



Natural Language Processing von klinischen Routinedaten zur Abrechnungskodierung: Weder trivial noch unerreichbare Magie

Vortrag

Datum	Zeit	Ort
21. April 2020	15:00 - 15:15 Uhr	_Stage A, Halle 6.2

Mit zunehmender Menge an elektronischen Routinedaten wird es für das Medizincontrolling schwieriger, die Kodierung nicht nur schnell und korrekt, sondern auch nachträglich nachvollziehbar durchzuführen. Das MDK-Reformgesetz verschärft die Problematik nun weiter, da Abrechnungen nicht mehr nachträglich revidiert werden können. So erscheint der Einsatz von intelligenten Systemen unumgänglich.

Da ein Großteil der Dokumentation in natürlichsprachlichen Freitexten enthalten ist, z.B. Arztbriefe, OP-Berichte, radiologische Befunde, steigt der Bedarf an Expertise im Bereich "Natural Language Processing" (NLP). In einem Projekt zur KI-gestützten Qualitätssicherung in der Abrechnungskodierung hat die Westpfalz-Klinikum GmbH (WKK) Anforderungen und Erwartungen an ein NLP-System evaluiert.

Hier werden für eine automatische Prüfung pseudonymisierte Daten (P21, Labor, Befunde...) von einer Datenplattform in ein NLP-System der Empolis Information Management GmbH geladen. Auf Basis von Faktenwissen aus 50.000 Falldaten pro Jahr von WKK wird die ICD-Klassifikation dem System antrainiert.

Da es sich um z.T. falsche und unvollständige Trainingsdaten handelt, wird Abrechnungswissen aus 1.000 Kundenaufträgen der MIA Medical Information Analytics GmbH (MIA) in der Form von Entscheidungsbäumen hinzugefügt. Durch sog. Active Learning überprüft MIA zudem gezielt für Fälle mit erwartungsgemäß hohem Lerneffekt die Kodierung per VPN-Fernzugriff und generiert Feedback an das System. Mit dem Feedback trainiert das System kleinteilige Aufgaben des NLP, z.B. der Klassifikation und Segmentierung von Dokumenten, der Negationserkennung und der Konzepterkennung.

Die Korrektheit als Micro-F1-Score über alle ICD-Kodes hinweg lag bei 0,76 (n = 200, 95% Wilson Score Konfidenzintervall [0.70;0.81]). Seltene Kodes mit wenigen positiven Beispielen sind noch eine Herausforderung. Bei häufigen Kodes wurde trotz hoher Heterogenität der Fälle ausreichende Güte mit akzeptablem Trainingsaufwand erreicht.

Akteure

Speaker:



[Dr. Benedikt Kämpgen](#), Leiter SU Healthcare Analytics, Empolis Information Management GmbH