

Mesterséges Intelligencia

Algoritmus specifikáció

Feladat: Egy választott író munkásságai alapján eldönteni egy idegen írásról, hogy mekkora az esély, hogy ugyanaz az író írta.

Algoritmus:

1. Adott egy szövegbázis, melyről tudjuk, hogy az adott író munkássága. Ezt tudásbázisnak nevezzük.
2. A tudásbázisban megkeressük az N legtöbbet használt szót. ($N=50$)
3. Megnézzük, hogy ez az N szó az egyes szövegekben hányszor fordul elő és leosztjuk a ($<\text{az összes szó száma a szövegben}> / 100$)-al. (100 azért van, hogy pontosabb számokat kapjunk)
4. Ezek után elvégezzük ezt a műveletet az idegen szövegre is, majd ezekből a vektorokból csinálunk egy $(M+1) \times N$ -es mátrixot (\mathbf{M}_{orig}), ahol az M a tudásbázisban lévő szövegek száma. A $+1$ az idegen vektor.
5. Kiszámoljuk az \mathbf{M}_{orig} első M (N dimenziós) pont tömegközéppontját (minden pont 1 súlyú), melyet C -nek nevezünk
6. Kiszámoljuk C és az első M pont átlagtávolságát (Dist_{AVG}) és azt állítjuk, hogy a Dist_{AVG} távolságra lévő pont 85%-os egyezést jelent.
7. Ezen két értékből kiszámoljuk az 1% távolságot: $\text{Dist}_{\text{AVG}} / (100 - 85)$. Ezt elnevezzük $\text{Dist}_{1\text{PRC}}$
8. Kiszámoljuk a C és az $M+1$. pont távolságát (Dist_{new})
9. A végleges értéket a következő képlet adja meg: $\text{Dist}_{\text{new}} / \text{Dist}_{1\text{PRC}}$. Alsó/felső határ: 0/100.