Continuous Classification using Deep Neural Networks

Nick Strayer

2017-12-08

Contents

Chapter 1

Introduction

1.1 Continuous Classification

Imagine you are watching a movie. A friend walks in late and asks "what did I miss?" You tell them the main character has just escaped from a nasty predicament and has defeated the antagonist. What you have done is classification on a sequence. The sequence in this case is the frames of the movie and your classification was what was occurring in the movie at that moment. You *could* have given the same answer if you just saw a single frame, but most likely your assessment of the state of the movie depended on events you saw before and the context in which they placed the most recent frame.

Continuous classification in the context of statistics and machine learning is training models to observe data, like you watched the movie, and be able to report the status of the generating system at any given point. Sometimes seeing the most recent data is all that is needed, but more interesting and challenging problems need the algorithm to be able to make decisions about a current time while leveraging context from previous history to do so.

This report is a brief run through past attempts at continuous classification and a deeper explanation of the current state of the art methods.

1.2 Potential applications of continuous classification models

The following are just a few examples of biomedical applications made possible with effective continuous classification models.

1.2.1 Activity Prediction

With the advent of wearable devises such as fitbits and apple watches the amount of high temporal resolution data we have streaming from individuals is exploding and showing no sign of letting up.

Continuous classification models could be used to to assess the state of the wearer at any moment. Simple applications of this could be for things like detecting different exercise types (e.g. running vs. weightlifting) which is implemented by unpublished methods internally by companies such as fitbit.

More advanced and potentially impactful methods could be extended to predicting more subtle but medically relevant states such as dehydration or sleep apnea (sleep apnea citation). Preliminary work in these areas has shown surprising success with data as limited as heart-rate and motion indication being enough to predict sleep apnea and various cardiovascular risk states with _____ accuracy compared to invasive gold standards.

1.2.2 EHR monitoring

With more and more information on patients being accrued in governmental and hospital databases we have a clearer than ever picture of a patient's health over long periods of time. Unfortunately, due to a combination of overwhelming quantities and noise levels in the data our ability to make use of these data has not kept up with their quantity.

Sequential models can help ease the burden on health practitioners in making use of these data. For instance, a model could be trained on a patient's records to predict the likelihood of cardiovascular events. This model could then alert a doctor of potential risk in order to facilitate timely interventions. This could be especially helpful in large clinical settings where personal doctor-patient relationships may not be common.

(Insert reference to applications on these data already performed)

1.2.3 Hospital Automation

Patient monitoring systems already have alarms to alert staff of occurring anomaly for a patient. Continuous classification methods could extend these methods to more subtle actions (patient is experiencing pain and needs a change in the medications administered by their IV) or to give staff more heads-up time before an event occurs (model predicts patient has a high change of going into afribulation in the next five mins). These methods, if successfully implemented could help hospitals more efficiently allocate resources and potentially save lives.

1.3 History of methods

While sources of data to perform it on have recently greatly expanded, the interest in performing continuous classification is not a new topic. Many methods have been proposed for the task to varying degrees of success. Below is a brief review of some of the more successful methods and their advantages and limitations.

1.3.1 Windowed regression

Perhaps the most intuitive approach to the problem of incorporating context from previous time points into your prediction is to use a windowed approach. Broadly, in these approaches a window of some width (in previous observation numbers or time length) is sequentially run over the data and then the model uses information from within that window to predict.

The data obtained from the window may have some form of summary applied to it. This could be a mean, median, or any other function which is then used to predict with. By summarizing the multiple data-points into a single (or few) values noisy data can be made slightly cleaner but at the cost of potentially throwing away useful information captured by the interval (such as trajectory.)

If the data are kept intact mode advanced methods are available. These include dynamic time warping (Time warping citation) or kernel methods (see next section). This allows much more information to be retained in the sample but at the cost of setting a limit on how far back your model can learn dependencies in the data. For instance if your window is one hour long but an activity lasts two hours your model will have a very hard time recognizing it. This is equivalent to an infinitely strong prior on the interaction timeline (window prior assumptions citation).

1.3.2 Transformation methods

As mentioned before, when a window is scanned across the time dimension of data, one of the ways of extracting information is by performing some transformation on the data. Common examples include wavelet