Computers & Geosciences 57 (2013) 124–132

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Pareto-based evolutionary algorithms for the calculation of transformation parameters and accuracy assessment of historical maps

F. Manzano-Agugliaro [a,*], C. San-Antonio-Gómez [b], S. López [c], F.G. Montoya [a], C. Gil [c]

[a] Department of Engineering, University of Almería, 04120 Almería, Spain
[b] Department of Cartographic Engineering, Geodesy and Photogrammetry, Polytechnic University of Madrid, 28040 Madrid, Spain
[c] Department of Computer Science, University of Almería, 04120 Almería, Spain

ARTICLE INFO

Article history:
Received 10 October 2012
Received in revised form
18 February 2013
Accepted 9 April 2013
Available online 29 April 2013

Keywords:
Pareto front
Evolutionary algorithm
Historical cartography
Accuracy assessment
Transformation parameters
Coordinate transformation

ABSTRACT

When historical map data are compared with modern cartography, the old map coordinates must be transformed to the current system. However, historical data often exhibit heterogeneous quality. In calculating the transformation parameters between the historical and modern maps, it is often necessary to discard highly uncertain data. An optimal balance between the objectives of minimising the transformation error and eliminating as few points as possible can be achieved by generating a Pareto front of solutions using evolutionary genetic algorithms. The aim of this paper is to assess the performance of evolutionary algorithms in determining the accuracy of historical maps in regard to modern cartography. When applied to the 1787 Tomas Lopez map, the use of evolutionary algorithms reduces the linear error by 40% while eliminating only 2% of the data points. The main conclusion of this paper is that evolutionary algorithms provide a promising alternative for the transformation of historical map coordinates and determining the accuracy of historical maps in regard to modern cartography, particularly when the positional quality of the data points used cannot be assured.

1. Introduction

Historical maps are an important part of our cultural heritage (Jenny and Hurni, 2011). These maps not only represent valuable physical artefacts but also provide an important information source for historians and geographers, who frequently incorporate historical data into Geographical Information System (GIS) (Weir, 1997; Audisio et al., 2009). The scales, coordinate systems, projections, and surveying and mapping techniques used in historical maps vary widely (Podobnikar, 2009). The different reference systems employed in different historical maps often require coordinate transformations between them (Tierra et al., 2008). As historical maps typically exhibit higher degrees of inaccuracy and uncertainty compared to contemporary cartographic databases, it is not surprising that these two issues are of particular concern in historical cartography studies and historical GIS applications (Plewe, 2003). Accuracy analysis of early maps is therefore an important topic in historical cartography (Harley, 1968).

The positional accuracy of a point on a map is defined as the difference between its recorded location on the map and its actual location on the ground or location on a source of known higher accuracy (Tucci and Giordano, 2011).

The coordinate method calculates the correlation between two sets of map coordinates of points identified by modern latitudes and longitudes (Tobler, 1966). The deviation between each computer-generated point (digitised from an early map) and the corresponding point on the modern map can be displayed as a vector indicating the direction and magnitude of the error (Ravenhill and Gilg, 1974).

A related problem is the transformation of coordinates between two different geodetic reference systems. A set of points with known coordinates in both reference systems is used to obtain the transformation parameters. The Helmert transformation is one example of a commonly used method for geodesic transformations between two reference frames (Vanícek and Krakiwsky, 1986; Vanícek and Steeves, 1996). The quality of a transformation between two sets of coordinates depends on the positional quality of the points used to calculate the transformation parameters.

However, a problem arises when there is substantial spatial uncertainty in the historical data points. In this case, certain points with gross errors will not be used in the calculation of the accuracy of the entire map. When the inaccuracy of a measurement is not objectively known, as is frequently the case for historical maps, the measured feature is defined as uncertain (Hunter and Goodchild, 1993). Although uncertainty and inaccuracy are ontologically distinct concepts, it is often difficult to measure the two separately in practice, particularly in the context of historical maps (Hu, 2010).

* Corresponding author. Tel.:+34 950015396; fax: +34 950015491.
E-mail address: fmanzano@ual.es (F. Manzano-Agugliaro).

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2013.04.010

In previous papers (Manzano-Agugliaro et al., 2012), the coordinates obtained from the historical map were displaced by the average latitude and longitude error to correct the absolute displacement error in the historical map, i.e., to correct the georeferencing error. Any points whose displacements from the corresponding points in the modern map exceeded a specified distance were then eliminated from the analysis. These points were considered to contain gross errors, either because of incorrect identification with the current points or due to a gross error in the historical map. The final estimate of the map accuracy is obtained based on the root mean square (RMS) displacement after the removal of the points with gross errors.

In this historical cartography, we cannot be certain that any given point is more accurate than another. To displace the coordinates on a historical map to positions such that the final (RMS) errors are minimal, many combinations must be performed, each time discarding the points that exceed a fixed maximum RMS for a gross or unacceptable error and then calculating the new errors. A historical map may exhibit rotation (due to projection effects) as well as horizontal and vertical scale errors in addition to latitudinal and longitudinal displacement, making the number of possible combinations that must be considered even higher. If we also aim to discard as few points as possible from the historical map, then the problem becomes multi-objective.

The multi-objective nature of these mapping problems makes the decision-making process complex. Fortunately, the increase in computational resources in recent years has allowed researchers to develop efficient computational algorithms for handling complex optimisation problems. In particular, multi-objective evolutionary algorithms (MOEAs) are known for their ability to optimise several objective functions simultaneously to provide a representative Pareto front, which is a set of problem solutions representing a trade-off between the objectives (Márquez et al., 2011). The aim of this paper is to assess the performance of evolutionary algorithms in determining the accuracy of historical maps in regard to modern cartography.

2. Concepts in multi-objective optimisation

Many of the problems faced in engineering and other disciplines are optimisation problems. Many optimisation problems are difficult to solve because of features such as non-linear formulations, constraints, and NP-hard complexity.

The techniques for solving optimisation problems generally fall into two categories. Exact techniques provide the optimal solution to a given problem but are impractical for handling NP-hard problems because of prohibitive computation time and/or memory requirements. Non-exact techniques, such as metaheuristic methods (Glover and Kochenberger, 2003), provide satisfactory (though not necessarily optimal) solutions to complex problems in a reasonable amount of time.

Most computational optimisation research has been focused on solving single-objective problems, including constraints in some cases. Nevertheless, many real-world problems require the simultaneous optimisation of several competing objectives. Several authors have proposed multi-objective algorithms based on Pareto optimisation (Baños et al., 2009, 2011) to solve these multi-objective optimisation problems (MOPs).

In contrast to single-objective optimisation problems, the solution to a MOP consists of a set of *non-dominated* solutions known as the *Pareto optimal set* rather than a single solution. A solution that belongs to this set is said to be a *Pareto optimum* and when the solutions of this set are plotted in objective space, they are collectively known as the *Pareto front*. Obtaining the Pareto front is the main goal in multi-objective optimisation.

Evolutionary algorithms are particularly desirable in the solution of multi-objective optimisation problems because they simultaneously handle a set of possible solutions, yielding an entire set of Pareto optimal solutions in a single run of the algorithm (Coello, 1999).

While a single-objective optimisation problem may have only one optimal solution, a multi-objective optimisation problem may have an uncountable set of solutions. When evaluated, these solutions produce vectors whose components represent a trade-off between the various objectives of the problem. At this point, an expert in the problem, referred to as the "decision maker" (DM), must implicitly choose one (or several) solutions by selecting one or more of these vectors.

For the sake of generality, we can assume that although in the following definitions (Talbi, 2009) we use the term *minimisation* for all of the objectives, there are problems in which several (or all) of the objectives must be maximised to generate the optimum solutions.

A multi-objective optimisation problem can be informally defined (Osyczka, 1985) as the problem of finding: "a vector of decision variables that satisfies the constraints and optimises a vector function whose elements represent the objective functions. These functions form a mathematical description of the performance criteria, which are usually in conflict with each other. Hence, the term 'optimise' means finding a solution that would give the values of all of the objective functions that are acceptable to the decision maker."

Definition 1. Multi-objective optimisation problem.

A MOP is defined as

$$MOP \triangleq \min F(x) = (f_1(x), f_2(x), \ldots, f_M(x)), x \in S$$

where $M \geq 2$, is the number of objectives, $x = (x_1, \ldots, x_k)$ is the vector representing the decision variables and S represents the set of feasible solutions associated with equality and inequality constraints and explicit bounds. $F(x) = (f_1(x), f_2(x), \ldots, f_M(x))$ is the vector of objectives to be optimised.

Since in real-world MOPs the criteria are usually in conflict, there is a need to establish other concepts to consider optimality. In that sense, a partial order relation, known as *Pareto-dominance relation* can be defined.

Definition 2. Pareto dominance. An objective function vector is defined as a dominating vector (denoted by $F(x) \prec F'(x)$) if and only if no component of it is smaller than the corresponding component of and at least one component of it is strictly smaller, that is,

$$\forall i \in \{1, \ldots, M\} : f_i(x) \leq f_i^{'}(x) \wedge \exists i \in \{1, \ldots, M\} : f_i(x) < f_i^{'}(x)$$

Definition 3. Pareto Optimality. A solution $x^* \in S$ is Pareto optimal if for every, does not dominate, that is $F(x) \prec F(x^*)$.

The concept of *Pareto optimality* is directly related to the dominance concept and was initially proposed by Edgeworth (1881) and extended by Pareto (1896).

Definition 4. Pareto optimal set. For a given MOP, the Pareto optimal set is defined as $\wp^* = \{x^*\}$.

Definition 5. Pareto front. For a given MOP and its Pareto optimal set $\wp^*$, the Pareto front is defined as $\wp\mathfrak{I}^* = \{\mathfrak{I}(x^*)\} \wp^* = \{x^*\}$.

Thus the Pareto front is the image of the Pareto optimal set in the objective space, and as such it constitutes the main goal of multi-objective optimisation. As a consequence, all the solutions belonging to the Pareto front are *indifferent* because any change in any objective x_i that would improve it degrades any amount of other objectives. The concept of indifference is illustrated in Fig. 1.