

Fetal Health Classification

Machine Learning and Artificial Intelligence

Academic year 2020/2021

Filippo Grotto VR460638

August 19, 2021

Contents

1	Motivation and rationale	2
2	Problem definition and Dataset	2
3	Methodology	2
4	Experiments and Results	3
4.1	Dataset exploration	3

1 Motivation and rationale

This project is heavily inspired by a kaggle task [2] about fetal health classification.

Reduction of child mortality is reflected in several of the United Nations' Sustainable Development Goals and is a key indicator of human progress. The UN expects that by 2030, countries end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce under-5 mortality to at least as low as 25 per 1,000 live births.

Parallel to notion of child mortality is of course maternal mortality, which accounts for 295 000 deaths during and following pregnancy and childbirth (as of 2017). The vast majority of these deaths (94%) occurred in low-resource settings, and most could have been prevented.

In light of what was mentioned above, Cardiotocograms (CTGs) are a simple and cost accessible option to assess fetal health, allowing healthcare professionals to take action in order to prevent child and maternal mortality. The equipment itself works by sending ultrasound pulses and reading its response, thus shedding light on fetal heart rate (FHR), fetal movements, uterine contractions and more.

In this context this project is not only an application of machine learning techniques but it also have some meaningful application into real world scenarios thanks to the real data provided.

2 Problem definition and Dataset

The dataset comes from UCI Machine Learning Repository [3] and it is composed by

2126 fetal cardiotocograms (CTGs) automatically processed with the respective diagnostic features measured. The CTGs are were classified by three expert obstetricians and a consensus classification label assigned to each of them. Classification was both with respect to a morphologic pattern (A, B, C. ...) and to a fetal state (N=normal; S=suspect; P=pathologic). Therefore the dataset can be used either for 10-class or 3-class experiments.

We will address the 3-class classification problem so we will try to classify the data according to normal, suspect or pathologic.

3 Methodology

The classification problem is addressed using the following techniques:

- Naive Bayes [4]
- SVM (with different kernels linear and RBF) [4]
- KNN (with different k based on preliminary analysis) [4]
- ANN [5]

Moreover, PCA and Fisher(LDA) dimensionality reduction techniques are considered as well as the effects of data scaling and/or normalization. We don't expect to see improvements of the state of the art where other techniques (XGBoost, Random forests, Decision tree models) are used.

4 Experiments and Results

4.1 Dataset exploration

References

- [1] Ayres de Campos et al. (2000) SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms
- [2] Fetal Health Classification <https://www.kaggle.com/andrewmvd/fetal-health-classification>
- [3] Cardiocotography dataset <https://archive.ics.uci.edu/ml/datasets/cardiotocography>
- [4] Hoodbhoy, Zahra et al. "Use of Machine Learning Algorithms for Prediction of Fetal Risk using Cardiotocographic Data."
- [5] Yılmaz, E. Fetal State Assessment from Cardiotocogram Data Using Artificial Neural Networks.