Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge

Abstract:

The vocabulary gap between health seekers and providers has hindered the cross-system operability and the interuser reusability. To bridge this gap, this paper presents a novel scheme to code the medical records by jointly utilizing local mining and global learning approaches, which are tightly linked and mutually reinforced. Local mining attempts to code the individual medical record by independently extracting the medical concepts from the medical record itself and then mapping them to authenticated terminologies. A corpus-aware terminology vocabulary is naturally constructed as a byproduct, which is used as the terminology space for global learning. Local mining approach, however, may suffer from information loss and lower precision, which are caused by the absence of key medical concepts and the presence of irrelevant medical concepts. Global learning, on the other hand, works towards enhancing the local medical coding via collaboratively discovering missing key terminologies and keeping off the irrelevant terminologies by analyzing the social neighbors. Comprehensive experiments well validate the proposed scheme and each of its component. Practically, this unsupervised scheme holds potential to large-scale data.