Machine Learning methods for Thin Vessel Segmentation

in Fundus Images

Kushal Khandelwal

Birla Institute of Technology and Science,Pilani

This thesis is submitted in requirements of the course BITS C421T

May 2015

I would like to dedicate this thesis to my loving parents . . .

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

<div align="right">

Kushal Khandelwal
May 2015

</div>

Acknowledgements

And I would like to acknowledge ...

Abstract

This is where you write your abstract ...

Table of contents

List of figures

List of tables

Chapter 1

Introdcution

1.1 What is loren ipsum? Title with math σ

Lorem Ipsum is simply dummy text of the printing and typesetting industry (see Section 1.3). Lorem Ipsum [3] has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum [1, 4, 5].

The most famous equation in the world: $E^2 = (m_0 c^2)^2 + (pc)^2$, which is known as the **energy-mass-momentum** relation as an in-line equation.

A *LaTeX class file* is a file, which holds style information for a particular LaTeX.

$$CIF: \quad F_0^j(a) = \frac{1}{2\pi\iota} \oint_\gamma \frac{F_0^j(z)}{z-a} dz \tag{1.1}$$

1.2 Why do we use loren ipsum?

It is a long established fact that a reader will be distracted by the readable content of a page when looking at its layout. The point of using Lorem Ipsum is that it has a more-or-less normal distribution of letters, as opposed to using 'Content here, content here', making it look like readable English. Many desktop publishing packages and web page editors now use Lorem Ipsum as their default model text, and a search for 'lorem ipsum' will uncover many

web sites still in their infancy. Various versions have evolved over the years, sometimes by accident, sometimes on purpose (injected humour and the like).

1.3   Where does it come from?

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more obscure Latin words, consectetur, from a Lorem Ipsum passage, and going through the cites of the word in classical literature, discovered the undoubtable source. Lorem Ipsum comes from sections 1.10.32 and 1.10.33 of "de Finibus Bonorum et Malorum" (The Extremes of Good and Evil) by Cicero, written in 45 BC. This book is a treatise on the theory of ethics, very popular during the Renaissance. The first line of Lorem Ipsum, "Lorem ipsum dolor sit amet..", comes from a line in section 1.10.32.

The standard chunk of Lorem Ipsum used since the 1500s is reproduced below for those interested. Sections 1.10.32 and 1.10.33 from "de Finibus Bonorum et Malorum" by Cicero are also reproduced in their exact original form, accompanied by English versions from the 1914 translation by H. Rackham

"Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum."

Section 1.10.32 of "de Finibus Bonorum et Malorum", written by Cicero in 45 BC: "Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur?"

1914 translation by H. Rackham: "But I must explain to you how all this mistaken idea of denouncing pleasure and praising pain was born and I will give you a complete

account of the system, and expound the actual teachings of the great explorer of the truth, the master-builder of human happiness. No one rejects, dislikes, or avoids pleasure itself, because it is pleasure, but because those who do not know how to pursue pleasure rationally encounter consequences that are extremely painful. Nor again is there anyone who loves or pursues or desires to obtain pain of itself, because it is pain, but because occasionally circumstances occur in which toil and pain can procure him some great pleasure. To take a trivial example, which of us ever undertakes laborious physical exercise, except to obtain some advantage from it? But who has any right to find fault with a man who chooses to enjoy a pleasure that has no annoying consequences, or one who avoids a pain that produces no resultant pleasure?"

Section 1.10.33 of "de Finibus Bonorum et Malorum", written by Cicero in 45 BC: "At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti quos dolores et quas molestias excepturi sint occaecati cupiditate non provident, similique sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinctio. Nam libero tempore, cum soluta nobis est eligendi optio cumque nihil impedit quo minus id quod maxime placeat facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum hic tenetur a sapiente delectus, ut aut reiciendis voluptatibus maiores alias consequatur aut perferendis doloribus asperiores repellat."

1914 translation by H. Rackham: "On the other hand, we denounce with righteous indignation and dislike men who are so beguiled and demoralized by the charms of pleasure of the moment, so blinded by desire, that they cannot foresee the pain and trouble that are bound to ensue; and equal blame belongs to those who fail in their duty through weakness of will, which is the same as saying through shrinking from toil and pain. These cases are perfectly simple and easy to distinguish. In a free hour, when our power of choice is untrammelled and when nothing prevents our being able to do what we like best, every pleasure is to be welcomed and every pain avoided. But in certain circumstances and owing to the claims of duty or the obligations of business it will frequently occur that pleasures have to be repudiated and annoyances accepted. The wise man therefore always holds in these matters to this principle of selection: he rejects pleasures to secure other greater pleasures, or else he endures pains to avoid worse pains."

Chapter 2

Background and Literature Review

This chapter aims to provide the basic background knowledge to understand the work of thesis. Clustering algortighm and Dictionary Learning algortihms are discusses. We further discuss the various evaluation measures.

2.1 Machine Learning

Machine learning is the study dealing with the process to learn characteristic information from data. Given some data, X , a machine learning algorithm learns a function $f(x)$ which maps the input to an output variable y . The learnt model than can be used to make predictions on previously unseen data X'. The aim of a machine learning algorithm is to learn a generalized model which can make the best possible prediction on the new previously unseen data.

To train the algorithm, it is fed with data points (known as training observations/instances). The data points in the set are each described by a set of features, know as feature vectors. These descriptors can be learned from the data or can be computed for each instance. As an example, supppose we have a set of patches from an image as our data points. Descritors denote the key properties of data points.

Let $X = \{x_1, x_2, ...x_n\} \in R^{mxn}$ be a set of images patches with dimesnsions mxn. The feature vector describing these data points can than be denoted as

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{pmatrix}$$

Here each row of the matrix dentoes individual data points and $x_{i,j}$ are the gray scale intensities of each pixel of the given patch x_i.

Depending on whether the data is labelled or unlabelled, the machine learning algorithms can be divided into two types namely :

1. Supervised Learning Algorithms

2. Unsupervised Learning Algorithms

2.1.1   Supervised Learning

Give a training set of labelled data, supervised learning algorithims fit a model to the training set with the aim of predciting the unknown labels for the test instances (observations). During the training phase, the algorithms learns the unkown parameters for the model.

[Describe the process]

[Example of a learning algorithm]

2.1.2   Unsupervised Learning

In unsupervised learning settings, the task is to find group relations between instances of the unlabeled training dataset with a subsequent aim of categorizing or clustering the data. The algorithm aims to understand the general properties/structure in the dataset.

[Explain the method]

2.2   Clustering

A cluster is collection of data points grouped together on basis of some common properties.The data objects or points within a cluster are similar to each other, whereas the points in different groups are disimilar.

The process of partioning of data points into smaller groups (called as clusters) with the aim to minimize the intra cluster variance and maximize the inter cluster variance. The grouping of data points is based on the similarity or disimilarity of the objects as described by their properties or features. Clustering data objects create subspaces within the data.

[For example, given a toy dataset with 10 data points]

Similarity or disimilarity between two objectives can be very subjective and hence various measures are used to describe them quantitatively with distance measure being the most common.The distance measure is used to specify the distance between two objects and can

be used to create a distance matrix called as similarity/disimilarity matrix. The most used distance metric is the Minkowski Distances.

[Minkowski distances]

As the data labels are unknown,cluster analysis is known as an unsupervised method of data partionining.This is contast to classification where the data can be partitioned on the basis of their class labels. Thus, clustering segments the data on the basis of the properties of the objects within the dataset and finds previously unkown grouping within the data.

The distance based clustering algorithms can be divided into two types namely Partioning based and Hiearchical methods.

We look into these different types of Clustering methods.

2.2.1   K-Means Clustering

In this section we look into the K-Means clustering algorithm. We assume that the number of clusters 'K' is given and we use it to initiate our clustering algorirthm.

We are give a dataset D with n objects. Each n object in the dataset is described by a feature vector of length m. The feature vector can be anything which can represent the datapoint. For examples, for an image the feature vector can be grayscale values at every pixel. The aim of the clustering algorithm is to partition the the dataset D[m x n] into 'K' clusters while optimizing an objective function. The clusters are formed so as to the objects within the same cluster are similar and disimilar to objects in other clusters.

K-Means is a centroid based partioning method I.e the centroid of the cluster is the representative data point for the cluster. The centorid of the cluster is defined by the mean of all the data points within the cluster.

Give a dataset D with n objects in the Eucledian space, the aim is to partition the dataset into k cluster C1,C2...Ck susch that Ci subset of D and Ci Cj is nil. That is each point withing the dataset is exclusive to one cluster. Each of the clusters Ci is represented by its centroid ci. Each cluster is consists of p points and the distance between the point p and the cluster representative ci is defined as dist(p,ci). If the data points are in m dimensional , then the distance is defined as [eucledian distance]

The clustering algorithm optimizes the inter cluster variation, this is the Sum of Squared Error(SSE) between the data points in the cluster and the centroid. The SSE is defined as

Error =

Algorithm

Algorithm

2.3   Decompostion Methods

2.3.1   PCA

2.3.2   Matrix Factorization

2.4   Dictionary Learning

2.4.1   Sparse Coding

2.5   Preprocessing Methods

2.5.1   Normalization

2.5.2   CLAHE

2.6   Evaluation

Chapter 3

My third chapter

3.1 First section of the third chapter

And now I begin my third chapter here . . .

And now to cite some more people Ancey et al. [2], Read [6]

3.1.1 First subsection in the first section

. . . and some more

3.1.2 Second subsection in the first section

. . . and some more . . .

First subsub section in the second subsection

. . . and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it . . .

3.1.3 Third subsection in the first section

. . . and some more . . .

First subsub section in the third subsection

. . . and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it and some more and some more and some more and some more and some more and some more and some more . . .

Second subsub section in the third subsection

. . . and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it . . .

3.2 Second section of the third chapter

and here I write more . . .

3.3 The layout of formal tables

This section has been modified from "Publication quality tables in LATEX*" by Simon Fear.

The layout of a table has been established over centuries of experience and should only be altered in extraordinary circumstances.

When formatting a table, remember two simple guidelines at all times:

1. Never, ever use vertical rules (lines).

2. Never use double rules.

These guidelines may seem extreme but I have never found a good argument in favour of breaking them. For example, if you feel that the information in the left half of a table is so different from that on the right that it needs to be separated by a vertical line, then you should use two tables instead. Not everyone follows the second guideline:

There are three further guidelines worth mentioning here as they are generally not known outside the circle of professional typesetters and subeditors:

3. Put the units in the column heading (not in the body of the table).

4. Always precede a decimal point by a digit; thus 0.1 *not* just .1.

5. Do not use 'ditto' signs or any other such convention to repeat a previous value. In many circumstances a blank will serve just as well. If it won't, then repeat the value.

A frequently seen mistake is to use '\begin{center}' . . . '\end{center}' inside a figure or table environment. This center environment can cause additional vertical space. If you want to avoid that just use '\centering'

Table 3.1 A badly formatted table

	Species I		Species II	
Dental measurement	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

Table 3.2 A nice looking table

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

Table 3.3 Even better looking table using booktabs

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

References

[1] Abramovich, Y. A., Aliprantis, C. D., and Burkinshaw, O. (1995). Another characterization of the invariant subspace problem. *Operator Theory in Function Spaces and Banach Lattices.* The A.C. Zaanen Anniversary Volume*, Operator Theory: Advances and Applications*, 75:15–31. Birkhäuser Verlag.

[2] Ancey, C., Coussot, P., and Evesque, P. (1996). Examination of the possibility of a fluid-mechanics treatment of dense granular flows. *Mechanics of Cohesive-frictional Materials*, 1(4):385–403.

[3] Aupetit, B. (1991). *A Primer on Spectral Theory*. Springer-Verlag, New York.

[4] Conway, J. B. (1990). *A Course in Functional Analysis*. Springer-Verlag, New York, second edition.

[5] Ljubič, J. I. and Macaev, V. I. (1965). On operators with a separable spectrum. *Amer. Math. Soc. Transl. (2)*, 47:89–129.

[6] Read, C. J. (1985). A solution to the invariant subspace problem on the space l_1. *Bull. London Math. Soc.*, 17:305–317.

Appendix A

How to install LaTeX

Windows OS

TeXLive package - full version

1. Download the TeXLive ISO (2.2GB) from
   https://www.tug.org/texlive/

2. Download WinCDEmu (if you don't have a virtual drive) from
   http://wincdemu.sysprogs.org/download/

3. To install Windows CD Emulator follow the instructions at
   http://wincdemu.sysprogs.org/tutorials/install/

4. Right click the iso and mount it using the WinCDEmu as shown in
   http://wincdemu.sysprogs.org/tutorials/mount/

5. Open your virtual drive and run setup.pl

   or

Basic MikTeX - TeX distribution

1. Download Basic-MiKTeX(32bit or 64bit) from
   http://miktex.org/download

2. Run the installer

3. To add a new package go to Start » All Programs » MikTex » Maintenance (Admin)
   and choose Package Manager

4. Select or search for packages to install

TexStudio - TEX editor

1. Download TexStudio from
   http://texstudio.sourceforge.net/#downloads

2. Run the installer

Mac OS X

MacTeX - TEX distribution

1. Download the file from
   https://www.tug.org/mactex/

2. Extract and double click to run the installer. It does the entire configuration, sit back
   and relax.

TexStudio - TEX editor

1. Download TexStudio from
   http://texstudio.sourceforge.net/#downloads

2. Extract and Start

Unix/Linux

TeXLive - TEX distribution

Getting the distribution:

1. TexLive can be downloaded from
   http://www.tug.org/texlive/acquire-netinstall.html.

2. TexLive is provided by most operating system you can use (rpm,apt-get or yum) to get
   TexLive distributions

Installation

1. Mount the ISO file in the mnt directory

   ```
   mount -t iso9660 -o ro,loop,noauto /your/texlive####.iso /mnt
   ```

2. Install wget on your OS (use rpm, apt-get or yum install)

3. Run the installer script install-tl.

   ```
   cd /your/download/directory
   ./install-tl
   ```

4. Enter command 'i' for installation

5. Post-Installation configuration:
   http://www.tug.org/texlive/doc/texlive-en/texlive-en.html#x1-320003.4.1

6. Set the path for the directory of TexLive binaries in your .bashrc file

For 32bit OS

For Bourne-compatible shells such as bash, and using Intel x86 GNU/Linux and a default directory setup as an example, the file to edit might be

```
edit $~/.bashrc file and add following lines
PATH=/usr/local/texlive/2011/bin/i386-linux:$PATH;
export PATH
MANPATH=/usr/local/texlive/2011/texmf/doc/man:$MANPATH;
export MANPATH
INFOPATH=/usr/local/texlive/2011/texmf/doc/info:$INFOPATH;
export INFOPATH
```

For 64bit OS

```
edit $~/.bashrc file and add following lines
PATH=/usr/local/texlive/2011/bin/x86_64-linux:$PATH;
export PATH
MANPATH=/usr/local/texlive/2011/texmf/doc/man:$MANPATH;
export MANPATH
```

```
INFOPATH=/usr/local/texlive/2011/texmf/doc/info:$INFOPATH;
export INFOPATH
```

Fedora/RedHat/CentOS:

```
sudo yum install texlive
sudo yum install psutils
```

SUSE:

```
sudo zypper install texlive
```

Debian/Ubuntu:

```
sudo apt-get install texlive texlive-latex-extra
sudo apt-get install psutils
```

Appendix B

Installing the CUED class file

LaTeX.cls files can be accessed system-wide when they are placed in the <texmf>/tex/latex directory, where <texmf> is the root directory of the user's TeXinstallation. On systems that have a local texmf tree (<texmflocal>), which may be named "texmf-local" or "localtexmf", it may be advisable to install packages in <texmflocal>, rather than <texmf> as the contents of the former, unlike that of the latter, are preserved after the LaTeXsystem is reinstalled and/or upgraded.

It is recommended that the user create a subdirectory <texmf>/tex/latex/CUED for all CUED related LaTeXclass and package files. On some LaTeXsystems, the directory look-up tables will need to be refreshed after making additions or deletions to the system files. For TeXLive systems this is accomplished via executing "texhash" as root. MIKTeXusers can run "initexmf -u" to accomplish the same thing.

Users not willing or able to install the files system-wide can install them in their personal directories, but will then have to provide the path (full or relative) in addition to the filename when referring to them in LaTeX.