desenvolupament de models i algorismes per processar millor les imatges.

La investigació proposada se centra en la formulació i l'anàlisi matemàtica de principis variacionals per al processament d'imatges i vídeo, i en l'anàlisi de la teoria de la mesura geomètrica i equacions en derivades parcials subjacents en aquests models. En concret, un dels principals objetius és unificar la visió local *versus* la visió no local del problema, així com les corresponents expressions matemàtiques de les idees de regularitat i autosimilitud. El treball d'investigació estarà guiat per la investigació del problema d'*inpainting* (en imatges, vídeos, vídeo monocular i vídeo estèreo).

Malauradament, a causa de la trista pèrdua del professor Vicent Caselles, recentment traspassat al mes d'agost, l'execució efectiva del projecte ha quedat truncada.

Referències

[1] P. Arias, V. Caselles i G. Facciolo. «Analysis of a variational framework for exemplar-based image inpainting». *Multiscale Modeling & Simulation*, 10(2):473–514, 2012.

[2] J.-F. Aujol, S. Ladjal i S. Masnou. «Exemplar-based inpainting from a variational point of view». *SIAM J. Math. Anal.*, 42(3):1246–85, 2010.

[3] L. Demanet, B. Song i T. Chan. «Image inpainting by correspondence maps: a deterministic approach». *Applied and Computational Mathematics*, 1100:217–50, 2003.

[4] P. Pérez, M. Gangnet i A. Blake. «Poisson image editing». *ACM Transactions on Graphics*, 22:313–318, 2003.

[5] R. Sadek, G. Facciolo, P. Arias i V. Caselles. «A variational model for gradientbased video editing». *International Journal of Computer Vision*, 103(1):127–162, 2013.

Grup de Processant d'Imatges
Universitat Pompeu Fabra

Big data: la revolució de les dades

Cada dia generem una quantitat immensa de dades sobre nosaltres: amb qui parlem per telèfon, quant de temps mirem la televisió, a quines hores comprem, quins són els llibres que més ens agraden, quins medicaments prenem, quines pàgines web visitem, etc. Totes aquestes dades, que anys enrere no es podien guardar ni, pràcticament, tractar, avui dia es poden guardar i analitzar amb facilitat, i permeten l'obtenció d'informació molt valuosa. De les grans quantitats de dades i les tècniques per emmagatzemar-les i gestionar-les se'n diu *big data*, i de la disciplina que s'encarrega de la seva modelització i anàlisi, *machine learning*.

El naixement del món *big data* ha estat possible gràcies a dos factors principals. D'una banda, les millores tecnològiques, que han incrementat de manera dràstica la capacitat de generar i guardar grans quantitat d'informació a un cost molt més baix. Aquest fet ha anat acompanyat de la descentralització de la computació vers el núvol i de la substitució dels servidors centrals, cada cop més potents i cars, per xarxes d'ordinadors utilitaris, més barats i flexibles. D'altra banda, l'augment de capacitats tecnològiques ha propiciat l'aparició de nous paradigmes de tractament de dades: nous sistemes de fitxers i bases de dades, com les bases de dades NoSQL; i nous models de programació en paral·lel, com la tècnica *map reduce*. Les bases de dades NoSQL es basen en un sistema molt flexible d'emmagatzematge i permeten guardar dades no estructurades o tabulades. Això fa possible guardar dades heterogènies ràpidament i facilita la reutilització de la informació per casos que no s'havien previst inicialment. El model *map reduce* ofereix un *framework* senzill pensat per executar programes altament paral·lelitzables sense que l'usuari hagi de ser un programador

expert. D'aquesta manera, els analistes poden processar moltes dades i obtenir resultats molt més ràpidament que amb programes seqüencials. Aquests factors han obert la gestió de *big data* a molts sectors, ja que permeten fer-ne un tractament més ràpid, senzill i, a la vegada, menys costós.

Pel que fa al *machine learning*, hi ha dues metodologies principals segons l'estil de problema que es tracti: l'aprenentatge supervisat i el no supervisat. En el primer es disposa, per a cada individu estadístic (pot ser un país, una persona, una botiga, etc.) d'unes dades amb què es pretén explicar un determinat fenomen, la variable resposta, del qual també es tenen dades. Per exemple, podem pensar a determinar la probabilitat de patir diabetis a partir de conèixer les dades de l'historial mèdic (edat, pes, sexe, malalties que pateixen, presència d'antecedents familiars de diabetis, etc.) d'un grup de pacients dels quals se sap si han desenvolupat la malaltia o no. En el segon, no es tracta d'explicar una variable en funció d'altres, sinó de conèixer-ne, determinar-ne i inferir-ne les propietats explicatives, com per exemple determinar si es poden modelitzar d'acord amb una determinada llei de probabilitat. Una de les aplicacions més importants de l'aprenentatge no supervisat és el *clustering*, que consisteix en l'agrupació d'individus estadístics en grups o clústers d'individus similars.

Hi ha moltes maneres a partir de les quals l'anàlisi de dades pot tenir impactes sorprenents en tots els sectors de l'economia. Per exemple, en sanitat, són moltes les empreses que estan desenvolupant *softwares* de previsió i tractament de malalties a partir dels historials mèdics dels pacients, o recerca sobre els efectes secundaris no desitjables en la combinació de medicaments. Recentment, la companyia americana Heritage Provider Network va finançar un concurs premiat amb tres milions de dòlars per a la previsió dels dies que un pacient estaria hospitalitzat durant l'any. En el món de les telecomunicacions, els models de fuga de clients aconsegueixen resultats tan sorprenents com el de la noruega Telenor, que va reduir la pèrdua de clients un 36 % amb una reducció de costos de les estratègies de màrqueting de retenció de clients del 40 %. Altres exemples molt rellevants es troben en la gestió del risc. Els models de propensió al pagament de deutes permeten incrementar els recobraments més d'un 10 %, amb reduccions de costos superiors al 15 %.

Tots aquests exemples són només el començament d'una nova revolució que transformarà l'economia i la societat en els propers anys. L'estudi més important de l'impacte del món *big data*, elaborat per McKinsey & Company, parla d'oportunitats de més de 500.000 milions ens els pròxims deu anys (més de dos cops el PIB d'Espanya!) en els mercats europeu i nord-americà, i apunta els professionals amb formació matemàtica i estadística com a imprescindibles per poder capturar tot aquest valor i sobrepassar les fronteres de la innovació i la productivitat.

Ferran A. Mazaira Font, llicenciat en Matemàtiques (McKinsey & Company)
Joaquim Castellà, màster en Intel·ligència Artificial (Code & Conquer, Nomorecode)