Expose on the SIAM Article Improved prediction of antibiotic resistance: learning from the data

Darja Strahlberg

August 16, 2019

1 Introduction

1.1 Motivation

Antibiotic Resistance is a rising problem in the modern world. Bacterial infections, for example like pneumonia, meningitis, are commonly treated through antibiotic therapy applied to individuals. Knowing the frequency of the spread may help to creat a reaction scheme. The antimicrobial-drug resistance has many concerns like its economic impact on physicians, patients, health-care administrators, pharmaceutical producers, and the public. [1] Machine Laerning models and Big Data offer new approach on the prediction of the antibiotic resistance. Machine Learning methods can be integrated into the solving high dimensional models, where standard mathematics cannot help.

The classical SIR Model (math) has a disadvantage, that it does not consider randomness in the prediction, so making the model less robust. AR (Autoregressive) models use history data to make a future predictions. It usually requires very little data, so it is very useful in the epidemiological models.

The current research, using the ML methods, is concentrated on the on Antimicrobial resistance gene detection. So it is the resistome prediction. We are asking a question is it possible to predict the spread pattern of the antibiotic resistance? Is it important?

The central hypothesis states, that machine learning method will provide more accurate prediction of the spread of antibiotic resistance. It has a randomness factor (although we do not know how the machine learning), additionally, we can use additional data, like the amount of sick people, time of spread of the resistance, the amount of consumed antibiotics?

Figure 1: Different levels to predict the spread of antibiotic resistance: withing a patient, a hospital or a community.

There are different ways to investigate the development of the antibiotics resistance. For example, withing the patient, a hospital or a community. We have a goal to look at the spread of the resistance in the community globally. [2] The current research, using the ML methods, is concentrated on the on Antimicrobial resistance gene detection. So it is the resistome prediction. We are asking a question is it possible to predict the spread pattern of the antibiotic resistance? Is it important?

1.2   Historical Overview

There was observed a shift in the developed world, that the mortality from infectious diseases had been replaced by mortality from chronic illnesses such as heart disease, cancer and stroke. It lead to the fact that pharmaceutical companies stopped spending resources on the development of the new antibiotics, but instead on the medicine to cure other diseases. By the end of the twentieth century, in most of the developed world, mortality from infectious diseases had been replaced by mortality from

chronic illnesses such as heart disease, cancer and stroke [3] Table? It is importnat to show the statistics on death rate from infectious deseases in Germany and the EU

1.3   Goal

We can try to answer the question how the frequency of resistance changes over time. We want to provide a scenario analysis.

The big data analysis can help answering following questions: surveillance, hypothesis-generating research, and causal inference. [4]

Antibiotic consumption has a direct influence on the frequency of resistance in communities. It is seen as difficult due to the lack of the recorded data. So the machine learning approach can give a new life to the prediction of the antibiotic resistance. [2] The temporal nature of epidemiology data and the need for real-time prediction by the system makes the problem residing in the category of time-series forecasting or prediction. [5] There is also an issue with the time delay in the epidemiological data. So the good prediction can make a difference [6]

Additionally it is important to include the number of sick people of the disease. Pay attention to the correlation. Antibiotic consumption has a direct influence on the frequency of resistance in communities. It is seen as difficult due to the lack of the recorded data. So the machine learning approch can give a new life to the prediction of the antibiotic resistance. [2] The idea is to include additional data to broader analysis of bacterial fitness, infection dynamics, horizontal gene transfer and other factors. [7, S. 689] Big Data should satisfy three characteristics: high volume and high variety data, high velocity feedback mechanisms. [8]

- factors that influence the evolution of antibiotic resistance and discuss these factors in the context of early drug discovery - paths for a resistance to be developed: vertical evolution, horizontal evolution. **Acinetobacter baumannii,** carbapenem-resistant **Pseudomonas aeruginosa,** carbapenem-resistant **Enterobacteriaceae*,** carbapenem-resistant, 3rd generation cephalosporin-resistant [9]

2   Results and Discussion

2.1   Biological Overview: two ways of the spread

he traditional desease reporting istaking several steps. Public informs their Healthcare workers, they inform Laboratories, they Minister of Health and the World bodies, like WHO. [6] Resistance genes produce enzymes that either modify or eliminate an antibiotic. Resistance genes can be spread by vertical transmission from parent to
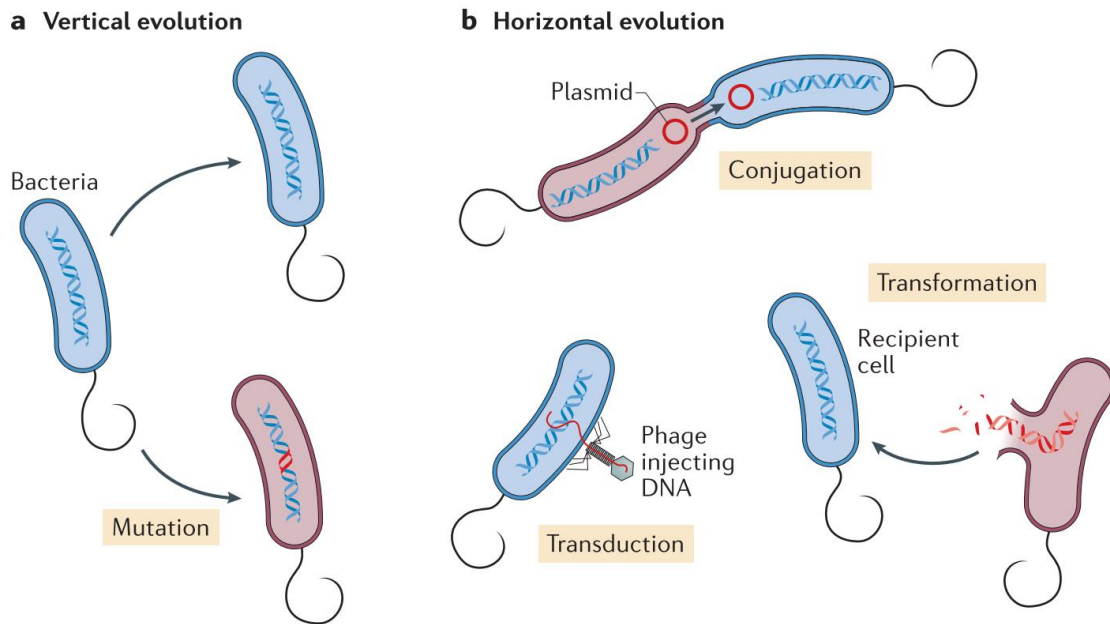
Figure 2: Evolution of resistance. Sommer, Munck et al. 2017 - Prediction of antibiotic resistance.jpg
 [7, S. 690]

offspring or by horizontal transfer between different strains and species of bacteria. Sensitive microbes become resistant to antibiotics either through the acquisition of plasmidborne resistance genes or through mutations that either upregulate the expression of a resistance gene or alter the binding or substrate specificity of an enzyme encoded by a resistance gene. [10, S. 1237]

2.2   Math Method of Linear Regression

Autoregressive (AR) models have been most popular for time series forecasting [6, 8]. The basic idea is to model the future state as a linear combination of past data points. [5] While being intuitive and popular, such models have limited prediction power due to the rather narrow function space, lack ofthe ability to model individual-level information, and do not embrace new developments in recent machine learning and data mining technologies. [5]

2.3 ML Method

Machine learning is an umbrella term for techniques that fit models algorithmically by adapting to patterns in data [4] We adopt Recurrent Neural Networks (RNNs) to capture the long-term correlation in the data and Convolutional Neural Networks (CNNs) to fuse information from data of different sources. A residual structure is also applied to prevent overfitting issues in the training process. [5] We employ an recurrent neural network (RNN) module to capture the temporal dependencies in the data. Specifically, we utilize an Gated Recurrent Unit (GRU) [2] in our framework. The input data are passed through a gate, which is then used to compute the new state in the memory cell ofGRU given the old value. [5] a CNN for capturing correlation between signals, a RNN for linking up the dependencies in the temporal dimension and the residual links for fast training and overfitting prevention. We carefully restrain the parameter space, making the total model have a similar size as AR. [5] We have measured the amount of people with the antibiotic resistance?

3 Conclusion

Hopefully, we have proved the hypothesis that Machine Learning approach gives a better prediction, and that more data contribute to it. Why does it make a difference?

References

[1] John E. McGowan JR., "Economic impact of antimicrobial resistance," *Emerging Infectious Diseases, Vol. 7, No. 2, March–April 2001*, 2001.

[2] D. J. Austin and R. M. Anderson, "Studies of antibiotic resistance within the patient, hospitals and the community using simple mathematical model," *The Royal Society*, vol. 1999.

[3] B. Daşbaşı and İ. Öztürk, "Mathematical modelling of bacterial resistance to multiple antibiotics and immune system response," *SpringerPlus*, vol. 5, p. 408, 2016.

[4] S. J. Mooney and V. Pejaver, "Big data in public health: Terminology, machine learning, and privacy," *Annual review of public health*, vol. 39, pp. 95–112, 2018.

[5] Y. Wu, Y. Yang, H. Nishiura, and M. Saitoh, "Deep learning for epidemiological predictions." [Online]. Available: http://arxiv.org/pdf/1406.1078v3

[6] Harvard, "Epidemiology with ml."

[7] M. O. A. Sommer, C. Munck, R. V. Toft-Kehler, and D. I. Andersson, "Prediction of antibiotic resistance: time for a new preclinical paradigm?" *Nature reviews. Microbiology*, vol. 15, no. 11, pp. 689–696, 2017.

[8] S. J. Mooney, D. J. Westreich, and A. M. El-Sayed, "Epidemiology in the era of big data," *Epidemiology (Cambridge, Mass.)*, vol. 26, no. 3, pp. 390–394, 2015.

[9] E. Tacconelli, "Who-priority pathogens list."

[10] M. Barlow and B. G. Hall, "Experimental prediction of the natural evolution of antibiotic resistance," *Genetics*, vol. 163, no. 4, pp. 1237–1241, 2003. [Online]. Available: https://www.genetics.org/content/genetics/163/4/1237.full.pdf