

NLP zur Unterstützung von SNOMED CT Codierung

Studiengang: BSc in Medizininformatik | Vertiefung: Advanced Data Processing
Betreuer: Prof. Dr. Murat Sariyar
Experte: Reto Mettler (Asparagus Engineering AG)
Industriepartner: SIWF Schweizerisches Institut für ärztliche Weiter- und Fortbildung, Bern

Natural Language Processing (NLP) ermöglicht die maschinelle Verarbeitung von medizinischen Dokumenten ohne manuelle Vorstrukturierung. Um das SIWF beim Mapping zwischen textuellen Beschreibungen von medizinischen Prozeduren und SNOMED CT zu unterstützen, wurde der Prototyp SNOMAST 2.0 entwickelt. SNOMAST 2.0 ist ein auf BioBERT-basierendes SNOMED CT Annotation-Tool, das Entity-Linking durchführt und in der Evaluation vielversprechende Ergebnisse erzielte.

Ausgangslage

Im Gesundheitswesen liegen 80 % der Daten in einer unstrukturierten Form vor [1]. Der Einsatz von SNOMED CT und NLP kann bei der Gewinnung von medizinisch relevanten Informationen aus diesen Daten helfen. Das SIWF ist die zentrale Anlaufstelle für ärztliche Weiter- und Fortbildung. Aufgrund von Weiterentwicklungsarbeiten einer Webanwendung sollen die Anforderungen der Weiter- und Fortbildungen in SNOMED CT erfasst werden. Um beim Mapping zwischen textuellen Beschreibungen von medizinischen Prozeduren und SNOMED CT zu unterstützen, sollte eine Applikation auf Basis von BERT entwickelt werden. BERT steht für Bidirectional Encoder Representations from Transformers und stellt für viele NLP-Aufgaben den State-of-the-art-Ansatz dar.

Umsetzung und Ergebnisse

Die Applikation wurde auf Basis von BioBERT entwickelt, ein BERT-Modell, das zusätzlich auf biomedizinischen Texten trainiert wurde. BioBERT wurde mittels Finetuning um einen Classifier zu einem Entity-Linking-Modell erweitert und trainiert (Abb. 1). Es sind drei Modelle entwickelt worden, die 2 bis 89'855 Klassen umfassen, wobei jede Klasse einem SNOMED CT Konzept entspricht. Die Modelle berechnen anhand des Textes, die Wahrscheinlichkeiten für

SNOMED CT Konzepte und erzielten in der Evaluation einen F1-Wert über 70 %. Für die Verwendung der Modelle ist ein GUI in Python programmiert worden (Abb. 2), das durch das SIWF mit einem ausgezeichneten SUS Score von 88.75 bewertet wurde.

Fazit und Ausblick

Die Vorbereitung des Trainingsdatensatzes war aufwendig und die Trainingsdauer nimmt mit der Datensatzgröße zu. Trotz eines geringen Trainingsdatensatzes erreichten unsere Modelle akzeptable Resultate. Zudem zeigte die Evaluation, dass SNOMAST 2.0 unseren Industriepartner beim Mapping unterstützt. Besser noch. Das SIWF möchte SNOMAST 2.0 weiterentwickeln und es wurden Bedürfnisse für neue Anwendungsfälle geweckt. Die Skalierbarkeit der Classifier ermöglicht es, Modelle für weitere Anwendungsfälle mit anderen Terminologien zu trainieren und zu verwenden. Somit könnten Modelle entworfen werden, die bei der Leistungscodierung aus Verlaufsberichten helfen oder die Metadaten anhand des Textes im Dokument beim Hochladen in das elektronische Patientendossier automatisch befüllen.

Referenz

[1] Kong H.-J. Managing Unstructured Big Data in Healthcare System. Healthc Inform Res 2019



Sebastian Benedikt Kunz
sbkunz85@gmx.net



Cyril Raphaël Zraggen
zraggen.cyril@gmail.com

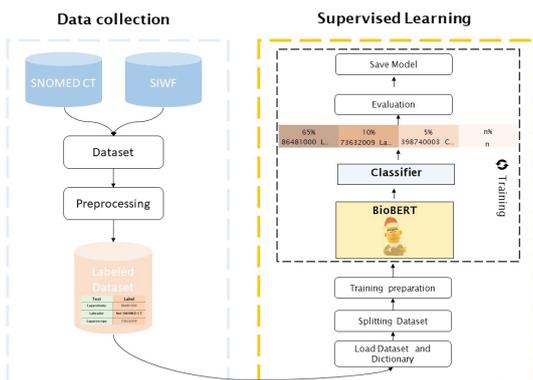


Abb. 1: Workflow Finetuning

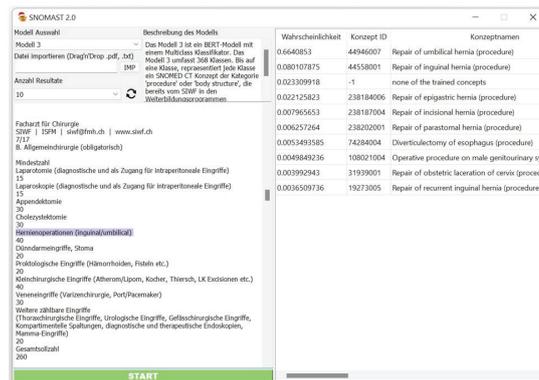


Abb. 2: SNOMAST 2.0 GUI – Outputscreen