
Resum

La detecció d'opinions falses o vertaderes al voltant d'un producte o servei s'ha convertit en un problema força rellevant de la nostra època. Segons estudis recents, fins el 80% de les persones han canviat la seua decisió final en base a les opinions revisades en la web. Algunes d'aquestes opinions poden ser falses positives, amb la finalitat de promoure un producte, o falses negatives per tal de desacreditarlo.

Per a ajudar a resoldre aquest problema es proposa en aquesta tesi un nou mètode de detecció d'opinions falses, anomenat PU-Learning*. Aquest mètode augmenta la precisió mitjançant un algoritme iteratiu i resol el problema de la falta d'opinions etiquetades.

Per al funcionament del mètode proposat, s'utilitzen un conjunt reduït d'opinions etiquetades com a falses i un altre conjunt gran d'opinions no etiquetades, del qual se n'extrauen les opinions que faltaven i, així, aconseguir una classificació de dues classes. Aquest tipus d'escenari s'ha convertit en una situació molt comuna en els corpus d'opinions de què es disposa.

Com una segona contribució es proposa una representació basada en n-gramas de caràcters. Aquesta representació té l'avantatge de capturar tant elements de contingut com a d'estil d'escriptura, permetent amb això millorar l'efectivitat del mètode proposat en la detecció d'opinions falses.

L'avaluació experimental del mètode es va dur a terme mitjançant tres experiments de classificació d'opinions utilitzant dues col·leccions diferents. Els resultats obtinguts en cada experiment permeten veure l'efectivitat del mètode proposat, així com també les diferències entre la utilització de varis tipus d'atributs.

Ja que la falsedat o veracitat de les opinions vessades pels usuaris es converteix en un paràmetre molt important en la presa de decisions, el mètode que ací es

presenta pot ser utilitzat en qualsevol corpus on es troben les característiques abans esmentades.