

# Vereenvoudiging LSS

2023-05-01

## Inleiding

- Peterson (2011) email leesbaarheid classificatie
- Collins (2005) voorspellen van gestileerde tekstgeneratie
- Hovy (1990) pragmatiek
- Inkpen (2006) bouwen
- Li (2018) verwijderen
- Prabhumoye (2018) stijloverdracht
- We presenteren een taalkundig gemotiveerd proces om een taakonafhankelijk begrip van stijl te ontwikkelen, dat niet gebonden is aan bovengenoemde taken en algemeen genoeg is om deze te omvatten
- Stilistische variaties in taal zijn weerspiegelingen van factoren zoals context, auteur-lezer dynamiek en de achtergronden van de betrokken partijen.

## Invloed op factoren

- In de laatste sectie concluderen we het werk en geven aan wat de mogelijkheden zijn voor de toekomst
- Stijlelementen in tekst kunnen worden geïdentificeerd op woordniveau (lexicaal), aan de manier waarop zinnen zijn gestructureerd (syntactisch) en door analyse van kenmerken van de kern boodschap (semantisch)
- Let op dat een stijlelement kan behoren tot één of meer van bovenstaande categorieën [dimarcoStyle]
- We beschrijven kort elk van de stijlelementen en geven ook voorbeelden om te tonen hoe de boodschap in tekst verweven is

- Lexicale elementen van stijl worden uitgedrukt op woordniveau , en de stijlvariantie kan ontstaan door toevoeging , verwijdering of vervanging van woorden
- Deze varianties kunnen leiden tot tekst die karakteristiek verschillend is in termen van gevoelens , formaliteit , opwinding etc .

## Counterparts

- Er zijn stijlvarianten die goed worden gevangen in de parsebomen die worden gegenereerd met behulp van een probabilistische contextvrije grammatica (PCFG) [booth1973applying].
- Voorbeelden hiervan zijn dat de parsebomen van losse zinnen dieper en ongelijkmatig zijn, terwijl die van periodieke relatief meer gelijkmatig en breder zijn [feng2012characterizing].
- Het is opmerkelijk dat sommige van deze syntactische stijlelementen zich over meerdere zinnen uitdrukken en niet beperkt zijn tot één zin.

## Losse zinnen

- De interessante aspecten die naar voren komen bij het oplossen van taken voor auteurschapstoewijzing en emotie-voorspelling, variëren in mate waarin stijlelementen op verschillende niveaus bijdragen aan de taak.
- Bijvoorbeeld, in Tabel 5 voegen de lexicale en syntactische stijlelementen significant meer waarde toe aan de taak emotie-voorspelling dan oppervlakte.
- Dit beweren wordt ondersteund, omdat we opmerken dat de originele handgemaakte functies die in de baseline worden gebruikt, vrij zijn van enige stijl.
- De voorgestelde structuur biedt meer holistische interpretatie bij het modelleren van stijl om gerelateerde taken op te lossen.
- Gerelateerd werk.

## Stand van zaken

- Recent er zijn pogingen gedaan om stijlkenmerken te meten om tekstcategorisatie op basis van stijl mogelijk te maken
- Er is een tendens om stijl en daarmee stijlkenmerken op taakspecifieke manier te definiëren

- De definitie van stijkenmerken is onder verschillende werken inconsistent
- Bijvoorbeeld, Brooke (2013) richt zich voornamelijk op lexicale stijkenmerken, terwijl Feng (2012) stijkenmerken op syntactisch niveau onderzoekt.
- 

## Discussie

- Mohammad et al. hebben word- en karakter-tf-idf uitgebreid gebruikt voor het voorspellen van emotie
- Akhtar et al. en Preoctiuc et al. hebben ook geprobeerd
- Kao heeft echter weinig gedaan om de invloed van stijkenmerken op emotie te analyseren
- We laten zien dat linguïstisch gemotiveerde stijkenmerken niet alleen gecorreleerd zijn met emotie, maar ook helpen bij het verbeteren van de prestaties van bestaande voorspellingsbenaderingen.

## Conclusie

- Begrijpen van stijl is een belangrijk aspect bij het modelleren van inherente subjectiviteit in tekst.
- We hebben een taalkundig gemotiveerd proces voorgesteld om stijl-gerelateerde linguïstische intuïties te begrijpen en te kwalificeren op lexicaal, syntactisch en semantisch niveau.
- We hebben bestaande methoden gebruikt om deze stijl-elementen te kwantificeren, waarmee we de schrijfstijl van 5 auteurs hebben geanalyseerd en taken zoals auteurschapattributie en emotie-voorspelling hebben opgelost.
- We hebben aangetoond dat de stijl van een auteur een samengesteld factor is van verschillende stijlelementen, wat de noodzaak bevestigt van een multi-niveau analyse.
- We hebben ook de prestaties van bestaande state-of-the-art benaderingen voor auteurschapattributie en emotie-voorspelling verbeterd door stijl op een gestructureerde, interpreteerbare manier te modelleren.