

SENTIMENT ANALYSIS

ehk

Meelsusanalüüs

Maali Tars ja Kerttu Talts

Meelestuse analüüs on automaatne viis tuvastada tekstis esinevaid arvamusi, meelestatust, hinnanguid, suhtumist ja vaateid, tundeid ja emotsioone. [1]

EESMÄRK

Projekti eesmärk oli kirjutada programm, mis üritab ennustada teksti meelsust. See tähendab, et programm määrab, milline on teksti või selle kirjutaja peamine emotsioon.

1.

Valitud **andmebaas** koosnes 5112 tweedist, millele oli juba meelsus käsitsi määratud.

3.

Alustasime programmi treenimist tweetide mitmikuteks ehk **n-grammideks** tegemisega (üks- kuni neligrammid). N-grammid jagasime nelja rühma nende meelsuse järgi: **positiivsed**, **negatiivsed**, **neutraalsed**, **ebaolulised** (määramatud). Seejärel leidsime iga rühma kohta n-grammide sagedused ning valisime välja **enim esinenud n-grammid**. Nendest omakorda võtsime igale meeleolule eraldi **unikaalsed n-grammid**, mis on programmi **parameetriteks**. Neile lisame vastavalt n väärtusele **kaalud** (üksgramm saab kaalu 1, kaksgramm kaalu 2, ...).

5.

Rakendasime parameetreid testhulgale (ehk *test set*) ning arvutasime tulemuste põhjal välja kui täpselt määrab programm teksti meelsust. Seda tegime **f-score**'i meetodiga. Lõpptulemus varieerub igal programmi käivitamisel, sest sama sagedusega võib olla mitu n-grammi.

Autoritest: Maali Tars ja Kerttu Talts on loodus- ja täppisteaduste valdkonna arvutiteaduse instituudi I aasta bakalaureuse informaatika eriala tudengid.

GitHub: <https://github.com/maalitars/Sentiment-Analysis>

PROGRAMMI ETAPID

1. Andmebaas käsitsi määratud meelsustega tweetidest
2. Nendest koostatud kolm erinevat hulka: treenimishulk, silumishulk, testhulk
3. Programmi treenimine
4. Programmi silumine
5. Programmi testimine

2.

Jagasime tweedid juhuslikult **kolmeks osaks:**

4500 treenimishulka, 400 silumishulka ja ülejäänud 212 testhulka.

4.

Programmi **silumine** toimus silumishulga peal. Rakendasime saadud parameetreid silumishulgale (ehk *development set*) mitu korda ning iga korra järel **eemaldasime osa parameetreid**, mis määrasid teksti meelsust enim valesti. See muutis programmi meelsuse määramise täpsemaks. Silumise tulemusena saime **lõplikud parameetrid**, mida rakendasime testhulgale.

PROJEKTI TULEMUS

Tulemusena valmis programm, kuhu kasutaja saab sisestada soovitud teksti. Seejärel programm analüüsib seda. Vastuseks väljastatakse, kas teksti peamine emotsioon või vastav teade, kui selle meelsust polnud võimalik määrata.

[1] Timo Petmanson. Esitlus "Meelestuse analüüs". Keeletehnoloogia aine (MTAT.06.045). Tsiteeritud: [09.02.2017]. <https://courses.cs.ut.ee/2016/nlp/spring/Main/KeeletehnoloogiaLoeng14>

