

Evaluating ATC-ICD: Assessing the relationship between selected medication and diseases with machine learning

Weibrecht, N^{1*}, Endel, F¹, and Zechmeister, M¹¹Verein dexhelpp

Introduction

Coded diagnoses (ICD-9, ICD-10) are only available in routine data of the Austrian Health-Care system in connection with sick leave or inpatient hospital stays. Therefore, they only cover a small part of the population. Coded diagnoses from the outpatient sector are not documented.

The aim of the project is to estimate diagnoses based on filled prescriptions reimbursed by a public health insurance institution. The result is a model that can provide probable diagnoses (ICD-10 coding) based on individual medication (ATC coding).

Method

Beginning in 2008 / 2009, the project ATC->ICD-9 has been developed by means of a statistical procedure. Here, hospital and sick leave diagnoses, as well as data on received medication are used to determine assignment probabilities.

In this project, we developed a new method to derive diagnoses from medications. Our method is based on the word2vec-algorithm: Patient histories are used as input phrases, so that low-dimensional embeddings of medications and diseases are learned. In the learned vector space, similar medications and diseases are close to each other.

Results

To evaluate our model, we compute the vector representation for medications and look for nearby diseases. E.g., the closest diseases to typical diabetes medication are different kinds of diabetes and retina affections, while nearby gout medications, gout and kidney diseases are found.

Conclusion

For the given examples, our model provides reasonable results. It does not only yield typical diseases to a medication, but also common secondary symptoms. This motivates to apply the model on further use cases. For example, given an

*Corresponding Author:

Email Address: nadine.weibrecht@dexhelpp.at (N Weibrecht)