

Szakdolgozat kivonat

Ajánlórendszerek

Kerékgyártó Márk

Témavezetők: Benczúr András, Molontay Roland

Szakdolgozatomban egy gépi tanulást használó módszerrel, az ajánlórendszerrel, foglalkozom. Először bemutatom a két alapvető módszert, amit használ, a kollaboratív szűrő és tartalom alapú szűrő módszereket.

Ezután az ehhez szükséges adatok típusait vázolom, kétféle adattípusról írok, az explicit és implicit adatokról. Az egyik gyakran használt osztályozó algoritmuson, a k -legközelebbi szomszéd algoritmuson, keresztül mutatom be, hogyan is épül fel egy ajánlórendszer. Leírom az algoritmus alapgondolatát, és hogy hogyan valósul meg, milyen eszközök kellenek a használatához. A hasonlóság fontos tényezője, definiálok különböző hasonlósági mértékeket, amelyeket az algoritmus használhat a legközelebbi szomszédok kereséséhez.

Leírom, hogyan lehet az algoritmust az ajánláshoz használni, és mutatok rá két példát, a Kolokált előadók - legnépszerűbb zeneszámok, és egy asszociációs szabályokat megvalósító hasonlósági függvényt. Ezután írok egy sokszor előforduló problémáról, a nagy dimenzióban való távolság meghatározásáról, és hogy ez hogyan viszonyul a k -legközelebbi szomszéd algoritmushoz.

Erre két módszert vázolok, a multi-klaszterező módszert és az szinguláris érték szerinti felbontáson (SVD) alapuló módszert, amelyeknek a használatát a RecSys Challenge 2018 versenyen keresztül mutatom be. A versenyről, amit a Spotify és két egyetem rendez, tudatok néhány fontos információt, például a Million Song Dataset adathalmazról, amit külön erre készítettek. Leírom, hogy mi a verseny célja, és hogyan lesz kiértékelve az ajánlórendszer.

Készítek egy ajánlórendszert, először az adatok előkészítésével és feldolgozásával foglalkozom, majd a két említett módszert alkalmazom, ami a k -legközelebbi algoritmust valósítja meg, emellett alkalmazok egy harmadik fajtát is. A munkám során létrejövő eredményekből látszik, hogy még további kutatás szükséges a feladat teljesítéséhez, ezért kijelölök néhányat, hogy mivel lehet az elért eredményeken javítani.