

Localización basada en interacciones entre negocios mediante un algoritmo genético simple

Localization based on business interaction through a simple genetic algorithm

Sánchez RM¹, Galán JM, Santos JI, Izquierdo LR

Abstract: The goal of this work is to present a genetic algorithm as a decision support tool for the localization problem of retail stores in metropolitan areas. The methodology is based on the empirical estimation of positive and negative externalities that emerge as a consequence of the spatial interaction among businesses. This measure is carried out by means of complex networks modeling techniques.

Resumen El objetivo de este trabajo es presentar un algoritmo genético como herramienta de soporte a la decisión para el problema de localización de negocios minoristas en áreas metropolitanas. La metodología se basa en la obtención empírica de las externalidades negativas y positivas que surgen en la interacción espacial entre negocios. Esta cuantificación se lleva a cabo mediante técnicas de modelado de redes complejas.

Keywords: Genetic Algorithm, localization problem, complex networks, externalities, decision support tools

Palabras clave: Algoritmo genético, el problema de localización, redes complejas, externalidades, herramienta de soporte a la decisión

¹ Rosa M^a Sánchez Saiz (✉)
Escuela Politécnica Superior, Universidad de Burgos
Avda. Villadiego s/n, 09001 Burgos, Spain
e-mail: rsanchez@ubu.es

1. Introducción

Tradicionalmente tanto la literatura económica como la organización industrial han estudiado y postulado teorías sobre las relaciones que se establecen entre los diferentes tipos de negocios y cómo afectan a sus decisiones de localización. Muchos de estos modelos carecen de una formulación formal que permita medir y cuantificar estas interacciones. Recientemente, algunos autores se han atrevido a proponer metodologías que permitan empezar a entender de forma cuantitativa la naturaleza de estos fenómenos (Duranton and Overman, 2005; Jensen, 2009).

Los patrones de localización de las empresas no se pueden explicar únicamente como resultado de decisiones individuales independientes o de eventos puramente aleatorios, sino que parecen ser un resultado complejo de la actuación de una serie de fuerzas repulsivas, atractivas o de mutua indiferencia entre los distintos tipos de comercios (Hoover and Giarratani, 1984).

Las fuerzas repulsivas se manifestarían sobre todo en empresas orientadas al mercado de los factores de producción o al mercado de clientes, y también entre aquellas que no funcionan con fines lucrativos, como fundaciones, asociaciones, escuelas públicas, universidades o centros sanitarios. En este caso las empresas toman sus decisiones de precios y producción atendiendo a cómo están orientados sus clientes en el espacio, lo que condicionaría claramente la superficie del área de mercado de cada una de ellas, por las que competirían también sus rivales. La concurrencia de fuerzas repulsivas y procesos competitivos interdependientes dificulta significativamente las decisiones de localización de las empresas y genera una gran incertidumbre en las mismas. En algunos modelos de competencia espacial, como el de Hotelling (1929), podemos ver casos de concurrencia geográfica de unidades empresariales competidoras, motivados por el deseo de capturar ventas en las zonas centrales impugnadas por ambos. Sin embargo, este modelo ha sido muy criticado por su simplicidad y por no presuponer que estas fuerzas atractivas acabarían siendo compensadas, ya que la pura competición por estos clientes llevaría a unos precios más bajos e implicaría perder clientes consolidados por cada uno de ellos en sus respectivas periferias por el traslado a estas nuevas posiciones.

En cambio, existen otro tipo de actividades que se caracterizarían por fuerzas de cohesión o de agrupación, por ejemplo:

- a) Empresas que elaboran un producto diferenciado y que varía en algunos aspectos, por lo que se requiere que el comprador lo inspeccione personalmente. Es el caso de los vendedores de automóviles, distritos de zonas nocturnas, agencias de publicidad, comercios de prendas de vestir, etc., en los que los competidores se localizan agrupadamente.
- b) Empresas con un producto cambiante, con un proceso de producción en continuo cambio, con una utilización intensiva del factor de producción trabajo y un gran número de operaciones a realizar, algunas de las cuales requieren de una maquinaria específica muy cara. La concentración de

empresas en este caso, haría factible la adquisición de esa maquinaria específica ya que podrían utilizarla entre todas las empresas e incluso dar lugar a la creación de empresas especializadas en cada una de las operaciones a realizar, que podría tener de clientes a todas ellas. De esta forma se conseguiría una especialización y más eficiencia que llevaría a un abaratamiento de los costes de producción y demás.

Habría también ocasiones en las que encontraríamos una combinación de fuerzas de repulsión y de atracción entre empresas competidoras. Así, en actividades dominadas claramente por fuerzas de repulsión, que se caracterizan por un mosaico de áreas de mercado, algunos lugares tendrán más de una planta, tienda o unidad de este tipo. De igual forma en actividades muy clusterizadas como “confección de prendas de ropa” también podemos encontrar ejemplos de repulsión.

En este trabajo analizamos los fundamentos esenciales de algunas metodologías que capturan empíricamente los fenómenos de agregación y dispersión, y las fuerzas de atracción y repulsión en la interacción entre negocios. Además, aprovechando la formalización anterior, presentamos una función de optimización y un algoritmo genético para buscar localizaciones de tipos de negocio que maximice las externalidades positivas y minimice las negativas.

2. Metodología para el estudio de la localización

2.1 Análisis basado en la distancia de Duranton y Overman

Duranton y Overman (2005) desarrollan un método estadístico y, suponiendo el espacio como un continuo, comparan la distribución real de los negocios en el Reino Unido con los resultados de una distribución aleatoria en las localizaciones ya existentes, desarrollando una serie de indicadores que miden el porcentaje de concentración y dispersión industrial a lo largo de una distancia dada.

En su estudio, partieron de una base de datos de 176.106 localizaciones de negocios de manufacturas de Reino Unido. Para cada establecimiento se conoce la clasificación industrial de 5 dígitos, el tamaño de la planta y el código postal, a partir del cual, y con los datos código-punto de la Agencia Nacional de Mapas de Gran Bretaña (OS) se tienen las coordenadas geográficas.

El método toma como base una industria A con n establecimientos, lo que genera $\frac{n(n-1)}{2}$ distancias bilaterales únicas entre ellos, se aplica a todos los datos existentes y posteriormente se compara con los resultados aleatorios, utilizando un indicador desarrollado por ellos, la densidad de distancias:

$$\hat{K}(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^n f\left(\frac{d-d_{ij}}{h}\right) \quad (1.1)$$

Donde h es el ancho de banda, f la función de kernel y d_{ij} es igual a la distancia euclidiana entre los establecimientos i y j .

Para poder comparar la concentración y dispersión de una industria con la de una muestra distribuida al azar, se hacen 1000 simulaciones para cada industria, para cada simulación se prueban tantos sitios como plantas tenga la industria estudiada. Luego se calcula la densidad uniforme de distancias y posteriormente los intervalos de confianza local, entre distancias de 0 a 180 kilómetros, que es la distancia media entre parejas de establecimientos de manufacturas. Para cada kilómetro se clasifican las simulaciones en orden ascendente y elegimos los percentiles 5 y 95 para definir el intervalo de confianza más bajo y más alto, que se denotan por $\underline{K}_A(d)$ y por $\bar{K}_A(d)$ respectivamente. De esta forma se deduce que cuando la industria A tiene un $\hat{K}_A(d) > \bar{K}_{A,95}(d)$ la industria muestra localización a la distancia d con un nivel de confianza del 95%; mientras que si la industria A tiene un $\hat{K}_A(d) < \underline{K}_A(d)$ la industria mostraría dispersión a la distancia d y con un nivel de confianza del 95%. Basándose en esto, Duranton y Overman (2005) desarrollan una serie de indicadores como:

a) el índice de localización, que mide el grado de localización de la industria A a una distancia determinada d :

$$\gamma_A(d) = \max(\hat{K}_A(d) - \bar{K}_A(d), 0) \quad (1.2)$$

b) y el índice de dispersión, que mide el grado de dispersión de la industria A a una distancia d :

$$\psi_A(d) = \max(\underline{K}_A(d) - \hat{K}_A(d), 0) \quad (1.3)$$

El estudio presenta varios aspectos interesantes, como a) la idea de situar los fenómenos de aglomeración y dispersión dentro de las distancias en las que se producen; b) el considerar todo el espacio en conjunto, y no a través de unidades espaciales arbitrarias definidas por el hombre (que pueden variar de unos países a otros, como condados, estados, regiones, etc.); y c) ser el primer método desarrollado para cuantificar fenómenos de agregación y dispersión desde un punto de vista estadístico.

Como muestra, repasamos algunas de las conclusiones obtenidas por Duranton y Overman (2005). Para las industrias de 4 dígitos se comprobó que el 51% de ellas se localizan agrupadamente con un nivel de confianza del 5% y que dicha concentración tenía lugar entre los 0 y 50 kms.

También se encontró que no había marcadas diferencias entre las industrias de 4 y 5 dígitos, mientras que las de 3 tendían a mostrar patrones de localización distintos, ya que no sólo se agrupan a pequeñas escalas (0-50 Kms), sino también a nivel regional (80-140 kms). En algunas ramas industriales, la localización a nivel

industrial parece estar gobernada por los establecimientos grandes mientras que en otros sectores son los pequeños establecimientos los que tienden a agruparse.

En general la mayoría de los resultados obtenidos por Duranton y Overman (2005) confirman los hallazgos previos de la literatura, como la heterogeneidad de los modelos de localización entre industrias. No obstante, algunos de sus resultados difieren de los estudios previos, como que el hecho de que el número de industrias que tienden a agruparse no es tan grande como se pensaba y el que la tendencia a la dispersión es más frecuente de lo establecido.

Además, este método supuso un gran avance ya que se podía aplicar a unos datos con una extensión y un alcance mucho mayor de lo que se podía hacer con las herramientas anteriores. El análisis basado en la distancia puede ser aplicado con cualquier base de datos con información geográfica.

2.2 Análisis de la localización del comercio minorista utilizando herramientas de sistemas complejos.

Inspirado en el trabajo de Duranton y Overman (2005), Jensen et al (Gómez et al., 2009; Jensen, 2006; Jensen, 2009) desarrollan y ponen en práctica a lo largo de diferentes trabajos un interesante sistema para medir empíricamente la intensidad de las fuerzas de atracción o repulsión entre negocios. Supongamos un conjunto de sitios $T = \{T_i; i=1, \dots, N_T\}$ compuesto por dos subconjuntos: $A = \{A_i; i=1, \dots, N_A\}$ que representa los sitios ocupados por negocios del tipo A, y $B = \{B_i; i=1, \dots, N_B\}$ el subconjunto de sitios ocupados por negocios tipo B. Dado un sitio S, se definen las variables $N_{t(S,r)}$, $N_{A(S,r)}$ y $N_{B(S,r)}$, como el número de sitios totales, el número de sitios ocupados por negocios A y el número de sitios ocupados por B que están a una distancia menor de un radio r del sitio S, respectivamente. Para estudiar la distribución de sitios N_A consideramos la distribución puramente aleatoria que reparte los N_A y N_B negocios uniformemente en los N_T sitios. Teniendo en cuenta esto, se definen los siguientes coeficientes:

El coeficiente *intra*, que mide las relaciones de dependencia entre los negocios del tipo A. Para cada sitio del conjunto A podemos calcular el ratio local de ocupación de tipo A como $\frac{N_{A(A_i,r)}}{N_{T(A_i,r)}}$ y dividirlo por el ratio total $\frac{N_A-1}{N_T-1}$. Calculando el valor medio de estos ratios locales sobre todo el conjunto A obtenemos el coeficiente M_{AA} :

$$M_{AA}(r) = \frac{N_T-1}{N_A(N_A-1)} \sum_{i=1}^{N_A} \frac{N_{A(A_i,r)}}{N_{T(A_i,r)}} \quad (1.4)$$

La interpretación de este coeficiente resulta sencilla: existe concentración cuando $M_{AA} > 1$ (lo que indica que el número de negocios del tipo A es mayor que el valor esperado si se distribuyeran de forma aleatoria), mientras que existe dispersión cuando $M_{AA} < 1$.

Análogamente se define el coeficiente *inter*, que mide las relaciones de dependencia entre los negocios del tipo B con los del tipo A, como:

$$M_{AB}(r) = \frac{N_T - N_A}{N_A N_B} \sum_{i=1}^{N_A} \frac{N_B(A_i, r)}{N_T(A_i, r) - N_A(A_i, r)} \quad (1.5)$$

Si el valor observado del coeficiente *inter* es $M_{AB} > 1$, diremos que los comercios del tipo A tienen una tendencia a atraer a los comercios del tipo B; mientras que valores de $M_{AB} < 1$ indicarán una tendencia al rechazo entre estos negocios.

Jensen (2006;2009) aplicó esta formulación a la base de datos de comercios de la ciudad de Lyon. En el caso de las panaderías, encontró que el coeficiente *intra* mostraba dispersión solamente a una distancia de 150 metros, mientras que a distancias mayores la distribución real no se diferenciaba de la distribución aleatoria. Aplicando el coeficiente *inter* a un radio de 100 metros, se encontró que las actividades que más agregadas estaban eran las textiles, las dedicadas a artículos de segunda mano, joyerías, zapatos, mobiliario urbano y artículos del hogar.

Además pudo comprobar que estos patrones de localización se repetían igualmente en la ciudad de París. Estos resultados coincidían con la hipótesis de que situar comercios en la misma localización tiene dos efectos opuestos, por un lado incrementa el atractivo del vecindario, multiplicando las ofertas; por otro lado, divide la demanda generada entre sectores. Este efecto se compensa en aquellos casos en los que los comercios ofrecen productos diferenciados, mientras que en aquellos que no ofertan productos diferenciados (como las panaderías) sólo llevaría a una disminución de los beneficios. Entre las parejas de actividades que más se atraen entre sí estarían ropa-zapatos, joyería-artículos de cuero, artículos de segunda mano-artículos domésticos y comida-fruta y verduras.

La interacción medida mediante estos coeficientes ha sido además utilizada bajo la perspectiva de la teoría de análisis de redes para tratar de capturar de forma más agregada las relaciones que existen entre negocios (Gómez et al., 2009; Jensen, 2009). Una red es un conjunto de nodos que están relacionados entre sí a través de unos vínculos que identifican la relación (Newman, 2003). Modelando el sistema como una red, es posible representar cada nodo como un tipo de actividad y los distintos tipos de relaciones entre cada una de estas actividades como los vínculos de la red. Para calcular los pesos de los vínculos tomamos logaritmos a los coeficientes *intra* obtenidos empíricamente:

$$\alpha_{AB} = \log(M_{AB}) \quad (1.6)$$

Esta transformación de los coeficientes permite una interpretación intuitiva de los pesos de la red. Así, un valor positivo es sinónimo de una relación de atracción, y un valor negativo de una relación de repulsión.

3. Búsqueda de nichos mediante un algoritmo genético simple

Para la detección de nichos una vez cuantificados la interacción entre negocios, Jensen (2006) propone un índice matemático basado en la hipótesis de que los negocios actualmente localizados tienen una situación acorde a las externalidades tanto positivas como negativas que han sido capturadas empíricamente por la matriz de atracción/repulsión. Dado un punto geográfico x , y para una actividad de negocio A, la calidad del punto viene dada por:

$$Q_A(x, y) = \sum_B \alpha_{AB} (neA_{AB}(x, y) - \overline{neA_{AB}}) \quad (1.7)$$

Donde $neA_{AB}(x, y)$ representa el número de tiendas vecinas alrededor de (x, y) , y $\overline{neA_{AB}}$ representa el valor medio del número de establecimientos de actividades B en el entorno de una actividad A (dado un radio).

Basándonos en esta idea definimos una función de *fitness* a minimizar para localizar un nuevo establecimiento de un tipo de actividad A como:

$$F_A(x, y) = \sum_B (neA_{AB}(x, y) - \overline{neA_{AB}})^2 \quad (1.8)$$

Esta función es mínima en aquellos puntos en los que la localización se corresponde a la relación media encontrada empíricamente y aumenta cuanto más nos alejamos de esa configuración. Mientras que el indicador Q premia o penaliza las relaciones encontradas en la matriz α_{AB} , la función F asume que la relación media entre las interacciones de los negocios es óptima, por lo que penaliza las desviaciones tanto por exceso como por defecto.

Para buscar zonas atractivas de localización recurrimos a un algoritmo genético simple. Los algoritmos genéticos son algoritmos de optimización combinatoria basados en búsqueda por poblaciones (De Jong, 2006; Goldberg, 1989; Mitchell, 1998).



Fig. 1 Izquierda: mecanismo de reproducción del algoritmo. Derecha: detalle de búsqueda de localización idónea de un tipo de negocio en la ciudad de Turín (Italia). Las soluciones evaluadas muestran exploración general en el espacio de localización y fenómenos de intensificación de búsqueda en los nichos de atracción para tipo de negocio. Se ha utilizado $F_A(x, y)$ como función de fitness en este ejemplo.

En el algoritmo implementado se inicializa una población de soluciones de x,y aleatorias. En cada iteración se calcula $F_A(x,y)$ de cada solución y se aplica un mecanismo de selección por ruleta aleatoria. El algoritmo implementado no implementa cruce a través de dos progenitores, sino que a partir de cada solución elegida en el mecanismo de selección se genera una solución hija eligiendo de forma aleatoria un punto x',y' incluido en la circunferencia de radio r y centrada en el punto x,y de la solución padre (Fig. 1). El algoritmo muta con una tasa de probabilidad y termina tras un número fijo de iteraciones o de iteraciones sin mejora. El proceso de mutación supone la generación de una nueva solución candidata eligiendo de forma aleatoria un par x,y .

Agradecimientos

Queremos agradecer a Bruno Camarero Ciruelos y a Ignacio Plaza Lacalle su contribución en el desarrollo y aplicación de este trabajo. Este trabajo ha sido realizado con la financiación del Ministerio de Ciencia e Innovación a través de los proyectos CSD2010-00034 (SimulPast CONSOLIDER-INGENIO 2010), TIN2008-06464-C03-02 y DPI2010-16920 y por la Junta de Castilla y León mediante los proyectos BU034A08 y GREX251-2009.

Referencias

- De Jong, K.A., (2006). *Evolutionary computation. A unified approach*. Cambridge, Mass, MIT Press.
- Duranton, G.; Overman, H.G., (2005). Testing for localization using micro-geographic data: Review of *Economic Studies*, 72(4): 1077-1106, doi: 10.1111/0034-6527.00362.
- Goldberg, D.E., (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, Mass, Addison-Wesley Pub. Co.
- Gómez, S.; Jensen, P.; Arenas, A., (2009). Analysis of community structure in networks of correlated data: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 80(1), doi: 10.1103/PhysRevE.80.016114.
- Hoover, E.M.; Giarratani, F., (1984). *An Introduction to regional economics*. McGraw-Hill.
- Hotelling, H., (1929). Stability in competition: *Economic Journal*, 39(153): 41-57
- Jensen, P., (2006). Network-based predictions of retail store commercial categories and optimal locations: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74(3), doi: 10.1103/PhysRevE.74.035101.
- Jensen, P., (2009). Analyzing the Localization of Retail Stores with Complex Systems Tools: *Lecture Notes in Computer Science*, 5772: 10-20, doi: 10.1007/978-3-642-03915-7_2.
- Mitchell, M., (1998). *An introduction to genetic algorithms*. Cambridge, Mass, MIT Press.
- Newman, M.E.J., (2003). The structure and function of complex networks: *SIAM Review*, 45(2): 167-256, doi: 10.1234/12345678.