



DOCTORATS  
INDUSTRIALS

EL PLA DE  
DOCTORATS  
INDUSTRIALS

# PROJECTE DE DOCTORAT INDUSTRIAL EXPEDIENT 2013 DI 062

## DADES DE L'EMPRESA I DE L'ENTORN ACADÈMIC

### **Títol del projecte**

Detecció de Fraud a Plataformes de 'CrowdSourcing'

### **Empresa**

CA TECHNOLOGIES DEVELOPMENT SPAIN

### **Responsable de l'empresa**

Victor Muntés Mulero

### **Universitat**

Universitat Politècnica de Catalunya

### **Director/a de tesi**

Jordi Nin Guerrero

### **Treballador/a de l'empresa i doctorand/a**

David Sánchez Charles

## BREU DESCRIPCIÓ DEL PROJECTE DE RECERCA

CA Technologies treballa en la creació d'una plataforma digital que promogui la implicació de comunitats virtuals en els processos de localització de grans empreses. S'entén per procés de localització el procés que totes les empreses internacionals realitzen per traduir les seves guies d'usuari, interfícies d'usuari i textos de promoció dels seus productes a d'altres llengües. La idea principal és l'ús del crowd computing, una nova àrea d'investigació que fa recerca per entendre com es pot aprofitar la capacitat de treball de les persones connectades a les xarxes socials a través d'internet. L'objectiu del crowd computing, o crowdsourcing, és aprofitar la capacitat que tenen les persones connectades a la xarxa per realitzar tasques de manera remota i combinar-la amb la capacitat de càlcul dels computadors per crear un sistema híbrid que permeti resoldre problemes complexos. En aquest cas la tasca a resoldre és la traducció de qualitat entre dues llengües, i el model de negoci implica remunerar als treballadors de la crowd de forma justa i d'acord amb el volum de feina i l'experiència de cada individu. En aquest sentit ha internet es troben ja diverses plataformes que utilitzen alguns conceptes bàsics del crowd computing, com per exemple Mechanical Turk (mturk.com) o CrowdFlower (crowdfower.com), de manera genèrica, o algunes propostes que utilitzen el crowdsourcing per propòsits més específics com la traducció, com Smartling (smartling.com) i Gengo (gengo.com), el disseny, com Threadless, o moltes altres aplicacions.

A banda dels reptes tecnològics relacionats amb la qualitat o la gestió distribuïda de volums de dades molt grans, el crowdsourcing és sensible a possibles atacs o intents de frau perpetrats pels mateixos treballadors de la plataforma. Aquest aspecte és especialment important si tenim en compte que les activitats que es realitzen són remunerades. D'aquesta manera, subconjunts



Generalitat de Catalunya  
Departament d'Economia i Coneixement  
**Secretaria d'Universitats i Recerca**



Agència  
de Gestió  
d'Ajuts  
Universitaris  
i de Recerca



DOCTORATS  
INDUSTRIALS

## EL PLA DE DOCTORATS INDUSTRIALS

de treballadors que participen a la plataforma podrien arribar a organitzar-se per tal de lucrar-se indegudament. Per exemple, molts del sistemes d'avaluació de la qualitat es basen en la verificació de la qualitat de la feina feta per un individu per part d'altres individus de la plataforma. Grups organitzats podrien arribar a coordinar-se per tal d'avaluar-se positivament independentment de la qualitat real de la tasca feta. Cap de les plataformes esmentades anteriorment ha desenvolupat noves tècniques per detectar frau a la crowd.

Aquesta proposta vol desenvolupar una tesis doctoral que estudiï, formalitzi i proposi solucions per evitar activitats fraudulentes en sistemes de crowdsourcing. Per tal de realitzar aquesta recerca basarem el treball en tres aspectes fonamentals: la detecció eficient de patrons complexos indicadors de mal comportament, el processament d'esdeveniments complexos i l'extracció de coneixement d'alt nivell a partir d'aquest anàlisi.

El reconeixement de patrons [1] es basa en l'assignació d'etiquetes (categories) a un conjunt de valors d'entrada. Un bon exemple de reconeixement de patrons és la classificació automàtica, on s'assigna a cada valor d'entrada una etiqueta d'un determinat conjunt. Aquestes tècniques són molt utilitzades en sistemes de reconeixement de veu, on el problema és determinar quines paraules componen una frase. En aquesta proposta, volem estudiar la seva aplicació a la detecció de frau en les diferents interaccions que els usuaris realitzen dins d'una plataforma de crowdsourcing.

Detectar un patró determinat dins d'un conjunt d'interaccions és per si mateix un problema complex, però resoldre aquest problema no es suficient per a garantir la seguretat a una plataforma de crowdsourcing. Un problema diferent és el fet d'extreure coneixement (generalitzar) dels diferents patrons de frau un coneixement comú que permeti a un sistema reconèixer patrons de frau dinàmics, o fins i tot generar nous patrons de forma no-supervisada. Les xarxes bayesianes d'aprenentatge profund (deep learning) [2], són un nou tipus de xarxes neuronals jeràrquiques que aprenen de forma esglaonada diferents nivells de generalització. Aquesta característica, fa possible als sistemes informàtics extreure coneixement en diferents nivells de generalització de forma no-supervisada. Aplicar aquestes tècniques a la detecció de nous patrons de frau a plataformes de crowdsourcing és un dels objectius d'aquesta proposta.

Finalment, donat el caràcter industrial d'aquesta proposta dedicarem una part del treball al fet de modelar els patrons de frau com a esdeveniments complexos [3]. Els sistemes de processament d'esdeveniments complexos són sistemes informàtics per a la indústria basats en restriccions i implementats com a sistemes de programació amb sencers. El fet d'implementar els sistemes de detecció de frau amb aquestes tecnologies ens permetrà estendre el seu ús a d'altres plataformes.

Bibliografia:

[1] Richard O. Duda, Peter E. Hart, David G. Stork (2001) Pattern classification (2nd edition), Wiley, New York, ISBN 0-471-05669-3

[2] Marcus, Gary, Is "Deep Learning" a Revolution in Artificial Intelligence?. The New Yorker. Nov. 2012.