\{s\},A} > \beta_4$ (R_i is large enough), the area s is assigned to its optimal region, and step 2 is repeated. If that condition is not fulfilled, mutation is finished.

– *Mut8*: An exchange of areas between regions is performed. One area is randomly chosen and it is assigned to its optimal region. One area of that optimal region is then transferred to the source region.
– *Mut9*: This operator is similar to Mut2 in the sense that areas are assigned to their optimal regions. In this case, however, instead of a single area a group of them is transferred. Such a group is chosen so that the relationships among its component areas are high. The process is as follows:
  1. An area i is randomly selected.
  2. The k areas belonging to $R(i)$ with which area i has more interaction are also selected. k is chosen at random.
  3. All the selected areas are assigned to the optimal region for area i.
– *Mut10*: As, in some cases, there is a great interaction between regions, this operator tries to redistribute areas in such regions. The procedure is:
  1. A number $k \geq 2$ of regions to mute is randomly chosen.
  2. A region R_i is selected at random.
  3. The k-1 regions that have a higher degree of interaction with R_i are selected.
  4. These regions are then disintegrated into their constituting areas.
  5. k areas from this new group are selected at random. These areas act as seeds for the new regions.
  6. The rest of unassigned areas are individually taken at random and merged with their optimal region among those k new regions.

4 Experimentation

Our proposal has been implemented for the delineation of local labour markets in the Region of Valencia, Spain, using data about travel-to-work derived from the Spanish Census of Population [12]. This data allow us to build a 541x541 matrix (541 is the number of municipalities constituting the Region), where each cell represents W_{ij}.

Parameters employed in the following examples are: size population = 100, offspring size = 123 with the following individuals from the application of the different operators of recombination and mutation (5 for each recombination operator, 30 for mutations 5 and 6; and 6 for each one of the other mutation operator), iterations without changes in the best individual to stop the process = 1,000. Since one of the criteria stated in the introduction section of the paper is Detail, i.e. reaching the highest possible number of independent LLMs, division operators are over considered.

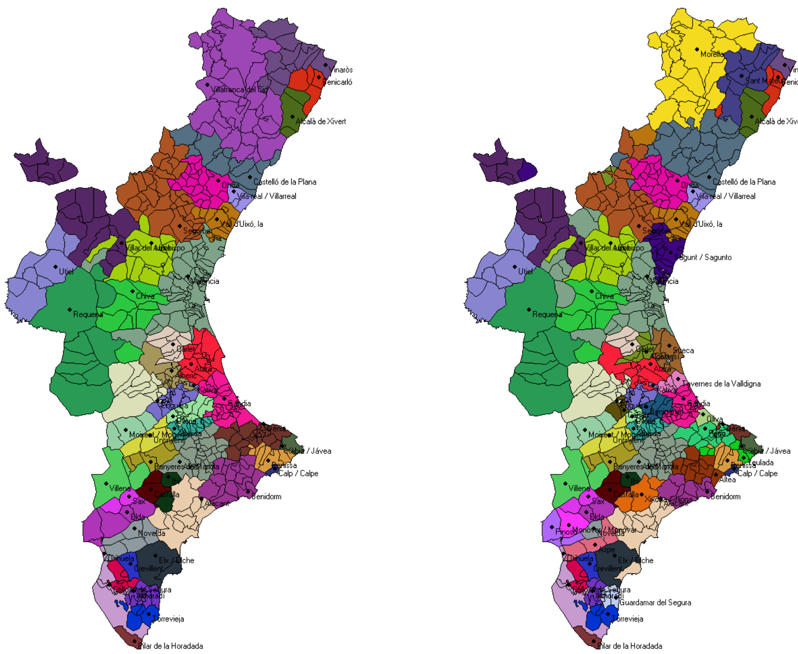

Fig. 1. Comparison between the delineation employing Coombes method (left) and our evolutionary approach (right)

The thresholds for the self-containment and minimum size conditions (equations 4 to 6) are: $\beta_1 = 0.7$, $\beta_2 = 0.75$, $\beta_3 = 20{,}000$ and $\beta_4 = 3{,}500$; that is, the levels used in the British procedure for the delineation of Travel-to-Work Areas. This allows the comparison of our results with those from that procedure. Parameter γ of "neighbourhood" is established to 5.

Our algorithm has been executed 100 times. Results depicted in Table 1 and Figure 1 are quite straightforward. The number of independent markets identified through the evolutive procedure is approximately 35 per cent larger compared to the results reached through the use of the British official method which has become the standard in the field, as already noted in Section 1. In this sense, this procedure performs clearly better according to one of the good practices criteria listed in Section 1, that of Detail. In territorial terms it is clear that the evolutive procedure manages to identify independent LLMs following a nested pattern in which LLMs identified in the TTWAs' method are divided into LLMs which keep on meeting the statistical requisites which are the same (notably self-containment, criterion 5, and minimum size, criterion 6), but with little variation of the external boundaries of such markets. Also criteria 2 and 3 are fully met by our procedure. Regarding criterion (1), the procedure proposed here is clearly based on statistical properties of the areas considered, and it is not subject to subjectivism (which is the main concern in that criterion), although as in any other genetic algorithm procedure, it is affected by a lack of determinism in the results that could at least potentially be relevant in a policymaking context. Assessing the degree of accomplishment of criterion (8) is difficult here due to space constraints, although it can be stated that the procedure proposed meets this criterion in a degree that is at least equal to that of the TTWAs procedure. Finally, and concerning criterion (4), the number of discontinuities is higher in our evolutionary approach (although the absolute number remains low considering that no information on geographical distance or contiguity between basic areas has been used in the procedure). It is important to state, however, that these are raw results, and that the observed discontinuities can be solved in any case, as it was in the TTWAs case, through the application of a final calibration stage in which residual areas are assigned to LLMs they share boundaries with through a decision rule based on an interaction measure.

Table 1. Comparative results between traditional method (Coombes) and our evolutionary proposal

	Coombes method	Our proposal		
		Best individual	**Mean**	**Standard deviation**
Number of labour markets	46	62	59.76	1.07
Fitness function	129.18	191.03	182.43	3.29
Non-contiguous regions	4	7	-	-

5 Conclusions and current works

The degree of success in the delineation, implementation and monitoring of public policies in different contexts (Statistics, labour markets, housing markets, transportation, urban planning…) heavily depends on the adequateness of the geographical reference. Official methods for the delineation of functional areas which serve as a reference for these purposes have until now rely on procedures that very frequently were designed some decades ago and that can now be improved through the use of new procedures such evolutionary computation that allow to deal with complex datasets in a different way so as to reach better results. In this piece of work we model the regionalisation problem as one of optimisation which is then solved through a genetic algorithm based on operators and strategies that have been designed to meet the specificity of the problem. The need for exhaustively covering the territory and avoiding overlapping is one of the more characteristic features of the regionalisation problems. The experimental results show that, once the respect of statistical constraints such minimum size or minimum separation between functional regions is granted, the proposed method performs better and identifies a number of LLMs that is significantly larger than that resulting from current official methods whilst it manages to meet all the criteria that has been included in the codes of good practices like such of Eurostat, the Statistical Office of the European Commission.

The major concern in this policy making context is undoubtedly the fact that the use of our evolutionary approach does not guarantee that the results of the regionalisation exercise would remain unaltered in different trials. Despite giving place to worse results in the referred terms, traditional methods are consistent through different applications. Further research is needed on the way the convergence can be assured in a reasonable time and on the reduction of uncertainty. Different solutions are to be explored in immediate research: statistics extracted from different independent executions [14], [15], parallel evolutionary algorithms [16], application of other evolutionary proposals for clustering as the Grouping Genetic Algorithms [17], [18] and multi-objective optimization [19].

Acknowledgements

José M. Casado-Díaz has received financial support from the Spanish Department of Education and Science (ref. BEC2003-02391) through a program partly funded by the European Regional Development Fund (ERDF). Lucas Martínez-Bernabeu acknowledges financial support from the Spanish Department of Education and Science, the European Social Fund (ESF) and the University of Alicante.

References

1. OECD: Redefining Territories. The Functional Regions. OECD Publications, Paris (2002)
2. Goodman, J.F.B: The Definition and Analysis of Local Labour Markets: Some Empirical Problems. British Journal of Industrial Relations, 8(2) (1970) 179-196.
3. EUROSTAT, Coombes, M.G.: Étude sur les zones d'emploi. Document E/LOC/20. Office for Official Publications of the European Communities, Luxembourg (1992)
4. Casado-Díaz, J.M., Coombes, M.G.: The Delineation of 21st Century Local Labour Market Areas (LLMAs). In Proceedings of the 8th Nectar Conference, Las Palmas C.G. (2005)
5. Coombes, M.G., Green, A.E., Openshaw, S.: An efficient algorithm to generate official statistical reporting areas: the case of the 1984 Travel-to-Work Areas revision in Britain. Journal of the Operational Research Society, 37 (1986) 943-953
6. ISTAT-IRPET: I mercati locali del lavoro in Italia, Milan, Angeli (1989)
7. ISTAT: I sistemi locali del lavoro 1991, Rome, Istat (1997)
8. ISTAT: I sistemi locali del lavoro 2001, available at http://www.istat.it (2005)
9. Casado-Díaz, J.M.: Local Labour Market Areas in Spain: A Case Study. Regional Studies, 34(9) (2000) 843-856
10. Papps, K., Newell, J.O.: Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data. Discussion Paper nº. 443. Bonn, Institute for the Study of Labor (IZA) (2002)
11. Watts, M.: Local Labour Markets in New South Wales: Fact or Fiction?, Centre of Full Employment and Equity, WP 04-12, The University of Newcastle, Australia. (2004)
12. Censo de población (2001): Detailed results available at http://www.ine.es
13. Tomassini, M.: A Survey of Genetic Algorithms. In: Stauffer (ed.): Annual Reviews of Computational Physics, III. World Scientific, Singapore (1995) 87-118
14. Szpiro, G.G.: Tinkering with Genetic Algorithms: Forecasting and Data Mining in Finance and Economics. In: Chen, S.H. (ed.): Evolutionary Computation in Economics and Finance, tudies in Fuzziness and Soft Computing, 100. Heidelberg: Physica-Verlag (2002) 273-285
15. Fogel, D.B.: Evolutionary Computation. Towards a New Philosophy of Machine Intelligence, Third Edition. John Wiley & Sons (2006)

16. Nowostawski, M., Poli, R.: Parallel Genetic Algorithm Taxonomy. In Proceedings of the Third International Conference on Knowledge-based Intelligent Information Engineering Systems. IEEE Computer Society (1999) 88-92
17. Falkenauer, E.: Genetic Algorithms and Grouping Problems. John Wiley & Sons (1998)
18. Brown, E.C., Sumichrast, R.T.: Evaluating Performance Advantages of Grouping Genetic Algorithms. Engineering Applications of Artificial Intelligence, 18. Elsevier (2005) 1-12
19. Coello, C., Lamont, G.B. (eds.): Applications Of Multi-Objective Evolutionary Algorithms. Advances in Natural Computation, vol. 1. World Scientific Publishing Company (2004)
20. Duque, J.C.: Design of homogeneous territorial units: A methodological proposal and Applications. Unpublished PhD dissertation, Universidad de Barcelona, Spain (2004)