Henri Mynttinen

Classification of emotions from functional connectivity graphs

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 30.9.2016

Thesis supervisor:

Prof. Lauri Nummenmaa

Thesis advisor:

D.Tech. Enrico Glerean

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Henri Mynttinen

Title: Classification of emotions from functional connectivity graphs

Date: 30.9.2016 Language: English Number of pages:7+39

Department of Neuroscience and Biomedical Engineering

Professorship: Computational Neuroscience Code: IL3003

Supervisor: Prof. Lauri Nummenmaa

Advisor: D.Tech. Enrico Glerean

Functional neuroimaging has proven to be a valuable tool in mapping local brain activation patterns corresponding to different perceptual and behavioural tasks. In functional magnetic resonance imaging the task-related changes in blood oxygenation level are inspected and inference about the involvement of brain locations is made based on statistical modelling and testing. The statistical models have traditionally been univariate allowing for inference on individual parts of the brain (voxels) and requiring correction for multiple comparisons. Recent applications of machine learning in neuroscience extended the univariate approach to multivariate methods that consider the simultaneous involvement of multiple voxels in modelling the brain activation. Particularly, classification algorithms have enabled brain state decoding in which the current task is predicted from the local activation pattern. However, the generalizability of the method in studies concerning emotional states has been poor due to the distributed nature of emotional information. In this thesis, connectivity patterns were used to discriminate between different emotional states. Based on functional connectivity between pairs of brain areas (nodes), the classifier was able to determine the corresponding emotional state by an accuracy significantly above the chance level. The classification was performed using three different sets of nodes and it was demonstrated that the choice of nodes does impact the classification accuracy. The results show that similarities exist among the connectivity patterns of multiple individuals and that discrimination between brain states is possible based on these patterns. The results also demonstrate that machine learning applications are powerful enough to extract underlying connectivity structure from the data even with moderately few samples. Further studies are required to investigate if increasing the sample size allows using more detailed node structures.

Keywords: Functional connectivity, Classification,
Feature selection

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Henri Mynttinen

Työn nimi: Emootioiden luokittelu funktionaalisen konnektiivisuuden perusteella

Päivämäärä: 30.9.2016 Kieli: Englanti Sivumäärä:7+39

Neurotieteen ja lääketieteellisen tekniikan laitos

Professuuri: Laskennallinen neurotiede Koodi: IL3003

Valvoja: Prof. Lauri Nummenmaa

Ohjaaja: TkT Enrico Glerean

Toiminnallinen aivokuvantaminen on osoittautunut arvokkaaksi työkaluksi kartoitettaessa eri koeasetelmissa esiintyvien aktivaatiokuvioiden anatomisia lokaatioita. Toiminnallisessa magneettikuvantamisessa tarkastellaan koeasetelman tehtäviin liittyviä muutoksia aivojen verenkierrossa, minkä pohjalta voidaan päätellä eri aivoalueiden osuus tilastollista mallintamista ja testejä hyödyntäen. Tilastolliset mallit ovat perinteisesti olleet yksimuuttujamalleja, jotka mahdollistavat tulkinnan paikallisella tasolla (vokselittain) ja vaativat useista vertailuista johtuvan korjauksen tilastollisen korjauksen. Viimeaikaiset koneoppimissovellukset ovat mahdollstaneet yksimuuttujamallien laajentamisen monimuuttujamalleiksi, jotka mahdollistavat useiden vokselien samanaikaisen aktivaation tarkastelun. Erityisesti luokittelualgoritmit ovat mahdollistaneet aivotilojen dekoodaamisen, jossa suoritettava tehtävä päätellään paikallisen aktivaation perusteella. Tämän menetelmän käyttö emootiotutkimuksissa on kuitenkin ollut rajoittunutta aktivaation hajautetusta luonteesta johtuen. Tässä työssä konnektiivisuuskuvioita käytettiin paikallisten aktivaatioiden sijasta emootiotilojen erottamiseksi toisistaan. Aivoalueparien (solmujen) välisen konnektiivisuuden perusteella luokittelija kykeni määrittelemään kyseessä olevan emootiotilan satunnaistasoa paremmin. Luokittelu suoritettiin käyttämällä kolmea solmujoukkoa. Solmujoukon valinnalla todettiin olevan vaikutusta luokittelutulokseen. Tulokset osoittivat, että eri yksilöiden konnektiivisuuskuvioissa on samankaltaisuuksia ja emootiotilojen erottelu näiden kuvioiden perusteella on mahdollista. Tulokset osoittivat myös, että koneoppimissovellukset kykenevät erottamaan konnektiivisuusrakenteita suhteellisen vähäisestä otoskoosta huolimatta. Tulevat tutkimukset osoittavat, onko suuremmalla otoskoolla mahdollista käyttää yksityiskohtaisempaa solmujoukkoa.

Avainsanat: Funktionaalinen konnektiivisuus, luokittelu, piirteiden valinta

Preface

Thanks to Lauri Nummenmaa for giving me this wonderful opportunity. Thanks to Enrico Glerean for his infinite patience in guiding me through it. Thanks to Olli for his friendship and for making studying fun since 2011. Finally, thanks to Philippians 4:4-9 for continually reminding me which way is up.

Otaniemi, 30.9.2016

Henri J. Mynttinen

Contents

Symbols and abbreviations

Symbols

$\mathbf{J}$	angular momentum
I	quantum number
ω	angular frequency
ω_0	Larmor frequency
γ	gyromagnetic ratio
B_0	external magnetic field
$\|\psi\rangle$	spin state
M	magnetization
T_1	time constant for longitudinal relaxation
T_2	time constant for transversal relaxation
T_2^*	time constant for combined transversal relaxation
TR	repetition time
TE	echo time
σ	summation
R_{ij}	Pearson's correlation coefficient between i and j

Opetators

$\frac{\partial}{\partial\omega}$	partial derivative with respect to ω
$\|\|\cdot\|\|$	norm of vector

Abbreviations

AAL	Automated Anatomical Labeling atlas
ATP	adenosine triphosphate
BOLD	blood oxygen-level dependent
EPI	echo planar imaging
fMRI	functional magnetic resonance imaging
HO	Harvard-Oxford atlas
MRI	magnetic resonance imaging
MVPA	multivoxel pattern analysis
NMR	nuclear magnetic resonance
SVM	support vector machine

1 Introduction

The quantity of data has increased dramatically in both scientific and non-scientific field in the past few decades. Large amount of data having multiple dimensions is an everyday challenge in today's scientific research and often requires novel ways to characterize it. Machine learning (also called statistical learning) has become a vital tool to extract information from such large data. Machine learning methods uncover underlying patterns and structure from the data and enable finding the most important features buried within the heaps of numbers.

The virtues of many machine learning methods are highlighted in studying the brain. The brain is an organ comprising of tens of billions of communication units called neurons, and of even more connections between them. The neurons are constantly sending countless of electric signals to each other during the simplest mental tasks. It has been shown that there is considerable neuronal activity even when the brain is 'at rest' ie. not performing a specific task. Powerful mathematical models are needed to characterize this flow of information.

Several non-invasive methods have been developed during the second half of the twentieth century to obtain measurements of brain activity. Functional magnetic resonance imaging is the most common tool used by both physicians and neuroscientists around the world. The method relies on an intrinsic physical property of certain atomic nuclei called spin. Many isotopes present in human body possess spin, the most common of them being the hydrogen atom. The controlled modifications of the spin system of human body allow to obtain the fMRI-signal. The signal is an indirect measure of neuronal activity in that it reflects changes in blood oxygenation caused by energy consumption related to mental tasks.

Decoding this activity related information presents a computational challenge that neuroscientists have been attempting to tackle for over two decades. The original solution was to model the activity on a single voxel level and calculate a metric describing how well the measured signal fitted the predicted time course. This method of regression analysis is known as the univariate analysis and it quickly became the standard way of inspecting fMRI-data. The mathematical foundation of the method was in random field theory, which allowed accounting for the inherent spatial correlation in the data and more efficient identification of statistically significant active brain regions during mental processing tasks.

The restriction of the univariate method was that it could not account for information encoded in the simultaneous activation of multiple voxels. This problem called for a multivariate method that would be able to identify spatial patterns of activation. The solution came in the form of a particular machine learning method, namely classification. Classification concerns the problem of categorizing observations into one of two or more classes and is accomplished through training a classifier algorithm. In a supervised learning paradigm the training process involves presenting the classifier with observations from known classes and letting the algorithm learn appropriate weight parameters based on the information structure of the data points. The classifier is then tested with a separate data set to see if the classes can be discriminated based on this information.

Classification is perfectly suited for investigating fMRI-data. As the algorithms are optimized further, the computational load is decreases allowing the classifiers are able to discern smaller details of structure in the data. This multivariate method of fMRI was named multivoxel pattern analysis (MVPA). Despite its success in recognizing detailed activation patterns on intrasubject level, the classifier often faces difficulties when trying to categorize intersubject data of complex functions such as emotions. It seems that the fine-grained patterns are at least to some extent subject specific. To overcome this limitation, the data has to be further processed before feeding it to the classification analysis. One such processing step is making of connectivity graphs from windowed time courses.

In this thesis the possibility of improving the classification accuracy of emotional states through connectivity analysis is investigated. Instead of training the classifier with images where the voxel value corresponds to the BOLD-signal, multiple images from a windowed time period are used to construct a connectivity matrix. In this matrix each element describes the similarity between the time series of two brain regions.

The emotional states that are investigated are the six basic emotions: surprise, happiness, disgust, fear, sadness and anger. Additionally the neutral state is included to represent the state on non-emotional activity. The six basic emotions refer to the work by Paul Ekman who, along with Wallace V. Friesen studied if isolated tribal cultures were able to identify a common set of emotional expressions (Ekman, 1992). Other scholars have argued for a dimensional emotion theory suggesting that all emotions can be constructed from a set of common dimensions by different weighting of the components (Reisenzein, 1992). In the classification paradigm the discrete categorization of emotions is the more useful one in modelling the brain responses but the it has been discovered that many brain areas are shared in processing different emotions (Saarimäki et al., 2015).

The second part of the thesis describes the physics behind fMRI-signal, the modelling of brain activity with networks and the mathematical background behind classification methods. The third part focuses on the data acquisition procedure and structure of the data analysis. In the next part the results of the work are presented. The results are discussed in the fifth part of the thesis and finally the conclusions of the work are outlined.

2 Background

2.1 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is widely used in medical imaging to study the anatomy of the body. MRI is a non-invasive imaging modality and it does not use ionizing radiation. It utilizes the intrinsic nuclear angular momentum called spin that is a property of many isotopes present in human body. In forming an image the nuclei of a body are first placed in an external magnetic field creating an equilibrium state among the nuclei at different spin energy levels. The equilibrium is then disturbed by a radio frequency pulse and a process of spin relaxation follows. The relaxing spins emit a decaying signal that is observed resulting in the image after a reconstruction process.

The basic mechanism in magnetic resonance imaging is exciting nuclei in the body that is studied (Purcell, 1946). But not all nuclei are visible in MRI. The nucleus is required to have an angular momentum ($\mathbf{J}$) and magnetic moment (μ) that together constitute the so called nuclear magnetic resonance (NMR) property. There are a few nuclei present in human body that possess the NMR-property, eg. ^{1}H, ^{1}3C, ^{1}9F, ^{2}3Na and ^{3}1P (Huettel et al., 2004, p. 59). Because of the large amount of water present in human body, hydrogen ^{1}H nuclei are the most common source of MR-signal in MRI and will be the focus of the following description of spin dynamics. The classical analogy of angular momentum would be the rotation of a particle around its axis. This property is also called the spin of the nucleus. In the case of ^{1}H nuclei, the particle in question is a proton. The rotation produces a circular current on the surface of the particle which in turn induces a magnetic source. When placed in a magnetic field, the particle experiences a torque due to the interaction of the field with its magnetic moment. The spin is denoted by the nuclear spin quantum number I. The spin quantum number determines the number of nuclear states $(2I+1)$ which are energetically degenerate under normal conditions. However, when an external magnetic field is applied, the degeneracy ceases and the different nuclear states have different energies. This breaking down of the degeneracy is called Zeeman splitting. A hydrogen ^{1}H nucleus has spin quantum number $I = \frac{1}{2}$ and so has two nuclear Zeeman sublevels (Levitt, 2001, p. 14). The spin of a nucleus containing multiple protons and neutrons is determined by the combining the spins of the individual particles

$$I_3 = \begin{cases} |I_1 - I_2| \\ |I_1 - I_2| \\ \vdots \\ |I_1 + I_2| \end{cases}, \tag{1}$$

where I_3 is the total spin of the resulting system consisting of parts I_1 and I_2 (Levitt, 2001, p. 8). The different energy states corresponding to the spins have three different energies. The state with the lowest energy is called the ground state.

When no external magnetic torque is acting upon the protons, the spins are oriented randomly. But when an external magnetic field is applied, the spins start to precess around an axis parallel to the field. The frequency of the precession is called the Larmor frequency and it is proportional to the magnitude of the field

$$\omega = -\gamma B_0, \tag{2}$$

where ω is the frequency of the precession, γ is the gyromagnetic ratio and B_0 is the magnitude of the field (Levitt, 2001, p. 29). The gyromagnetic ratio is a constant that is specific for each isotope. The sign of the precession frequency indicates the direction of the motion with respect to the field. Most nuclei have a positive resulting in negative frequency which means that the precession happens clockwise around the field axis as viewed from the end of the field vector.

There are two so called eigenstates of this precession corresponding to the two energy states: parallel and anti-parallel state denoted by $|\beta\rangle$ and $|\alpha\rangle$ respectively. In the parallel state the axis of precession is in the same direction as the external magnetic field and in the anti-parallel state it is the opposite. The energy of the anti-parallel state is greater than that of the parallel state. The two eigenstates are not the only allowed states for a spin. The state of the spin is usually a superposition of the two eigenstates expressed as

$$|\psi\rangle = c_1|\beta\rangle + c_2|\alpha\rangle \tag{3}$$

When the magnetic field is turned on, the spins start to precess at an angle to the field direction. Because of local field inhomogeneities, the angle changes over time and the spin executes a 'wandering' motion (Levitt, 2001, p. 31) driven towards the low energy state reaching an equilibrium which results in the net magnetization ($\mathbf{M}$) of the object. Magnetization is a vector quantity with a longitudinal component (parallel to the magnetic field) and a transverse component (perpendicular to the magnetic field) and its magnitude is proportional to the difference in the number of spins on the two energy states. Transverse components of the individual spins cancel each other out resulting in zero net transverse magnetization. Magnetization of the object and the energy difference between the two states increase as the magnitude of the external field increases. This is why strong external magnetic field (1.5—3 T) is used in MRI (Huettel et al., 2004, p. 63). The equilibrium is not reached instantly, but rather the magnetization curve behaves exponentially. The build-up of longitudinal magnetization component $M_z^{nuc}(t)$ over time is described with

$$M_z^{nuc}(t) = M_{eq}^{nuc}(1 - e^{-\frac{t-t_{on}}{T_1}}) \tag{4}$$

where M_{eq}^{nuc} is the magnitude of total magnetization in the equilibrium state, t_{on} is the field onset time and T_1 is the appropriate relaxation time constant.

Strong external magnetic field is an essential part of fMRI signal generation but to extract information the equilibrium state has to be perturbed. This is achieved by a radio frequency pulse. The frequency is chosen such that the energy carried by

a photon matches the energy difference between the two spin states. This frequency is the Larmor frequency (Huettel et al., 2004, p. 61). The pulse has the effect of rotating the entire spin population around an axis laying in the transverse plane by an angle that depends on the length of the pulse. This is called spin excitation (Huettel et al., 2004, p. 64). Before excitation the transverse components of the incoherent spins canceled each other out resulting in zero net magnetization in the transverse plane. On the other hand the imbalance between the number of spins in the different energy states resulted in the longitudinal component of the net magnetization. If electromagnetic waves oscillating at the Larmor frequency are delivered continuously for certain amount of time, there will be equally many coherent spins in both energy states. At this point the longitudinal component of magnetization disappears but the coherently precessing spins produced by the excitation now constitute a transverse component of the net magnetization. The pulse that produces this transition of the magnetization is called the 90 degree pulse (Huettel et al., 2004, p. 65).

When the transmission is turned off the net magnetization begins to change back to equilibrium. This process is called relaxation. The spins precessing in the same phase lose coherence and the excited spins return to the lower energy state restoring the longitudinal net magnetization component. These two mechanisms of relaxation are called the transverse and longitudinal relaxation. Both relaxations start at the same time but are controlled by different time constants T_2 and T_1 respectively. These time constants vary according to the tissue type. This fact is utilized to produce images where different tissue types are represented by different image intensities. The transverse relaxation or correspondingly the T_2 decay takes less time than the T1 recovery, usually only few tens of milliseconds whereas longitudinal recovery may take hundreds of milliseconds. The longitudinal and transverse relaxation processes can be expressed in terms of the three components of the magnetization as follows (Levitt, 2001, p. 35).

$$M_z^{nuc}(t) = M_{eq}^{nuc} e^{-\frac{t}{T_1}} \tag{5}$$

$$M_y^{nuc}(t) = -M_{eq}^{nuc} \cos(\omega t) e^{-\frac{t}{T_2}} \tag{6}$$

$$M_x^{nuc}(t) = M_{eq}^{nuc} \sin(\omega t) e^{-\frac{t}{T_2}} \tag{7}$$

In order to construct the three-dimensional image, the contribution of each source location to the total signal intensity needs to be determined. This is accomplished by using three gradient magnets. The Lamour frequency is determined by the nucleus-specific gyromagnetic ratio and the magnitude of the external field. With the first gradient magnet, the magnitude of the field is varied in the longitudinal z-direction. This process is also called slice selection because by applying the RF pulse on a specific frequency range, only a portion or a slice of the object can be excited. After slice selection it remains to encode the two-dimensional location information within the selected slice to the signal using the x- and y-gradients. This process can be divided into frequency and phase encoding. In frequency encoding the precession frequencies of the spins are varied in the x-direction. In phase encoding the y-

gradient adds phase difference to the similarly precessing spins. The measured signal will depend on the magnitude and duration of the two gradients applied. By measuring with different combinations of the gradients, a so-called k-space describing the spatial content frequency of the image is filled. By recognizing that the image space and k-space are Fourier pairs, a Fourier transformation can be performed on the filled k-space resulting in the image (Huettel et al., 2004, p. 109).

2.2 Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) is a popular non-invasive neuroimaging method. It produces an indirect measure of neuronal activity in a brain location by detecting changes in the blood oxygenation level. The fact that the magnetic properties of oxygenated and deoxygenated hemoglobin differ from each other is a fundamental concept in fMRI theory. Oxyhemoglobin is diamagnetic and has no magnetic moment while deoxyhemoglobin is paramagnetic. This means that hydrogen nuclei near deoxyhemoglobin experience a distorted field which consequently alters their transverse magnetization decay.

Neuronal activity comprises of populations of neurons communicating with each other by sending axon potentials. The propagation of an axon potential involves depolarization of the cell membrane potential that is produced by the cell through anion concentration manipulation. This membrane potential is restored by Na^+/K^+-ATPases acting to transfer Na^+ ions out of the cell and K^+ ions into the cell. As its name implies, ATPase requires energy in the form of ATP to accomplish this task. The ATP is ultimately provided through the metabolism of glucose, a cellular process that consumes oxygen. The cerebral blood flow adjusts in response to the oxygen consumption by directing blood containing oxygenated hemoglobin towards the active region. Because of the surge of oxyhemoglobin, the T_2^*-contrast signal from the active region increases through what is called the blood-oxygenation-level dependent (BOLD) contrast.

2.2.1 Preprocessing

Discrepancies in fMRI data appear because of features of the imaging method and the motion of the subject. Additionally, the anatomical details of same brain regions may differ greatly between subjects. Consequently, the data is usually preprocessed to remove artefacts, allow higher level data analysis across subjects and to smooth both temporal and spatial noise.

Slice timing correction

An fMRI volume is constructed by imaging a number of two-dimensional slices and then combining these to produce the final three-dimensional image. Imaging one slice takes 50—150 ms and so the time interval between two consecutive volumes, the so-called repetition time (TR), ranges from hundreds of milliseconds up to few seconds. The delay between two spatially adjacent slices is typically half the repeti-

tion time. The delay causes temporal shifts between slices of the same volume and can impair the analysis of the time series (Sladky et al., 2011).

Slice timing correction is applied in order to compensate for the delay between two spatially adjacent slices. Most common method is to use interpolation. Linear, cubic and sinc interpolation have all been proposed, but sinc interpolation is normally preferred (Calhoun et al., 2000).

Motion correction

Although movement of subjects head during the scan is highly undesirable, it cannot be completely prevented. As a consequence, the brain location corresponding to a particular voxel may change from volume to volume. The images need to be realigned in order to correct for this motion induced artefact. A template image is chosen and used as a reference. The differences between the reference image and all other images in the time series is calculated using a cost function (Jenkinson et al., 2002).

Coregistration

Images from different modalities are also realigned using similar rigid body transforms as in motion correction. However, the intensities for the same tissue may vary in different imaging modalities. For this reason a different cost function that estimates the mutual information between the images is used (Maes et al., 1997).

Spatial normalization

Analysis are often carried out on an inter-subject level. There are anatomical differences in the volume and shape of the brain between subjects, which requires the images to be normalized to some common atlas for the voxel intensities to be comparable (Jenkinson & Smith, 2001). Montreal Neurological Institute (MNI) space is commonly used.

Spatial filtering

Spatial filtering is performed in order to reduce the risk of false positive findings through high frequency components of the data. Because fMRI-data is inherently spatially correlated, it is reasonable to assume that a potentially detected activation includes a diffused area. Thus, by constructing the spatial filter based on the estimated correlation, the signal-to-noise ratio of the data can be improved (Worsley et al., 1996). This increases the chance of detecting a true positive finding in the analysis.

Temporal filtering

The fMRI-data often contains low-frequency noise component called the scanner

drift. This noise is the result of the imaging instruments heating during the scanning. Scanner drift is removed by filtering the signal with a high-pass filter at a cut-off frequency of 0.01 Hz. When functional connectivity analysis are performed, the signal is additionally band-passed filtered, typically with a 0.01-0.08 Hz filter. This is done to reduce the unwanted effects of the high-frequency components on the connectivity analysis (Purdon & Weisskoff, 1998).

2.3 Brain data as networks

Brain function involves patterns that can be characterized in multiple scales. Cognitive processes require the interplay of individual neurons that form circuits and neuronal pathways. These combine to form different brain areas and finally the whole organ. The different scales operate together and are dependent on each other. Therefore it is challenging to study complex processes of the brain by focusing on different scales separately. The mathematical modelling of complex networks offers one solution to characterizing the organization of patterns in different scales.

Analysis of complex networks has penetrated many research fields from particle physics to social sciences and neuroscience. Complex networks describe system consisting of hundreds, thousands or millions of elements forming subunits and groups eventually leading to a large entity called a network. The study of these entities explores the interplay of the elements and communication between the nodes. The goal is to discover how this interplay leads to the emergent network-level phenomena that can be quantified but not broken down.

A network is defined as a collection of nodes and edges. The edges connect the nodes forming clusters and subregions resulting in the structure of the whole network. Mathematically a network can be expressed as a symmetrical square matrix $\mathbf{A}$ where each element $\mathbf{A}_{i,j}$ represents a connection between nodes i and j of the network. If the values $\mathbf{A}_{i,j}$ are binary, ie. zeros and ones, they describe a network where the connection between the nodes either exists or does not exist. The values can also lie in some range, eg. $[0, 1]$, when they also contain information about the strength of the connection between the two nodes.

2.3.1 Functional connectivity with fMRI

Bullmore and Sporns (2009) reviewed findings in brain connectivity research obtained by different imaging modalities. They also considered how these relate to non-human studies on the subject. Graph theoretical methods have been extensively used to characterize the networks. There appeared to be convergence in these findings with respect to some topological characteristics of the connectivity architectures discovered through different modalities. The networks consistently showed small-world characteristics, short path length and clustering. The network structure seemed to be the result of a balance between two driving factors: effective information transfer and low wiring cost. Minimization of wiring cost drove the network towards a highly regular structure whereas if efficiency was the only criterion the network would be random.

Later, Bullmore and Sporns (2012) suggested that the development of brain network architecture is not the result of an optimization exercise but an economical trade-off process. According to them, the brain aims to deliver maximal value with minimal cost. The trade-offs are related to producing the most adaptive network topology with the minimal physical cost.

2.4 Machine learning

In many scientific fields the amount and dimensionality of data available for analysis is growing at an increasing speed. Traditional statistical analysis methods face challenges in extracting the characteristic features of the data. Machine learning tools and algorithms are an attempt to meet this challenge by searching for underlying structure and patterns in the data. Machine learning tools can be divided into supervised, unsupervised and reinforced learning methods.

In supervised learning the data is labelled and an error function is constructed to give feedback on the output of the algorithm. The labelled data is used as training data, a "teacher", for the algorithm. It adapts in response to the feedback and the error is reduced until convergence (Mohri et al., 2012, p. 7). Unsupervised learning methods are applied to unlabelled data. These methods search for structures and patterns without a "teacher" and they have no means to evaluate the resulting output (Hastie et al., 2009, p. 486). The reinforced learning model includes an environment which the algorithm can interact with to collect information. For each action there is an assigned reward and the method aims to maximize this reward (Sutton & Barto, 1998).

2.4.1 Classification

Classification is one of the most popular and widely applicable machine learning methods. Classification is defined as the process of categorizing an observation into one of the multiple discrete classes. Classification is related to the more commonly known method, regression. In regression, a scalar value for a dependent variable y is estimated based on a set of values of independent variables $x = \{x_1, x_2, ..., x_n\}$. In classification, the observation corresponds to a set of features x, but the predicted value y is discrete instead of continuous (Alpaydin, 2004). Classifier can be thought of as a function $f(.)$ that takes as a parameter the observation x and determines its class based on the functions output $y = f(x)$. The elements of the observation vector are also called features. In the context of fMRI-data analysis the observation corresponds to the brain activation at a given time point and features represent the BOLD-signal intensity at particular voxels (Norman et al., 2006).

2.4.2 Multivoxel pattern analysis

The main method of statistically analysing the magnitude of the activation has been the use of regression analysis. In these analysis, a model of predicted shape of activation has been fitted to the activation timeseries of each voxel. Measuring the match between the predicted and observed time series is then used to determine

weather there is activation present in the voxel. These methods are called univariate because they consider the activation of each voxel separately. However, recently the focus of analysis has shifted towards so called multivariate methods. These take into account the simultaneous activation of multiple voxels and are able to extract neural information encoded in the activation patterns (Norman et al., 2006).

Multivariate pattern analysis (MVPA) is the name used in fMRI-literature to describe analysing different brain states with classification. The goal of the analysis is to train a chosen classifier, test its generalizability by test data and evaluate its performance by an accuracy measure (Haxby, 2012). The three-dimensional voxel data is usually vectorized before feeding it to the classifier. The classifier input is a two-dimensional matrix with each column vector representing a scan that belongs to some category. Each element of the vector is the activation for the voxel. The total number of columns is divided into five parts, four of which are used for training the classifier and one for testing. This procedure is repeated five times using each one of the five parts as the testing data. The general procedure for MVPA is illustrated in Figure 1.

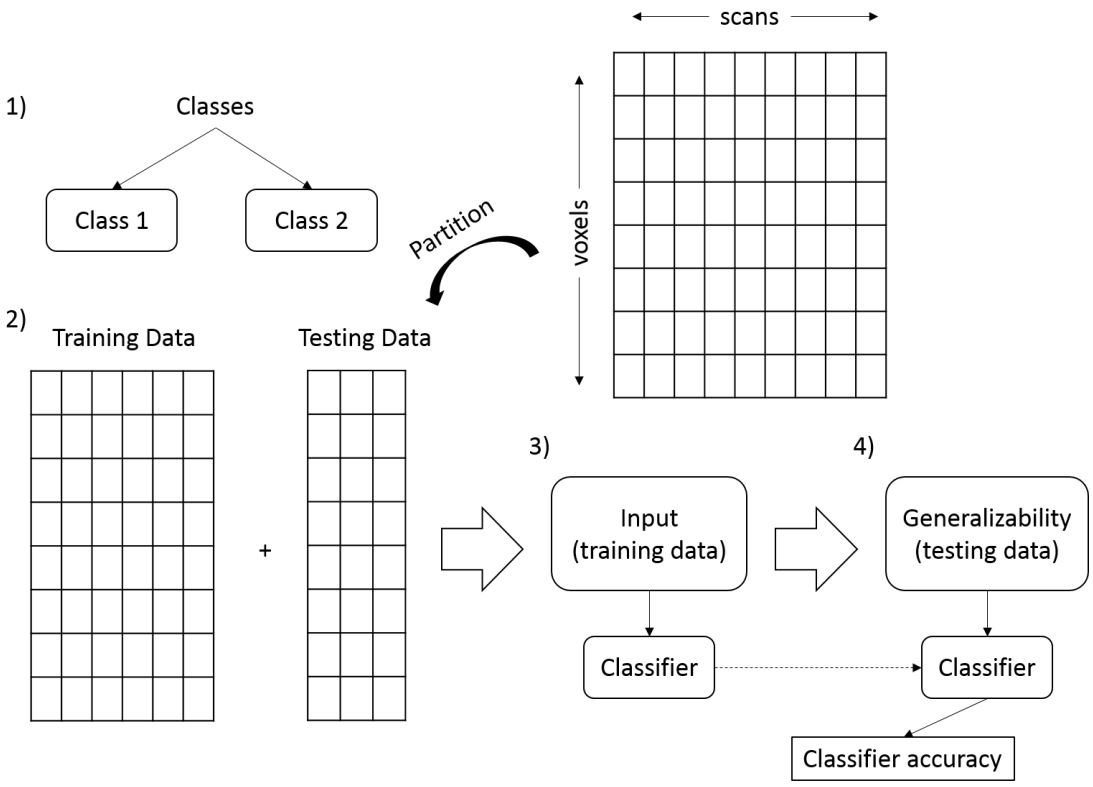


Figure 1: Illustration of the MVPA pipeline. Adapted from Norman et al. (2006)

Visualization of the classification results in neuroimaging can be divided into inspection of the classifier performance and inference on neural activation. The former can be effectively visualized using a confusion matrix. Confusion matrix is a square matrix in which rows represent the actual class {1,2,...,n} and columns

represent the predicted class {1,2,...,n}. The number in matrix element (i,j) then tells, how many samples belonging to class i were classified into class j. The diagonal, elements i=j, give the number of correctly classified samples for each class. The larger the number on the diagonal, the better the performance. Confusion matrix also gives information about which classes were confused with one another.

However, perhaps the more interesting result is which features contributed most to the classification result. This is interpreted from the classifier weights. Typically used classification algorithms determine a separation boundary (hyperplane) between the classes. The coefficients for each feature indicate how much variability between the classes there is in this dimension. Plotting the magnitude of the classifier weights might give an idea where the regions critical for the studied functions are located.

2.4.3 Linear classifiers

Linear classifiers derive their name from the characteristics of the boundary they produce to solve the classification problem. The boundary is a hyperplane (a straight line in two dimensions) and has the familiar mathematical form of

$$f = w_1 x_1 + w_2 x_2 + ... + w_n x_n \tag{8}$$

The learning algorithm computes the weights w_i of the plane, typically minimizing the value of some loss function. The input x is classified to class $y = \{0, 1\}$ by calculating the weighted linear combination of the features and thresholding the result using some mapping function f.

$$y = f(w \cdot x) \tag{9}$$

Linear classifiers can be further divided into generative and discriminative models. Generative models assume a conditional distribution for the samples given their class $f(x|class)$ and additionally often impose further assumptions on the shape of this distribution or the indepence of the observations. However, in the present application, the validity of these assumptions cannot be easily confirmed. For this reason, only discriminative models are considered. Examples of different discriminative models are described below.

Logistic regression

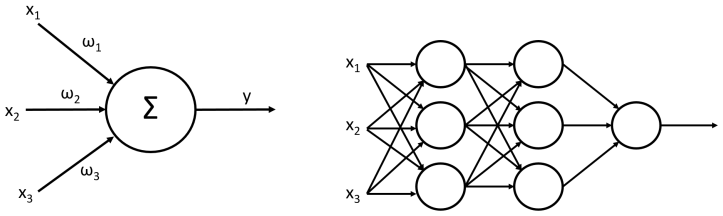
Logistic regression is a type of linear classifier that uses the sigmoid function as the means of mapping the weighted combination of features to the decision space. The mathematical form of the sigmoid function is given in ref and its shape is sketched in ref below.

$$\sigma(z) = 1/(1 + e^{-z}) \tag{10}$$

There are several arguments for using a mapping function that has this saturating quality. While training the classifier, the loss function is differentiable. The saturating shape also promotes a large margin between the classification boundary and the separated classes.

Multilayer perceptron model

The concept of multilayer perceptron model builds on the idea of combining the output of classification units such as the sigmoid function described above. The perceptron is a general term and has historical significance in the machine learning literature. It refers to supervised classification process in general, more specifically the operation of weighted combination of inputs and the assigning of the sample to one of two categories through thresholding. The idea of multilayer model derives from the notion of further combining the output of multiple perceptron units. Such a combination is actually a perceptron itself. This is demonstrated in Figure 2 below.



(a) Single perceptron. (b) Multilayer model.

Figure 2: Single perceptron and multilayer model illustration.

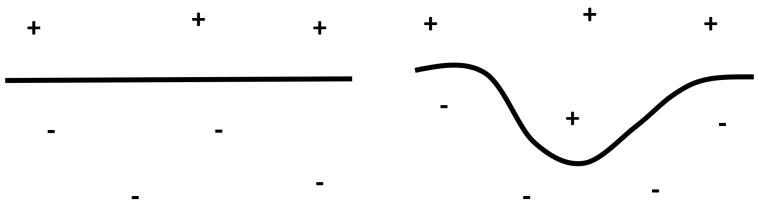
The weights of a single perceptron unit represent the contribution of the input features. In the multilayer model, the outputs are further weighted and combined, allowing for increased representional capability. This results in a powerfull classification mechanism. However, as the multilayer model increases in complexity, the association between the original feature of the data and the representational content of the weights gets more difficut to visualize. Also, the choice of model parameters, the number of layers and percpetron units per layer, may be somewhat artificial.

2.4.4 Support vector machine

Support vector machine (SVM) is one of the most popular and widely used classifiers. In fMRI-studies only, SVM has been applied in temporal classification, lie detection, classification of brain states and diagnosing deseases (LaConte et al., 2005; Davatzikos et al., 2005; Mourão-Miranda et al., 2005; Vemuri et al., 2008). Among the many machine learning tools developed for classification, SVM is a relatively recent invention (Vapnik, 1995).

The derivation of the decision boundary for an SVM begins with an assumption of a linearly separable data that consists of samples $\{x_i, y_i\}$, where $i = 1, ..., n$. The data points belong to one of two groups. If data point x_i belongs to group one,

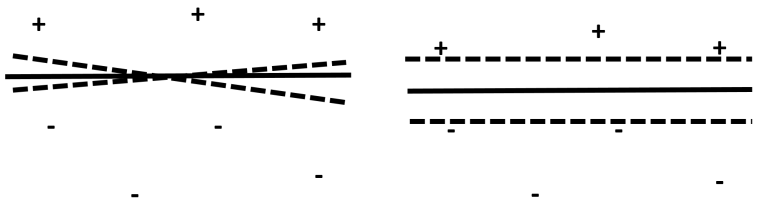
$y_i = 1$, otherwise $y_i = -1$. The assumption regarding linear separability means that it is possible to draw a line in the data space so that the line separates the data points belonging to the two classes. This is illustrated in Figure 3 below. The assumptions are made only for the derivation presented here. The classifier can be extended to multiple classes and also applied to non-separable data through the kernel method described below.



(a) Linearly separable data. (b) Linearly non-separable data.

Figure 3: Illustration of linearly separable and non-separable data.

The decision boundary can be placed in multiple orientations while still separating the two classes correctly. Different classifiers try to solve the optimal orientation for the boundary in varying ways. The SVM places the boundary based on a large margin criterion. It's orientation is such that the distance between the boundary and the data points closest to it is maximized. This is illustrated in Figure 4.



(a) Multiple possible boundaries. (b) Large margin criterion.

Figure 4: Illustration of large margin criterion.

The slope of the boundary is defined by two weights ω_1 and ω_2. These are the elements of the normal vector ω of the boundary. Given a data point u with unknown class, it needs to determined, on which side of the boundary it lies by calculating the magnitude of its projection on the normal vector as illustrated in Figure 5(a). If this magnitude is greater than or equal to some constant $\omega \cdot u \geq c$, the data point is classified in the first class and otherwise in the second one.

$$\omega \cdot u + b \geq 0 \quad \text{then } y_i = 1 \tag{11}$$

Here $c = -b$. The parameters ω and b are unknown and need to be solved by adding some constraints. To introduce the idea of a margin, the decision rule is required to give a value greater than one for positive samples and less than negative one for negative samples. These two expressions can be combined by taking advantage of the class indicator variable y_i associated with each sample.

$$\omega \cdot x_i + b \geq 1 \quad \text{for } y_i = 1 \tag{12}$$

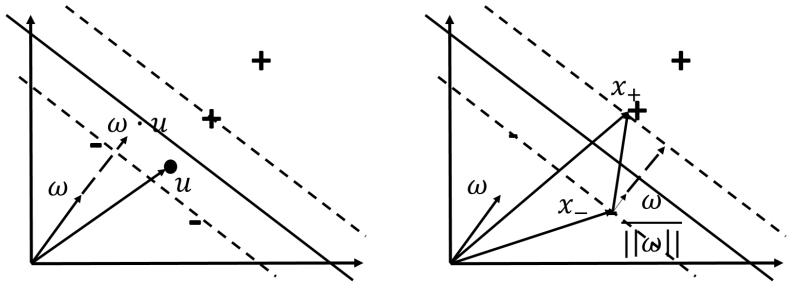
$$\omega \cdot x_i + b < -1 \quad \text{for } y_i = -1 \tag{13}$$

$$y_i(\omega \cdot x_i + b) \geq 1 \quad \text{for all } \{x_i, y_i\} \tag{14}$$

This means that for the samples lying on the edge of the margin, the following expression holds

$$y_i(\omega \cdot x_i + b) = 0 \tag{15}$$

These samples are called the support vectors. These are denoted by x_+ and x_- in Figure 5(b).



(a) Decision rule. (b) Width of the total margin.

Figure 5: Illustration of the decision rule and the width of the total margin.

The expression for the width of the total margin separating the positive and negative samples can be obtained by projecting the difference between a positive and a negative support vector on the unit normal vector $\frac{\omega}{||\omega||}$.

$$(x_+ - x_-) \cdot \frac{\omega}{||\omega||} = (x.\omega + -x_- \cdot \omega) \cdot \frac{1}{||\omega||} = (1 - b + 1 + b) \cdot \frac{1}{||\omega||} = \frac{2}{||\omega||} \tag{16}$$

This is the quantity that is being maximized. This is equivalent of minimizing the expression $\frac{1}{2}||\omega||^2$. Minimization is done using Lagrange multipliers.

$$L = \frac{1}{2}||\omega||^2 - \sum \alpha_i[y_i(\omega \cdot x_i + b) - 1] \tag{17}$$

$$\frac{\partial L}{\partial \omega} = \omega - \sum_i \alpha_i y_i x_i = 0 \rightarrow \omega = \sum_i \alpha y_i x_i \tag{18}$$

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0 \rightarrow \sum_i \alpha_i y_i = 0 \tag{19}$$

The resulting normal vector is a weighted sum of the samples. The optimization problem is quadratic because of the second power of the quantity ω that is being optimized. There are a number of numerical methods available to solve the optimization problem, Logo and Cplex algorithms among them.

But the expression 17, that is minimized, can be modified further by plugging in the constraints obtained in 18 and 19

$$\frac{1}{2}\left(\sum_i \alpha_i y_i x_i\right)\left(\sum_i \alpha_i y_i x_i\right) - \sum_i \alpha_i y_i x_i \left(\sum_j \alpha_j y_j x_j\right) - \sum_i \alpha_i y_i b + \sum_i \alpha_i, \quad (20)$$

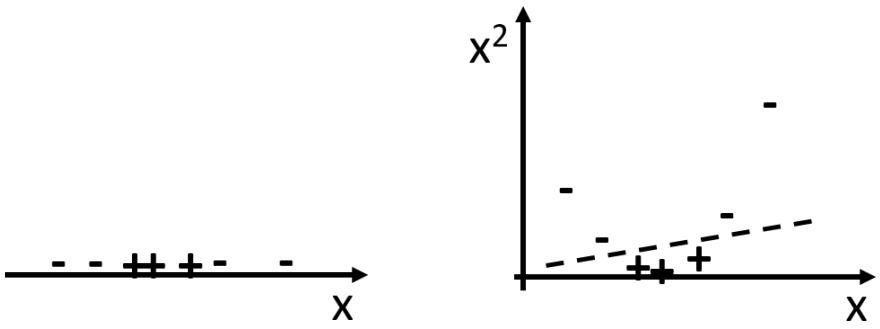
which can be simplified into

$$\sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i, \cdot x_j \quad (21)$$

using the constraint 19 and combining the first two sums (Vapnik, 1995). This form of the expression, that is being minimized, is the starting point for the discussion of applying the kernel method to the SVM.

2.4.5 Kernel Method

Kernels are the feature that make SVMs appealing and popular. The kernel trick is a widely used method to expand the feature space of the data. This is done in order to make the data more linearly separable. This is illustrated in Figure 6.



(a) Data linearly unseparable. (b) Data linearly separable.

Figure 6: Increased linear separability by feature expansion.

In the examples of the previous section the data was always linearly separable in order to more clearly demonstrate the concept of classifier margin. However, in Figure 6(a) the data cannot be separated using a straight line. Hence, when using

feature expansion by adding the feature x_i^2, it can be seen that the data becomes separable by a linear boundary as illustrated in Figure 6(b). Next, the effect of feature expansion on the SVM computation is investigated.

In the last section, the SVM primal and dual forms were introduced and in the end the following form of the optimization problem for calculating the weights was obtained.

$$\sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i \cdot x_j \tag{22}$$

This expression depends on the pairwise dot products of the data points, namely x_i and x_j. The matrix formed by the product elements is called the Gram matrix and can be viewed as measuring the angular similarity between the data points. If the points are orthogonal, the product is zero. If the angle between the data points is small, the product increases.

Suppose that feature expansion was performed on the pair x_i and x_j using the polynomial features.

$$\Phi(x) = (1 \ \sqrt{2}x_1 \ \sqrt{2}x_2 \ ... \ x_1 \ x_2 \ ... \ \sqrt{2}x_1x_2 \ \sqrt{2}x_1x_3 \ ...) \tag{23}$$

The dot product between the two data vectors x_1 and x_2 now becomes

$$\Phi(x_i)^T \Phi(x_j) = 1 + \sum_j 2x_{1j}x_{2j} + \sum_j x_{1j}^2 x_{2j}^2 + \sum_j \sum_{k>j} 2x_{1j}x_{1k}x_{2j}x_{2k} + ... \tag{24}$$

which can be simplified into

$$\Phi(x_i)^T \Phi(x_j) = \left(1 + \sum_j x_{1j}x_{2j}\right)^2 = K(x_1, x_2) \tag{25}$$

Here the effectiveness of kernel functions in feature extension is revealed. This simple expression is much easier to calculate than transforming the data vectors into the new feature space and computing the dot product between them in the transformed space.

2.5 Decoding functional connectivity graphs

Decoding brain states through MVPA paradigm has yielded valuable information about the spatial distribution of activity patterns during various brain states. However, performing successful intersubject decoding has proven to be difficult when analysing complex states, such as emotions Saarimäki et al. (2015). Although the classifier recognizes state-specific patterns in the intrasubject data, the method fails to discriminate between emotional states when the data is pooled from multiple

subjects. This implies that the neural signature of the activation pattern is different for individual subjects. Consequently, these fine-grained differences prevent the classifier from recognizing the common larger scale features of the states. The idea behind introducing the further preprocessing step of generating functional networks before the classification is to reduce these signature differences between the subjects and enhance the intersubject classification accuracy. The high frequency aspects of the data that are supposed to confuse the classifier are reduced.

It can also be argued, that connectivity features express brain function better than the local activation. The ability of brain to integrate multimodal sensory information and execute cognitive processes relies on the simultaneous interplay of various brain regions (Van Den Heuvel & Pol, 2010). Emotions have been shown to be no exception to this (Saarimäki et al., 2015), and it is therefore reasonable to use connectivity information in decoding the affective brain processes. However, accurate connectivity measurement requires temporal stability over a relatively long time period. The effects being estimated by correlation coefficients are small by default and thus require a wide temporal window in order to manifest. This places additional demand on data acquisition and conclusion can often be drawn only from large quantities of data.

Shirer et al. (2012) used classification of functional graphs to successfully decode cognitive states. They constructed the graphs using independent component analysis on group-level resting state data measured from a subset of the subjects. After thresholding, this yielded 14 components which were then used to identify 90 regions of interest. Shirer et al. (2012) were able to classify with reasonable accuracy four cognitive states based on functional connectivity graphs. The accuracy level was around 80% even when time windows of 0.5-1 min were used to construct the graphs.

Further, Richiardi et al. (2011) were able to discriminate the brain state of viewing a movie and resting. They also created an atlas of 90 regions of interest based on structural MRIs of the subjects. Richiardi et al. (2011) used an ensemble classifier which was based on classification trees.

In this study the connectivity based classification of fMRI-data was used to decode the representation of emotions in the brain. The goal was to see if connectivity based classification would be able to perform reasonably when classifying intersubject data. First, MVPA-analysis was conducted to confirm that the data agreed with previous findings. Then, the connectivity based analysis were performed and their statistical significance was inspected. Finally, the major cortical components of the connectivity features were visualized.

3 Materials and methods

3.1 Subjects

Sixteen female volunteers with ages 20-39 (mean 25 years) took part in the experiment. The subjects were right-handed, neurologically healthy and with normal or corrected-to-normal vision and hearing. They also gave written informed consent according to the Declaration of Helsinki. The experimental protocol was approved by the Institutional Review Board of Aalto University.

3.2 Experimental Design

The stimuli consisted of thirty-five 60-second-long auditory narratives describing an event or experience with emotional content (Smirnov et al., In preparation). The narratives were designed to elicit primarily one out of six possible emotions (anger, fear, disgust, happiness, sadness, surprise) or a neutral emotional state. For each emotion state, five narratives were recorded. A female speaker recorded the narratives. She was given a list of 35 story topics (e.g., Happiness: I was lying on a blanket underneath a tree with my lover. We were kissing passionately, and I felt I was so much in love with him. Sadness: I was sitting with my mother on her hospital bed. She fell into a coma during an unsuccessful operation. She could no longer talk to me, and I knew she would never recover. Neutral: I was spending an afternoon at home. There was nothing particular to do, so I went to the kitchen. I opened the fridge and started wondering what I should eat for dinner.). The speaker then formulated them into one-minute short episodes that she could relate to and narrate as if it would have happened to her.

The narratives were spoken by the same female speaker. Speech was recorded during an fMRI scanning session (see Smirnov et al. for details) using an MR-compatible noise-cancelling recording system (FOMRI; Optoacoustics Ltd.). Another speech sample with high quality spectral shape outside of the fMRI scanner was recorded. In order to clean the noise in the original recording, the original recordings were equalized by matching their sound frequency spectra with a high quality control spectrum derived from the speech sample outside the scanner. Finally, the audio recording was cut into segments containing the narrated parts only.

In the fMRI experiment, the recorded stories were played back to the participants. The narratives were divided into five runs (seven narratives per run, one from each category). Each trial started with a visual presentation of a cue phrase, which indicated the episode that would be presented next. This was accompanied by the target emotion elicited by the episode (e.g. "happy"). The cue stayed on the screen for five seconds, and was followed by five seconds of white fixation cross presented on the screen. Then the story recording was presented with white fixation cross in the centre of the screen. The order of the narratives within the runs as well as their division into runs was same for all participants. Participants were instructed to listen to the narratives similarly as if the narrator would be telling them during a face-to-face conversation, and to try to get involved in the stories by imagining

the described events vividly. Auditory stimuli were delivered with Sensimetrics S14 insert earphones (Sensimetrics Corporation, Malden, MA, United States). Sound intensity was adjusted loud enough for each subject individually to be heard over the scanner noise. Visual cues were delivered using Presentation software (Neurobehavioral Systems Inc., Albany, CA, USA). Visual stimulation was back-projected on a semi-transparent screen using a 3-micromirror data projector (Christie X3, Christie Digital Systems Ltd., Mönchengladbach, Germany) and reflected via a mirror to the subject.

3.3 fMRI data acquisition

MRI scanning was performed with $3T$ Siemens Magnetom Skyra scanner at the Advanced Magnetic Imaging Centre, Aalto NeuroImaging, Aalto University, using a 20-channel Siemens head coil. Whole-brain functional images were collected using a whole brain T_2^*-weighted echo-planar imaging (EPI) sequence, sensitive to blood oxygenation level-dependent (BOLD) signal contrast, with the following parameters: 33 axial slices, TR = 1.7 s, TE = 24 ms, flip angle = 70°, voxel size = 3 x 3 x 4.0 mm^3. A total of 365 volumes were acquired in each run, and the first three volumes of each run were discarded due to the potential effects of temperature changes on the data. High-resolution anatomical images with isotropic 1 x 1 x 1 mm^3 voxel size were collected using a T_1-weighted MP-RAGE sequence.

3.4 Preprocessing

The runs were slice time corrected by temporal sinc interpolation followed by resampling the interpolated data to compensate for the acquisition delay. Motion correction was performed using MCFLIRT (Jenkinson et al., 2002). The scans were realigned by choosing a reference image and minimizing a cost function representing the dissimilarity between the reference image and the rest of the images. The scans were also registered to standard MNI space using FLIRT (Jenkinson & Smith, 2001). This was done in order to obtain standardized networks to be used in intersubject analysis. Spatial smoothing was left out because it could potentially reduce the high frequency information.

3.5 Data analysis

Both intrasubject and intersubject classification of the connectivity graphs were performed on the data. The intrasubject-analysis was done by usual MVPA-procedure to confirm the validity of the data. The connectivity networks were generated and fed to the same classifier used in the MVPA-scheme. Custom Matlab-scripts were used to generate the connectivity networks. Python package PyMVPA and the algorithms already implemented in the toolbox waere used for the classification.

3.5.1 Classification of whole-brain activity

The traditional intrasubject MVPA-analysis was performed in order to confirm the validity of the data. In case of a poor intrasubject classification accuracy, further analysis would have been pointless. The spatially non-normalized and unfiltered images were fed to the classifier along with the metadata concerning the emotional state of each image and the run during which the image was acquired. Images were left in the native space because the analysis was carried out on intrasubject level and spatial normalization process could introduce unnecessary noise into the data. The rest-periods of the data, that were not relevant for the analysis, were removed. A whole-brain gray matter mask was applied on the images. An SVM classifier using a linear kernel (LinearCSVMC) was used as the classifier for both intrasubject and intersubject analysis. The default parameters of the implementation were applied (C=-1, $\epsilon = 5 \cdot 10^{-5}$, $\nu = 0.5$) and the multiclass classification was handled according to one-versus-rest scheme. The partitioning for the cross-validation was based on the runs. The data was partitioned so that each run represented one segment and the cross-validation was performed according to the leave-one-run-out scheme.

3.5.2 Classification of ROI connectivity patterns

Connectivity graphs for each emotional state were generated from the preprocessed images. The functional regions of interest identified by Shirer et al. (2012) were used as the basis of the analysis. The ROIs are publicly available at `http://findlab.stanford.edu/research`. The ROI time series were created by averaging the time points from all voxels included in the ROI. The time series then needed to be parsed into blocks containing stimulus evoked activity relevant for the connectivity analysis. After the onset of the stimulus, the first five seconds were discarded because the assumed rise in the signal would distort the correlation estimation. Additional ten seconds of the wipe-out section were included because it was assumed to contain relevant signal due to the slow hemodynamic response. Altogether, the last 60 seconds of the obtained signal were used to generate the graphs. Shirer et al. (2012) demonstrated that correlation estimates using a 60 second window already contain sufficient information for decoding brain states. After creating the ROI time series and parsing them into blocks containing different emotional states, the connectivity graphs for each state were created. The Pearson's correlation between each ROI during one block was calculated.

$$R_{ij} = \frac{\sum_t [(x[t]_i - \bar{x}_i)(x[t]_j - \bar{x}_j)]}{\sqrt{\sum_t [(x[t]_i - \bar{x}_i)^2} \sqrt{\sum_t [(x[t]_j - \bar{x}_j)^2}} \tag{26}$$

The graphs were fed to the classifier. The same cross-validation scheme was used in graph classification as in the MVPA-pipeline. This resulted in one sample of each category per run.

3.6 Visualization

Hubmaps of the obtained emotion state networks were created in order to visually inspect the differences of the brain states. The maps were calculated by averaging the correlation values across subjects yielding one map for each emotion. The correlation values were first mapped from the range $[-1, 1]$ to $]-\infty, \infty[$ using hyperbolic arctangent function

$$\text{artanh} = \frac{1}{2}\log(\frac{1+x}{1-x}), \tag{27}$$

averaged and then mapped back by the inverse function

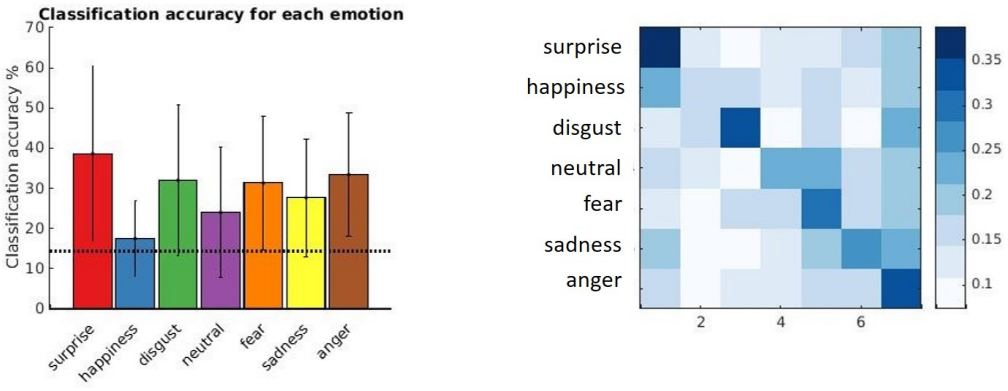
$$\text{arcoth} = \frac{1}{2}\log(\frac{1-x}{1+x}). \tag{28}$$

Finally, the strongest five percent of the links were chosen to be displayed in the resulting hub maps. The medial and lateral views of both hemispheres were plotted by Caret visualization tool (Van Essen et al., 2001).

4 Results

4.1 MVPA results

The results for regional within-subject classification of the activation patterns are shown in Figure 7. In the Figure on the left the averaged accuracy for each category (emotion) is shown. For each category the accuracy lies above the chance level although for categories happiness and neutral it is rather low. For these two categories and the category sadness the chance level is within the standard error.



(a) Classification accuracies

(b) Confusion matrix

Figure 7: Bar plot showing the mean within subject accuracy for the seven emotional states (surprise, happiness, disgust, neutral, fear, sadness, anger) averaged across all the subjects. The confusion matrix illustrating the classification results. The emotional state sadness has the highest average classification accuracy. The vertical lines depict the standard error of the mean. The accuracies of the emotional states surprise and disgust had relatively large variance across the subjects, while the accuracies for happiness showed consistence.

In the confusion matrix a relatively strong diagonal is discernible indicating reasonable classification accuracy for all categories. The confusion matrix supports the conclusion that the happiness condition is least recognized by the classifier. On the other hand the diagonal element for the conditions of surprise and disgust can be clearly discerned. It indicates that these conditions are most reliably recognized.

4.2 Classification accuracy for the emotions

The results for inter-subject classification of the connectivity patterns are shown in Figure 8. Overall, the best best classification accuracy when averaged across categories is obtained using the functional region of interest atlas as shown in Figure below. In this instance, the only category having an accuracy slightly below the chance level is happiness. Applying the AAL-atlas gives the lowest discriminative power.

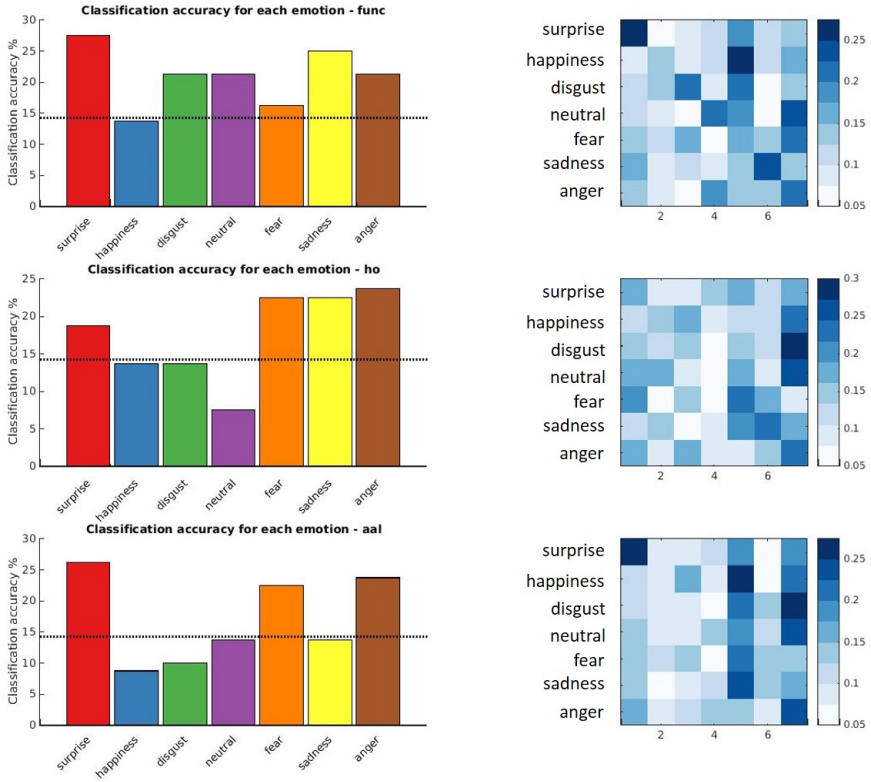


Figure 8: Bar plots showing the mean inter-subject accuracy and the corresponding confusion matrix for the three sets of nodes. The accuracies for individual categories vary greatly among the three choices. The best overall accuracy is achieved with the functional regions of interest while using the AAL-atlas gives relatively poor results. The standard error of the mean is not shown here because the accuracies could not be averaged across subjects; the data was pooled and the single accuracy obtained for the connectivity-based analysis. The significance of the results is considered below.

4.3 Connectivity maps for the emotions

The hubmaps for the seven emotional states are shown in Figure 9. The five percent of the averaged links are plotted for each state.

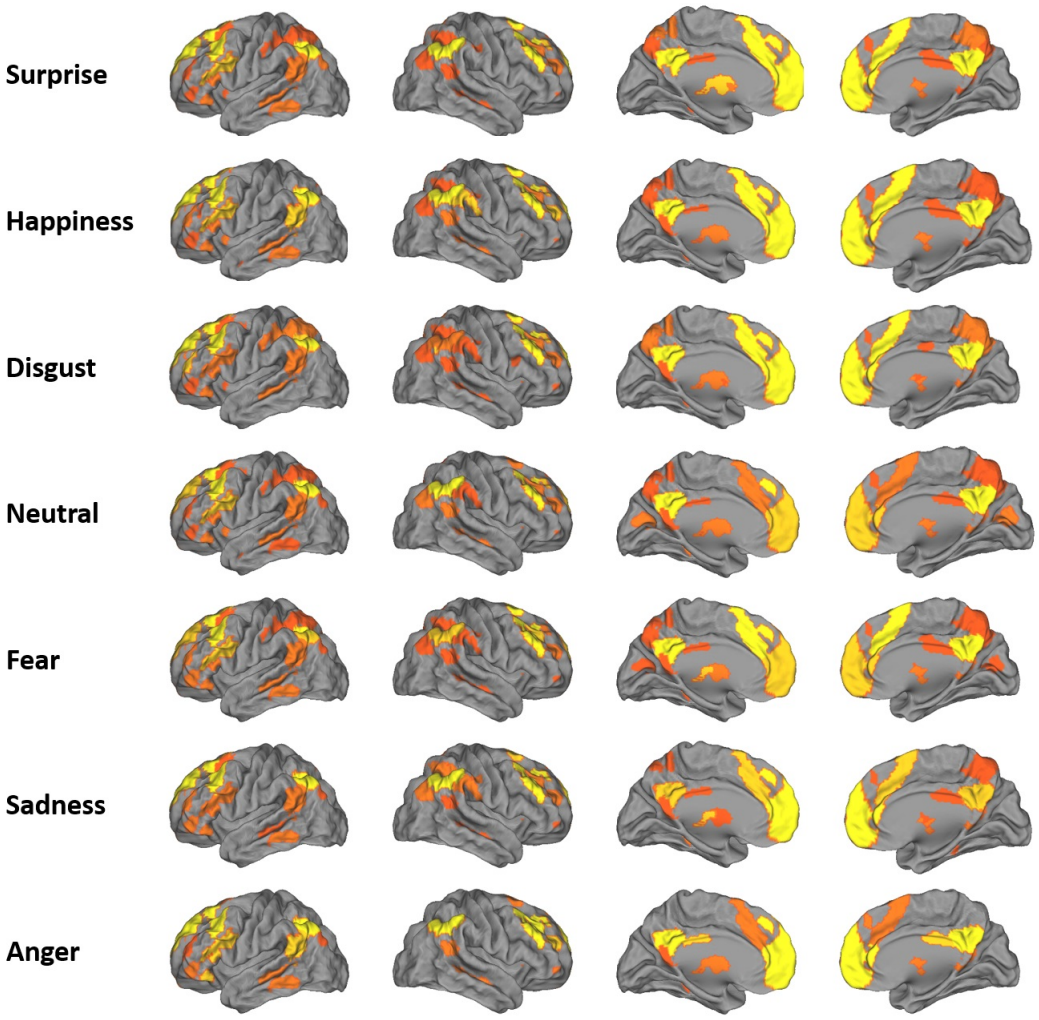
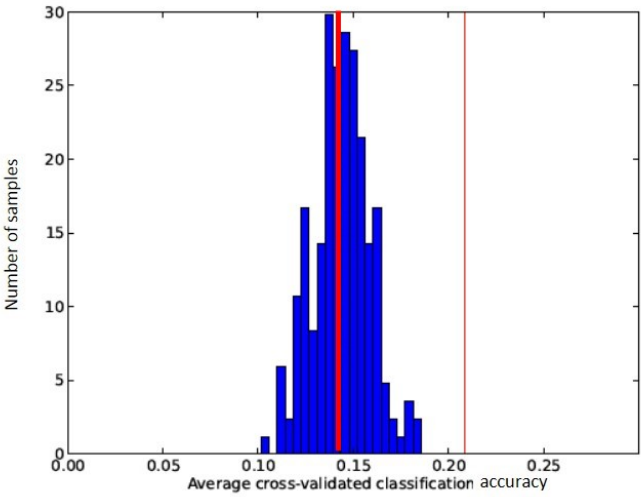


Figure 9: Hubmaps of the seven emotional states calculated from averaged node degree values and thresholded at the strongest five percent.

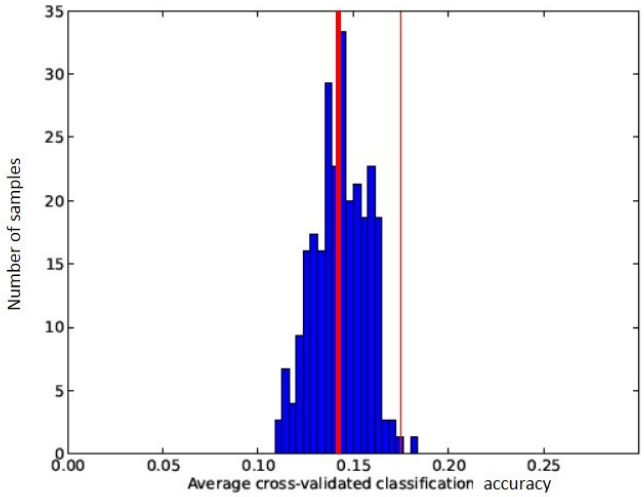
The hubmaps show similar regions in the frontal lobe are involved in multiple emotional states, although their individual strength varies slightly. Also, frontal medial areas are consistently among the strongest nodes. The hubmaps also show some salient network characteristics, including the anterior cingulate and lateral parietal cortices.

4.4 Significance testing

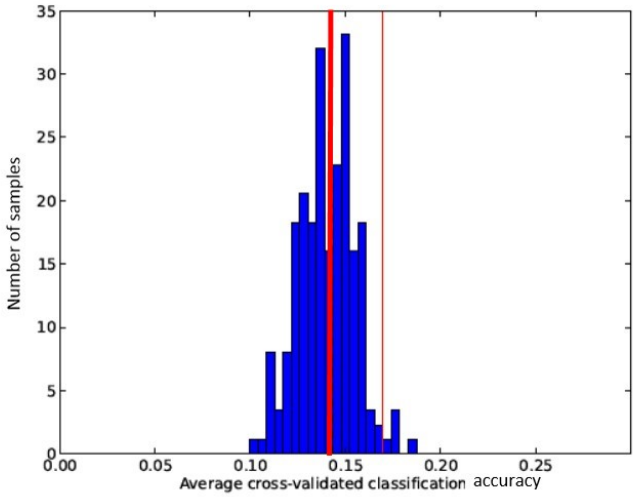
The significance testing results are displayed in Figure 10. The classification accuracies for the different atlases were: functional ROIs - 20.89, Harvard-Oxford - 17.5 and AAL - 16.96. The accuracy for the functional ROIs is clearly better than the accuracy for the two other atlases. The AAL-atlas had the lowest accuracy. The chance-level for seven classes $1/7 \approx 0.143$ is indicated by a vertical red line. The null distributions for the atlases center approximately at this point.



(a) Functional ROIs



(b) Harvard-Oxford



(c) AAL

Figure 10: Testing the significance of obtaining an accuracy above the chance-level. The null distributions, the chance-level and the obtained accuracy are plotted for each set of nodes. The average classification accuracies for the functional ROIs, Harvard-Oxford and AAL atlases were 20.89, 17.5 and 16.96 respectively. The accuracy was significantly higher for the functional ROIs than the other two atlases. The accuracies were significantly better ($p < 0.5$) than with random permutation for all three sets of ROIs. However, the mean accuracies across emotion states were clearly below that obtained with activity-based classification.

The shape of all null distributions is small-tailed. This indicates that running the classification analysis on randomized data robustly gives accuracies close to the chance level. The obtained average classification accuracies using all three atlases were significantly better ($p < 0.5$) than those obtained using random permutation testing.

5 Discussion

The main finding of the thesis was that emotional states can be differentiated on an intersubject level based on connectivity patterns. The classifier was able to discriminate between the different emotions and the significance of the classification result was confirmed with a permutation test.

In this chapter, the differences between classification based on activity and connectivity patterns is discussed and the impact of the choice of nodes is considered. Additionally, the significance testing results are inspected in more detail. Emotions as a complex brain function are difficult to study and characterize by computational methods. Whether emotions are best represented by activation or connectivity patterns is discussed by reviewing recent findings. Finally, methodological issues related to the experimental design and analysis tools are discussed along with some suggestions for future studies.

5.1 Comparison of activity and connectivity decoding

Multivoxel pattern analysis has been extensively used to investigate whether neural activation patterns are able to separate mental conditions from each other. Moving from MVPA procedure to using connectivity patterns for classification input has been a relatively new phenomenon (Richiardi et al., 2011; Shirer et al., 2012). The connectivity-based data can also be viewed as better describing the integrating nature of brain processes, especially in complex ones, such as emotions. In this study it was found that average classification accuracy was clearly higher for activity-based classification compared to connectivity-based classification. The activity-based classification accuracy was first calculated within each subject for each emotion. The average accuracy for each emotion was then calculated by averaging over the subjects. On the other hand, the connectivity-based accuracies were calculated by first pooling the data from all subjects. This was done because of the small number of samples per each subject.

Inter-subject classification of complex processes, such as emotion, has been shown to be difficult based on the mere activations. This may be due to the distributed nature of the activation. As Saarimäki et al. (2015) showed, emotions are governed by a widely distributed network and the local activation patterns of this network are possibly highly distinctive between subjects. Although the classifier is able to distinguish the neural pattern among samples from the same subject, it is not able to match the activation pattern of the same emotion from two different subjects.

However, connectivity as a metric is more distributed by nature (Bullmore & Sporns, 2009). Calculating the connectivity pattern of an emotion loses statistical learning power when the number of samples is decreased but it may better preserve the distributed nature of the activation. The local "salt-and-pepper" patterns that are individual to subjects are balanced by emphasizing the different areas that show constant activation during an emotional state. Thus it could be argued that the global connectivity patterns would show much greater resemblance between subjects. This would be due to the emphasis on the features of the data that are common

among subjects. Further, the role of the local, distinct activation patterns that confuse the classifier in the case of inter-subject would be diminished.

Classifying the brain states based on the connectivity patterns has been attempted before. Richiardi et al. (2011) were able to successfully discriminate between two brain states, watching a movie and resting state, using connectivity-based classification. The reasonable inter-subject discriminative power of the classifier between the two states is largely attributed to the similarity of the resting state network between the subjects. This assumption is supported by the fact that the weights within this network showed low variation between the subjects. In many respects, the classification problem studied in this thesis is more complex compared to that of Richiardi et al. (2011). In this study the aim was to discriminate between seven emotional states not including resting state. Moreover, classification of two states is made easier if the sates do not resemble each other whereas the seven emotional states very likely share some connectivity patterns. Also, discriminating between resting state and a task-related state on inter-subject level is made easier by the close resemblance of the connectivity architecture of the first condition between the subjects. The difference in the choice of classifier is also of significance when comparing the two results. While the use of an ensemble classifier combining the discriminative power of multiple frequency bands is a novel way to solve the problem of small sample size, the results are not strictly comparable with activation based MVPA results, where a linear classifier was used. In this study the possibility of intersubject classification was investigated in the context of previous work on classifying emotional states (Saarimäki et al., 2015). For these reasons, choosing a linear classifier was more in line with the previous trend.

The possibility of separating between multiple distinct brain states has also been investigated among others by Shirer et al. (2012) who were able to classify four distinct brain states based on their functional connectivity patterns. Consequently, the ROI map they constructed for the connectivity analysis was also chosen for this study, as discussed further in the next section. The brain states they studied included remembering past events, simple mathematical calculations and singing song lyrics in the subject's head. The classification procedure used by Shirer et al. (2012) was a novel one. Connectivity graphs were constructed for each of the states and the nodes that showed relevance for each of the states were identified. Thus, as in Richiardi et al. (2011), the results to the present study are not directly parallel. However, one of the key observations made by Shirer et al. (2012) was that discriminative connectivity information could be obtained from signals of the length of one minute. Consequently, signals of the length of approximately one minute were extracted for the analysis in the present study.

In general, the comparison between activity-based and connectivity-based decoding is made difficult by the varying methodological approaches taken in the choice of the classifier. Support vector machine has usually been the popular choice in MVPA-pipeline, but was not used in the previous studies described here and in the background. Also, the complexity of the stimulus and the number of categories vary greatly between studies. The results in this study provide a useful reference point in comparing the two classification approaches because the same classification

pipeline was used for both features. In the context of emotional processing, the results confirmed a previous finding of great intrasubject accuracy by the activity-based approach but poorer generalization in the intersubject setting (Saarimäki et al., 2015). The intersubject accuracy was significantly better when using the connectivity-approach but the intrasubject classification failed possibly due to lack of samples. Further studies should compare the approaches more thoroughly with a larger dataset.

5.2 Impact of regions of interest selection

In this section, the considerations leading to the ROI atlases used in this study are discussed. For carrying out effective but non-biased connectivity analysis, the choice of mapping out the regions of interest is critical. In the ideal situation, the feature space used in classification analysis should be of as high resolution as possible. However, in constructing the network for the connectivity analysis, the number of interactions between signal components leads to a dramatic expansion in the dimensionality of the data potentially confusing the classifier. Thus, the original feature space is compressed by averaging the signal on a collection of predetermined nodes which then form the basis for the network. However, the nodes need to be chosen carefully in order not to introduce unwanted bias to the classification result.

First, it should be noted that the issue of whether regions of interest are useful compared to having no "a priori" assumptions about the structure of the network is still open. Marrelec and Fransson (2011) inspected the effect of ROI selection method on functional connectivity analysis and concluded that their main results of interest were unaffected by the selection procedure. They used four different ROI selection methods and investigated whether the default mode network (DMN) would show decrease during a task when compared to resting state. DMN decreased in all four cases. They pointed out that the methods used for ROI selection are based on resting state data which could potentially lead to suboptimal ROIs for task related data. In conclusion, they remarked that different selection methods had relatively little impact on resting state connectivity.

There is a clear trend towards using voxel-wise networks (Stanley et al., 2013). Voxel-wise networks are defined by calculating the connectivity measure between each pair of voxels, possibly after some downsampling of the data space. The rational for using voxel-wise graphs is that networks that are constructed based on ROIs depend upon a priori knowledge. They are undoubtedly useful in confirming previous findings but also include a bias towards these results. Stanley et al. (2013) argued that voxel-wise networks are essential in discovering new principles of the brain.

In this thesis, the analysis was first started by attempting to classify voxel-wise networks covering the whole brain within individual subjects. The images were first downsampled to 6x6x6 mm voxel space decreasing the number of nodes of the eventual network. It was discovered that the number of features is enormously large compared to the number of samples making efficient discrimination of the classes practically impossible. Various feature reduction methods were also unable to direct the classification algorithm to the most relevant parts of the data. Thus the atlas

of ROIs was used to reduce the size of the feature space and pool the data from all subjects to increase the number of samples per condition. Moreover, the ROI atlases were chosen based on previous studies in order to be able to compare the findings with existing results.

Shirer et al. (2012) used a set of functional ROIs defined by applying independent component software FSL MELODIC (Beckmann & Smith, 2004) on the resting state data of the subjects. The ROIs were constructed independent of classification results and covered most of the grey matter. Using this atlas that was constructed completely independent of data used in this thesis guaranteed a minimal bias on the results. Besides, Shirer et al. (2012) had already demonstrated significant accuracies in classifying brain states based on connectivity patterns of this network. Richiardi et al. (2011) utilized AAL (Tzourio-Mazoyer et al., 2002) in successfully discriminating between resting state and task related connectivity patterns. Finally, Harvard-Oxford atlas was chosen because of its prominent status among the structurally defined ROI atlases.

The connectivity-based classification accuracies across all emotions were clearly highest using the functional ROI atlas. Only the accuracy for happiness was lower than chance-level. Surprise and sadness had exceptionally high accuracies. The superiority of functional ROI atlas is further demonstrated visually using the confusion matrices in figure 8 to display how the samples were misclassified. Only in the first confusion matrix one can discern the diagonal. When using the AAL atlas, only fear, sadness and anger are reasonably distinguished by the classifier. When using the Harvard-Oxford atlas, the classifier appears to recognize the surprise condition only.

The reasons for the success of the functional ROI atlas can only be speculated. The atlas does consist of anatomically relevant areas. Due to the way the components were obtained, they possibly have many features in common with the so called intrinsic network of the brain. This has been suggested to be a valid candidate for the common network underlying emotional processing, as discussed below. It is therefore possible, that some features of this atlas were able to guide the classifier to recognize some critical aspects of the data.

5.3 Significance of classification accuracies

In the pattern analysis the main result can be condensed into one value, the accuracy of the classifier. It tells how well the classifier could separate the studied conditions from each other. If there are only two conditions that the classifier needs to choose between, an accuracy that is well above 50% is usually considered significant. This means that using the classifier to label the samples is better than flipping a coin to decide the class at random.

When multiple conditions are considered, the chance level that the classifier should exceed is one over the number of conditions. Many studies are content to report only the chance level and demonstrate that the classification accuracy exceeds it. Also, it has to be noted that the overall accuracy across categories is not enough to demonstrate validity of the results because the classifier could simply recognize a

subset of the categories accurately while performing poorly on the others. The more elegant way to demonstrate that the data actually contains information about the conditions is to use significance testing. Here, the null distribution of classification accuracies was obtained under the null hypotheses that the samples contain no information about the conditions. By estimating the probability of getting the observed classification result under the null hypotheses, it was possible show whether the accuracy results were significantly above the chance level.

Figure 10 shows that the overall accuracy obtained when using the functional ROI atlas was significantly greater than the chance level. The accuracies obtained by using AAL and Harvard-Oxford atlases also exceeded the chance level. It is notable that the null distributions in all three cases were rather concentrated at the chance level. Moreover, the small tails of the null distribution indicated that accuracies exceeding the chance level cannot be obtained by classifying random data that does not contain condition-relevant signal.

There has been discussion in the literature about permutation when applied to significance testing in the context of the MVPA procedure (Etzel & Braver, 2013). The discussion has concentrated on the proper reporting of the used procedure, comparing two approches labeled the dataset-wise and fold-wise schemes. For the present analysis, the fold-wise scheme was chosen, because this is the implementation provided by pyMVPA. The main difference between the two schemes is that fold-wise permutation tends to give a more narrow variation for the null distributions, which can lead to smaller p-values. However, the implementation of a permutation analysis can be tedious and not in the range of this thesis. Despite the potentially underestimated p-value, pyMVPA is a widely used toolbox for carrying out the analysis, and the consistency of the results in this study give no reason to doubt the overall conclusions.

5.4 Activation and connectivity based representation of emotions

Central debate in psychological imaging studies concerning emotions is whether the functions under investigation can be mapped to some anatomical unit in the brain. In the beginning of imaging studies the assumption that distinct functions could be attributed to distinct brain regions was prevalent (Fodor, 1983). However, over time and after accumulating evidence, the opinion has shifted towards a more distributed nature of cognitive and affective processing.

Recent research argues that emotions cannot be attributed to only one brain location, but that they are a function involving and engaging multiple regions simultaneously (Barrett & Satpute, 2013). A potential candidate for the source of this emotion-related network would be the so-called intrinsic network. The intrinsic network is engaged with task-independent processes and is active during the resting state. However, Barrett & Satpute (2013) showed that emotion-specific networks do not exists within the intrinsic network. Instead, the overlap of emotion-specific discovery maps showed similarities to the so-called the salience network (Seeley et al., 2007). The visualization produced for the connectivity data used in this study

also supports this notion that the salient network is involved in emotional processing. As shown in Figure 9, the anterior cingulate cortex is present in all seven emotions. Areas of lateral parietal cortex seem to be consistently recruited across the categories.

Saarimäki et al. (2015) also confirmed that the six basic emotions are supported by multiple brain areas through activity based decoding. They used multivoxel pattern analysis to uncover neural signatures corresponding to the emotions. No one-to-one mapping could be found to connect each emotion to a single brain location. Instead, similar areas were found to activate during many emotions. Since emotions engage many distinct areas of the brain, it is reasonable to include the whole brain in the analysis. This thesis further explored the possibility of demonstrating that connectivity patterns indeed underlie affective processing. The results clearly indicate this and even hint that these patterns are shared between individuals.

5.5 Methodological issues

The construction of connectivity graphs has inherent technical difficulties related to head motion. Head motion is a persistent source of artifatcs in nearly all manner of analysis. In connectivity analysis, head motion has been shown to introduce artifacts into the network data by increasing the estimated correlation between voxels separated by a short distance and decreasing the correlation of voxels that are far apart (Power et al., 2012) . Power et al. (2012) investigated two methods to remove the head motion artefacts, excluding contaminated frames from the analysis and trying to separate the real signal from the noise eg. by regression. Both methods reduced the amount of head motion artefacts in the connectivity data but could not completely remove them. In this study, was used to remove the artefacts.

The difference between brain connectivity during rest and task has been researched extensively. An intrinsic network has been identified that is thought to be involved in brain processes during rest (Golland et al., 2007). Correspondingly an extrensic network responds to external stimulation and these networks appear to overlap considerably. Due to the major role played by the intrinsic network and the potentially minor effects introduced by the recruitment of the extrinsic network, the networks identified in this study for each emotional state naturally share many nodes. This is also consistent with previous findings suggesting that intrinsic and extrensic networks resemble each other (Cole et al., 2014; Betti et al., 2013). Further, the challenge faced by the classifier is to identity these potentially minor changes. The fact that in this study the identification was observed to be possible, is a testament to the capability of the multivariate method.

Ever since classification analysis were first applied to neuroimaging data, there has been discussion and misconceptions about the interpretability of the weights produced by the algorithm. Machine learning analysis can of course be used to simply produce numerical predictions that are as accurate as possible. This approach is often useful and adequate in the context of eg. brain-computer interfaces where the need for accuracy trumps interpretability. However, in the clinical context there

is a need for understanding what features of the data drive the algorithm. This has understandably led to the widespread misconception that large classifier weights correspond to significant features in the data. This was elegantly demonstrated to be false by Haufe et al. (2014).

5.6 Future work

From the beginning in designing the analysis for this thesis there were two main challenges that had to be accounted for. These are the capability of dealing with the spatial resolution of the data and the ability of the classifier to determine the part of the network that is essential in driving the classification.

The role of efficient utilization of big data has been pivotal in fields such as physics and genomics and its need in neuroscience has also been recognized (Sejnowski et al., 2014). From a computational point of view, which has been promoted in this thesis, this is the way forward in discovering the processes underlying brain functionality. However, presently the analysis tools often end up struggling with the high dimensionality of the data which has even led to researchers taking steps towards reducing the data complexity (Mwangi et al., 2014). Particularly in brain network analysis, the amount of complexity can increase dramatically as the analysis often inherently involve mapping the original features to higher dimensional space through modelling of the interactions between individual signals. The high dimensionality combined with the often relatively small number of samples makes it difficult for the machine learning tools to distinguish the signal out of the noise. Here the small number of samples was compensated by pooling the data from multiple subjects, thereby increasing the signal-to-noise ratio. The capability of the analytic tools remains a future challenge.

Another area that future research on the computational methods should focus on is the identification of the part of the network that is driving the classification. It is a challenge that can be placed in the wider context of connectivity analysis. The recognition of subnetworks that guide the emotional response based on the parameters of the obtained model would be an ideal tool for interpretation but has been demonstrated to be problematic (Haufe et al., 2014). The interpretation of algorithm weights is one aspect that has been discussed in this thesis.

Connectivity analysis is an essential tool in neuroscience in discovering how the interplay of various brain regions promotes different functionalities. The required stability of the cognitive state provides some difficulties when using fMRI as the imaging modality but connectivity analysis are also widely utilised in MEG analysis. This imaging modality provides data with much higher temporal resolution and thus allows more detailed analysis on the temporal scale. The trade-off is that the sources of MEG data are difficult to locate and often accurate source estimation can only be done on the cortical level. fMRI detects signals from both cortical and subcortical regions and allows the inspection of larger networks. This could be essential when analysing some particular brain processes that heavily rely on subcortical brain structures.

6 Conclusions

The goal of this work was to apply classification on brain connectivity data to see if emotional states can be distinguished on an intersubject level. The effectiveness of this analysis was then compared to more traditional MVPA procedures. First, the previous findings were confirmed by carrying out MVPA analysis on the activation data on intrasubject level. Then, the connectivity networks were created and used as an input for the classifier.

The main finding was that brain networks contain enough shared information for the classification algorithm to succesfully distinguish between different emotional states. This finding is interesting because the individual brain activation patterns tend to have enough unique features to confuse the classifier in MVPA analysis. The result gives reason to suggest that the further processing of the signal in constructing the networks somehow diminishes the distinctiveness of these features enabling the classifier to find the relevant structure shared by individuals. The results also clearly indicate that emotions are based on different patterns of underlying brain networks.

The analysis of complex natural stimuli is challenging the capabilities of the current methods, a fact that was encountered during this work. However, the results presented in this thesis also encourage to pursue this line of analysis in the future.

References

Alpaydin, E. (2004). *Introduction to machine learning.* MIT press.

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology*, *23*(3), 361–372.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, *23*(2), 137–152.

Betti, V., Della Penna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G. L., & Corbetta, M. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. *Neuron*, *79*(4), 782–797.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198.

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, *13*(5), 336–349.

Calhoun, V., Golay, X., & Pearlson, G. (2000). Improved fmri slice timing correction: interpolation errors and wrap around effects. In *Proceedings, ismrm, 9th annual meeting, denver* (p. 810).

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, *83*(1), 238–251.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D., Acharyya, M., Loughead, J., ... Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage*, *28*(3), 663–668.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169–200.

Etzel, J. A., & Braver, T. S. (2013). Mvpa permutation schemes: Permutation testing in the land of cross-validation. In *Pattern recognition in neuroimaging (prni), 2013 international workshop on* (pp. 140–143).

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* MIT press.

Golland, Y., Bentin, S., Gelbard, H., Benjamini, Y., Heller, R., Nir, Y., ... Malach, R. (2007). Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. *Cerebral Cortex*, *17*(4), 766–777.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Unsupervised learning*. Springer.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, *87*, 96–110.

Haxby, J. V. (2012). Multivariate pattern analysis of fmri: the early beginnings. *Neuroimage*, *62*(2), 852–855.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (Vol. 1). Sinauer Associates Sunderland, MA.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825–841.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, *5*(2), 143–156.

LaConte, S., Strother, S., Cherkassky, V., Anderson, J., & Hu, X. (2005). Support vector machines for temporal classification of block design fmri data. *NeuroImage*, *26*(2), 317–329.

Levitt, M. H. (2001). *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., & Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, *16*(2), 187–198.

Marrelec, G., & Fransson, P. (2011). Assessing the influence of different roi selection strategies on functional connectivity analyses of fmri data acquired during steady-state conditions. *PLoS One*, *6*(4), e14.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of machine learning*. MIT press.

Mourão-Miranda, J., Bokde, A. L., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: support vector machine on functional mri data. *NeuroImage*, *28*(4), 980–995.

Mwangi, B., Tian, T. S., & Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, *12*(2), 229–244.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends in cognitive sciences*, *10*(9), 424–430.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage*, *59*(3), 2142–2154.

Purcell, E. M. (1946). Spontaneous emission probabilities at radio frequencies. *Physical Review*, *69*, 681.

Purdon, P. L., & Weisskoff, R. M. (1998). Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fmri. *Human brain mapping*, *6*(4), 239–249.

Reisenzein, R. (1992). A structuralist reconstruction of wundt's three-dimensional theory of emotion.

Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., & Van De Ville, D. (2011). Decoding brain states from fmri connectivity graphs. *Neuroimage*, *56*(2), 616–626.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I., Lampinen, J., Vuilleumier, P., Hari, R., ... Nummenmaa, L. (2015). Discrete neural signatures of basic emotions. *Cerebral cortex (New York, NY: 1991)*.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., ... Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of neuroscience*, *27*(9), 2349–2356.

Sejnowski, T. J., Churchland, P. S., & Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nature neuroscience*, *17*(11), 1440–1441.

Shirer, W., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral cortex*, *22*(1), 158–165.

Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional mri. *Neuroimage*, *58*(2), 588–594.

Smirnov, D., SaarimÃ¤ki, H., Glerean, E., Hari, R., Sams, M., & Nummenmaa, L. (In preparation). Synchronized speaker-listener brain activity supports speech-driven emotional contagion.

Stanley, M. L., Moussa, M. N., Paolini, B. M., Lyday, R. G., Burdette, J. H., & Laurienti, P. J. (2013). Defining nodes in complex brain networks. *Frontiers in computational neuroscience*, *7*.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1) (No. 1). MIT press Cambridge.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, *15*(1), 273–289.

Van Den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fmri functional connectivity. *European Neuropsychopharmacology*, *20*(8), 519–534.

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of the American Medical Informatics Association*, *8*(5), 443–459.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer Science & Business Media.

Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., ... Jack, C. R. (2008). Alzheimer's disease diagnosis in individual subjects using structural mr images: validation studies. *Neuroimage*, *39*(3), 1186–1197.

Worsley, K. J., Marrett, S., Neelin, P., & Evans, A. (1996). Searching scale space for activation in pet images. *Human brain mapping*, *4*(1), 74–90.