

Verificación de Taxímetros usando mediciones GPS como conjuntos borrosos

José R. Villar
villarjose@uniovi.es
Dpto. de Informática
Universidad de Oviedo
Gijón Asturias

Luciano Sánchez
luciano@uniovi.es
Dpto. de Informática
Universidad de Oviedo
Gijón Asturias

Adolfo Otero
otero@uniovi.es
Dpto. de Informática
Universidad de Oviedo
Gijón Asturias

José Otero
jotero@uniovi.es
Dpto. de Informática
Universidad de Oviedo
Gijón Asturias

Resumen

Para proceder a la verificación de un taxímetro es necesario simular el recorrido de un trayecto. Los recientes avances en ayudas a la conducción como ABS, ESP, TCS, etc. dificultan el proceso de verificación, cuando se utiliza una bancada estática en la que las ruedas tractoras giran mientras que las del otro eje no. Adicionalmente, la medida es afectada por la presión de inflado de las ruedas. Todo ello conlleva sustituir el mencionado dispositivo por otro que utilice la tecnología GPS. Debido a la naturaleza imprecisa de la tecnología GPS, ha sido necesario un preproceso de los datos con el fin de evitar posibles impugnaciones de la verificación. Mostraremos como, usando un algoritmo genético y técnicas fuzzy proporcionamos una cota del error cometido por el taxímetro que permite rechazar de forma incontestable un taxímetro cuando la medición que proporciona es incorrecta.

Palabras Clave: GPS, metrología, ITV, taxi, algoritmo genético, fuzzy.

1 INTRODUCCIÓN

El servicio proporcionado por los Taxis tiene carácter concesional, con tarifas reguladas por algún organismo oficial. Anualmente las tarifas se actualizan, implicando ajustes en los taxímetros y la consiguiente verificación de los mismos en una ITV. La tarificación depende de la longitud del trayecto y de la velocidad con que se recorra. Si se desea simular el recorrido en un trayecto calibrado, un técnico de la ITV tendría que verificar constantemente la velocidad del vehículo para utilizar en cada momento la tarifa corres-

pondiente. Esto no es adecuado, requiriéndose algún dispositivo que permita medir la velocidad y la distancia simultáneamente. El grupo de Metrología y Modelos de la Universidad de Oviedo viene encargándose desde 1990 del diseño de los dispositivos que realizan esta tarea en Asturias. Actualmente, en la verificación de los taxímetros se utiliza una bancada estática que dispone de unos rodillos sobre los que se sitúan las ruedas tractoras, con los que se mide satisfactoriamente la velocidad y la distancia. Sin embargo, la sucesiva incorporación a los turismos de diversas ayudas a la conducción dificultan el proceso de verificación, pudiendo llegar a averiar el vehículo debido a la diferencia de velocidad entre las ruedas tractoras y las del otro eje. Es por ello que se ha diseñado un dispositivo de verificación portátil, autónomo -sin necesidad de agente de ITV- y que no genere averías. Se le ha dotado de GPS y antena, así como de software adecuado. Con él, se ha conseguido registrar la posición del vehículo y su velocidad simultáneamente, consiguiendo así abaratar tanto el coste del equipo como el de la mano de obra.

Existe un claro inconveniente relacionado con la imprecisión intrínseca de los datos obtenidos mediante GPS. El grado de ésta no es constante, variando en cada medida, por lo que es necesario proporcionar una cota superior del error cometido en la medida de la distancia en caso de que la verificación resulte negativa. En la figura 1 se ilustra esta situación. El margen legal del error en distancia en España es del 10%: si un taxi presenta un error del 7% sólo se puede rechazar con seguridad si el proceso de verificación se produce con una precisión en la medida de la distancia, como mucho, del 3%. Dado que la distribución real del error cometido por los taxímetros no está libre de sesgo (gausiana que aparece a la derecha, desplazada hacia el límite legal), pequeños cambios en la tolerancia producen un gran aumento de la tasa de rechazo. Por lo tanto es fundamental calcular la mencionada cota de la forma más precisa posible y además de modo que se beneficie al propietario del taxi. Por este motivo

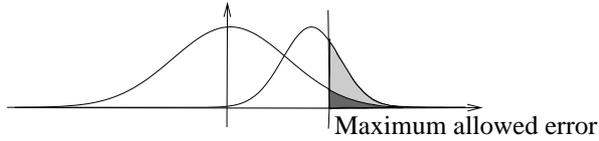


Figura 1: Distribución del error de los taxímetros: a la izquierda la esperada -libre de sesgo-, desplazada hacia la tolerancia máxima la real.

proponemos calcular la Menor Cota Superior (MCS) de todas las trayectorias compatibles con las medidas proporcionadas por el GPS.

En este artículo se presenta una nueva forma de calcular la MCS de una trayectoria a partir de datos imprecisos. Mediante un algoritmo genético multi objetivo se filtran los datos de entrada de modo que se obtenga el subconjunto de los mismos más pequeño que defina una multi poligonal que cubra los datos de entrada. La MCS se calcula a partir de esa multi poligonal mediante un heurístico determinista.

La estructura del trabajo es como sigue, en primer lugar se explica en la sección 2 como se obtienen las medidas GPS y la interpretación borrosa de las mismas. En la sección 3 se detalla el filtrado de los datos y el algoritmo determinista para el cálculo de la MCS. Posteriormente, en la sección 4 se muestran los experimentos realizados y los resultados obtenidos.

2 MEDIDAS GPS COMO DATOS FUZZY

El término Global Positioning System (GPS) se refiere a un sistema formado por un conjunto de dispositivos (satélites y receptor) cuya misión es determinar la posición del receptor. El receptor recibe de los satélites ciertas señales que le permiten calcular su longitud, latitud, altitud, número de satélites usados en la medida, fecha, etc. Cada señal que llega al receptor contiene información acerca del momento en el que partió del satélite el receptor.

Dado que se producen perturbaciones debidas a la atmósfera, desviaciones de los satélites, errores en el reloj de los satélites, errores del receptor y rebotes, es necesario disponer de un mayor número de satélites que los cuatro con los que en teoría bastarían para obtener una medida correcta.

Cuanto mayor sea el número de satélites, mejor será la precisión. Sin embargo, aún con un gran número de satélites disponibles, la configuración de la constelación de satélites es relevante a la hora de estimar la precisión de la medida. La influencia de este fac-

tor se mide con la magnitud denominada Dilution of Precision (DOP).

Los GPS comerciales proporcionan una indicación de su precisión utilizando una magnitud denominada Circular Error Probable (CEP), que informa de la probabilidad de que el receptor se encuentre en un entorno de cierto radio alrededor de la medida con determinada probabilidad. Así son comunes valores de CEP de 3 metros con un 50% de probabilidad o de 7 metros con un 90% de probabilidad.

2.1 HDOP Y PRECISIÓN DE MEDIDAS HORIZONTALES

Tal y como hemos dicho, en la precisión de las medidas realizadas con un GPS no sólo influye el número de satélites disponible sino que la geometría de la constelación de satélites también tiene impacto en dicha precisión. En el contexto de este artículo, la componente horizontal (HDOP) del DOP es la más importante, ya que está estrechamente relacionada con el CEP. A continuación se detalla como se relacionan estas magnitudes.

El RMS del error cometido en una serie de medidas GPS bajo un HDOP común se puede aproximar mediante la ecuación 1, donde A y B se pueden obtener de forma empírica para un receptor y antena específicos [17].

$$RMS_Error(HDOP) = \sqrt{(A \cdot HDOP)^2 + B^2} \quad (1)$$

La componente HDOP se relaciona con el número de satélites disponible mediante la ecuación 2, donde C , D son específicas de nuevo.

$$RMS_HDOP = \frac{C}{(\text{number of satellites})^2} + D \quad (2)$$

La probabilidad de que el error cometido en una medida sea menor que un CEP dado, condicionada a estar adquirida bajo un HDOP determinado sigue la distribución de Rayleigh, tal y como se muestra en la ecuación 3.

$$P(Err \leq CEP|HDOP) = 1 - e^{-\left(\frac{CEP}{RMS_Error(HDOP)}\right)^2} \quad (3)$$

De las ecuaciones 1 y 3 y despejando CEP, se puede obtener este en función de HDOP y la probabilidad deseada. De forma análoga se puede obtener el CEP en función del número de satélites, tal y como se observa en la ecuación 4.

$$CEP = -((A \cdot HDOP)^2 + B^2) \cdot \ln(1 - P(Err \leq CEP|HDOP))^{0.5} \quad (4)$$

2.2 INTERPRETACIÓN BORROSA DE DATOS GPS

En el contexto de las probabilidades imprecisas, tiene sentido entender un conjunto borroso como un conjunto de tolerancias, asignado a cada una de ellas un grado de confianza, de modo que la de menor confianza sea la más pequeña [10]. En concreto, se dice que dado un conjunto incompleto de intervalos de confianza para una variable aleatoria, podemos construir una variable aleatoria borrosa cuyos α -cortes son intervalos de confianza de grado $1 - \alpha$ [5], que contienen toda la información que conocemos sobre la variable aleatoria desconocida. En nuestro caso, el GPS proporciona dos intervalos de confianza, al 50% y al 90% (el mencionado CEP) y por lo tanto la representación borrosa de las coordenadas GPS es inmediata.

3 DETERMINACIÓN DE LA LONGITUD DE LAS TRAYECTORIAS USANDO DATOS BORROSOS

Los datos GPS se toman a intervalos regulares. Cada muestra es un conjunto borroso, como se ha mencionado, cuyos α -cortes son círculos. Por lo tanto cada círculo es un intervalo de confianza para las coordenadas del taxi en ese momento. Tomar los centros de esos círculos como coordenadas del taxi no es válido. Necesitamos calcular la MCS cuyos extremos están contenidos en los círculos y esta longitud será siempre mayor que el valor obtenido uniendo los centros. Este cálculo no es fácil, sin embargo. Si intentamos calcular la longitud máxima de todos los trayectos lineales a trozos cuyos extremos están en los círculos, es obvio que a menor tiempo de muestreo, más grande será la estimación de la MCS. Esto es inadecuado, la estimación debería de ser independiente del período de muestreo [14]. Hemos decidido preprocesar los datos borrosos y eliminar toda la información redundante utilizando un algoritmo genético, como se mostrará en la siguiente sección.

Si se usan datos nítidos, la simplificación de una multipoligonal se ha estudiado en [8, 12]. La alternativa más parecida a la que se presenta, a la vista de la bibliografía consultada, usa datos borrosos de una base de datos geográficos [1] para reconstruir imágenes tridimensionales mediante B-splines borrosas [3].

3.1 FILTRADO DE LOS DATOS DE ENTRADA MEDIANTE UN ALGORITMO GENÉTICO MULTI OBJETIVO

Los datos de entrada al algoritmo genético es un vector de puntos borrosos. La salida es el subconjunto de tamaño mínimo que define una trayectoria borrosa conteniendo el mayor número de puntos posible. Usando ese subconjunto de puntos y para cada α -corte, se calcula la MCS usando el heurístico que explicaremos más adelante.

Cada candidato a solución se evalúa de la forma siguiente: en primer lugar se construye la cadena de poligonales para cada α -corte de los datos seleccionados, usando las superficies tangentes a los puntos borrosos seleccionados¹. El objetivo es que la cadena contenga el máximo número de puntos y que ocupe el menor área posible. Ambos objetivos son números borrosos y definen un problema multicriterio [4] que resolveremos mediante el algoritmo NSGA-II [6, 7].

3.1.1 Codificación de cada individuo:

La codificación utiliza un vector de booleanos indicando si un punto está incluido en o ha sido eliminado de la trayectoria, de longitud arbitraria. Para generar un individuo, cada punto no extremo se incluye en el vector con probabilidad p . Los puntos extremos del camino nunca serán eliminados.

3.1.2 Operadores Genéticos:

Las definiciones de las operaciones de cruce y mutación deben reducir el número de vértices y por lo tanto no poseen sesgo. Dados dos progenitores A y B , los descendientes son dos cadenas C y D tales que $A \cap B \subseteq C$ y $A \cap B \subseteq D$; un vértice $v \in A - B$ tiene una probabilidad p^+ de estar en C , un vértice $v \in B - A$ tiene una probabilidad p^- de estar en C , donde p^- es mucho más pequeña que p^+ . El conjunto D se construye de la misma forma. La mutación se define como la eliminación aleatoria de un punto no extremo de la cadena.

3.1.3 Fitness borroso multiobjetivo:

Los objetivos utilizados en esta propuesta son el número de puntos borrosos sin cubrir y el área total de la cadena poligonal, ambos números borrosos. Es por ello que la dominancia de Pareto debe ser extendida para valores borrosos. Una manera de realizar la extensión es la α -dominancia descrita en [18]. En

¹Esta cadena puede contener algunos puntos adicionales no cubiertos por los datos de entrada, pero esto beneficiaría al usuario del servicio y por lo tanto es legal

[13] se propone obtener un rango de dominancia a partir de agregar las reglas borrosas para determinar el *grado de dominancia de x sobre y* y el *grado en que x es dominado por y* . En [11] se define la comparación entre números borrosos, de modo que la dominancia de Pareto puede usarse tal cual. En [9] se propone una generalización del concepto de dominancia de Pareto. En este trabajo se propone el uso de la α -dominancia, minimizando ambos criterios.

3.2 CALCULO DETERMINISTA DE LA LONGITUD DE LA TRAYECTORIA MAS LARGA

Una vez que los datos se preprocesan mediante el algoritmo genético, obteniendo el subconjunto más pequeño de vértices que contiene el recorrido real, necesitamos calcular su MCS.

Para cada α -corte de la B -spline borrosa que contiene la trayectoria del taxi, tenemos una cadena poligonal construida a base de trapecoides, como se ve en la figura 2. El sentido de la marcha se indica mediante una flecha. Cada vértice de un trapecoide se identifica mediante una pareja de enteros. Los que están a la izquierda de la flecha, según el sentido de la misma, tienen un cero como primer valor de ese par, los que están a la derecha un uno. El otro número es el índice del tramo. En cada tramo, la MCS pasa por $(0, i)$ o por $(1, i)$. El conjunto de vértices que forman parte de la MCS puede ser determinado por exploración exhaustiva, pero este cálculo es muy costoso e impracticable para cualquier trayectoria real del orden de 500 puntos. Este problema ha sido estudiado previamente, relacionado con el conocido como “Longest Path With Forbidden Pairs” [2] el cual es NPO-PB completo.

Debido a estas cuestiones y a que, en la práctica, los cambios en dirección y en velocidad (que implica una cierta distancia entre vértices) están limitados por la dinámica del taxi, la geometría del recorrido y el comportamiento del GPS, usamos un heurístico que es lineal con el número de vértices. El heurístico se basa en la selección de los vértices convexos: cuando un vehículo gira, el trayecto más largo que puede recorrer pasa por el exterior de las curvas que tome.

El heurístico es como sigue: el primer segmento de la MCS une los vértices $(x, 0)$ e $(y, 1)$ siendo x e y los valores escogidos de entre 1, 0 que maximizan la distancia entre esos vértices. Desde el vértice del paso 1 hasta al penúltimo, si sólo hay un vértice convexo, la MCS pasa por ese vértice. Si hay dos vértices convexos, la MCS pasa por el vértice más alejado del anterior. Si no hay vértices convexos, la MCS pasa por el vértice más alejado del anterior. El último tramo de la MCS termina

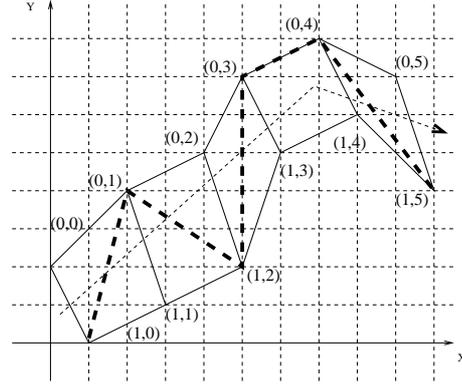


Figura 2: Ejemplo del cálculo de la MCS (línea gruesa a trazos) mediante el heurístico determinista.

en el vértice más alejado del vértice seleccionado en la etapa anterior. En la figura 2 se muestra un ejemplo de una MCS calculada mediante este heurístico, indicada por una línea gruesa a trazos.

4 EXPERIMENTOS Y RESULTADOS

En los experimentos correspondientes a este trabajo, los parámetros del algoritmo NSGA-II son los siguientes: 400 generaciones, población de 100 individuos, probabilidades de mutación y cruce 0.1 y 0.7 respectivamente, $p^+ = 0.7$, $p^- = 0.01$. Cada individuo debe cubrir un mínimo del 85% de los puntos para ser incluido en el frente de Pareto.

Hemos decidido evaluar nuestro algoritmo en un recorrido realista que cubra las situaciones encontradas habitualmente en la verificación de un taxímetro, con varias curvas, aceleraciones y deceleraciones, cambios en HDOP, CEP y número de satélites disponibles. Las longitudes y latitudes obtenidas se convirtieron en un par Northing/Easting del sistema Universal Transverse Mercator [16]. De este modo se puede utilizar la norma Euclídea para calcular distancias entre mediciones, al proyectarse la superficie terrestre en unos ejes cartesianos.

La trayectoria se ha muestreado cada segundo, la longitud total es de 21273.21 metros. En cada punto el número de satélites disponibles es una variable aleatoria uniforme que varía entre 4 y 9. A partir de estos datos, construimos un conjunto de mediciones GPS realizadas cada una en el mismo instante en el que se ha muestreado la trayectoria. El proceso de simulación se realiza de la siguiente forma: con una probabilidad igual a 0.95, se selecciona un punto que dista del correspondiente de la trayectoria menos que

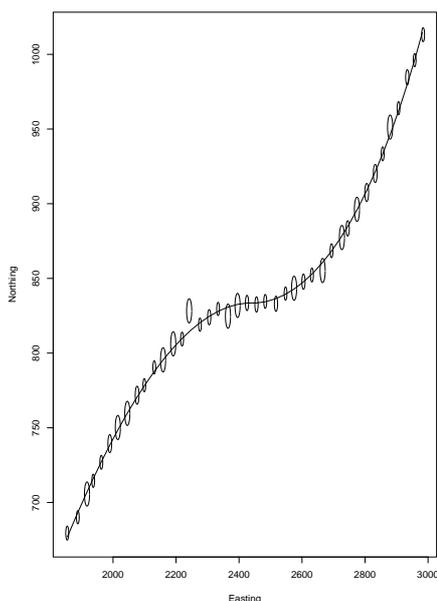


Figura 3: Parte de los datos usados en los experimentos. Los círculos (elipses por el escalado) representan las medidas GPS. La línea continua la trayectoria real. La mayor parte de los puntos de la trayectoria están en el CEP al 95%.

el CEP correspondiente al número de satélites adjudicado y a la mencionada probabilidad. Con probabilidad 0.05 el punto se toma más allá del correspondiente CEP. Este procedimiento remedia la incertidumbre que ocurre cuando se usa un GPS y los datos obtenidos sirven para averiguar como de ajustadas son las MCS obtenidas con nuestro algoritmo. En la figura 3 se muestra parte de los datos generados. Las medidas GPS se representan como círculos (elipses en realidad debido al escalado) con radio igual al CEP al 95% y la trayectoria original como una línea continua. Como se puede observar, la mayor parte de los círculos (de hecho el 95%) cortan a la trayectoria, es decir, el punto correspondiente de la trayectoria esta en el círculo de radio CEP centrado en la medida.

Hemos realizado dos experimentos con dos subconjuntos de los datos generados de tamaño 120 cada uno, que se pueden ver en la parte izquierda y derecha de la figura 4.

La longitud real de la primera trayectoria es de 3228.574 metros. La longitud de la MCS estimada, compatible con el 85% de los puntos es de 3418.81. Si el taxímetro devuelve una longitud un 10% mayor que esa cota superior, debe rechazarse, ya que incluso en el peor caso el taxímetro está fuera de tolerancia. La

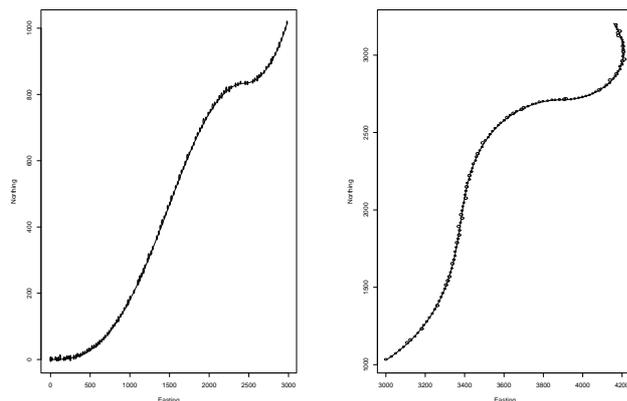


Figura 4: Trayectorias utilizadas en los experimentos.

distancia medida a través de las mediciones GPS es de 3238.521, que está mucho más cerca de los datos reales, pero el usuario del servicio puede aducir que la medida es imprecisa. Si proporcionamos una cota superior de la medida, esta argumentación no se sostiene. La longitud de la segunda trayectoria es de 2741.306 metros. La longitud de la MCS en los mismos términos que en el caso anterior es de 3250.78. En este caso, la cota es menos ajustada ya que la trayectoria tiene giros más pronunciados y por lo tanto existen MCS de mayor longitud compatibles con ella.

5 CONCLUSIONES Y TRABAJO FUTURO

Durante el desarrollo de este trabajo hemos visto que existen problemas legales si usamos directamente los datos obtenidos mediante el GPS. Los propietarios de los taxis podrían impugnar fácilmente cualquier verificación que resulta negativa esgrimiendo ese argumento. Sin embargo, debido a los problemas comentados en la sección 1, la única alternativa factible para verificar los taxímetros hoy en día es el uso de GPS. Por ello hemos decidido modificar la forma en la que una ITV presenta la información, dando una estimación de la MCS de modo que si el taxímetro está un 10% por encima de ella, entonces no hay duda de que el taxímetro está mal ajustado. Además esta alternativa es menos restrictiva, lo cual se ajusta mejor a la realidad de la distribución del error cometido por los taxímetros, sesgado hacia la tolerancia máxima. Hemos constatado que nuestro algoritmo funciona mejor cuando la trayectoria no presenta giros pronunciados, en futuras implementaciones, haciendo hincapié en la limitación de los cambios de dirección por cuestiones físicas, imitando el comportamiento dinámico de un conductor real.

El trabajo futuro incluye el uso de distintos algoritmos evolutivos como recocido simulado con operadores genéticos que es más rápido y produce mejores resultados en problemas multiobjetivo [15]. En el mismo sentido se deben probar distintos conceptos de dominancia borrosa, con el fin de estimar de forma más ajustada la MCS.

Agradecimientos

Este trabajo ha sido subvencionado por el Ministerio de Educación Español, en el marco del proyecto TIN2005-08386-C05.

Referencias

- [1] A.M. Anile, B. Falcidieno, G. Gallo, M. Spagnuolo and S. Spinello, *Modeling uncertain data with fuzzy B-splines*, Fuzzy Sets and Systems 113, 397–410, 2000.
- [2] Berman, P., and Schnitger, G. (1992), “On the complexity of approximating the independent set problem”, Inform. and Comput. 96, 77-94.
- [3] C. de Boor, *On Calculatin with B-splines*, Journal of Approximation Theory 6, 50-62, 1972.
- [4] C. A. Coello, *An Updated Survey of Evolutionary Multiobjective Optimization Techniques : State of the Art and Future Trends*, 1999 Congress on Evolutionary Computation, IEEE Service Center, 1999.
- [5] I. Couso, S. Montes, P. Gil The necessity of the strong alpha-cuts of a fuzzy set International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9-2, 249-262, 2001
- [6] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan, *A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II*, In Marc Schoenauer, Kalyanmoy Deb, Günter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, and Hans-Paul Schwefel, editors, Proceedings of the Parallel Problem Solving from Nature VI Conference, 849-858, Springer. LNCS, 2000.
- [7] K. Deb and T. Goel, *Controlled Elitist Non-dominated Sorting Genetic Algorithms for Better Convergence*, In Eckart Zitzler, Kalyanmoy Deb, Lothar Thiele, Carlos A. Coello Coello, and David Corne, editors, First International Conference on Evolutionary Multi-Criterion Optimization, 67-81. Springer-Verlag. LNCS No. 1993, 2001.
- [8] R. Estkowski and J. S. B. Mitchell, *Simplifying a polygonal subdivision while keeping it simple*, SCG '01: Proceedings of the seventeenth annual symposium on Computational geometry, 40–49, ACM Press, New York, NY, USA, 2002.
- [9] M. Farina and P. Amato, *Fuzzy Optimality and Evolutionary Multiobjective Optimization*, in Carlos M. Fonseca y otroseds., Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003, pp. 58–72, Springer. LNCS. Volume 2632, Faro, Portugal, April 2003.
- [10] Goodman, Nguyen. *Uncertainty Models for Knowledge-based Systems*. North-Holland. 1985
- [11] M. Hapke, A. Jaszkievicz and R. Slowinski, *Pareto Simulated Annealing for Fuzzy Multi-Objective Combinatorial Optimization*, Journal of Heuristics, 6(3), 329-345, August 2000.
- [12] J. Hershberger and J. Snoeyink, *Speeding up the douglas-peucker line simplification algorithm*, in Proceedings of the 5th International Symposium on Spatial Data Handling, 134–143, Charleston, United States, 1992.
- [13] M. Köppen, K. Franke and B. Nickolay, *Fuzzy-Pareto Dominance Driven Multiobjective Genetic Algorithm*, In Proceedings of the 10th IFSAWorld Congress (IFSA 2003), pages 450-453, Istanbul, Turkey, June, 2003.
- [14] N. Meratnia and R. A. de By, *Trajectory representation in location-based services : problems and solutions*, in Proceedings of the 3rd IEEE Workshop on Web and Wireless Geographical Systems (W2GIS 2003) in conjunction with the Fourth International Conference on Web Information Systems Engineering (WISE), Rome, Italy, 2003.
- [15] L. Sánchez, J. Otero and J. R. Villar, *Boosting of fuzzy models for high-dimensional imprecise datasets*, in Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU06, Paris, 2006.
- [16] Snyder, J. P., ”Map Projections Used by the U. S. Geological Survey”, 2nd edition, Geol. Survey Bulletin 1532, 313 p., U. S. Government Printing Office, Washington, D. C., 1982.
- [17] Wilson, D., David L. Wilson’s GPS Accuracy Web Page, <http://users.erols.com/dlwilson/gps.html>.
- [18] J. Zhang, B. Pham, and P. Chen, *Fuzzy Genetic Algorithms Based on Level Interval Algorithm*, In Kazmierczak, E, Eds. Proceedings The 10th IEEE International Conference on Fuzzy Systems, pages pp. 1424-1427, Melbourne, Australia, 2001.