Learning

Subtitle

Mathias Kirkerød

Thesis submitted for the degree of
Master in Informatics: Language and Communication
60 credits

Department of Informatics
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

Spring 2018

Learning

Subtitle

Mathias Kirkerød

Learning

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

Contents

List of Figures

List of Tables

Preface

Part I

Introduction

Chapter 1

Introduction

1.1 Background and Motivation

1.1.1 Introduction REM

Cancer Cancer is, today, the second leading cause of death in the world, only behind cardiovascular diseases.
It is one of the leading causes of mortality worldwide, with approximately 14 million new cases in 2012. It is defined as a disease that has an abnormal cell growth with the potential to spread into other parts of the body.In contrast to many other diseases cancer does not start from a foreign entity (such as a bacteria or virus), but it is often from a malfunctioning cell that starts dividing rapidly. This can happen when a cell is damaged, by for instance by radiation, and the resulting damage causes the cell to uncontrollably divide. Especially in the later part of life everyone has the chance of getting cancer, and in fact everyone does. Our own body is designed to detect and remove cells that are prone to divide uncontrollably. Unfortunately this system is not perfect, and the immune system can in some cases overlook cells that are cancerous.

1.1.2 Statistics on cancer REM

The western (or modern) world has been in a battle against cancer, and despite a lot of new cures/innovations it is still one of the deadliest killers in the world.

The most common types of cancer in males are lung cancer, prostate cancer, colorectal cancer and stomach cancer.Stewart and Wild 2014

1.1.3 colorectal cancer REM

You can get cancer in every major organ, but some types of cancer are more common than others. For instance cancer in the gastrointestinal tract (GI) is one of the more common places to get cancer. This is just behind x, and it has a mortality rate of x in the first y years. We often call this 5 year survival rate for z. This is the standard way to measure the life expectancy of a patient diagnosed with cancer.

1.1.4 polyps REM

The colorectal cancer often starts in polyps. Polyps are, polyps do.

1.1.5 preventative matters and early detection REM

-colonoscopy
-mri
-pillcam
A good way to fight cancer is to detect and remove it early, or some times remove areas with a high chance of getting cancer. We classify cancer in to x stages, and the stage the patient are in often determines the chance you have for survival. In general, the earlier you find the cancer, the more likely it is that the patient will survive. And as mentioned above, the colorectal cancer often starts in these polyps. A crucial stage to prevent cancer lies in the early removal of there polyps. Reports shows x about this

*4 stages maybe? *early detection *survival rate

Because of this the ability to find, and remove colorectal polyps is great for preventing cancer in the GI tract.

colonoscopy/Ontonoscopy In the most common way to look for polyps in the GI tract is to use a medical team, and perform a colonoscopy or Ontonoscopy colonoscopy is preformed with a camera.....
Onoskopy is the same procedure, only the camera is inserted orally.
Advantages * Accurate * **Disadvantages** *expensive

MRI

pillcam

In the last 3-4 years there have been testing and development on the pillcam project EIR. Machine learning has, through many of the earlier projects, got the detection rate for the polyps up to x%

1.1.6 Simulas contribution to the pillcam project REM

Simulas EIR

* CAD ACD (computer aided diagnosis, Automated computer diagnosis)

1.2 Goal / Problem

1.2.1 pillcam project has lots of data, can be used to train an unsupervised network REM

1.2.2 Use Unsupervised learning as a pre-processing tool

1.2.3 use Unsupervised-NN/GAN for image enhancements so that a NN can train better

* Now that we got a lot of tests, why not unsupervised As mentioned, simula research centre has done a lot of testing on the pillcam project.

* We know that we can get some results using a neural network * Can this be done unsupervised? * Can it be done in a fashion that is better than S-ML

REM

1.3 Scope and Limitations

1.3.1 Use Unsupervised NN to find polyps

1.3.2 Use Unsupervised NN for pre-processing

* Something about earlier research already got far, so the scope is mainly unsupervised deep learning. * (and how to generalise it?)

1.4 Research method

1.5 Related work

1.6 Outline

The rest of the thesis is structed as follows:

Chapter 2 - Background

t REMalk about cancer *talk about machine learning. *how to use ML on the pillcam video? **Chapter 3 - Me doing stuff**

Chapter 4 - Me got and present result
Chapter 5 - Me saying result was good A+

Chapter 2

Background

2.1 Cancer and polyps

2.1.1 What we are looking for REM

2.1.2 images from pillcam, and what we are looking at/for REM

2.2 Machine Learning

Machine learning is a very broad term, but can i short be summarised by:

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with the experience E. Mitchell 1997

Here we have a couple of parameters:
E text about p
T text about p
P text about p

From this we see that the goal of machine learning is to improve some performance P with experience. **might here talk about different tasks ML can do?**

2.2.1 Supervised & Unsupervised machine learning

We often divide machine learning in to two (diffuse) categories: supervised and unsupervised.
Supervised learning: is the act of training with data that has an answer. The learning algorithm cant get supervision while training on the task.

Unsupervised learning: is the act of training without any supervision, on the sense that we do not give the algorithm the answer to the training data set.

In the description of supervised vs unsupervised we looked at a specific branch of machine learning: Classification. Classification is, as the name implies, the task of getting data sorted in to groups of similarity.

- subsfication
- r to the pillcam projec REMegression
- transcription/translation
- de-noising /finding missing inputs

Now that we have the definintion of machine learning we focus on the task at hand; finding polyps. In an ideal world we have a Classification problem with only two classes: Non-polyp and polyp.

- SVM
- CNN
- random forests
- knn

2.2.2 CNN

UCNN?

2.2.3 Tasks (other better word goes here)

2.2.4 The rate of success

What is a good result, how to measure?
FP,TN,FN,TP

2.3 supervised vs unsupervised

What it means to be S/US.
Something about the kind of experience allowed during the learning process.

2.4 Unsupervised

noe med å dele i grupper? Experience the dataset containing many features, and finds useful properties of the structures. ***Unsupervised learning algorithms*** *experience a dataset containing manyfeatures, then learn useful properties of the structure of this dataset. In the contextof deep learning, we usually want to learn the entire probability distribution thatgenerated a dataset, whether explicitly, as in density estimation, or implicitly, fortasks like synthesis or denoising. Some other unsupervised learning algorithmsperform other roles, like clustering, which consists of dividing the dataset intoclusters of similar examples.* Goodfellow, Bengio, and Courville 2016

2.4.1 Approaches to unsupervised learning

look at the subsection 2.2.3 to see what applies to the unsupervised.

2.4.2 Deep Unsupervised learning

2.4.3 more

2.5 Related work

Part II

The project

Chapter 3

Planning the project

3.1 Using Generative Adversarial Networks for enhancement and prediction

Generative Adversarial Networks (GAN) was proposed by Goodfellow in 2014. **GoodfellowGAN** GANs is a specialised network consisting of generally two neural networks working against each other. The two networks are often called discriminator and generator. The generators job is to make data similar to the training data, and discriminator has the job if finding fake (generated) data.

The GAN presented by Goodfellow 2014**GoodfellowGAN** is a simple adversarial network, and it needs to be built upon to be used. Since we are working with images it is more suitable to use DC-GAN instead.

3.1.1 DC-GAN

DC-GAN is a type of an Adversarial Network that still uses the adversarial approach, like the original GAN, but the two networks are deep convolutional networks. DC-GANs.

3.1.2 CC-GAN

removing green squares

3.2 Training an autoencoder to help the GAN

An autoencoder (AE) is another form of a machine learning network. The autoencoders task is to

Part III

Conclusion

Chapter 4

Results

Bibliography

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press.

Mitchell, Tom M (1997). *Machine learning*. eng. New York.

Stewart, B.W. and C.P. Wild (2014). *World Cancer Report 2014*. International Agency for Research on Cancer. International Agency for Research on Cancer. ISBN: 9789283204299. URL: https://books.google.no/books?id=0QHbngEACAAJ.