1.

Introducere

Sunt descrise cateva tehnici de pre-filtrare, urmate de o metoda de a cauta imagini care creste precizia. In continuare, este descris cum se anticipeaza scorul de credibilitate si se propune o abordare bazata pe filtrarea de imagini de catre user. De asemenea, este aratat si cum se imbunatateste diversitatea prin procedeul de clustering.

2.Tintind spre precizie

2.1 Pre-filtrare initiala

Sunt utilizati doi pasi de filtrare cu scopul de a elimina zgomotul din lista de imagini.

2.2 Cautarea de imagini

A fost folosit Caffe, o puternica componenta bazata pe CNN, pentru a extrage reprezentatii ale imaginilor din colectie, cat si a exemplelor de imagini de pe Wikipedia. Imaginilie au fost apoi ordonate pentru fiecare topic dupa similaritatea dintre imaginea gasita si toate exemplele de imagini. Un dezavantaj al acestei metode este pierderea mare de precizie.

3.Ascultarea aluziilor sociale

3.1 Predictia credibilitatii user-ului

A fost exploatata multimea credibilitatii pentru a forma un model regresiv ce prezice scorul de credibilitatea al unui user din componentele puse la dispozitie. Performanta predictiilor a fost evaluata utilizand coeficientul de corelatie al rangurilor Spearman.

3.2 Selectia user-ului

Pentru fiecare topic, este pastrata o submultime de useri care au adus contirbutii in topul celor n imagini gasite in ierarhia imaginilor. Apoi, ca o filtrare suplimentara, in ierarhia finala sunt retinute doar acele imagini ce provin de la userii din multimea selectata. Fiind data buna precizie a imaginilor gasite, exista certitudinea ca imaginile gasite in topul ierarhiei sunt relevante.

4.Imbunatatind diversitatea

4.1 Clustering

In primul rand este efectuat un k-Means clustering pe multimea completa de imagini. Pentru a asigura o distributie stabila a clusterului, centrele de greutate sunt initializate prin selectarea uniforma a imaginilor din ierarhia formata dupa cautare.

4.2 Ierarhizarea clusterelor

Clusterele sunt ordonate dupa media scorului de credibilitate a userilor care contribuie cu imagini in cluster. Pentru rularile care nu permit utilizarea credibilitatii , ierarhizarea clusterelor a fost facuta in functie de numarul de useri unici reprezentati in fiecare cluster. In cazul unei egalitati, este preferat clusterul care are cea mai bine pozitionata imagine in top dupa cautarea imaginilor.Ierarhia finala este obtinuta selectand din fiecare cluster, pe rand, imaginea care este ce amai bine plasata in top.

5.Rezultate si discutii

Au fost facute 5 rulari diferite: prima folosind descriptorul vizual LBP3x3, a doua a fost pur textuala, a treia a fost om combinatie intre primele doua, a patra a fost la fel ca prima, dar a fost folosita credibilitatea userului in ierahizarea clusterelor, iar pentru ultima rulare a fost utilizat descriptorul vizual Caffe.