Advancing Computer-Assisted ICD Coding for Enhanced Clinical Data Annotation

The International Classification of Diseases (ICD) has long been used for clinical data annotation, but accurate coding is challenging and time-consuming, leading to low data quality. Computer science and health informatics researchers have developed various methods, primarily neural networks (NNs), to automate this process using both structured and unstructured data. However, the adoption of these methods in real-world settings is limited, with coders preferring semi-automatic systems. Additionally, existing studies often overlook leveraging the inherent hierarchy of ICD codes and rarely address class imbalance. In a [paper recently published in the Journal of Biomedical Informatics](link), the authors propose performance metrics tailored to clinical coders, compare different evaluation metrics, explore using the entire set of ICD codes, and investigate the impact of hierarchical properties on NNs. They find that optimising certain metrics improves ranking performance, utilising the entire set of codes maintains performance, but leveraging the hierarchy does not consistently improve results. Their study also promotes broader applicability through data standardisation.

Challenges in Clinical Data Annotation and the Need for Improved Methods

Authors delve into several strategies aimed at enhancing computer-assisted coding (CAC) systems for the International Classification of Diseases (ICD) codes. The importance of selecting appropriate performance evaluation metrics tailored for clinical coders is emphasised. The authors introduce a set of such metrics derived from existing literature and discussions with health informatics professionals. They demonstrate that filtering ICD codes can be detrimental to code retrieval without providing any positive effect on the model's performance within the cherry-picked subset. Furthermore, the paper explores the efficacy of various evaluation metrics, comparing traditional metrics like F1 score with ranking-based metrics such as normalised Discounted Cumulative Gain (NDCG) or Precision-Recall curve-derived metrics. They find that ranking-based metrics are more effective in selecting superior models compared to traditional ones.

Tailoring Performance Evaluation Metrics for Clinical Coders

The study investigates multitask learning strategies in generic neural network models, observing minor ranking improvements and significant reduction in computational load. However, methods like hierarchical multilabel classification and class imbalance correction do not show improvements in their neural network models. The authors stress the importance of performance metric selection and CAC system design, underlining the need for further research to validate the proposed evaluation metric's usage in real-world settings. They compare their results with previous studies and demonstrate superior classification performance using medication data alone. Despite utilising a single dataset (MIMIC-III), the authors argue that their findings have broader applicability to other clinical settings. They conclude by suggesting future research directions, such as validating the proposed evaluation metric and developing more sophisticated architectures to leverage medication information for improved prediction quality in CAC systems.

Exploring Multitask Learning Strategies for Enhanced CAC Systems

The study explores the development of more effective computer-assisted coding (CAC) systems using medication data and a simple neural network architecture. They find that the practice of cherry-picking ICD codes reduces overall retrieval performance without improving performance within the selected subset, posing challenges for reproducibility in research. Introducing a novel metric tailored for the ICD coding task, they show that optimising for metrics like NDCG and AUPRC leads to superior ranking performance compared to traditional F1-based metrics. Multitask learning, where neural networks are trained simultaneously on different levels of the ICD hierarchy, provides minor benefits for ranking and runtime gains. However, hierarchical multilabel classification (HMC) methods and class imbalance correction techniques do not improve performance in their experiments.

These findings offer insights for researchers and healthcare practitioners interested in developing or evaluating CAC systems. Despite using a straightforward neural network model, they demonstrate that medical prescriptions offer a rich data source for CAC systems, achieving competitive ICD code retrieval capabilities with lower computational load compared to text-based models. The study suggests that future research should validate these findings in a production setting and explore more sophisticated architectures to fully leverage medication information for improved CAC system performance.

Published on : Mon, 15 Apr 2024