

Un approccio sistematico per l'analisi di mobilità tramite l'uso di immagini geolocalizzate

Monica Meini (*), Rossella Nocera (*), Gianfranco Spinelli (**)

(*), Dipartimento Bioscienze e Territorio (DiBT) dell'Università del Molise, Laboratorio MoRGaNA,
Via Duca degli Abruzzi snc, 86039 Termoli (CB), Tel. +39 0874 404924 – 404928, Fax +39 0874 404814,
monica.meini@unimol.it, rossella.nocera@unimol.it

(**) Dipartimento di Studi per l'Economia e l'Impresa (DiSEI) dell'Università del Piemonte Orientale,
Via Perrone 18, 28100 Novara, Tel. +39 0321 375423, gianfranco.spinelli@eco.unipmn.it.

Riassunto

I siti di condivisione di fotografie con sistemi di localizzazione geografica, quali Panoramio o Flickr, hanno una libreria in continua crescita. L'alto numero di fotografie presenti sul nostro territorio nazionale ha motivato gli autori del presente lavoro alla ricerca di metodologie per ricavare da questi dati nuove informazioni utili nel settore delle analisi sul comportamento spaziale legato al tempo libero.

Le fotografie diventano, così, un importante punto di partenza per lo sviluppo di cluster analysis che possono portare nuovi risultati alle imprese e alla gestione del territorio nel suo complesso. Scopo principale del lavoro è quello di introdurre un approccio sistematico per uno studio di mobilità tramite l'uso integrato di algoritmi statistici, sistemi informativi geografici e geo-visualizzazione delle immagini. Si tratta di una ricerca in itinere che vuole presentare considerazioni preliminari in merito alla elaborazione e validazione di metodologie di analisi quantitative e qualitative a partire da questo tipo di banche dati.

Abstract

The use of photo-sharing services based on georeferencing systems, such as Panoramio and Flickr, is growing rapidly in recent times. The very large number of photos of our territory uploaded on such systems motivates the development of methodologies that allow one to extract useful information in order to analyze the behavior of people during leisure.

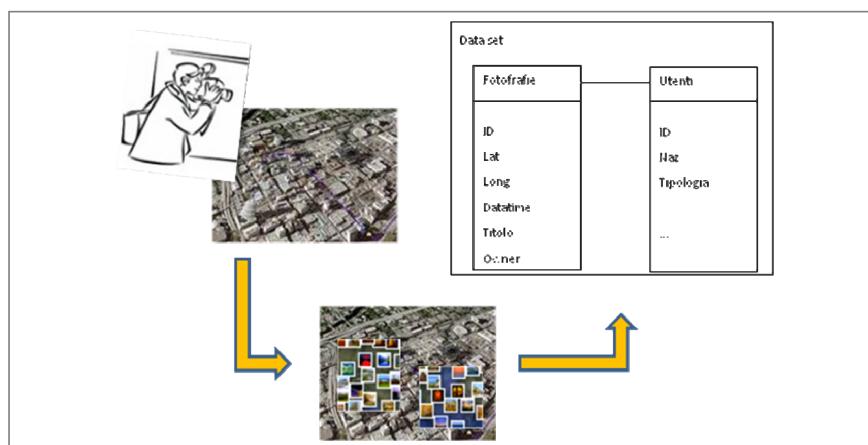
Uploaded photos thus represent a good starting point for developing cluster analysis techniques that can provide new data to territory management sectors. The main goal of this work is the introduction of a systematic approach to study people's mobility with the integration of statistical algorithms, geographical information systems and geo-visualization of the images.

Introduzione

Negli ultimi anni le informazioni presenti sulla rete crescono rapidamente e il *web* è diventato un'importante fonte per diverse attività legate al territorio. Tra queste informazioni sono particolarmente utili quelle geolocalizzate, che forniscono la posizione di fatti e fenomeni nello spazio geografico. E' il caso dei siti di condivisione di fotografie quali Panoramio o Flickr, che hanno il pregio di fornire un contenuto informativo quantitativamente rilevante costituendosi come vere e proprie banche dati territoriali. Si tratta di enormi quantità di dati per i quali si richiedono sofisticate tecniche computazionali e capacità di analisi per l'individuazione di pattern spaziali e temporali. Le informazioni ottenute dalle fotografie non rappresentano solo un evento nello spazio e nel tempo ma individuano anche traiettorie ed itinerari degli utenti.

I dati

La nostra ricerca si basa sulle fotografie rilevate sul territorio italiano e caricate in Panoramio, piattaforma molto popolare per la condivisione ed organizzazione delle foto geolocalizzate. Gli utenti acquisiscono fotografie durante i loro viaggi ed escursioni. Successivamente le caricano sulla piattaforma associandovi manualmente un riferimento geografico oppure, in casi di utenti abituati all'uso di fotocamere con GPS incorporato, l'inserimento preciso delle coordinate.



Le fotografie diventano quindi un importante punto di partenza per lo sviluppo di analisi che possono portare nuovi risultati utili alle imprese e alla gestione del territorio nel suo complesso.

Il numero totale di foto contenute nel nostro dataset è di 941651, ossia le foto caricate fino ad aprile 2009. L'alto numero di fotografie presenti sul nostro territorio nazionale stimola la ricerca di metodologie per ricavare nuove informazioni da queste enormi quantità di dati ed integrare così le classiche analisi di mercato sul comportamento spaziale dei turisti.

Al fine di studiare la tipologia dell'utente risulta molto utile il titolo delle foto. Generalmente esso contiene elementi caratteristici quali:

- 1) Data - Il titolo contiene una data (p.es. "Alberobello_16.7.2006"). Il confronto con la data di caricamento su Internet permette di assicurarsi che si tratta della data relativa allo scatto. In tale caso, ciò fornisce il mezzo principe per individuare un itinerario, tenendo conto che gli utenti tendono ad essere sistematici e coloro che mettono le date lo fanno per tutte le foto.
 - 2) Lingua usata - La lingua usata per intitolare le foto può dare indicazioni utili per la provenienza del turista; va considerato tuttavia l'uso dell'inglese come lingua veicolare internazionale.
 - 3) Accuratezza - Numero di parole contenute nel titolo (p.es. "Alberobello_16.7.2006" è meno accurato di "Alberobello chiesa parrocchiale"). L'accuratezza è uno degli elementi che permettono di stabilire la tipologia di utente o distinguere diverse categorie tipologiche di riferimento, sulle quali torneremo in seguito.
 - 4) Dettaglio - Alcune parole chiave permettono di capire se l'utente scatta foto occasionali (la rappresentazione del territorio è un "effetto collaterale" di altre attività, p.es. vacanze al mare), oppure mirate (il territorio è un elemento importante se non la ragione di viaggi/vacanze/escursioni).
 - 5) Riferimenti personali - La presenza/assenza di riferimenti personali come "noi", "io", "Alberto", "pic-nic" denota l'approccio generale delle foto (legato a sé in forma di "diario" se ricco di riferimenti personali, legato alla comunicazione del territorio se ricco di dettagli).
- Tutte queste informazioni risultano utili ai fini della selezione del campione.

Analisi e classificazione dei dati

Dal punto di vista dell'analisi spaziale, la *cluster analysis* offre un primo strumento di lavoro per giungere ad una classificazione degli utenti sulla base della distribuzione geografica delle foto.

L'analisi effettuata ha portato a individuare quattro tipi di distribuzione:

- 1) forte concentrazione in un solo cluster, che definiremo "cluster principale" (CP);
- 2) presenza di due cluster, che delineano una distribuzione bipolare;
- 3) presenza di un CP e di una serie di punti non appartenenti al CP distribuiti abbastanza omogeneamente in una macroarea (p.es. Centro-Nord);
- 4) le foto presentano una distribuzione sparsa sul tutto il territorio italiano oppure concentrazioni di lieve entità. Sono i casi in cui non esiste il cluster principale o comunque esso non supera una determinata dimensione (espressa in percentuale sul numero totale di punti) e i punti rimanenti sono sparsi abbastanza omogeneamente su tutto il territorio senza nessuna concentrazione evidente o preferenza tipologica o regionale (p.es. solo mare o esclusione evidente Nord/Sud).

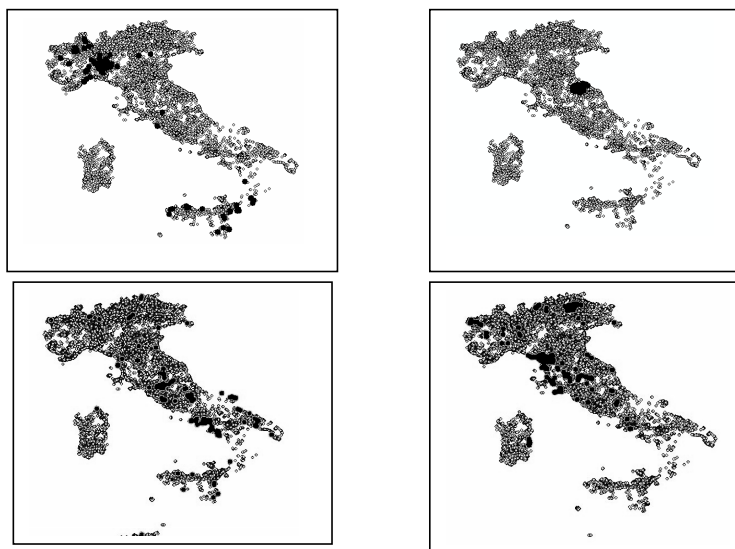


Figura 2. Alcune tipologie di distribuzione spaziale dei dati

Fonte: ns. elaborazione dati Panoramio.com

L'analisi ha permesso di individuare per la quasi totalità degli utenti il cluster più denso (CP, quello la cui media delle distanze dei punti contigui è minima). Questo cluster principale non risulta avere relazioni con l'insieme degli altri punti, per cui è necessario che i punti di questo cluster non siano analizzati assieme agli altri, in quanto presumono una diversa modalità di spostamento.

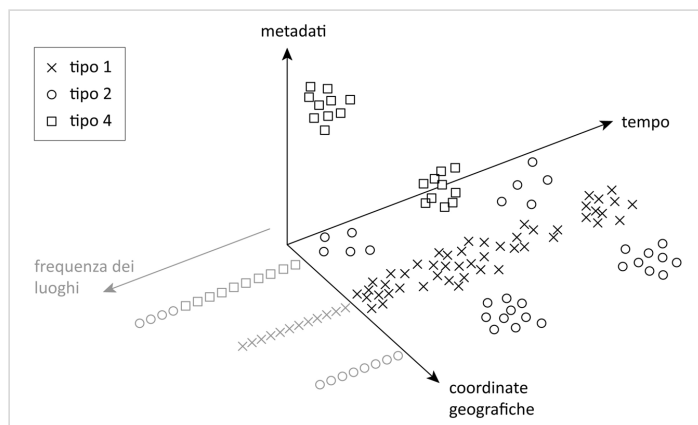
Riguardo al CP possiamo formulare varie ipotesi sul motivo della sua localizzazione:

- 1) luogo di residenza;
- 2) luogo ricorrente;
- 3) altri motivi di legame territoriale.

L'esame della dimensione del cluster principale rappresenta un primo indicatore della tipologia dell'utente; vi sono CP relativamente piccoli, altri che occupano un'intera regione. Inoltre, l'analisi dettagliata del cluster principale ne permette una prima classificazione in diverse categorie tipologiche di riferimento.

Tuttavia, per comprendere meglio l'interazione con il territorio da parte degli utenti occorre analizzare i dati in uno spazio generato, non solo dalle coordinate geografiche, ma anche dal tempo e dai metadati deducibili dalle fotografie.

In linea di principio lo scopo può essere raggiunto mediante *cluster analysis* di questo spazio opportunamente definito da coordinate numeriche. La principale difficoltà di questa operazione consiste nella conversione dei metadati in formato numerico.



Una rappresentazione schematica di questo spazio è data in figura 3, sulla base di alcuni esempi individuati nei nostri dati. Il primo caso (tipo 1) riguarda l'utente che si sposta nel tempo in un'area relativamente limitata. Si tratta, spesso, di un utente che ha un forte legame con quel territorio, egli risiede in quell'area e riporta le gite/escursioni nei dintorni, oppure, può anche trattarsi di un utente che frequenta quei luoghi abitualmente pur non essendo residente. Il secondo caso (tipo 2) riporta un'altra tipologia molto diffusa: è l'utente che si muove tra due luoghi in determinati periodi dell'anno. Uno rappresenta il luogo di residenza e l'altro il posto abituale delle vacanze. Emerge quasi sempre una traiettoria nord-sud. L'importanza di conversione dei metadati si nota soprattutto nel 3° caso (tipo 3); esso rappresenta l'utente che nel tempo si sposta su tutto il territorio in modo sparso, per cui risulta rilevante, molto più che nei casi precedenti, l'informazione dei metadati per l'ulteriore classificazione di questa tipologia. Infine, di particolare difficoltà è l'analisi di fotografie scattate nello stesso luogo ma in circostanze e confini diversi (tipo 4, esempio: Roma_chiese, Roma_compleanno); in questo caso le differenze sono riconoscibili esclusivamente con l'analisi dei metadati.

I raggruppamenti e le definizioni delle varie tipologie si definiscono, quindi, combinando le informazioni sui luoghi (distribuzione geografica) e sui criteri (deducibili prevalentemente dai metadati).

Di particolare interesse per i nostri obiettivi di lavoro è la possibilità di identificare specifici itinerari turistici a partire dalle foto scattate durante un viaggio. Occorre perciò integrare metodi matematici di cluster analysis con altri dati utilizzando, stavolta, categorie spaziali di riferimento¹:

1. "Multicluster concentrato" - Preponderanza di cluster la cui estensione massima (massima distanza tra punti) non eccede una gita giornaliera a piedi; esempi tipici: turismo urbano, mare sedentario.
2. "Cluster con satelliti" - Cluster centrati sui luoghi di pernottamento con satelliti raggiungibili in una gita giornaliera con un mezzo; esempi tipici: turismo urbano con puntate in centri limitrofi, mare con gite entroterra.
3. "CP lungo un percorso" - L'utente si muove tra due località abbastanza distanti individuando elementi di interesse tra origine e destinazione. Il territorio circostante il percorso definisce il suo CP.

¹ Le categorie individuate rimandano almeno in parte al modello di circolazione turistica di Campbell (1967).

Conclusioni

La tipologia dei dati esaminati è in continua crescita e nel tempo possono divenire vere e proprie banche dati significative. La prospettiva è quella di verificare le potenzialità che esse possono riservare per un'analisi spaziale che permetta di quantificare e catalogare con nuovi criteri la mobilità dei visitatori sul territorio nazionale. Studiare soprattutto la sequenza degli spostamenti sul territorio tra specifiche località da parte di determinati gruppi di utenti. Ciò comporta un'analisi quantitativa e qualitativa delle variabili dell'utente (affinamento di metodi e strumenti di analisi statistiche di testo) ed un'analisi spaziale e territoriale tramite un'analisi geografica dei luoghi visitati.

Riferimenti bibliografici

- Campbell C.K. (1967), "An Approach to Research in Recreational Geography", *British Columbia Occasional Papers*, 7, pp. 85-90.
- Girardin, F., Dal Fiore, F., Blat, J. & Ratti, C. (2007), "Understanding of Tourist Dynamics from Explicitly Disclosed Location Information". Proceedings of the 4th Intl Symposium on LBS & TeleCartography, Hong Kong, 8-10 November 2007.
- Goodchild M.F. (2007), "Citizens as Sensors: The World of Volunteered Geography," *GeoJournal*, 69, n. 4, pp. 211-221.
- Graham M. (2010), "Neogeography and the Palimpsests of Place", *Tijdschrift voor Economische en Sociale Geografie*, 101, n. 4, pp. 422-436.
- Kellerman A. (2002), *The Internet on Earth: A Geography of Information*, London and New York, John Wiley.
- Meini M., Nocera R., Spinelli Gf. (2012), "Flussi turistici e regionali di destinazione: come misurare la mobilità turistica", *Memorie Geografiche della Rivista Geografica Italiana*, Firenze, Florence University Press, (in corso di stampa).
- Turner A. (2006), *Introduction to Neogeography, Sebastopol - CA*, O'Reilly.
- Zook M., Devriendt L., Dodge M. (2011), "Cyberspatial Proximity Metrics: Reconceptualizing Distance in the Global Urban System", *Journal of Urban Technology*, 18, n. 1, pp. 93-114.