

Tecniche automatiche di derivazione cartografica

Sandro Savino, Massimo Rumor, Sergio Congiu

sandro.savino@dei.unipd.it, rumor@dei.unipd.it, congiu@dei.unipd.it, Università degli Studi di Padova,
Dipartimento di Ingegneria dell'Informazione, via Gradenigo 6/B, 35131 Padova

Abstract

Il laboratorio di Sistemi Geografici e Real Time (GIRTS) del Dipartimento di Ingegneria dell'Informazione dell'Università di Padova è da tempo impegnato nella ricerca, implementazione e sperimentazione di tecniche di derivazione automatica di database a scala medio-alta.

Questo articolo si propone di illustrare alcune delle più recenti tecniche sviluppate all'interno del laboratorio, che da un lato riguardano algoritmi di derivazione di alcuni particolari temi, dall'altra riguardano algoritmi per la valutazione automatica dei risultati di derivazione. In particolare sono stati sviluppati recentemente nuovi algoritmi che hanno migliorato la derivazione di alcuni temi specifici quali l'idrografia ed è stata messa a punto una nuova tecnica che, basandosi su analisi statistiche dei dati, permette di rilevare disomogeneità nel dato derivato ed è in grado, in alcuni casi, di fornire un feedback agli algoritmi di derivazione stessi al fine di raffinare l'output finale del processo.

Abstract

Since some years the Geographic Information and Real Time Systems (GIRTS) laboratory at the Department of Information Engineering of the University of Padua is active in the research and development of algorithms for the automatic generalization of cartographic databases.

In this article we will show some of our most recent developments: one is a new technique that improves our algorithms for the generalization of hydrographical networks and the other is a new process aimed at the automatic evaluation of the generalized data that is able, using statistical analysis, to detect areas where generalization errors may have occurred and, in some cases, to use this data as a feedback for the generalization algorithms in order to improve the generalized results.

Introduzione

Nel campo della geomatica la generalizzazione cartografica è uno dei temi principali di ricerca. L'obiettivo di tale ricerca è sviluppare un processo automatico per la produzione di carte derivate, ovvero un processo in grado di produrre una nuova cartografia a partire da dati cartografici esistenti a scala maggiore.

La ricerca di un traguardo così ambizioso è spinta in modo deciso dai grandi vantaggi che sarebbero portati da un sistema di derivazione automatica. Un primo vantaggio riguarda i costi: se da un lato infatti la produzione di una carta derivata eseguita a mano permette di risparmiare rispetto alla produzione di una carta ex-novo, questo processo risulta ancora molto costoso in termini di risorse necessarie (finanziarie, di personale, di tempo), mentre queste sarebbero invece ridotte drasticamente con l'utilizzo di un processo automatico.

Un secondo vantaggio riguarda l'aggiornamento dei dati: un processo di derivazione automatica permette cicli di aggiornamento più brevi e, nel caso di un ente cartografico che produce cartografie a varie scale, permetterebbe inoltre di aggiornare in maniera automatica e sincrona tutte le cartografie, propagando in modo automatico le modifiche effettuate sui dati più ad alta scala ai dati a scala minore.

A fronte di questi vantaggi, la ricerca sulla derivazione automatica è un tema molto complesso.

Una delle più grandi sfide che la ricerca in questo campo deve affrontare è il fatto che la generalizzazione è da sempre dominio di esperti cartografi, che operano integrando specifiche e linee guida con la loro abilità e sensibilità personale. Questo tipo di operatività, per la sua stessa natura di comprendere una componente quasi artistica oltre che tecnica, non è facilmente traducibile in un processo software ben definito e rende molto difficile cercare di raggiungere con una elaborazione automatica gli stessi risultati di altissima qualità raggiunti da un operatore umano.

Negli ultimi anni il consolidamento delle teorie e l'ampliamento delle tecniche disponibili ha portato un costante miglioramento dei risultati ottenuti e, benché il campo di ricerca non si possa che considerare ancora molto aperto, l'introduzione di tecniche di generalizzazione automatica nelle linee di produzione di vari enti cartografici (si veda Foerster et al, 2010) e lo sviluppo dei primi processi di derivazione completamente automatici sono un chiaro segno che la ricerca è sufficientemente matura per portare risultati concreti.

Una caratteristica che rende la generalizzazione cartografica un tema di grande fascino è la sua complessità; questa complessità è ben intuibile se si pensa in grande al processo nella sua totalità, ma si palesa anche nel piccolo, pensando alla eterogeneità e alla molteplicità di casistiche che devono essere affrontate nella derivazione di anche solo un singolo tema cartografico.

In questo articolo verranno brevemente illustrate due tecniche che bene esemplificano le complessità insite nella ricerca di un processo automatico di derivazione. Entrambe queste tecniche sono applicate al tema dell'idrografia, ma mentre una riguarda la trattazione di una particolare casistica, l'altra riguarda un approccio più generale che si applica al processo di derivazione nel suo complesso.

Queste tecniche sono state sviluppate presso il Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova e si inseriscono all'interno della ricerca per lo sviluppo di processi di derivazione automatica di dati in scala 1:25000 e 1:50000 IGM a partire dai dati in scala 1:5000 del Database Topografico Nazionale.

Il trattamento di fiumi a rami intrecciati

L'idrografia è uno dei temi principali che compongono una carta topografica. Non sorprende quindi che la derivazione dei corsi d'acqua sia uno degli argomenti più trattati nella ricerca sulla generalizzazione cartografica.

La derivazione di corsi d'acqua naturali ed artificiali è stata oggetto di numerosi lavori (si vedano ad esempio Touya, 2007, Savino et al, 2011); la maggior parte degli autori utilizzano tecniche di *data enrichment* per classificare i corsi d'acqua che compongono una rete idrografica e successivamente usano questa informazione per sfoltire la rete, rimuovendo i fiumi meno importanti con un processo di *selection*. Per il processo di *data enrichment* si utilizzano varie metriche, ad esempio sulla connessione delle aste fluviali (Horton, 1945) o sull'area drenata (Stanislawski et al, 2006) o sulla loro direzione (Thomson e Brooks, 2002) molte delle quali sono basate sul verso di scorrimento dell'acqua.

Ci sono però dei casi in cui queste tecniche non sono in grado di offrire una soluzione ottimale: è questo il caso dei fiumi a rami intrecciati. Un fiume a rami intrecciati, detto anche a canali intrecciati o pluricursale, possiede un particolare tipo di morfologia fluviale, consistente in una rete di canali d'acqua intrecciati fra loro all'interno di un alveo ghiaioso molto ampio.

Questo tipo di morfologia, a causa della scarsa pendenza delle aste fluviali e dei loro diversi andamenti, rende inefficaci le tecniche standard di *data enrichment*; inoltre proprio la speciale morfologia, fa sì che in questi casi sia da preferire l'uso dell'operatore di generalizzazione di *typification* (si veda Regnauld e McMaster, 2007) a quello di *selection* in quanto il primo è più adatto a mantenere nel dato derivato pattern spaziali caratteristici.

Per generalizzare questa particolare tipologia di fiumi a rami intrecciati, è stato sviluppato un algoritmo nuovo, che invece di analizzare le ramificazioni del fiume, analizza le isole tra essi. Il funzionamento di base dell'algoritmo è accorpare tra loro le isole più piccole al fine di formare isole più grandi. Ogni volta che si accorpano due isole vicine, di fatto si rimuove il ramo del fiume tra esse, ottenendo il risultato di semplificare il corso d'acqua.

Come primo passo l'algoritmo costruisce le isole trovando gli spazi compresi tra i rami del fiume e ne calcola la dimensione. Successivamente, partendo dalla isola più piccola, individua le isole più vicine e valuta come ognuna di queste possa essere amalgamata all'isola più piccola. Questa valutazione tiene conto di due fattori: l'isola prodotta deve essere compatta (ovvero con un alto valore del rapporto area/perimetro) e l'eliminazione del ramo fluviale non deve creare angoli acuti nel reticolo idrografico rimanente. La prima misura serve a sfavorire la formazione di isole strette e lunghe, che possono risultare di scarsa leggibilità (o rischiano di scomparire) nel caso la simbologia applicata al corso d'acqua utilizzi un tratto di spessore elevato. La seconda misura serve ad evitare che l'andamento del fiume risulti innaturale, con curve troppo brusche. Una volta trovato il candidato migliore, le due isole vengono amalgamate, riempiendo lo spazio tra esse e rimuovendo il ramo fluviale tra le due isole; il reticolo idrografico viene aggiornato di conseguenza e viene calcolata l'area della nuova isola, che sarà inserita nella lista delle isole in attesa di essere processate.

Il processo si ferma quando l'area dell'isola più piccola è superiore ad una certa soglia; il valore di questa soglia dipende dalla scala desiderata per il dato generalizzato.

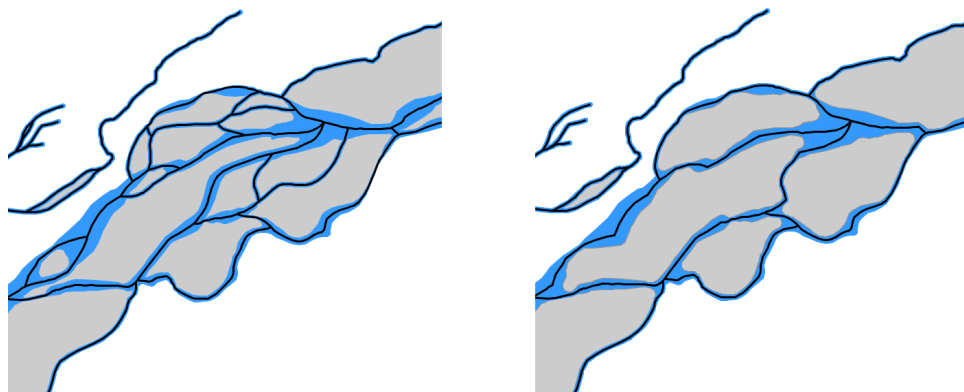


Figura 1. La generalizzazione di una sezione di un fiume a rami intrecciati: a sinistra i dati originali, a destra i dati generalizzati.

La valutazione automatica dei risultati

Uno degli aspetti più critici di un sistema automatico è la qualità dei risultati. Per poter massimizzare i vantaggi ottenuti dai sistemi automatici, infatti, bisogna ridurre il controllo umano dei risultati, molto costoso in termini di risorse. Nel campo della generalizzazione cartografica ad esempio è facile intuire che controllare manualmente ogni singolo oggetto prodotto richiederebbe un tempo paragonabile a quello necessario a derivare tali oggetti a mano, vanificando i vantaggi di un sistema automatico di generalizzazione.

Data l'importanza di questo argomento, in letteratura è possibile trovare vari lavori che riguardano la valutazione automatica della qualità dei dati generalizzati, sia nel complesso a livello di processo globale, sia nel dettaglio, analizzando vari aspetti del dato generalizzato come la topologia, la leggibilità, la classificazione (si veda ad esempio Bard, 2004, Mackaness e Ruas, 2007).

All'interno del laboratorio GIRTS si stanno sviluppando varie tecniche per la valutazione della correttezza dei dati topografici; tra le varie caratteristiche del dato generalizzato in questo articolo ci occuperemo solo di un particolare aspetto, che riguarda la rappresentazione grafica dei dati, ovvero la loro distribuzione.

Uno degli aspetti più importanti di un buon sistema di generalizzazione è essere in grado di individuare gli aspetti particolari che caratterizzano i dati di ingresso e riuscire a mantenerli nei dati in uscita; la distribuzione spaziale degli oggetti è uno di questi aspetti in grado di caratterizzare il territorio: ad esempio l'assenza o presenza di corsi d'acqua indica quanto un'area è arida, la distribuzione degli edifici può distinguere un centro storico da una periferia e la presenza di strade può identificare un'area urbanizzata da una non antropizzata (vedi Borruso, 2003).

Nonostante la distribuzione degli oggetti sia un aspetto importante da considerare (vedi Zhang et al, 2008), gli algoritmi di generalizzazione non sempre tengono conto della distribuzione, che è un valore che tipicamente si misura su spazi di un ordine di grandezza maggiore di quelli su cui lavorano gli algoritmi più semplici (ad es. la semplificazione di una linea); come risultato può capitare che applicando algoritmi a soglia fissa su *dataset* abbastanza estesi si produca un risultato troppo omogeneo, dove caratteristiche locali sono perse (si veda ad es. Stanislawski e Savino, 2011)

È stato quindi sviluppato un algoritmo apposito per gestire le variazioni di distribuzione dei dati: l'idea di base dell'algoritmo è identificare le variazioni locali nella distribuzione degli oggetti sia nel dato di ingresso che in quello in uscita e verificare se queste variazioni locali sono state mantenute o perse.

Per fare questo l'algoritmo calcola la densità locale degli oggetti usando una griglia regolare che divide in celle il territorio da generalizzare; per ogni cella viene calcolata la densità degli oggetti contenuti e i valori di densità locale sono poi usati per calcolare un valore medio di densità per la mappa totale. Successivamente viene calcolata, per ogni cella, la variazione percentuale della densità locale rispetto a quella media. Questo processo viene eseguito sia sui dati originali che sui dati generalizzati.

Poiché il processo di generalizzazione tende a ridurre il numero di dati sulla mappa, i valori di densità saranno mediamente più bassi nei dati derivati che in quelli originali; ciononostante le variazioni tra cella e cella, in linea teorica, dovrebbero essere rimaste e dovrebbero corrispondere con quelle presenti nel dato originale.

L'algoritmo confronta quindi cella per cella il valore percentuale di variazione locale rispetto alla media dei dati di input e di quelli generalizzati, segnalando in quali celle la differenza tra i due valori sia superiore ad una soglia.

Al termine del processo le celle segnalate indicano le aree in cui l'algoritmo di generalizzazione può avere cancellato troppi o troppo pochi elementi; in particolare nel caso in cui il valore della cella di input sia maggiore del valore in quella di output, i dati in quella cella sono stati probabilmente generalizzati troppo, perdendo troppa informazione; viceversa se il primo valore è più basso del secondo, è stata mantenuta troppa informazione.

Questa informazione può essere usata per richiamare l'attenzione di un operatore, che potrà controllare manualmente le celle segnalate. Le informazioni ricavate dall'analisi però possono anche essere usate per raffinare i dati derivati: è possibile infatti fornirle come *feedback* all'algoritmo di generalizzazione che, rimuovendo ulteriori elementi nelle aree segnate come sotto-generalizzate, o ripristinando elementi eliminati nelle aree segnalate come sovra-generalizzate, può ricostituire le differenze nella distribuzione degli elementi.

Questo processo è stato implementato e testato su dati idrografici in scala 1:5000 e generalizzati in scala 1:25000; la densità è stata calcolata come la somma della lunghezza delle aste fluviali contenute in ogni cella.

I risultati dell'analisi sono stati utilizzati per effettuare un feedback sull'algoritmo di generalizzazione, che è andato a generalizzare ulteriormente le celle sotto-generalizzate e a re-inserire nelle celle sovra-generalizzate alcune aste fluviali eliminate. In figura 2 è possibile vedere i risultati del processo.

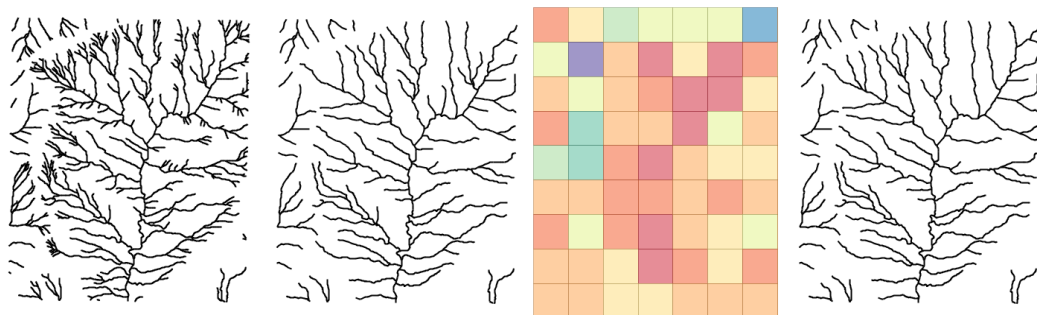


Figura 2. Nella sequenza, da sinistra a destra, si vedono: i dati originali, i dati generalizzati, il confronto tra le variazioni percentuali, il dato raffinato dopo il feedback.

Conclusioni e sviluppi futuri

In questo articolo sono state presentate due delle tecniche più recenti sviluppate all'interno del laboratorio GIRTS del Dipartimento di Ingegneria dell'Informazione dell'Università di Padova.

La tecnica di generalizzazione dei fiumi pluricursali è stata testata su alcuni dati della Regione Veneto con buoni risultati; in futuro è in programma di confrontare questa tecnica con tecniche differenti usate per la generalizzazione di corsi d'acqua con questa particolare morfologia (es. Anderson-Tarver et al, 2012) e di testare l'approccio in un contesto diverso quale l'area della foce di un fiume a delta.

Il processo di valutazione automatico della distribuzione è stato sviluppato e testato ed ha fornito risultati interessanti su reti idrografiche generalizzate; un ulteriore lavoro di ricerca è necessario per valutare i parametri di funzionamento dell'algoritmo (ad es. la forma e la dimensione della cella); uno degli sviluppi futuri previsti è di utilizzare questo processo per valutare dati generalizzati di temi diversi, quali ad esempio l'edificato e la rete stradale.

Bibliografia

- Anderson-Tarver C, Gleason M, Battenfield BP, Stanislawski LV, 2012, Automated Centerline Delineation to Enrich the National Hydrography Dataset. Proceedings GIScience 2012 Columbus, Ohio: 748, 15-28
- Bard S, 2004, Quality assessment of cartographic generalisation, Transactions in GIS 8(1), 63-81
- Borruso G (2003) Network density and the delimitation of urban areas. Transactions in GIS, 7(2), 177-191
- Foerster T, Stoter J, Kraak M-J (2010) Challenges for automated generalisation at European Mapping Agencies: a qualitative and quantitative analysis The Cartographic Journal, 47 (1), 41-54
- Horton RE, 1945, Erosional Development of Streams and Their Drainage Basins; Hydrophysical Approach to Quantitative Morphology. in Bulletin of the Geological Society of America, 56, 275-370
- Liu X, Ai T, Liu Y., 2009, Road density analysis based on skeleton partitioning for road generalization, Geo-spatial Information Science, 12 (2), 110-116
- Mackness W, Ruas A, 2007, Evaluation in the map generalisation, in: Mackness W. A., Anne Ruas A., Sarjakoski L. T. (Editors), Generalisation of geographic information: Cartographic modelling and applications, Elsevier Science, 89-111

- Merwade VM, Maidment DR, Hodges BR, 2005, Geospatial Representation of River Channels. *Journal of Hydrologic Engineering* 10(3), 243-251
- Regnauld N, McMaster RB, 2007, A Synoptic view of generalisation operators in: Mackaness W. A., Anne Ruas A., Sarjakoski L. T. (Editors), *Generalisation of geographic information: Cartographic modelling and applications*, Elsevier Science, 37-66
- Savino S., Rumor M., Zanon M., 2011, Pattern Recognition and typification of ditches. In A. Ruas (Ed), *Advances in Cartography and GIScience*. 1, 425-437
- Silverman BW, 1986, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, 22pp
- Stanislawski LV, Savino S., 2011, Pruning of Hydrographic Networks: A Comparison of Two Approaches, 14th ICA Workshop, 2011, Paris
- Stanislawski LV, Finn M, Starbuck M, Utery EL, Turley P, 2006, Estimation of Accumulated Upstream Drainage Values in Braided Streams Using Augmented Directed Graphs. *Proceedings AutoCarto 2006*, Vancouver, Washington
- Stoter J, Burghardt D, Duchêne C, Baella B, Bakker N, Blok C, Pla M, Regnauld N, Touya G, Schmid S, 2009, Methodology for evaluating automated map generalization in commercial software, in *Computers, Environment and Urban Systems*, 33(5), 311-324
- Thomson RC, Brooks R, 2002, Exploiting perceptual grouping for map analysis, understanding and generalization: The case of road and river networks. In *Graphics Recognition Algorithms and Applications*, 148-157
- Touya G, 2007, River network selection based on structure and pattern recognition. In: *Proceedings 23rd International Cartographic Conference*, A Coruña, Spain
- Xiang Z, Tinghua A, Stoter J, 2008, The evaluation of spatial distribution density in map generalization, in *The XXI Congress of the International Society for Photogrammetry and Remote Sensing*, Beijing, China, 181-188