Aalto University
School of Electrical Engineering
Degree Programme in Information Technology

Niklas Strengell

Transfer Learning in Emotion Recognition

Bachelor's Thesis
Espoo, May 5, 2017

Supervisors: University Lecturer Kalle Ruttik

Advisor: Professor Stephan Sigg

Aalto University
School of Electrical Engineering

Aalto University
School of Electrical Engineering
Degree Programme in Information Technology

ABSTRACT OF
BACHELOR'S
THESIS

Author:	Niklas Strengell		
Title:	Transfer Learning in Emotion Recognition		
Date:	May 5, 2017	**Pages:**	49
Major:	-	**Code:**	-
Supervisors:	University Lecturer Kalle Ruttik		
Advisor:	Professor Stephan Sigg		

The purpose of this thesis is to review and discuss automated emotion recognition and transfer learning. Firstly, emotional theories are discussed and the various modalities from which emotions can be recognized are introduced. Secondly, transfer learning, which is a machine learning technique for transferring previously learned knowledge, is discussed and explained in detail. Lastly, new research which exploits transfer learning techniques in automated emotion recognition is reviewed and the future of research is discussed.

Keywords:	emotions, affective computing, emotion recognition, machine learning, transfer learning
Language:	English

Aalto-yliopisto
Sähkötekniikan korkeakoulu
Informaatioteknologian koulutusohjelma

KANDITAATINTYÖN TIIVISTELMÄ

Tekijä:	Niklas Strengell		
Työn nimi:	Siirto-oppiminen tunteiden tunnistamisessa .		
Päiväys:	5. toukokuuta 2017	**Sivumäärä:**	49
Pääaine:	-	**Koodi:**	-
Valvojat:	Luennoitsija Kalle Ruttik		
Ohjaaja:	Professori Stephan Sigg		

Tämän työn tarkoituksen on tutkia ja käydä läpi automaattista tunteiden tunnistamista sekä siirto-oppimista. Ensimmäisessä osassa esitellään erilaisia tunneteorioita ja keinoja tunnistaa tunteita. Toiseksi esitellään siirto-oppimista, joka on koneoppimisen metodiikka, jossa ennen opittua voidaan hyödyntää tulevissa oppimistehtävissä. Lopuksi tutkitaan uusimpia tieteellisiä julkaisuja, joissa siirto-oppimisen keinoja on hyödynnetty tunteiden tunnistamisessa, sekä pohditaan tutkimuksen tulevaisuutta.

Asiasanat:	tunteiden tunnistaminen, tunneäly, koneoppiminen, siirto-oppiminen, tunteet
Kieli:	Englanti

Acknowledgements

Finally, my studies at Aalto University are turning to their later years: this bachelor thesis is the first step into graduation. Many thanks for professor Stephan Sigg, who provided me with such an inspiring topic.

I would also like thank of course my sponsors (i.e. (grand)parents), who've always trusted and supported me. And give lots of love for Anna, who truly makes me happy.

P.S. The late and long hours spent writing this thesis also proved once again my long-standing hypothesis correct: coffee truly is the greatest invention of mankind.

Espoo, May 5, 2017

Niklas Strengell

Abbreviations and Acronyms

-

.

Contents

Chapter 1

Introduction

Emotions have always been the core of the human experience and the raw material for thousands of books and stories, but now they have been in the interest of science and technology too.

It was Rosalind Picard in her essay "Affective Computing" who first proclaimed the importance of reading and understanding of emotions in the modern technological world [79]. In the time of ubiquitous computing, technology must move from computer-centered to human-centered design [14][17][74][79][76]. Emotional states are a fundamental concept in human-to-human communication and so should they be in human-to-computer interaction as well. Affective states motivate our actions and enrich our social interactions. If computing ignores these aspects, it also loses a considerable amount of information received from the user in the interaction. The paradigm of *affective computing* suggests that the user interfaces should not only respond to the users commands but also to their emotions [73][79].

The technological applications of emotionally intelligent are numerous: they range from personalized tutors [78] and homes [31] to games and entertainment [43][67]. Research shows that it could also provide cures and therapies for modern plights of autism [71], stress [42] and depression [79][97][55].

The ability to measure and recognize emotions is also in the interest of many scientific research in fields such as psychology, psychiatry, neuroscience and behavioral sciences. Automated systems with precise measuring can greatly enhance the quality and speed of modern affective research, where much of the data is stilled processed manually [38][86].

The problem is that identifying emotions comes naturally and automatically for humans, but quantitative methods to measure them have only recently been developed. Paul Ekman was one of the first ones to quantify his studies of human emotional expression and his *Facial Action Coding System* is still widely used [34]. No automated help was available at the time and he

had to rely on training human experimenters to go manually through the photographs and videos from the experiments [34]. Compared to these methods, the modern researcher now has a wide range of different precise measuring devices and he can apply a wide range of machine learning technologies to extract meaningful patterns from this data [101].

However, even though some emotional axioms such as universal facial expression have been found [36] and the computational algorithms developed to measure these corpora have greatly increased in accuracy and efficiency [5], challenges still remain.

A major assumption in many machine learning and data mining algorithms is that the training and testing data must be in the same feature space and have the same distribution [72]. However, emotions are complex phenomena and thus emotion recognition as machine learning problem is not simple. Emotions arise from countless different situations and the emotional reactions vary from person to person. Thus the data we can use to recognize emotions varies as well: it can be for example auditive, visual or physiological [101]. Documenting and data-labeling all these possibilities of emotional states would be costly and time-consuming, if not impossible, and one of the biggest problems in emotion recognition remains the huge variability in the data [16][55].

First automated emotion recognition methods have also received criticism about their usability in natural and spontaneous situation and the development of multimodal (i.e. audiovisual) systems that work in real-time and naturalistic settings has been a great interest in research [101].

However, recent advancements in computational technology and the huge amount of data available over Internet has brought automated emotion recognition technology a great leap forward [29].

In these new technologies a new approach to machine learning, called transfer learning or knowledge transfer, is giving promising results. Transfer learning is a technique to use previously learned knowledge from other tasks, which are related to the target task like human learning [72]. While traditional machine learning concentrates on solving a problem using training data in a domain, the data from various domains can be used in transfer learning, which makes it an interesting possibility for complex data [72], e.g. emotion recognition [53].

In this thesis emotion recognition and transfer learning are reviewed and the possibility of combining these two methods as well the future of research is discussed.

Chapter 2

Emotions

2.1 What is an emotion?

In daily life, we use the word *emotion* to describe an astonishing diversity of phenomena. A good cup of coffee in the morning can be emotional and the death of a loved one certainly is. The early psychologist William James famously asked in 19th century, "what is an emotion?" [52]. And quite rightly he did as to date the scientific community has yet to give an answer.

But to study emotions a certain degree of semantic coherence must be established. What's more, to understand the different terms and concepts used in various fields helps to understand the theories and their applications. In affective sciences the terms have been used and changed back and forth, but they are now quite coherently used in the following manner [44]:

- *Affect* is a wide umbrella term covering a variety of internal states, such as moods, emotions and attitudes.
- *Attitudes* are relatively stable beliefs about the goodness or badness of something or someone. They bias how a person will think about, feel towards or behave regarding a person or a thing [41].
- *Moods* are less stable than attitudes but longer lasting than emotions. Feeling tones are prevalent and moods bias cognition but not necessarily action [94].
- *Emotions* are the briefest in duration of affective processes. They are reactions to situations that are relevant for the individual's current desires. Emotions are a way of analyzing a situation and this process leads to loosely coordinated physiological and physical changes in the

individual [63]. Recent research confirms that emotions are crucial for decision making [62][6][90].

In machine learning, the word *sentiment* is also used often. It is similar to the word attitude, but sentiment refers to the valence (like/dislike). *A feeling* is the conscious perception of an emotion, but it is rarely used in science, except for it's verb form i.e. *feeling an emotion* (it could of course be argued that the answers acquired from questionnaires are not about emotions but about feelings).

In computer science there is also usually no differentiation made with the words emotion or affect [79] and I also use these words as synonyms for each other. However, this thesis is specifically interested in **emotions**, because these are the affective phenomena that can be measured.

2.2 Emotion theories

There is no doubt that the progress of automated affective recognition is tied to our understanding on the nature of human emotions. The more psychologists and linguistics understand about emotions and their linked expressions, the better we can measure them [88][37]. But how emotions are classified and distinguished from each other is still a contested issue in science [3][45]. The classification of emotions has been researched from two fundamental viewpoints: one, that emotions are discrete and fundamentally different constructs; or two, that emotions can be characterized on a dimensional basis in groupings [45]. An example of how the classification of emotional stimuli could be done can be seen in the figure 2.3.

2.2.1 Discrete emotion theories

In discrete emotion theories, humans (and to some extent animals too) are thought to have innate, basic emotions that are recognizable also in a cross-cultural context. A highly influential basic emotion theory by Paul Ekman proposes a limited number of basic emotions (for example, happiness, sadness, anger, fear, disgust, and surprise), each with their own distinctive physiological and neurological signs. The facial expressions corresponding to these basic emotions can be seen in figure 2.1. According to Ekman all other emotional states could then be considered as combinations of some number of these basic emotions [32]. For example, melancholy can be thought as an emotion that is a complex mixture of love and sadness [9].

Another influential discrete theory of emotion is *wheel of emotions* by Robert Plutchik [80]: a wheel-like diagram of emotions visualising eight basic

Figure 2.1: Examples of Paul Ekman's proposed six basic universal emotions that can be read from facial actions [32].

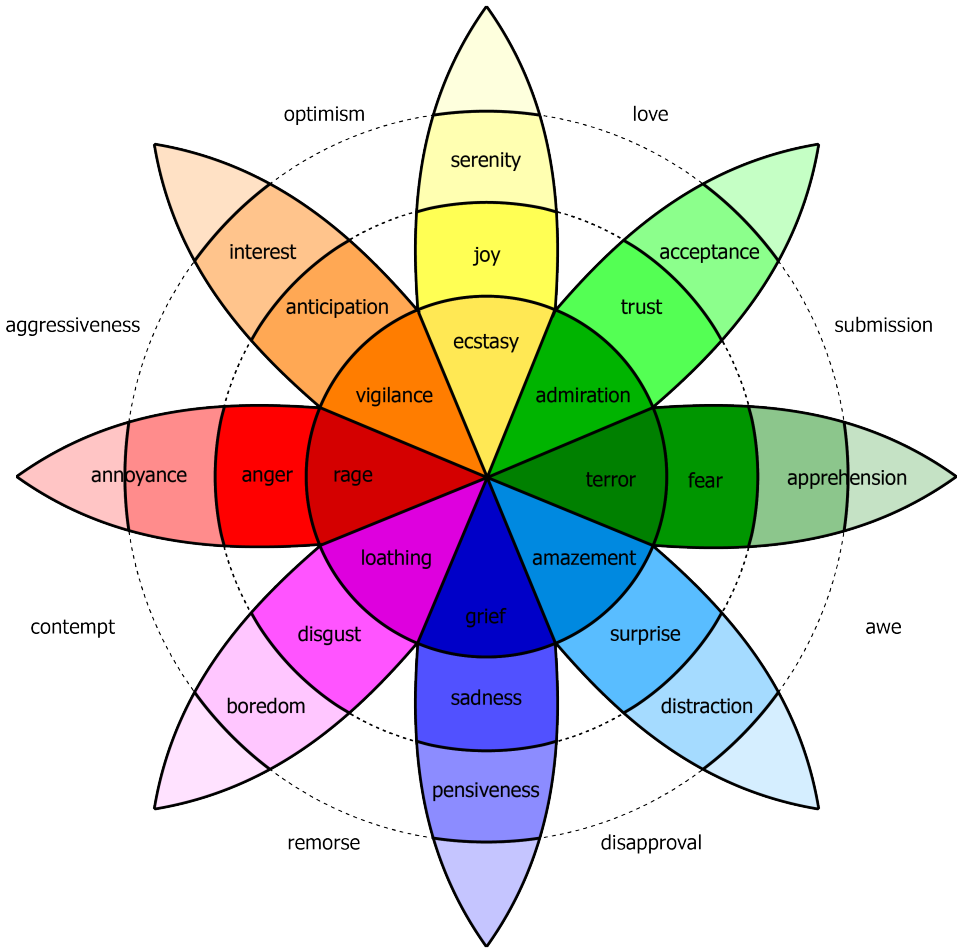

Figure 2.2: Visualization of Robert Plutchik's Wheel of Emotions. *Picture released by User Machine Elf 1735 on Wikimedia. Released under Public Domain.*

emotions: Joy, Trust, Fear, Surprise, Sadness, Disgust, Anger and Anticipation. The diagram can be seen in figure 2.2.

The wheel combines the ideas of circles representing emotions and a color wheel. Similar emotions, such as serenity and acceptance, in the wheel are adjacent and antagonist emotions are opposite to each other. Emotions are arranged in pairs according to behavioural and evolutionary mechanisms [81] and they also come in a variety of intensities. For example, Distraction is a mild form of Surprise, and Rage is an intense form of Anger. Weaker emotions lay among the outer circles and stronger emotions bloom in the middle. As with Ekman's theories, emotions can also be mixed to form new ones, e.g. love is a combination of joy and trust.

2.2.2 Dimensional emotion theories

By contrast, in dimensional emotion theories, emotions are thought in terms of dimensions, such as valence (degree of pleasantness) and arousal (intensity of the emotional state). For example, in this model fear could be thought as an emotion with high arousal and negative valence [3].

Although two-dimensional model with valence and arousal was the most common in research [45], new research argues that a two-dimensional representation is not enough and more dimensions should be added [39]. A common addition is the third dimension of *dominance* (the degree of control exerted by a stimulus) [98][49]. Fontaine *et al.* [39] further suggest that dimensional space should be four-dimensional with the axes being evaluationpleasantness, potency-control, activation-arousal, and unpredictability.

2.3 Problems

Recent research [4][64][100] suggests that basic emotion categories, such as proposed by Ekman [35] and Plutchik [82], are not by itself emotions, but merely prototypical modal responses which cannot capture the full range of human emotion. Evidence for a dimensional representation of emotion in the human brain has also not been found, but dimensional approach has been proven to be empirically powerful, successfully accounting for a wide range of emotion effects [45].

To date no unified consensus exists whether dimensional or discrete approach is better. However, brain imaging research has identified consistent neural correlates associated with basic emotions and other emotion models. Simple one-to-one mappings between emotions and brain regions have been

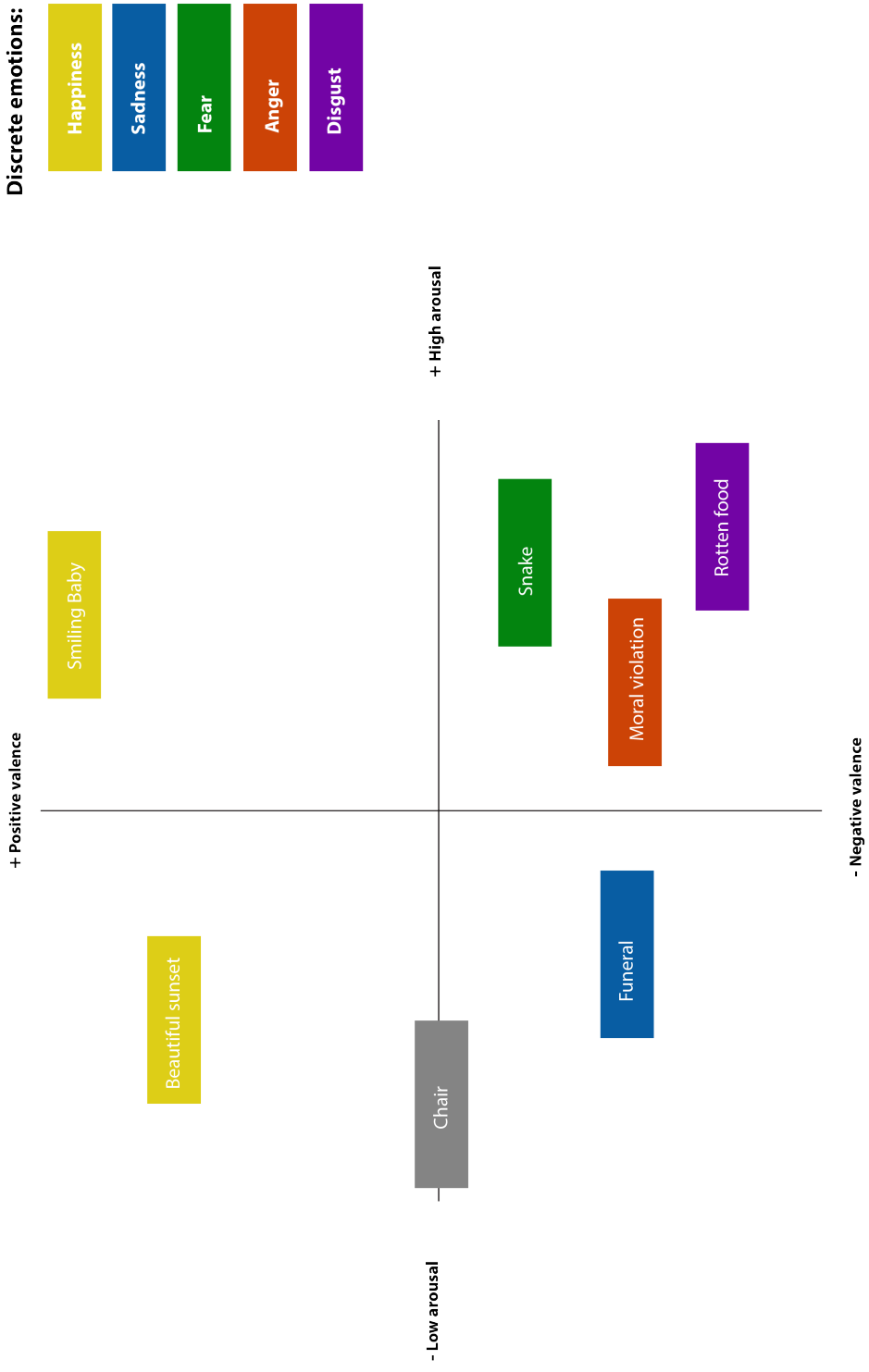

Figure 2.3: An example visualization of how emotions and emotional stimuli could be represented in dimensional or discrete emotion spaces. *Adopted from Hamann [45].*

ruled out and research points out to the need for more complex, network-based representations of emotion in the brain [45].

Chapter 3

Recognizing emotions

3.1 How to recognize emotions?

The ability to recognize emotions comes naturally to us humans. We're able to distinguish our own emotions and the people around us - even animals'. The process is intuitive: we understand emotions instinctively without the need for conscious reasoning [48]. But in this intuitive process lies the biggest problem of affective sciences: how can we measure or model something, when we don't even know what to measure?

The father of evolution theory, Charles Darwin, was one the first scientists to systematically study emotions. In his book "The Expression of Emotions in Man Animals" his concern was to show, how humans link their movements with emotional states and how they are genetically determined and derived from animal actions [18]. Even though the modern emotion theories are more sophisticated then back in Darwin's day, it is exactly this problem that automated affect recognition tries to solve: how emotional states can be read from visible and or otherwise measurable cues such as bodily movements.

3.1.1 Facial gestures

When talking about emotional cues, the first thing that crosses the mind is the face. We associate upward lips with happiness; downward with sadness. We simulate facial expression with "smileys" in written text to make-up for the lack of emotional cues in text. Thus it is no surprise, that the face is the first place the scientists looked when starting to study and quantify emotions: by far the most extensive body of data in the field of human emotions is that on facial expressions of emotion [70].

Paul Ekman found in his ground breaking study evidence for universally recognizable facial cues for emotions. They found high agreement on both

Western and Eastern literate cultures in selecting the emotional term for facial expressions. The findings were later extended to a Papua New Guinean preliterate tribe, whose members could not have learned the emotional cues from media.[34] Ekman found six basic emotions that can be read from facial expressions: surprise, anger, sadness, disgust, happiness and fear [37]. One or two more are sometimes added to this list: contempt and interest [33].

Ekman et al. [35] suggested a sign judgment method to labeling facial behavior to emotions, Facial Action Coding System (FACS). It lists all visually distinguishable facial action behaviour in terms of action units (AU). These AUs can be used for simple emotion recognition as well as complex such as pain [65] and depression [38].The beauty of FACS lies in the ability to describe all possible facial behaviour as combinations of 27 different AUs. It is thus suitable to be used in studies of naturalistic human facial expression [101]. It is also no surprise that in the recent years quite a few facial emotion recognizing sofware development kits (SDKs) have popped out such as Affectiva, Kairos, OpenCV, Amazon Rekognition, Google Vision API, Microsoft Face API, IBM Watson Visual Recognition API, Cognitec, Face++ and NEC Face Recognition.

3.1.2 Body gestures

However, facial gestures might not provide a full and comprehensive view on emotional cues. Recent studies suggest that body cues, not facial expressions, discriminate between intense positive and negative emotions [2]. Body expressions could also contribute to automatic depression analysis [55].

3.1.3 Auditive cues

Speech is a crucial part of human communication. In speech, affective information is delivered through explicit means (linguistic), i.e. which words are used, and implicit means (paralinguistic), i.e. how the words are spoken.

One of the biggest challenges is to distinguish between the relevant cues that can be linked to emotional states and the ones that are only part of a normal conversation. [27]. At the acoustic level the suggested features to measure are prosodic (fundamental frequency, duration, energy), spectral (MFCC) and voice-quality [11]. Also lexical and dialogical cues [26], speech disfluency [25] and non-verbal speech such as laughter [20] can help to classify emotions. The most widely used strategy is to compute as many of this features as possible [27].

Emotion categorization is still difficult from auditive cues and studies have mostly focused on binary classification such as positive vs. negative

emotions [60]. Trying to combine the information from paralinguistic and linguistic domains remains a research challenge [61][27].

3.1.4 Physiological signals

Although they have been in the shadow of audiovisual cues, physiological signals also provide a variety of interesting cues to emotion. The hypothesis that the feeling of basic emotions is associated with distinct patterns of somatic activity is receiving increasing support from research [84].

3.1.4.1 Skin conductance

Skin conductance or electrodermal activity (EDA) can be measured from the palms or feet of a subject with a way of two electrodes. The idea is that as sweating increases or decreases so does the the electrical conductivity in the hand. This change can then be measured.

Because sweating is controlled by the sympathetic nervous system [92] which controls the physiological or psychological arousal levels, skin conductance can be a measure of emotional response, namely alertness or arousal.

This technology has been widely developed and used by Rosalind Picard and her team in the MIT Affective Computing Lab and they've successfully used skin conductance as a measurement method in studies for stress [47], depression[77] and epilepsy [71]. The most interesting development in this field are the new wearable devices such as Empatica and finnish Moodring, which could be used to gather data in real-life situations.

3.1.4.2 Heart rate variability

Heart-rate variability (HRV) is the differentiating time interval between heartbeats. HRV is related to emotional arousal and research shows that basic emotions are associated with disctint patterns of cardiorespiratory activity [84].

The problem with HRV measurement is that it is usually done with electrocardiogram (EKG or ECG), which can be usually used only in clinical settings. However, recent developments in consumers electronics [13] and in a very fascinating video magnification technique called Eulerian Video Magnification[99] can make HRV an viable measurement for emotion recognition.

3.1.4.3 Other biomarkers

There are also more biomarkers for emotion. For stress we can measure e.g. blood pressure, heart rate, cortisol levels and pupil diameter [91]. A Finnish research team also mapped emotions to bodily sensations, which further suggests that there might be novel biomarkers for emotions still to be found [69].

3.1.5 Brain imaging

The most sophisticated and exact ways of recognizing affects are various neuroimaging techniques such as fMRI, MEG or PET. These brain imaging methods have brought great experimental data to support or contradict purely psychological models [45]. However, they are usually only used in laboratory experiments and not suited very well for naturalistic or spontaneous emotion recognition "in the wild". These methods however give us great insight into how emotions are mapped in the brain and can thus indirectly gives us hints, how other biomarkers can give us information about the emotional states.

3.2 Problems

The problem with all emotional content, is that it is usually taken at face value, even in research. Thus, an objective method is needed to assess the "real" emotional state of a human being [58].

Although there are many ways to measure the physiological signs linked to emotions, problems remain and no "holy grail" of measurement has been found.

We know from animal research that emotional processing in the brain is done by combining cues from different senses [19]. It would be natural to use this same principle in our technology too, but now most of the affect recognizing systems are only unimodal, focusing to measure only one cue, such as facial gestures. Multimodal systems combining cues from many sources are rightly now a current interest in research [101][50].

An intuitive assumption in emotion recognition since Darwin' days has been that emotions map to expressions and could thus be causally linked. However, recent research disagrees with this, and concludes it might not be that simple: emotion expressions may not be expressions and may not be related to emotions in any simple way [88]. Emotion recognition is not just a matter of "decoding a message", but a dynamic problem of ever changing

states, where important is to find out what is relevant and what is not.

Emotions are always context dependent and the responses vary from person to person. For example, in a study by Hoque *et al.* [50], they found out that people can smile during two very different kinds of affective state: delight or frustration.[1] And spotting the difference between the two can be hard. In fact, computational models did better than human subjects when asked to predict the emotional state of the subject from a still picture.

[1]"Is that smile real or fake?", a video published by the team can be found at `https://www.youtube.com/watch?v=MYmgCQjgXQU`

Chapter 4

Machine Learning

4.1 What is machine learning?

Machine learning is a subfield of computer science, that according to the definition given in 1959 by IBM researcher Arthur Samuel gives "computers the ability to learn without being explicitly programmed" [89]. Machine learning can be also thought as pattern recognition, a field where it evolved from, and it is closely linked to statistics and mathematical optimization [7]. It is also related to data mining, where the latter subfield focuses more on exploratory data analysis and is also known as unsupervised learning [66].

Machine learning tries to overcome the strictly static instructions of traditional programming by building models from previous learning data and using these learned models to make new data-driven predictions on future testing data [1]. It is thus well suited for tasks where designing and programming explicit algorithms with good performance is difficult or impossible; example applications include email filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition (OCR), learning to rank and computer vision [93] and - of course - emotion recognition [101].

4.2 Emotion recognition as a machine learning problem

Recognizing emotional states can be considered as dynamic pattern recognition problem. It requires the extraction of meaningful information from gathered data, which can be for example video for the analysis of facial expressions and body gestures [12]. Depending on the emotional theory used,

the problem can be considered as a regression task (for dimensional emotion theory) or a classification task (for discrete emotion theory).

Many machine learning algorithms have already been used to automatically recognize emotions (e.g. SVM, decision trees, linear discriminant analysis, Bayesian networks, naive Bayes, neural networks) and their usability has been proved and many state of the art affective computing SDKs use at least some machine learning in their code [101].

Emotion recognition as a machine learning problem is also gaining interest in the form of contests such as Emotion Recognition in the Wild organized annually since 2013[1] [29], Av+EC since 2014 [85], FERA2015 [95] and Conference Workshops and Special Sessions (CVPR2015-2016, ICCV2015, ECCV2016) .

4.3 The ideal emotion recognition method and it's problems

An ideal emotion recognition method would compose of adaptive classifiers, which could cope with high number of features of different types, i.e. not only combining visual cues, but combining audiovisual data. It should also be able to improve its effectiveness with increasing amounts of training data continuously recorded from users [59].

However, almost all of the automated emotion recognition systems available at the moment are unimodal, and few combine features from multiple cues (i.e. audiovisual fusion, linguistic and paralinguistic fusion, head and body gesture fusion) [101]. Another problem is the usage of artificially induced emotional data rather than spontaneous expressions recorded in naturalistic settings [101]. The variability in emotional stimuli data is also tremendous [16].

As a result, capturing a faithful and detailed record of human emotion as it appears in real action and interaction is not an easy task [16][28]. Nevertheless, the payoff is large [44][79] and new methods are thus required and encouraged to be developed [28].

[1]For more detailed information, readers are encouraged to visit `https://sites.google.com/site/emotiwchallenge/home`

Chapter 5

Transfer learning

5.1 What is transfer learning?

As noted, traditional data mining and machine learning methods have many uses in engineering and science and they are a valuable asset for automated emotion recognition. However, most of the techniques assume that the test and the training data have the same distribution and feature space. *Transfer learning* addresses this problem. It allows the domains and tasks to be different in training and testing data [72].

Humans are especially good at transfer learning. Recognizing apples helps us recognize other edible foods, such as pears, and learning to play the piano makes it easier to play the electric organ. Transfer learning is sometimes pinpointed the hallmark of human intelligence and even young children transfer knowledge on the basis of deep structural principles rather than perceptual features [10].

The ability to rely on prior learning to facilitate new learning is a great cognitive asset, for humans and machines alike. It is this *learning to learn* mindset that became the fundamental motivation for transfer learning in the field of machine learning in a NIPS-95 workshop. The workshop focused on the need for lifelong machine-learning methods that retain and reuse previously learned knowledge.[1]

[1]Neural Information Processing Systems 1995. Post-conference workshop: "Learning to Learn: Knowledge Consolidation and Transfer in Inductive Systems". `http://plato.acadiau.ca/courses/comp/dsilver/NIPS95_LTL/transfer.workshop.1995.html`

5.2 Why to use transfer learning?

Consider a problem, where out task is to classify the reviews of a product according to their sentiment, i.e whether the review had a positive or a negative opinion. Using traditional machine learning techniques for this classification task, we would first collect reviews and then manually label with correct sentiments. Then we would use this labeled data to train our classifier. However, the distribution of review data can be very different among the different kinds of products. To achieve good classification results, we would have to label huge amounts of data for each product. This would be impractical if not even impossible. In the following case using transfer learning methods can save us significant amount of time and resources spent on labeling [8].

The same impracticalities apply for every emotion recognition problem: measuring and labeling every emotional state for each individual in every context would be a massively challenging task [16].

5.3 Notations and definitions

In machine learning problems, we speak of *domains* and *tasks*. For example in the previous example, the reviews for products are the domain and the labeling is the task.

A *domain* D consists of two components: a feature space X and a marginal probability distribution $P(X)$, where $X = \{x_1, ..., x_N\} \in X$.

Given a domain, $D = \{X, P(X)\}$, a *task* consists of two components: a label space Y and a objective predictive function $f(\cdot)$. A task is thus denoted by $T = \{Y, f(\cdot)\}$.

Objective predictive function can not be observed, but it can be learned from the data, which consist of pairs x_i, y_i, where $x_i \in X$ and $y_i \in Y$. In the sentiment classification example x_i would be one of the reviews and y_i would be its class label: "negative" or "positive".

Then the objective predictive function $f(\cdot)$ can be used to predict the corresponding label, $f(x)$, for each new instance of x. From a probabilistic viewpoint $f(x)$ can also be marked as $P(y|x)$. In other words, using our example of sentiment analysis, the objective predictive function is our *model*, that tells us that *given these kind of words* this is *the probability* that *this review* has *a positive or a negative opinion*.

5.3.1 Unified definition of transfer learning

Definition 1. Given a source domain D_S and learning task T_S and a target domain D_T and learning task T_T, *transfer learning* aims to improve the learning of the target predictive function $f_T(\cdot)$ in D_T using the knowledge available in D_S and T_S, where $D_S \neq D_T$, or $T_S \neq T_T$ [72].

As defined above, a domain is a pair $D = \{X, P(X)\}$. Thus the condition $D_S \neq D_T$ implies that either $X_S \neq X_T$ or $P_S(X) \neq P_T(X)$. Using our sentiment classification example this means that either the term features (i.e. X) are different between two sets (e.g. reviews are in different languages) or their marginal distributions (i.e. $P(X)$) are different (e.g. reviews are for a different product).

Similarly, a task is defined as a pair $T = \{Y, P(Y|X)\}$. Thus the condition $T_S \neq T_T$ implies that either $Y_S \neq Y_T$ or $P(Y_S|X_S) \neq P(Y_T|X_T)$. Using the same sentiment classification example this means that the label spaces (i.e. Y) for tasks are different (e.g. the source task is a binary "positive/negative" sentiment classification, but target task has 10 different classes) or the conditional probability distributions (i.e. $P(Y|X)$) are different (e.g. the source and target reviews are very unbalanced in terms of the user-defined classes).

In addition, if there exists a relationship (explicit or implicit) between the feature spaces of the two domains, it is said that the source and target domains are *related* [72].

5.4 Categorization of transfer learning

As stated by Pan *et al.* [72], there are three major research questions to ask in transfer learning: (1) What to transfer; (2) How to transfer; (3) When to transfer.

"What to transfer" asks, what part of the knowledge can be transferred across the domains and tasks. Part of the knowledge can be specific only to an individual task or domain and some knowledge may be common between domains and can thus be helpful in increasing the performance for the target domain or task. After discovering the relevant information, algorithms need to be developed, which addresses the **"how to transfer"** question.

However, not all knowledge is relevant and when transferred it can be even harmful and decrease the performance of a model. This is called *negative transfer* and the questions **"when to transfer"** revolves around this issue.

Using the unified definition of transfer learning given above, transfer learning can further be divided into three sub-settings: (1) *inductive transfer*

Learning Settings	Source and Target Domains	Source and Target Tasks
Traditional machine learning	the same	the same
Inductive transfer learning	the same	different but related
Transductive transfer learning	different but related	the same
Unsupervised transfer learning	different but related	different but related

Table 5.1: Relationship between machine learning and various transfer learning settings. *Adopted from Pan et al. [72].*

Setting	Related Area	Source Domain Labels	Target Domain Labels	Task
Inductive transfer learning	Multi-task learning	Available	Available	Regression, classification
	Self-taught learning	Unavailable	Available	Regression, classification
Transductive transfer learning	Domain adaptation, sample selection bias, co-variate shift	Available	Unavailable	Regression, Classification
Unsupervised transfer learning		Unvailable	Unavailable	Clustering, dimensionality reduction

Table 5.2: Different settings of transfer learning. *Adopted from Pan et al. [72].*

learning, (2) *transductive transfer learning* and *unsupervised transfer learning*, depending on how the source and target domains and tasks are related.

See table 5.1 to see the relationship between traditional machine learning and transfer learning and table 5.2 for a quick summary of different settings of transfer learning.

5.4.1 Inductive transfer learning

Definition 2. Given a source domain D_S and a learning task T_S, a target domain D_T and a learning task T_T, *inductive* transfer learning aims to help to improve the learning of a target predictive function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S, where $T_S \neq T_T$ [72].

Based on the above definition, in this setting the target task is different from source task, but the domains are the same.

In this setting, a few labeled data in the target domain are required as a training data to *induce* the target predictive function $f_T(\cdot)$. Depending on the amount of available labeled data in the target domain, we can further divide this setting into two sub-settings.

One, where there is a lot of labeled data available in the source domain.

This is similar to *multi task learning*, but transfer learning only aims at achieving better performance in the target task - not learning source and target tasks at the same time.

And another, where there is no labeled data available in the source domain. This is similar to *self-taught learning* setting proposed by Raina *et al.* in [83].

5.4.2 Transductive transfer learning

Definition 3. Given a source domain D_S and a corresponding learning task T_S, a target domain D_T and a corresponding learning task T_T, *transductive* transfer learning aims to improve the learning of a target predictive function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S, where $D_S \neq D_T$ and $T_S = T_T$. In addition, some unlabeled data in target domain must be available [72].

Based on the above definition, in this setting the source and target tasks are the same, while the source and target domains are different.

In this situation, no labeled data are available in the target domain while a lot of of data are available in the source domain. Depending on the situations between the domains, we can further divide this setting into two sub-settings.

One, where the feature spaces between the target and source domain are different, $X_S \neq X_T$.

And another, where the feature spaces are the same, but the marginal probability distributions are different, $P(X_S) \neq P(X_T)$. This sub-setting is similar to domain adaptation, sample selection bias and co-variate shift [72].

5.4.3 Unsupervised transfer learning

Definition 4. Given a source domain D_S with a learning task T_S, a target domain D_T and a corresponding learning task T_T, *unsupervised* transfer learning aims to help to improve the learning of the target predictive function $f_T(\cdot)$[2] in D_T using the knowledge in D_S and T_S, where $T_S \neq T_T$ and Y_S and Y_T are not observable [72].

In this situation, no labeled data are available in the target domain nor the source domain.

[2]In unsupervised transfer learning, the predicted labels are latent variable, e.g. clusters or reduced dimensions.

	Inductive transfer learning	Transductive transfer learning	Unsupervised transfer learning
Instance-transfer	✓	✓	
Feature-representation-transfer	✓	✓	✓
Parameter-transfer	✓		
Relational-knowledge-transfer	✓		

Table 5.3: Different approaches used in different settings. *Adopted from Pan et al. [72].*

5.5 Approaches to transfer learning

Using these three sub-settings, we can now answer our question "what to transfer" and summarize them under four cases [72].

The first case is *instance-based transfer learning* or *instance-transfer* approach, which assumes the parts of the data in the source domain can be reused in the target domain. For example by *re-weighting* some labeled data in the source domain.

The second case is *feature-representation-transfer* approach, where the idea is to learn a "good" feature representation, which reduces the difference between target and source domains and the error of classification and regression models.

The third case is *parameter-transfer* approach, where the idea is to discover shared parameters or priors between the source domain and target domain models.

The fourth case is *relational-knowledge-transfer* approach, where the assumption is that some relationship in the source and target domain data are similar and this relationship is to be transferred.

In research, the three different sub-settings of transfer learning have been used in differing manner for these four different approaches. A quick summarization can be found in table 5.3. For a more extensive overview, readers are referred to the outstanding survey [72] by Pan and Jang.

Chapter 6

Transfer learning in emotion recognition

Combining these two hallmarks of human intelligence - emotion recognition and transfer learning - for emotionally intelligent technology and a better understanding of our own human condition is a lucrative opportunity, but to date the field still remains quite unresearched.

As of today, May 5, 2017, Google Scholar yields 461 hits for the search terms "'emotion recognition' AND 'transfer learning'". The papers are relatively new, mostly published in the last three years and the most researched and cited fields are acoustic phenomena and facial expression recognition from video.

From these papers it can be seen that transfer learning has been researched at least for predicting the emotional content of generalized sound events [68] and normal speech [75][24][22][46] as well as whispered speech [21][23], for spontaneous facial expression recognition [96], for sentiment analysis of reviews[8] and social media images [51], as well as for emotion recognition from video[57][100]. One study also included the fusion of facial expression recognition and audio emotion recognition subsystems [30]. Another one was focused on the relationship of sentiment analysis and human psychology [53]. Idiosyncrasy in face and body expressions was also researched [87].

6.1 Research so far

In this section, some of the newest research is examined in detail. However, the interested reader is suggested to read newest journals and check Google Scholar, as new research is being added daily. Some of the research papers used in this thesis were published during it's writing. A great number

of the research papers in this field are also submitted to various Emotion Recognition contests.

6.1.1 Generalized sound events

The research lead by Ntalampiras *et al* [68] achieved much greater accurary in automatically recognizing the emotions evoked by generalized sound events by transferring perceptual similarities between music and sounds.

The research used k-medoids algorithm for regression and dimensional emotion space with valence and arousal. Echo state networks (ESN) were used to transfer knowledge from music feature space to the generalized sound feature space.

The error rates presented surpass the so far best published results on the dataset used (IADS-2). The team thus encouraged further research in exploiting transfer learning for construnting a shared emotional space for improved prediction of valence and arousal.

A similar study was done by Coutinho *et al.* [15] in 2014, where they tried to transwer knowledge between music and speech. The results indicated a good cross-domain generalization performance.

6.1.2 Whispered speech

The study by Deng *et al* [21] was motivated by the fact that automated emotion recognition in speech focuses solely on normal speech and research for whispered speech is non-existent and and applications even impossible because of the fundamental differences between these two stimuli. The study was further encouraged by success in feature transfer learning.

They proposed methods based on denoising autoencoders, shared-hidden-layer autoencoders, and extreme learning machines autoencoders. The findings suggest that these feature transfer learning methods can significantly enhance the prediction accuracy on a range of emotion tasks (i.e. whispered speech) without reducing performance on source task (i.e. normal speech).

They further found that autoencoder-based feature transfer learning not only can aim to alleviate the mismatch between the training set and test set by discovering common features, but also can greatly improve the learning performance of a target task by transferring useful information in one source task to the target task in an unsupervised way.

6.1.3 Video emotion recognition

The study by Xu *et al.*[100] provides the study of knowledge transfer for both supervised and zero-shot emotion (i.e. emotions in the test set are completely unseen during training time) recognition. The study was motivated by the inability of basic discrete emotion theories to capture the full range of spontaneous human emotions.

They proposed a novel Image Transfer Encoding (ITE) process to encode and generate video representation. They also investigated the effectiveness of features from different convolutional neural network architectures and layers in the task of video emotion recognition and knowledge transfer. They also explored the complementarity of deep features with the existing visual and audio hand-crafted features.

Their framework improved upon the previous state-of-the-art results by 7.7% absolute percentage points on YouTube dataset.

A similar study that was also sent to the Emotion Recognition in The Wild 2016 contest by Ding *et al.* [30]. They also used deep CNNs, but also included fusion of facial expression recognition and audio emotion recognition subsystems at score level. Their experiments showed that show that both subsystems individually and as a whole can achieve state-of-the art performance on the datasets.

6.1.4 On-line Emotion Detection

The paper by Kollias *et al.* [57] presents a new methodology for retraining of deep neural networks when detecting emotion in video, using a mechanism for drift detection and a retraining algorithm. The team researched on-line emotion recognition based on facial expression analysis.

They utilized Deep Convolutional Neural Networks (DCNNs), Deep Belief Networks (DBNs) and Recurrent Neural Networks (RNN). The emotional space onto which predictions were classified was two-dimensional with valence and arousal.

Forthcoming work from the team includes application of the methodology to real-life human computer interaction scenarios, especially in interactive applications, where user behavior analysis plays a very important role.

6.1.5 Visual Sentiment Analysis

Islam and Zhang [51] propose a novel visual sentiment analysis framework using transfer learning approach to predict sentiment. The study was motivated by the fact that people are uploading millions of images in social

networks such as Twitter, Facebook, Google Plus, and Flickr. These images play a crucial part in expressing emotions of users in online social networks and thus image sentiment analysis has become important in the area of online multimedia big data research.

In their research they used hyper-parameters learned from a very deep convolutional neural network to initialize their network model to prevent overfitting.

They conducted extensive experiments on a Twitter image dataset and proved that their model achieved better performance than the current state-of-the-art.

6.1.6 The Relationship between Computational Models and Psychological State

The interdisciplinary study by Jo *et al.* [53] enlightens the relationship between computational models and psychological measurements.

The team first examined psychological state of 64 participants and asked them to summarize the story of a book, Chronicle of a Death Foretold by Gabriel Garcia Márquez. The team then trained their models with movie review data and evaluated participants' summaries using the pretrained model as a concept of transfer learning.

They compared different deep neural network algorithms such as CNN, LSTM and GRU. CNN had the best accuracy, however, it's predictions did not reflect the true psychological state of the participants. Rather, GRU shows more explainable results depending on the psychological state.

They state quite rightly that recent sentiment analysis models should be able to explain not only whether the data is negative or positive but also whether the person is negative or positive. For example, a depressed person can have a tendency of saying positively. The team thus emphasizes the importance of defining the meaning of scores correctly and rigorous and critical analysis in interpreting the results.

Chapter 7

Conclusions

Although a very new field, using transfer learning for automated emotion recognition has proved so far a fruitful venture. However, the field is still highly unexplored, and although promising, it needs more research.

Analyzing emotions is very much human-dependent in the sense that different humans express their emotions through different ways. Transfer learning can provide the needed adaptability. The state-of-the-art in emotion detection is currently based on using pre-trained deep neural networks which are adapted to new environments, where only a small amount of training data is available, through transfer learning, without overfitting. Many new competitions in emotion recognition are also being held, which further develop the techniques.

New technology has also brought automated emotion recognition nearer and anyone with a smart-phone is theoretically carrying an affective computer with them. Future applications for emotionally intelligent applications are numerous and transfer learning will play a role in their code.

The problems arise from not only technological challenges, but also from scientific theories: emotional archetypes, be they dimensional or discrete, seem to be too simplified to be real representations of how emotions are realized in the brain. The future development of automated emotion recognition will go hand in hand with our understanding of human emotions, and thus advancements in cognitive neuroscience and psychology will also play a great role.

Some new research also points out to this problem with highly statistical models: they are predicting the emotional content of the data, not the actual emotional state of a person.

Chapter 8

Discussion

The old phallacy of "you only see the things you are looking for" might be one of the biggest concerns in automated emotion recognition. If recognizing emotions has to based on labeling emotions, can we see anything else other than the labeled emotions?

Transfer learning can of course alleviate this problem by not making any assumptions about any emotional theories, but only by founding out the statistically best representation and comparing it to other data.

But can a statistical representation ever give the philosophical insight of a theory? To me, this same dilemma can be seen all over in sciences. For example in psychology, statistical personality theories such as "The Big Five"[54] are without any argument better at predicting human behavior than the esoteric and categorizing theories of Freud[40] or Jung[56]. However, saying that someone is "37% introvert" lacks all the insight given by categorical theories. It simply does not answer the underlying *whys*.

Of course from engineering viewpoint the statistical representation is interesting. If it work, it works. No need to know the laws behind it. But from scientific viewpoint, the idea of statistically mimicking and representing data without any analysis or synthesis, is problematic.

There is a famous quote made by the linguistic Noam Chomsky against scientist who use purely statistical methods to produce behavior that mimics something in the world, but who don't try to understand the meaning of that behavior.

> It's true there's been a lot of work on trying to apply statistical models to various linguistic problems. I think there have been some successes, but a lot of failures. There is a notion of success ... which I think is novel in the history of science. It interprets success as approximating unanalyzed data.

This speech he gave sparked quite a debate, and the interested readers should read more about it in the counterpart Peter Norvig wrote at `http://norvig.com/chomsky.html`.

I agree with Chomsky, but as Norvig puts it, I also do think, that engineering success usually implies at least a right direction - and transfer learning has engineering success. But then of course if we choose to only focus on statistical mimicking, how do know if our predictions are correct? Is our prediction concurrent with emotional state of the person? With labels maybe, but then we are back at the original problem.

The field of both automated affect recognition and transfer learning are very much in their infancy, but the results are promising. I think the biggest argument for transfer learning in emotion recognition comes from the fact, that sensory information seems to fuse in the brain on *feature level* and *not* on decision level. Traditional machine learning based automated emotion recognition fuses information on decision/score level, but rightly done transfer learning could provide feature based fusion.

But then the biggest challenge will be the problem of *negative transfer.* Human behaviour and humans in gerenal are full of idiosyncrasies. If we are trying to find common patterns and combining modalities in emotions, then we should find out which patterns are correlated but NOT related. After all, it's not that humans have one emotion active all the time, but there could be many complex emotional states at any given time.

For future research, I'd suggest that emotion recognition and transfer learning should not only be considered as a machine learning problem, but as a interdisciplinary problem. The research teams should include not only computer scientists, but also psychologists, neuroscientists, and even artists and teachers. By combining expertise from many fields, the future research could at the same time do statistical mimicking and analysis while also qualitatively analyzing and testing the findings in the wild for more robust theories.

Bibliography

[1] Alpaydin, E. *Introduction to machine learning*. MIT press, 2014.

[2] Aviezer, H., Trope, Y., and Todorov, A. Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science 338*, 6111 (2012), 1225–1229.

[3] Barrett, L. F. Discrete emotions or dimensions? the role of valence focus and arousal focus. *Cognition & Emotion 12*, 4 (1998), 579–599.

[4] Barrett, L. F. Are emotions natural kinds? *Perspectives on psychological science 1*, 1 (2006), 28–58.

[5] Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., and Movellan, J. Recognizing facial expression: machine learning and application to spontaneous behavior. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (2005), vol. 2, IEEE, pp. 568–573.

[6] Bechara, A., Damasio, H., and Damasio, A. R. Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex 10*, 3 (2000), 295–307.

[7] Bishop, C. M. Pattern recognition. *Machine Learning 128* (2006), 1–58.

[8] Blitzer, J., Dredze, M., Pereira, F., et al. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL* (2007), vol. 7, pp. 440–447.

[9] Brady, E., and Haapala, A. Melancholy as an aesthetic emotion. *Contemporary Aesthetics 1* (2003).

[10] Brown, A. L. Analogical learning and transfer: What develops. *Similarity and analogical reasoning* (1989), 369–412.

[11] Campbell, N., and Mokhtari, P. Voice quality: the 4th prosodic dimension. In *15th ICPhS* (2003), pp. 2417–2420.

[12] Castellano, G., Kessous, L., and Caridakis, G. Emotion recognition through multiple modalities: face, body gesture, speech. In *Affect and emotion in human-computer interaction*. Springer, 2008, pp. 92–103.

[13] Chou, C.-C., Tseng, S.-Y., Chua, E., Lee, Y.-C., Fang, W.-C., and Huang, H.-C. Advanced ecg processor with hrv analysis for real-time portable health monitoring. In *Consumer Electronics-Berlin (ICCE-Berlin), 2011 IEEE International Conference on* (2011), IEEE, pp. 172–175.

[14] Cohn, J. F. Foundations of human computing: facial expression and emotion. In *Proceedings of the 8th international conference on Multimodal interfaces* (2006), ACM, pp. 233–238.

[15] Coutinho, E., Deng, J., and Schuller, B. Transfer learning emotion manifestation across music and speech. In *Neural Networks (IJCNN), 2014 International Joint Conference on* (2014), IEEE, pp. 3592–3598.

[16] Cowie, R., Douglas-Cowie, E., and Cox, C. Beyond emotion archetypes: Databases for emotion modelling using neural networks. *Neural networks 18*, 4 (2005), 371–388.

[17] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine 18*, 1 (2001), 32–80.

[18] Darwin, C. *The Expression of the Emotions in Man and Animals*. John Murray, 1872.

[19] De Gelder, B. Why bodies? twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 364*, 1535 (2009), 3475–3484.

[20] Dellaert, F., Polzin, T., and Waibel, A. Recognizing emotion in speech. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (1996), vol. 3, IEEE, pp. 1970–1973.

[21] Deng, J., Fruhholz, S., Zhang, Z., and Schuller, B. Recognizing emotions from whispered speech based on acoustic feature transfer learning. *IEEE Access* (2017).

[22] Deng, J., Xia, R., Zhang, Z., Liu, Y., and Schuller, B. Introducing shared-hidden-layer autoencoders for transfer learning and their application in acoustic emotion recognition. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* (2014), IEEE, pp. 4818–4822.

[23] Deng, J., Zhang, Z., Eyben, F., and Schuller, B. Autoencoder-based unsupervised domain adaptation for speech emotion recognition. *IEEE Signal Processing Letters 21*, 9 (2014), 1068–1072.

[24] Deng, J., Zhang, Z., Marchi, E., and Schuller, B. Sparse autoencoder-based feature transfer learning for speech emotion recognition. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on* (2013), IEEE, pp. 511–516.

[25] Devillers, L., and Vasilescu, I. Reliability of lexical and prosodic cues in two real-life spoken dialog corpora. In *LREC* (2004).

[26] Devillers, L., Vasilescu, I., and Lamel, L. Annotation and detection of emotion in a task-oriented human-human dialog corpus. In *proceedings of ISLE Workshop* (2002).

[27] Devillers, L., Vidrascu, L., and Lamel, L. Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks 18*, 4 (2005), 407–422.

[28] Dhall, A., Goecke, R., Gedeon, T., and Sebe, N. Emotion recognition in the wild. *Journal on Multimodal User Interfaces 2*, 10 (2016), 95–97.

[29] Dhall, A., Ramana Murthy, O., Goecke, R., Joshi, J., and Gedeon, T. Video and image based emotion recognition challenges in the wild: Emotiw 2015. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (2015), ACM, pp. 423–426.

[30] Ding, W., Xu, M., Huang, D., Lin, W., Dong, M., Yu, X., and Li, H. Audio and face video emotion recognition in the wild using deep neural networks and small datasets. In *Proceedings of the*

18th ACM International Conference on Multimodal Interaction (2016), ACM, pp. 506–513.

[31] Dormann, C. Affective experiences in the home: measuring emotion. In *HOIT* (2003), vol. 3.

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

[33] Ekman, P. Facial expressions of emotion: New findings, new questions. *Psychological science 3*, 1 (1992), 34–38.

[34] Ekman, P. Facial expression and emotion. *American psychologist 48*, 4 (1993), 384.

[35] Ekman, P., and Friesen, W. V. Facial action coding system.

[36] Ekman, P., and Friesen, W. W. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology 17*, 2 (1971), 124–129.

[37] Ekman, P., and Oster, H. Facial expressions of emotion. *Annual review of psychology 30*, 1 (1979), 527–554.

[38] Ekman, P., and Rosenberg, E. L. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[39] Fontaine, J. R., Scherer, K. R., Roesch, E. B., and Ellsworth, P. C. The world of emotions is not two-dimensional. *Psychological science 18*, 12 (2007), 1050–1057.

[40] Freud, S., and Strachey, J. *The ego and the id*. WW Norton & Company, 1962.

[41] Frijda, N. H., et al. Varieties of affect: Emotions and episodes, moods, and sentiments.

[42] Ghosh, A., Danieli, M., and Riccardi, G. Annotation and prediction of stress and workload from physiological and inertial signals. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (2015), IEEE, pp. 1621–1624.

[43] Gilleade, K., Dix, A., and Allanson, J. Affective videogames and modes of affective gaming: assist me, challenge me, emote me. *DiGRA 2005: Changing Views–Worlds in Play.* (2005).

[44] Gross, J. J. The future's so bright, i gotta wear shades. *Emotion Review 2*, 3 (2010), 212–216.

[45] Hamann, S. Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in cognitive sciences 16*, 9 (2012), 458–466.

[46] Hassan, A., Damper, R., and Niranjan, M. On acoustic emotion recognition: compensating for covariate shift. *IEEE Transactions on Audio, Speech, and Language Processing 21*, 7 (2013), 1458–1468.

[47] Healey, J. A., and Picard, R. W. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems 6*, 2 (2005), 156–166.

[48] Hebb, D. O. Emotion in man and animal: an analysis of the intuitive processes of recognition. *Psychological review 53*, 2 (1946), 88.

[49] Hihn, H., Meudt, S., and Schwenker, F. Inferring mental overload based on postural behavior and gestures. In *Proceedings of the 2nd workshop on Emotion Representations and Modelling for Companion Systems* (2016), ACM, p. 3.

[50] Hoque, M., and Picard, R. W. Acted vs. natural frustration and delight: Many people smile in natural frustration. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* (2011), IEEE, pp. 354–359.

[51] Islam, J., and Zhang, Y. Visual sentiment analysis for social images using transfer learning approach. In *Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom), 2016 IEEE International Conferences on* (2016), IEEE, pp. 124–130.

[52] James, W. What is an emotion? *Mind 9*, 34 (1884), 188–205.

[53] Jo, H., Kim, S.-M., and Ryu, J. What we really want to find by sentiment analysis: The relationship between computational models and psychological state. *arXiv preprint arXiv:1704.03407* (2017).

[54] John, O. P., and Srivastava, S. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research 2*, 1999 (1999), 102–138.

[55] Joshi, J., Goecke, R., Parker, G., and Breakspear, M. Can body expressions contribute to automatic depression analysis? In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on* (2013), IEEE, pp. 1–7.

[56] Jung, C. G. *Psychological types*. Routledge, 2014.

[57] Kollias, D., Tagaris, A., and Stafylopatis, A. On line emotion detection using retrainable deep neural networks. In *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on* (2016), IEEE, pp. 1–8.

[58] Kragel, P. A., Knodt, A. R., Hariri, A. R., and LaBar, K. S. Decoding spontaneous emotional states in the human brain. *PLoS Biol 14*, 9 (2016), e2000106.

[59] Landowska, A., Szwoch, M., Szwoch, W., Wróbel, M., and Kołakowska, A. Emotion recognition and its applications.

[60] Lee, C. M., Narayanan, S., and Pieraccini, R. Recognition of negative emotions from the speech signal. In *Automatic Speech Recognition and Understanding, 2001. ASRU'01. IEEE Workshop on* (2001), IEEE, pp. 240–243.

[61] Lee, C. M., Narayanan, S. S., and Pieraccini, R. Combining acoustic and language information for emotion recognition. In *INTERSPEECH* (2002).

[62] Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S. Emotion and decision making. *Annual Review of Psychology 66* (2015), 799–823.

[63] Levenson, R. W. The intrapersonal functions of emotion. *Cognition & Emotion 13*, 5 (1999), 481–504.

[64] Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., and Barrett, L. F. The brain basis of emotion: a meta-analytic review. *Behavioral and brain sciences 35*, 03 (2012), 121–143.

[65] Littlewort, G. C., Bartlett, M. S., and Lee, K. Faces of pain: automated measurement of spontaneousallfacial expressions of genuine and posed pain. In *Proceedings of the 9th international conference on Multimodal interfaces* (2007), ACM, pp. 15–21.

[66] Mannila, H. Data mining: machine learning, statistics, and databases. In *Scientific and Statistical Database Systems, 1996. Proceedings., Eighth International Conference on* (1996), IEEE, pp. 2–9.

[67] Nacke, L. E., and Mandryk, R. L. Designing affective games with physiological input. In *Workshop on Multiuser and Social Biosignal Adaptive Games and Playful Applications in Fun and Games Conference (BioS-Play)* (2010).

[68] Ntalampiras, S. A transfer learning framework for predicting the emotional content of generalized sound events. *The Journal of the Acoustical Society of America 141*, 3 (2017), 1694–1701.

[69] Nummenmaa, L., Glerean, E., Hari, R., and Hietanen, J. K. Bodily maps of emotions. *Proceedings of the National Academy of Sciences 111*, 2 (2014), 646–651.

[70] Oatley, K., and Jenkins, J. M. Human emotions: Function and dysfunction. *Annual review of psychology 43*, 1 (1992), 55–85.

[71] Onorati, F., Regalia, G., Caborni, C., and Picard, R. Improvement of a convulsive seizure detector relying on accelerometer and electrodermal activity collected continuously by a wristband. In *Epilepsy Pipeline Conference* (2016).

[72] Pan, S. J., and Yang, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering 22*, 10 (2010), 1345–1359.

[73] Pantic, M., Pentland, A., Nijholt, A., and Huang, T. S. Human computing and machine understanding of human behavior: A survey. In *Artifical Intelligence for Human Computing*. Springer, 2007, pp. 47–71.

[74] Pantic, M., and Rothkrantz, L. J. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE 91*, 9 (2003), 1370–1390.

[75] Peng, S., Yun, J., Li, Z., and Minghai, X. Speech emotion recognition using transfer learning. *IEICE TRANSACTIONS on Information and Systems 97*, 9 (2014), 2530–2532.

[76] Pentland, A. Socially aware, computation and communication. *Computer 38*, 3 (2005), 33–40.

[77] Picard, R. W., Fedor, S., and Ayzenberg, Y. Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review 8*, 1 (2016), 62–75.

[78] Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., and Strohecker, C. Affective learning—a manifesto. *BT technology journal 22*, 4 (2004), 253–269.

[79] Picard, R. W., and Picard, R. *Affective computing*, vol. 252. MIT press Cambridge, 1997.

[80] Plutchik, R. A general psychoevolutionary theory of emotion. *Theories of emotion 1*, 3-31 (1980), 4.

[81] Plutchik, R. The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist 89*, 4 (2001), 344–350.

[82] Plutchik, R., and Kellerman, H. *Emotion: theory, research and experience*, vol. 3. Academic press New York, 1986.

[83] Raina, R., Battle, A., Lee, H., Packer, B., and Ng, A. Y. Self-taught learning: transfer learning from unlabeled data. In *Proceedings of the 24th international conference on Machine learning* (2007), ACM, pp. 759–766.

[84] Rainville, P., Bechara, A., Naqvi, N., and Damasio, A. R. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International journal of psychophysiology 61*, 1 (2006), 5–18.

[85] Ringeval, F., Schuller, B., Valstar, M., Jaiswal, S., Marchi, E., Lalanne, D., Cowie, R., and Pantic, M. Av+ ec 2015: The first affect recognition challenge bridging across audio, video, and physiological data. In *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge* (2015), ACM, pp. 3–8.

[86] Roisman, G. I., Tsai, J. L., and Chiang, K.-H. S. The emotional integration of childhood experience: physiological, facial expressive, and self-reported emotional response during the adult attachment interview. *Developmental psychology 40*, 5 (2004), 776.

[87] Romera-Paredes, B., Aung, M. S., Pontil, M., Bianchi-Berthouze, N., Williams, A. C. d. C., and Watson, P. Transfer learning to account for idiosyncrasy in face and body expressions. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on* (2013), IEEE, pp. 1–6.

[88] Russell, J. A., Bachorowski, J.-A., and Fernández-Dols, J.-M. Facial and vocal expressions of emotion. *Annual review of psychology 54*, 1 (2003), 329–349.

[89] Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of research and development 3*, 3 (1959), 210–229.

[90] Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. The neural basis of economic decision-making in the ultimatum game. *Science 300*, 5626 (2003), 1755–1758.

[91] Sano, A., and Picard, R. W. Stress recognition using wearable sensors and mobile phones. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on* (2013), IEEE, pp. 671–676.

[92] Scanlon, V. C., and Sanders, T. *Essentials of anatomy and physiology*. FA Davis, 2014.

[93] Schapire, R. Cos 511: Theoretical machine learning. *FTP: http://www. cs. princeton. edu/courses/archive/spr08/cos511/scribe notes/0204. pdf* (2008).

[94] Siemer, M. Mood-congruent cognitions constitute mood experience. *Emotion 5*, 3 (2005), 296.

[95] Valstar, M. F., Almaev, T., Girard, J. M., McKeown, G., Mehu, M., Yin, L., Pantic, M., and Cohn, J. F. Fera 2015-second facial expression recognition and analysis challenge. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on* (2015), vol. 6, IEEE, pp. 1–8.

[96] Wan, S., and Aggarwal, J. Spontaneous facial expression recognition: A robust metric learning approach. *Pattern Recognition 47*, 5 (2014), 1859–1868.

[97] Ward, N. G., Doerr, H. O., and Storrie, M. C. Skin conductance: A potentially sensitive test for depression. *Psychiatry Research 10*, 4 (1983), 295–302.

[98] Warriner, A. B., Kuperman, V., and Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods 45*, 4 (2013), 1191–1207.

[99] Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. Eulerian video magnification for revealing subtle changes in the world.

[100] Xu, B., Fu, Y., Jiang, Y.-G., Li, B., and Sigal, L. Video emotion recognition with transferred deep feature encodings. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* (2016), ACM, pp. 15–22.

[101] Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence 31*, 1 (2009), 39–58.

Appendix A

First appendix

A.1 Suomenkielinen tiivistelmä johdannosta

Tieteen filosofian suurimpia kysymyksiä on se, mitä voidaan tietää. Ja erityisesti luonnontieteissä siihen liittyy se, mitä voidaan mitata. Havaintoihin halutaan sovittaa matemaattisia malleja, eikä niinkään ilmiöiden laadullista kuvailua.

Kenties tästä samasta syystä luonnontieteet ovat pitkään karttaneet myös tunteita ja niiden tutkimista. Niitä on pidetty epätieteellisinä ja epävarmoina. Tunteet ovat kuitenkin aina olleet merkittävä osa ihmisyyden kokemusta. Tuhansia kirjoja ja lukemattomia runoja on kirjoitettu niistä ja niiden innoittamana, mutta luonnontieteiden kiinnostuksen kohteena ne eivät ole olleet kuin vastikään.

Evoluutioteorian isä, Charles Darwin, oli todennäköisesti ensimmäinen ihminen, joka tutki tunteita systemaattisesti. Kirjassaan *The Expression of Emotions in Man and Animals* hän halusi näyttää, kuinka ihmiset viestivät kehonliikkeillään tunnetilojaan ja kuinka nämä ovat perinnöllisesti määrättyjä ja peräisin vastaavista eläinten tavoista toimia. Vaikka ykypäivän tunneteoriat ovat huomattavasti hienostuneempia, on tämä teesi myös oman kanditaatin tutkielmani lähtökohta: mitä tunteista voidaan mitata ja sitä kautta myös tietää? Erityisen mielenkiinnon kohteena on kuinka tunnetilojen mittaamisen apuna voidaan käyttää uusia koneoppimisen metodeja, kuten siirto-oppimista.

A.1.1 Tunteiden mittaaminen ja niiden ymmärtämisen hyödyt

Kuten todettu, länsimainen tiede on pitkään väistellyt tunteita ja niiden tutkimista, mutta Rosalind Picard oli tiedeyhteisön ensimmäisiä tutkijoita, joka haastoi tämän näkemyksen. Kirjassaan *Affective Computing* hän julisti tunteiden mittaamisen ja ymmärtämisen merkityksestä sekä tärkeydestä tulevaisuuden tietotekniikassa. Tunnetilojan mittaaminen ja ymmärtäminen on on ongelma, jota neurotiede on vältellyt pitkään. välttämätöntä, jos haluamme ymmärtää ja tutkia ihmisyyttä tieteellisesti. Teknologian sovellukset ovat lukemattomat: mahdollisuuksiin lukeutuvat niin käyttäjätutkimus, henkilökohtaiset palvelukodit, kuin nykyajan vitsausten autismin ja masennuksen hoitaminen.

Tunteiden tunnistaminen on meille ihmisille luontaista, mutta kvantitatiiviset ja laskennalliset mittausmenetelmät ovat tuoreita. 1900-luvun alussa psykologit pystyivät lähinnä tarkkailemaan potilaitaan tai kuuntelemaan heidän omia kertomuksiaan. Ensimmäiset kvantitatiiviset menetelmät tunteiden tutkmiseen kehittänyt Paul Ekmankin joutui käyttämään ihmisiä, jotka pitkän koulutuksen jälkeen onnistuivat tunnistamaan ihmisten tunteita pienimmistäkin merkeistä. Näihin verrattuna on nykyisillä kognition tutkijoilla käytössään valtava määrä erilaisia tarkkoja mittausmenetelmiä.

A.1.2 Koneoppimisen hyödyntäminen

Lisääntyneen laskentakapasiteetin takia, yksi suurimmista harppauksista tieteessä viimeaikoina on tapahtunut nimenomaan koneoppimisen saralla. Puheentunnistus ja itseohjautuvat autot ovat pian arkipäivää.

Samaa oppivaa tilastollista mallintamista on hyödynnetty jo onnistuneesti tunteiden tunnistamisessakin mm. kuvaamalla ihmisten ilmeitä tai nauhoittamalla heidän puhettaan ja syöttämällä tästä saatu tieto koneoppimisalgoritmeihin. Ongelma on kuitenkin tunteiden monimuotoisuudessa.

Koneoppimisessa suuri oletus on, että algoritmin kouluttaminen ja ennustusten tekeminen tapahtuvat samassa avaruudessa ja samoilla todennäköisyysjakaumilla. Toisinsanoen omenoiden tunnistamiseen koulutettua algoritmiä ei voi hyödyntää päärynöiden tunnistamiseen - se tunnistaa vain omenoita. Jokaisen mahdollisen tunnetilan mallintaminen ja tallentaminen veisi kuitenkin ikuisuuden. Eikä niiden kaikkia mahdollisia kombinaatioita voitaisi samalla aikaa verrata toisiinsa.

Tähän ongelmaan ratkaisu saattaisi olla *siirto-oppimisessa.* Siirto-oppiminen on ihmisille hyvin luontaista ja se liittyy myös vahvasti *oppimaan oppimiseen.* Yksinkertaisuudessaan siirto-oppiminen tarkoittaa vanhan opetun hyödyntämistä uudessa

tilanteessa.

Tunteiden tunnistamista voidaan koneoppimisen parissa tutkia *dynaamisena kuvion tunnistusongelmana*. Nykyiset algoritmit onnistuvat tunteiden tunnistamisessa jo toisinaan paremmin kuin ihmiset, kun kysessä on eristettyä dataa (kuten kasvojen ilmeet). Kun data käsittelee isompia joukkoja, joissa dataa on saatavilla esimerkiksi kasvoista ja kehon asennosta, onnistuvat ihmiset paremmin tunnistamisessa. Tähän ongelmaan siirto-oppiminen voisi tuoda ratkaisun.