

Enunciat del projecte de PROP Quadrimestre de tardor, curs 25/26

Extractor de prototipus de comportament o perfils

Hem fet una enquesta i volem fer un programa per estudiar els diferents prototipus de comportament o perfils de les persones que l'han contestada. Més concretament, voldríem fer-nos una idea de quants comportaments/perfils hi ha, i com seria l'individu representatiu de cada comportament o perfil existent. Considerem que dues persones tenen un comportament similar si les respostes a l'enquesta també són similars, de manera que podem "agrupar" les persones per la semblança en les seves respostes (fem un clustering).

Les dades de l'enquesta les tindrem usualment en un fitxer de text. El programa ha de poder tractar enquestes on les preguntes siguin del tipus "escolliu una d'aquestes opcions o com a màxim n opcions, o bé donar una resposta com un string en format lliure, o bé donar una quantitat numèrica". Tot i així, ens podem trobar respostes no contestades.

Volem una aplicació que permeti, com a mínim, les següents funcionalitats:

- Creació interactiva d'enquestes i importació d'enquestes
- Resposta interactiva i importació de respostes
- Gestió completa de enquestes i respostes (esborrar, modificar, etc.)
- Anàlisi d'enquestes (obtenció dels diferents comportaments o perfils, obtenció dels representants dels diferents grups, llistat de tots els individus de cada comportament o perfil, etc.). S'ha de poder avaluar d'alguna forma la qualitat del clustering obtingut.

L'assoliment d'aquestes funcionalitats bàsiques és necessari per aprovar el projecte. Es valorarà la qualitat de la implementació proposada així com les eventuais funcionalitats addicionals.

Funcionalitats principals a entregar al primer lliurament:

- Algorisme de clustering *k-means*

Ampliació de funcionalitat als equips de 5 persones:

- S'ha d'implementar un segon algorisme de clustering, *k-medoids*

Dates dels lliuraments:

- Primer: dilluns, 17 de novembre
- Segon: dilluns, 15 de desembre
- Tercer: dilluns, 22 de desembre (lliuraments interactius a partir del 7 de gener)

Informació addicional:

Es tracta d'un problema de clustering. S'ha de resoldre el problema plantejat implementant els algorismes *k-means* i *k-means++*. *k-means* requereix una inicialització que pot ser trivial (random) o més sofisticada. Una d'aquestes inicialitzacions és l'algorisme *k-means++*, que s'ha d'implementar també. S'ha de tenir en compte que *k-means++* es pot utilitzar com a inicialització de *k-means* o com un algorisme de clustering "subòptim". En qualsevol cas, el disseny del projecte ha de ser tal que la implementació de nous algorismes o l'ampliació dels existents no impliqui un canvi en les classes preexistents.

Es publicarà documentació complementària. Tanmateix, les següents referències contenen informació al respecte:

https://en.wikipedia.org/wiki/Cluster_analysis
https://en.wikipedia.org/wiki/K-means_clustering
<https://en.wikipedia.org/wiki/K-medoids>