

Z Wikipedie, otevřené encyklopedie

Word2Vec je způsob [zpracování přirozeného jazyka](#), kdy jsou slova reprezentována vektory.^[1] Bezkontextový model publikovali v roce 2013 společně výzkumníci [Tomáš Mikolov](#), Kai Chen, Greg Corrado, [Ilya Sutskever](#) a [Jeff Dean](#), kteří pracovali pro Google. Algoritmus word2vec využívá model [neuronové sítě](#) k učení slovních asociací z velkého [korpusu](#) textu. Po natrénování může takový model odhalit [synonyma](#) nebo navrhnut další slova pro další část věty. Word2vec reprezentuje každé odlišné slovo určitým seznamem čísel, kterému se říká [vektor](#). Vektory jsou pečlivě vybrány tak,

ké a syntaktické vlastnosti slov; jednoduchá

aby zachycovaly sémantické

ous

kontextové vlastnosti slov

kos

sémantické podobnosti

Glove

novou reprezentaci, příjemz [BERT](#) bere v uvanu daného slova. Zatímco vektor pro slovo anglické "run" má v word2vec stejnou reprezentaci z word2vec pro oba jeho výskytů v větách "He is running a company" a "He is running a marathon", BERT poskytuje vektor pro "run", který je výrazně rozdílný, než je v tomto kontextu pro každý výskyt slova "run". Nově se využívají novější modely jako [BERT](#) či [GPT-3](#).^[2]

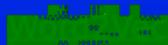
Slovo ve slovníku je umístěno v kontextu pro každý výskyt. Slovo "run" bude mít stejnou reprezentaci v různých kontextech, než je v tomto kontextu pro každý výskyt slova "run". Nově se využívají novější modely jako [BERT](#) či [GPT-3](#).

Odkazy

Reference

[Překlad textu z článku Wikipedia](#)

V tomto článku byly použity všechny citace z článku [Word2Vec](#) na [Wikipedia](#).



- [Oficiální stránka projektu](#)

Citováno z „<https://cs.wikipedia.org/w/index.php?title=Word2Vec&oldid=24823261>“

Kategorie:

- [Počítačová lingvistika](#)
- [Zpracování přirozeného jazyka](#)
- [Umělé neuronové sítě](#)

[Monitoring](#):[Články](#) přel.[češtiny](#) z [enwiki](#)

- [Monitoring:Autoritní kontrola s 0 identifikátory](#)

Hledání

Speciální:Hledání

Word2Vec

18 jazyků

[Přidat téma](#)