

HET POTENTIEEL VAN NATURAL LANGUAGE PROCESSING VOOR RADIOLOGEN

GPT in de Twentse radiologie



Allard Olthof



Onno Vijlbrief

Natural Language Processing (NLP) is de tak van kunstmatige intelligentie die zich bezighoudt met het begrijpen en verwerken van menselijke taal door computers. Sinds dit jaar staat NLP, en dan in het bijzonder ChatGPT, volop in de belangstelling. De vraag is dan ook wat dat voor de radiologie betekent.

Binnen de NLP is een heel spectrum aan methoden beschikbaar die variëren in de mate van benodigde input van de gebruiker en in de mate van benodigde hoeveelheid data. Zo moet de gebruiker bij *rule-based* NLP een hele set aan regels opstelt om daarmee teksten in categorieën in te delen. Dit kost dit veel tijd, omdat met alle synoniemen, schrijfwijzen, uitzonderingen en ontkenningen rekening moet worden gehouden. Bij *machine learning*

dig zijn in een trainingsset, zoals blijkt uit eigen onderzoek¹, is dit nog best veel werk voor de gebruikers. In de geneeskunde en in de radiologie in het bijzonder zijn de benodigde experts hiervoor (lees artsen en radiologen) echter schaars.

Transformer modellen zijn geavanceerde deep learning modellen die getraind zijn op grote hoeveelheden ongelabelde tekst. Dergelijke *large language models* (LLM's) zijn voor verschillende NLP-taken ge-

klinische kennis en expertise inbrengen, terwijl datawetenschappers en ICT-specialisten de technische knowhow hebben om de NLP-modellen te ontwikkelen en implementeren. Het proces van het toepassen van NLP in de radiologie omvat verschillende stappen. Dit begint met het opzetten van een programmeeromgeving en het verzamelen van de relevante data, zoals radiologische verslagen. Vervolgens moeten deze gegevens worden gecureerd en geannoteerd om ze geschikt te maken voor training van het NLP-model. Daarna is het model te trainen en testen op de verzamelde gegevens. Zodra het model is geoptimaliseerd, is het in te zetten op nieuwe gegevens om bruikbare inzichten en resultaten te genereren.

'Transformer modellen zijn getraind op grote hoeveelheden ongelabelde tekst en geschikt voor veel taken'

draagt de gebruiker zogenaamde features aan (bijvoorbeeld *n-grams*, groepjes van n aantal woorden), die het model vervolgens gebruikt om de gelabelde categorie van een tekst te leren en te voorspellen. Bij *deep learning* bedenkt het model zelf welke karakteristieken van de tekst handige *features* zijn om te gebruiken in het trainingsproces. Tot zover heeft de gebruiker dus een steeds kleinere rol met betrekking tot de manier hoe het model (eigenlijk een stuk code in een computerprogramma) om gaat met de tekst.

Transformer modellen

De output – datgene wat het model moet leren voorspellen – moet de gebruiker nog wel steeds definiëren. Dat kan door labelen of annoteren, waarbij op woord of tekstniveau per casus een handmatig label toegewezen wordt. Omdat bij deep learning toch wel 1.000 – 2.000 casus no-

schikt: bijvoorbeeld het classificeren van documenten of het beantwoorden van vragen. Voor specifieke taken kan zo'n model op een relatief kleine hoeveelheid teksten worden verfijnd. Een LLM wordt dan eigenlijk ge(sub)specificeerd voor bijvoorbeeld de radiologie. Transformer modellen bestaan in allerlei soorten en maten, zoals BERT (*Bidirectional Encoder Representations from Transformers*) en de GPT-familie (*Generative Pre-trained Transformers*).

Doe-het-zelf

Het gebruik van NLP in de radiologie vereist een multidisciplinaire aanpak. Het omvat samenwerking tussen radiologen, AI-onderzoekers, technisch geneeskundigen en ICT-specialisten. In Twente werken radiologen van MRON samen met informatie-specialisten van ZGT en studenten en onderzoekers van de UT. Radiologen en aios kunnen domeinexperts zijn die de

NLP, GPT-4 of ChatGPT?

Methoden van NLP in de radiologie variëren van zogenaamde rule-based technieken tot large language models. In alle stappen van de radiologie workflow waar tekst voorkomt, kan in principe NLP toegepast worden. Dat begint al bij de indicatiestelling voor radiologie.² Bij het kiezen van het juiste onderzoek en het juiste protocol op basis van informatie op de aanvraag met behulp van GPT-4 werd in deze publicatie een mate van overeenstemming van 84 procent bereikt.³ De ongestructureerde informatie uit het radiologieverslag kan met hulp van NLP gestructureerd worden en tot waardevolle inzichten leiden, zoals bijvoorbeeld bij het bepalen van de incidentie van pulmonale noduli in een periode van 10 jaar.⁴

Een andere belangrijke toepassing van NLP is tekstgeneratie. Een voorbeeld hiervan is het automatisch maken van de conclusie van het radiologieverslag. In een relatief kleine studie bleek dat GPT-4 conclusies

niet zo nauwkeurig zijn als die van de radioloog en dat er soms ongefundeerde zaken in staan.⁵ Een ander voorbeeld van tekstgeneratie is het maken van patientvriendelijke ‘vertalingen’ van radiologieverslagen. GPT-4 kan het ‘taalniveau’ van een radiologieverslag verlagen en daarmee het verslag begrijpelijker maken voor patienten.⁶ Het bijzondere van GPT-4 is dat dit LLM niet specifiek getraind is op radiologische data, maar toch erg goede resultaten laat zien. GPT-4 is bijvoorbeeld veel beter in het maken van een radiologie-examen dan GPT-3.5. Vragen over kennis doen beide modellen goed, maar vragen waar kennis ook toegepast moet worden daar scoort GPT-4 beter.⁷

Kritische geluiden

Het valt op dat de toename van nieuwe NLP-publicaties in de radiologie exponentieel is en dat het tempo versnelt waarin nieuwe technieken toegepast worden. Na de introductie van BERT in 2018 volgden in de jaren daarna geleidelijk de eerste publicaties over toepassing in de radiologie. Na de introductie van GPT-4 op 14 maart 2023 volgden als snel de eerste publicaties. Een artikel over het automatisch omzetten van vrije tekst verslagen naar gestructureerde verslagen is op 23 maart ontvangen door *Radiology* en verscheen 4 april 2023 al on-

Onderzoek met een Bertje

Om de rol van NLP en data science in de radiologie te onderzoeken en toe te passen zijn we gestart met het MRON-Datalab. Met behulp van Bertje (een BERT-model gebaseerd op Nederlandstalige data) hebben we de diagnostische opbrengst van radiologie bij patienten met buikpijn onderzocht over een periode van 10 jaar. Recent hebben we *rule-based* NLP gebruikt om te inventariseren welke woorden we gebruiken om ‘zekerheid’ aan te duiden in CT-verslagen in de oncologie. We gebruiken zo NLP bij de start en straks ter evaluatie van een nieuwe werkafpraak die moet leiden tot uniform gebruik van woorden als ‘mogelijk’ en ‘waarschijnlijk’ in radiologieverslagen. Verder werken we aan een project om bij foto’s en MRI’s van de LWK uit de eerste en tweede lijn te evalueren of en in welke mate het aanvraagpatroon volgens richtlijnen gebeurt, en wat de diagnostische opbrengst van deze diagnostiek is. Zo kan NLP bijdragen aan het vormgeven van de rol van de radioloog bij het bevorderen van gepast gebruik van radiologie.

Beeld én tekst

Kortom, er zijn dit jaar veel interessante ontwikkelingen op het gebied van NLP. Ook al associëren we ons vak al snel met beelden, een groot deel van onze work-

‘NLP kan bijdragen aan het vormgeven van de rol van de radioloog bij het bevorderen van gepast gebruik van radiologie’

line.⁹ Halverwege 2023 zijn er al bijna net zoveel NLP-publicaties op het gebied van radiologie verschenen als in heel 2022 en ChatGPT draagt daaraan bij. Terecht zijn er ook veel kritische geluiden. Zo kan de informatie die ChatGPT bijvoorbeeld geeft over radiologische procedures incompleet en inaccuraat zijn.¹⁰

Ook verschijnen er *editorials* waarin gereflecteerd wordt op de snelle opkomst en de uitgebreide mogelijkheden van ChatGPT en wat dat bijvoorbeeld voor de opleiding tot radioloog.¹¹ Zelf geven we hier vorm aan door op het heilig uur niet alleen casus te bespreken, maar ook te discussieren over recente ontwikkelingen op het gebied van bijvoorbeeld ChatGPT. Verder bieden we aios te mogelijkheid om actief betrokken te zijn bij AI/NLP projecten. Dat heeft onder andere geresulteerd in de eerste prijs van de NVvR best-abstract sessie dit jaar.¹⁰

flow is gebaseerd op tekst. Large language models als GPT-4 kunnen hierom veel betekenen voor ons vak. Wil je meer informatie of samenwerken? Neem dan contact met ons op. ■

Allard Olthof
radioloog MRON
Onno Vijlbrief
radioloog MRON

Over de auteurs:

De MRON-radiologen Onno Vijlbrief en Allard Olthof hebben in ZGT verschillende Computer Science en Technische Geneeskundestudenten van de Universiteit Twente en aios radiologie begeleid bij projecten op het gebied van Natural Language Processing in de radiologie.

Referenties

1. A. W. Olthof, P. M. A. van Ooijen, and L. J. Cornelissen, “Deep Learning-Based Natural Language Processing in Radiology: The Impact of Report Complexity, Disease Prevalence, Dataset Size, and Algorithm Type on Model Performance,” *J Med Syst*, vol. 45, no. 10, Oct. 2021, doi: 10.1007/S10916-021-01761-4.
2. E. T. Varney and C. I. Lee, “The Potential for Using ChatGPT to Improve Imaging Appropriateness,” *J Am Coll Radiol*, Jul. 2023, doi: 10.1016/J.JACR.2023.06.005.
3. R. Johannes Gertz et al., “GPT-4 for Automated Determination of Radiological Study and Protocol based on Radiology Request Forms: A Feasibility Study,” <https://doi.org/10.1148/radiol.230877>, vol. 307, no. 5, Jun. 2023, doi: 10.1148/RADIOL.230877.
4. W. Hendrix et al., “Trends in the incidence of pulmonary nodules in chest computed tomography: 10-year results from two Dutch hospitals,” *Eur Radiol*, vol. 1, pp. 1–10, Jun. 2023, doi: 10.1007/S00330-023-09826-3/FIGURES/5.
5. Z. Sun et al., “Evaluating GPT4 on Impressions Generation in Radiology Reports,” *Radiology*, vol. 307, no. 5, Jun. 2023, doi: 10.1148/RADIOL.231259.
6. H. Li et al., “Decoding radiology reports: Potential application of OpenAI ChatGPT to enhance patient understanding of diagnostic reports,” *Clin Imaging*, vol. 101, pp. 137–141, Sep. 2023, doi: 10.1016/J.CLINIMAG.2023.06.008.
7. R. Bhayana, R. R. Bleakney, and S. Krishna, “GPT-4 in Radiology: Improvements in Advanced Reasoning,” *Radiology*, May 2023, doi: 10.1148/RADIOL.230987.
8. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *arXiv*, p. 1810.04805v2, 2018.
9. L. C. Adams et al., “Leveraging GPT-4 for Post Hoc Transformation of Free-Text Radiology Reports into Structured Reporting: A Multilingual Feasibility Study,” *Radiology*, Apr. 2023, doi: 10.1148/RADIOL.230725.
10. C. J. McCarthy, S. Berkowitz, V. Ramalingam, and M. Ahmed, “Evaluation of an Artificial Intelligence Chatbot for Delivery of Interventional Radiology Patient Education Material: A Comparison with Societal Website Content,” *Journal of Vascular and Interventional Radiology*, vol. 0, no. 0, Jun. 2023, doi: 10.1016/J.JVIR.2023.05.037.
11. A. P. Lourenco, P. J. Slanetz, and G. L. Baird, “Rise of ChatGPT: It May Be Time to Reassess How We Teach and Test Radiology Residents,” *Radiology*, May 2023, doi: 10.1148/RADIOL.231053.