

Enaitz Ezpeletak tesia irakurri du

2016/10/21

Doktoretza programa: INGENIARITZAKO DOKTORETZA PROGRAMA

Epaimahaia:

-Lehendakaria: Manel Medina Llinas dokt. (Universitat Politecnica de Catalunya)

-Bokala: Igor Santos Grueiro dokt. (Deustuko Unibertsitatea)

- Bokala: Per Magnus Almgren dokt. (Chalmers University of Technology)

- Bokala: José Ramón Méndez Reboredo dokt. (Vigoko Unibertsitatea)

-Idazkaria: Iñaki Garitano Garitano dok. (Mondragon Unibertsitatea)



Tesiaren laburpena:

Hartzailearen baimenik gabe bidalitako mezuak egunean milioika erabiltzailereri eragiten dien mehatxua dira. Nahiz eta spam detekzio tresnek gero eta emaitza hobekoak lortu, arazoa konpontzetik oso urruti dago oraindik, batez ere spam kopuruari eta erasotzaileen estrategia berriei esker.

Hori gutxi ez eta azken urteetan sare sozialek izan duten erabiltzaile gorakadaren ondorioz, non milioika erabiltzailek beraien datu pribatuak publiko egiten dituzten, gune hauek oso leku erakargarriak bilakatu dira erasotzaileentzat. Batez ere bi arlo interesgarri eskaintzen dituzte webgune hauek: perfiletan pilatutako informazio guztiaren ustiapena, eta erabiltzaileekin harreman zuzena izateko erraztasuna (perfil bidez, talde bidez, orrialde bidez...). Ondorioz, gero eta ekintza ilegal gehiago atzematen ari dira webgune hauetan.

Spam mezuen helburu nagusienak zerbait saldu, alarma soziala sortu, sentzibilizazio kanpainak martxan jarri, etab. izaki, mezu mota hauek eduki ohi duten idazketa mezua berauen detekziorako erabilia izan daiteke.

Lan honen helburu nagusiak ondorengoak dira: alde batetik, sare sozialetako informazio publikoa erabiliz egungo detekzio sistemak saihestuko dituen spam pertsonalizatua garatzea posible dela; eta bestetik hizkuntza naturalaren prozesamendurako teknikak erabiliz, testuen intenzionalitatea atzeman eta spam-a detektatzeko metodologia berriak garatzea. Gainera, sistema horiek sare sozialetako spam mezuekin lan egiteko gaitasuna ere izan beharko dute.

Lehen helburu hori lortzekolan honetan spam pertsonalizatua diseinatu eta bidaltzeko sistema bat aurkeztu da. Era automatikoan erabiltzaileen informazio publikoa ateratzen dugu sare sozial ospetsu batetik, ondoren informazio hori aztertu eta txantilo ezberdinak garatzen ditugu erabiltzaileen iritzia kontuan hartuaz. Behin hori egindakoan, hainbat esperimentu burutzen ditugu spam normala eta pertsonalizatua bidaliz, bien arteko emaitza ezberdintasuna alderatzeko.

Tesiaren bigarren zatian hiru spam atzemate metodologia berri aurkezten ditugu. Berauen helburua tribialak ez den intentzio

komertziala atzeman ta hori baliatuz spam mezuak sailkatzean datza. Intentzionalitate hori lortze aldera, analisi sentimentala eta pertsonalitate detekzio teknikak erabiltzen ditugu. Modu honetan, hiru sistema ezberdin aurkezten dira hemen: lehenengoa analisi sentimentala soilik erabiliz, bigarrena lan honetarako pertsonalitate detekzio teknikak eskaintzen dutena aztertzen duena, eta azkenik, bien arteko konbinazioa.

Tresna hauek erabiliz, balidazio esperimentalak burutzen da proposatutako sistemak eraginkorrak diren edo ez aztertzeko, hiru mota ezberdinetako spam-ekin lan eginez: email spam-a, SMS spam-a eta sare sozial ospetsu bateko spam-a.