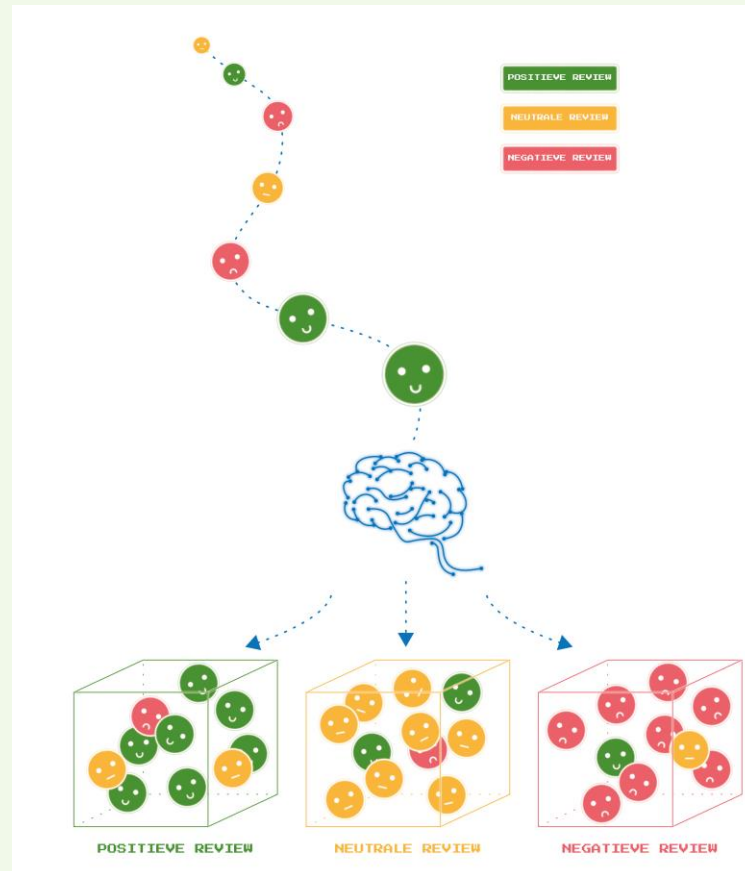


START2STEM - Activiteit uit het project Chatbot over taaltechnologie

Regelgebaseerde sentimentanalyse

Context

Bij automatische sentimentanalyse zoekt men met taaltechnologie, dus met artificiële intelligentie, naar gevoelens en opinies in teksten, zoals klantenreviews, tweets en blogs, vaak op sociale media. Naast expliciete meningen en gevoelens kan men met de technologie ook informatie uit een tekst halen die er niet letterlijk in staat. Is een klant tevreden over een bepaald product? Gaat de auteur van een tweet akkoord met een bepaalde politieke stellingname?



Artificiële intelligentie

Zo'n AI-systeem voor sentimentanalyse wordt ontwikkeld door het programmeren van algoritmes waarin expertenkennis vervat is (de kennisgebaseerde of regelgebaseerde systemen) of aan de hand van lerende algoritmes (datagebaseerde systemen).

Het AI-systeem zal aan een gegeven tekst automatisch een polariteit toekennen, een score die het sentiment in die tekst representeert. Een bepaalde klantenreview, bijvoorbeeld, wordt dan gecategoriseerd als een review met een positieve, neutrale of negatieve polariteit. Het toekennen van de sentimentscore kan gebeuren met een kennisgebaseerd AI-systeem dat gebruikmaakt van lexicons en regels, of met een datagebaseerd AI-systeem, een ML-systeem dat getraind is met gelabelde voorbeelden of een AI-systeem dat gebaseerd is op een Large Language Model (LLM).

In de eerste opdracht gaan de leerlingen aan de slag met regelgebaseerde sentimentanalyse. In de tweede opdracht experimenteren ze met een applicatie die werd ontwikkeld aan de UGent en die zowel regelgebaseerde als datagebaseerde sentimentanalyse demonstreert.

Unplugged activiteit: regelgebaseerde sentimentanalyse

In het lexicon staan woorden gekoppeld aan hun polariteit (positief, negatief of neutraal); de polariteit wordt weergegeven door een reëel getal tussen -2 en 2. Hieronder is een deel van zo'n lexicon van sentimentwoorden afgebeeld.

De polariteit van een tekst wordt gegeven door de som van de polariteiten van de sentimentwoorden in die tekst.

Lexicon

```
retorisch {'postag': ['ADJ'], 'polarity': [0.0]}
gezwind {'postag': ['ADJ'], 'polarity': [0.6]}
evenwichtig {'postag': ['ADJ'], 'polarity': [1.25]}
modaal {'postag': ['ADJ'], 'polarity': [0.4]}
zijn {'postag': ['VERB'], 'polarity': [0.0]}
digitaal {'postag': ['ADJ'], 'polarity': [0.0]}
fout {'postag': ['ADJ', 'NOUN'], 'polarity': [-0.5, -2.0]}
onverdeeld {'postag': ['ADJ'], 'polarity': [0.1]}
wulps {'postag': ['ADJ'], 'polarity': [0.7]}
bemoeiziek {'postag': ['ADJ'], 'polarity': [-1.35]}
grootscheeps {'postag': ['ADJ'], 'polarity': [0.2]}
vorig {'postag': ['ADJ'], 'polarity': [0.0]}
creatief {'postag': ['ADJ'], 'polarity': [1.2]}
tegenoverliggend {'postag': ['ADJ'], 'polarity': [-0.2]}
ijsbreker {'postag': ['NOUN'], 'polarity': [0.0]}
schalks {'postag': ['ADJ'], 'polarity': [0.0]}
kurkdroog {'postag': ['ADJ'], 'polarity': [0.0]}
spel {'postag': ['NOUN'], 'polarity': [1.0]}
kwistig {'postag': ['ADJ'], 'polarity': [-0.1]}
verknipt {'postag': ['ADJ'], 'polarity': [-1.2]}
tiranniek {'postag': ['ADJ'], 'polarity': [-0.3]}
mondig {'postag': ['ADJ'], 'polarity': [0.3]}
deels {'postag': ['ADJ'], 'polarity': [0.0]}
verrassend {'postag': ['ADJ'], 'polarity': [0.8]}
tomeloos {'postag': ['ADJ'], 'polarity': [-0.1]}
schadelijk {'postag': ['ADJ'], 'polarity': [-0.75]}
luguber {'postag': ['ADJ'], 'polarity': [-1.4]}
tof {'postag': ['ADJ'], 'polarity': [0.8]}
protestants {'postag': ['ADJ'], 'polarity': [0.0]}
uiterst {'postag': ['ADJ'], 'polarity': [0.5]}
eerloos {'postag': ['ADJ'], 'polarity': [-0.5]}
loodgrijs {'postag': ['ADJ'], 'polarity': [-0.1]}
zogeheten {'postag': ['ADJ'], 'polarity': [-0.1]}
harmonieus {'postag': ['ADJ'], 'polarity': [1.3]}
fijnbesnaard {'postag': ['ADJ'], 'polarity': [0.4]}
plechtig {'postag': ['ADJ'], 'polarity': [0.3]}
gemaakt {'postag': ['ADJ'], 'polarity': [-0.6]}
ondubbelzinnig {'postag': ['ADJ'], 'polarity': [1.1]}
overijverig {'postag': ['ADJ'], 'polarity': [0.0]}
psychisch {'postag': ['ADJ'], 'polarity': [0.0]}
vast {'postag': ['ADJ'], 'polarity': [0.0]}
schoon {'postag': ['ADJ', 'NOUN'], 'polarity': [0.466666666667, 1.0]}
```

- Gegeven zin: De spelletjes waren toffe ijsbrekers.
- Om de woorden van de gegeven zin (de data) te kunnen matchen met de sentimentwoorden in het lexicon, moet je de zin zo aanpakken dat je de gewenste informatie kunt opzoeken in het lexicon.
- Met welke zaken heeft het systeem het moeilijk?

Gegeven zin: De spelletjes waren toffe ijsbrekers.

Lowercasing: de spelletjes waren toffe ijsbrekers.

Tokenisering: de spelletjes waren toffe ijsbrekers .

→ Tokens: 'de', 'spelletjes' 'waren' 'toffe' 'ijsbrekers' '.'

Lemmatisering:

→ Lemma's: 'de', 'spel', 'zijn', 'tof', 'ijsbreker', '.'

Part-of-speech:

'de': lidwoord;

'spel': substantief;

'zijn': werkwoord;

'tof': adjectief;

'ijsbreker': substantief;

'.': leesteken (symbool).

Polariteit: De polariteiten van de lemma's worden opgezocht in het lexicon; lidwoorden en leestekens zijn daarbij niet van belang.

→ 'spel' heeft polariteit 1, 'zijn' heeft polariteit 0, 'tof' heeft polariteit 0,8 en 'ijsbreker' 0.

De polariteit van de gegeven zin is de som van deze polariteiten, dus 1,8.

Besluit: De zin roept een positief sentiment op.

Het systeem heeft het moeilijk met zaken zoals: 'niet goed', ironie, humor ...

Opdracht: Experimenteer met de tool 'Sentiment Demo'

- Surf naar de tool 'Sentiment Demo' van de onderzoeksgroep LT3 van de UGent (<https://lt3.ugent.be/resources/sentiment-demo>). Met deze tool bepaal je de sentimentwaarde van een ingegeven tekst op twee manieren: eens met een regelgebaseerd AI-systeem en eens met een ML-systeem dat getraind is met gelabelde voorbeelden.
- Schrijf zelf enkele reviews, bijvoorbeeld over een film, een gekocht kledingstuk, een boek ... of zoek er enkele op het wereldwijde web.
- Bied ze een voor een aan aan de tool. (Je hoeft geen e-mailadres in te vullen). Vergelijk de antwoorden van het regelgebaseerde en het datagebaseerde systeem.

Je merkt dat het datagebaseerde systeem soms beter presteert dan het regelgebaseerde systeem, maar zeker niet altijd. Hier een voorbeeld van 6.10.2024 met de tool 'Sentiment Demo'.

Original text

Amai, dat is weer een heel goed restaurant. Het eten trok op niets."

Lexical



Amai, dat is weer een heel goed restaurant. Het eten trok op niets.

Machine Learning





Opdracht: Experimenteer met sentimentanalyse gebaseerd op een LLM

- Doe nog eens hetzelfde met een chatbot, zoals ChatGPT of Google Gemini.
- Vergelijk de prestaties met die van de systemen in de demo.

Sentimentanalyse op basis van een LLM haalt er vaak wel ironie uit. Hier een voorbeeld van 6.10.2024 met Google Gemini en erna met ChatGPT 4o. Je kan de chatbots ook vragen om uit te leggen waarop zijn conclusie gebaseerd is.

N Wat is het sentiment van deze review: "Amai, dat is weer een heel goed restaurant. Het eten trok op niets."

Concepten tonen 

 **Het sentiment van deze review is uiterst negatief, ondanks de positieve opening.**

Wat is het sentiment van deze review: "Amai, dat is weer een heel goed restaurant. Het eten trok op niets."

 De review begint positief met "Amai, dat is weer een heel goed restaurant," maar wordt vervolgens direct negatief met "Het eten trok op niets." Het contrast tussen beide zinnen maakt de algehele toon van de review negatief, omdat de opmerking over het eten zwaar weegt in de beoordeling van een restaurant.

Lees meer over sentimentanalyse en waarom en door wie het ingezet wordt in het leerpad 'Sentimentanalyse' op de webpagina van het Chatbot-project (<https://dwengo.org/chatbot/>) en in de hoofdstukken 5 en 7 van de Leerkrachtenhandleiding van het project Chatbot (https://dwengo.org/assets/files/chatbot/Chatbot_handleiding_eerstedruk.pdf).

Alle informatie over het project 'Chatbot' over taaltechnologie en de bijbehorende minimumdoelen vind je via <https://www.istem.be/inspiratiebron/chatbot/> en <https://dwengo.org/chatbot/>.