Institut für Maschinelle Sprachverarbeitung
University of Stuttgart
Pfaffenwaldring 5b
D-70569 Stuttgart

Bachelorarbeit

Emotion Classification Based on the Emotion Component Model

Amelie Heindl

Course of Study:	Informatik
Examiner:	Dr. Roman Klinger Prof. Dr. Sebastian Padó
Supervisor:	Dr. Roman Klinger
Commenced:	May 8, 2020
Completed:	October 26, 2020

Abstract

The term emotion is, despite its frequent use, still mysterious to researchers. This poses difficulties on the task of automatic emotion detection in text. At the same time, applications for emotion classifiers increase steadily in today's digital society where humans are constantly interacting with machines. Hence, the need for improvement of current state-of-the-art emotion classifiers arises. The Swiss psychologist Klaus Scherer published an emotion model according to which an emotion is composed of changes in the five components cognitive appraisal, physiological symptoms, action tendencies, motor expressions, and subjective feelings. This model, which he calls Component Process Model (CPM) gained reputation in psychology and philosophy, but has so far not been used for Natural Language Processing (NLP) tasks. With this work, we investigate, whether it is possible to automatically detect the CPM components in social media posts and, whether information on those components can aid the detection of emotions. We create a text corpus consisting of 2100 Twitter posts, that has every instance labeled with exactly one emotion and a binary label for each CPM component. With a Maximum Entropy classifier we manage to detect CPM components with an average F1-score of 0.56 and average accuracy of 0.82 on this corpus. Furthermore, we compare baseline versions of one Maximum Entropy and one Convolutional Neural Network (CNN) emotion classifier to extensions of those classifiers with the CPM annotations and predictions as additional features. We find slight performance increases of up to 0.03 for the F1-score for emotion detection upon incorporation of CPM information.

Contents

List of Figures

List of Tables

Acronyms

A Accuracy. 22

CNN Convolutional Neural Network. 3

CPM Component Process Model. 3

FN False Negatives. 22

FP False Positives. 22

IMS Institut für Maschinelle Sprachverarbeitung. 26

NLP Natural Language Processing. 3

P Precision. 22

POS Part-Of-Speech. 30

R Recall. 22

SVM Support Vector Machine. 23

TEC Twitter Emotion Corpus. 5

TFIDF Term Frequency-Inverse Document Frequency. 28

TN True Negatives. 22

TP True Positives. 22

1 Introduction and Problem Statement

Most people have an intuitive understanding of their own and other's emotions. But, as they are based on intuition instead of deterministic rules, these understandings can also differ greatly, inducing difficulties in finding an appropriate official definition of the term emotion and mark the boundaries to related terms like feeling, mood, and affect. For such a definition one needs to identify the way emotions are evoked, perceived, and displayed, and which 'states' are qualified to be classified as an emotion. Addressing these problems, a wide range of researchers has tried to establish a universally applicable emotion theory. There exist, for example, physiological theories like the James-Lange-theory which states that emotions occur as a result of bodily reactions to a stimulus [Can27], and cognitive theories like the Schachter-Singer-theory that argues emotions are triggered following the cognitive labeling of an experienced sensation [SS62]. Those theories are also subject to criticism concerning their insufficiency. In particular, they tend to oversimplify the development of emotions and do not take into account all factors that are necessary for an outright definition of emotions. Klaus Scherer published a more complex theory which he calls the Component Process Model of Emotion or CPM for short. In [Sch82] Scherer defines five components that contribute to emotions, namely cognitive appraisal, physiological symptoms, motor expression, motivational tendencies, and subjective feelings. The appraisal component consists of a process of several steps with further influence on the other components. These appraisal processes regard relevance, implications, the capability of coping with the stimulus, and significance [BS08]. The resulting set of changes in the aforementioned components is then defined to be an emotion according to Scherer [Sch05]. Humans have the ability to distinguish between the distinctive components of emotions and perceive them individually, for example, being capable to identify a change in their facial expression without the need to be aware of their feelings. This fact leads to the assumption, that people also express the experiences concerning the different components of emotions individually in speech and text. In some cases, humans might even deliberately choose to only hint at their emotion through one of the CPM components instead of openly saying how they feel. For example, someone, who is in a professional environment where the disclosure of personal feelings is scorned, could report a cognitive appraisal of a situation as a substitute for his feeling. An instantiation of this phenomenon would be to declare "This task was not scheduled for today.", which evaluates the unexpectedness of the event ¸'Somebody wants to do something that is not according to the plan." instead of announcing "I am surprised and angry", which explicitly communicates the emotional state.

With today's ever-growing amount of unstructured data in the World Wide Web and the need to find ways for extracting its information, research in the field of information retrieval is more important than ever. With the ability to derive information regarding personal feelings, opinions, or reactions it is possible to extend the knowledge about human behavior. This data can then be used by companies for building chatbots [AU17], detecting the feeling of their customers that send messages [GGF13] and similar use cases. Applications like these are one of the benefits that motivate emotion classification. The goal of emotion classification is to construct programs that determine which emotions a piece of text accommodates. The assignment of the correct emotion leads to the capability of interpreting the text in the correct way and comprehend its meaning and implications. In Chapter 2 it becomes clear that several researchers recognized the relevance of emotion extraction but despite great improvements in the past decades, current state-of-the-art emotion recognition mechanisms are still subject to misinterpretations [AM20]. While it is already possible to detect explicitly mentioned emotions, like in the sentence 'He was angry.' with great exactness, there are shortcomings in the detection of emotions in sentences where the emotion itself is not stated, but the situation gives some indications on what could be felt. If we look at sentences like 'He narrowed his lips and drew his eyebrows together.' or 'I did not expect that.', human readers would probably see some reference to the emotions of anger and surprise respectively (which are both components of the CPM), whereas computer programs would not, as they do not know the relations between facial expressions or appraisals and emotions. Thus, the hypothesis emerges that evidence of the individual components of Scherer's CPM can contribute to detecting currently missed utterances about emotions and thereby improve the recall of emotion detection mechanisms. This hypothesis is a motivating reason for building a classifier that discovers CPM components in text. With the output of such a classifier, it is possible to give additional information to an emotion classifier. This work focuses on exploring the applicability of the CPM to text classification and more precisely on answering the following two research questions.

Problem statement 1: Is it possible with an acceptable performance to automatically classify text into references to emotions according to the component process model? What are particular difficulties?

Problem statement 2: Can this emotion component process model be utilized to improve the classification of fundamental emotion categories?

Regarding the type of text for applying the developed methods on, social media posts are chosen as they are known to contain a great amount and variety of emotion-related utterances. The expected resulting improvement of the emotion classifier is based on the assumption that emotions are not equally distributed over the occurrences of the five components in text. In the following chapters we first present background knowledge and existing research papers on the inspected topics, then we define the used materials and methods, followed by a display of the results, and lastly, we discuss this paper's contribution and give an outlook to future work.

2 Background and Related Work

This chapter introduces the topics and ideas this paper is based on, starting with the definition of emotions and different emotion theories. Furthermore, the tasks of text and emotion classification as well as classification methods and performance metrics are explained. Lastly, the social media platform Twitter and the TEC are presented.

2.1 Emotions and Emotion Models

The term emotion is widely used, in scientific contexts as well as in everyday life, while its definition is still not entirely agreed upon. The question, what exactly identifies an emotion is largely being investigated for a long time but without definite consent [Köv12]. The Cambridge Dictionary defines emotion as "a strong feeling such as love or anger, or strong feelings in general" [McI13]. From a scientific viewpoint, this definition seems rather vague, hence psychologists and philosophers working in the area of emotion theory are required to refine this. Two of them, Mulligan and Scherer, propose a working definition of the term emotion in [MS12] to facilitate interdisciplinary discourse about this subject. They conceptualize emotion as a class and their definition of its instances can be expressed as

$$
\begin{aligned}
x \in \text{emotion} \Leftrightarrow & x \text{ is an affective episode} \\
& \wedge x \text{ has the property of intentionality} \\
& \wedge x \text{ contains bodily changes that are felt} \\
& \wedge x \text{ contains a perceptual or intellectual episode, } y\text{, which has the property} \\
& \text{of intentionality} \\
& \wedge \text{ the intentionality of } x \text{ is inherited from the intentionality of } y \\
& \wedge x \text{ is triggered by at least one appraisal} \\
& \wedge x \text{ is guided by at least one appraisal}\,.
\end{aligned}
$$

The authors go on to explain that emotions have a beginning and an end, are directed to an object and co-occur with evaluations of something's value.

Moreover, there is a need for boundaries between emotion and related terms that are sometimes falsely interchanged in language. One possible partition is to say that "Feelings are personal and biographical, emotions are social, and affects are prepersonal." [Sho05]. What is meant by that statement is that affects occur unconscious and determine the intensity of the sensation, feelings are the internal sensation that arises as a result of the meaning of something perceived, and expressions are the exposed part of the sensation. This is largely compatible with Scherer's views. He states in [Sch05] that feeling is a part of an emotion that is concerned with the evaluation of subjective emotional experiences. He further sees a difference between emotion and moods, because emotions are evoked by an internal or external event whereas moods are not. Moods are also of longer duration and lower intensity. Concerning attitudes, the main contrasts to emotion are lower intensity, lower impact on behavioral actions and less response to the eliciting stimulus. After defining the term emotion, there are still various possibilities how to divide this general term into actual instances. In the following we present some directions from which approaches to resolve this inadequacy exist.

2.1.1 Categorical and Dimensional Emotion Models

There are researchers who regard emotions as a bounded set of possible values and construct their emotion concepts out of values which are needed in order to cover every emotional state. These concepts can be further distinguished by whether they have a discrete or continuous value domain [PG17]. Concerning the former, the most famous categorization is by Paul Ekman, who declares {anger, disgust, fear, joy, sadness, surprise} as the finite set of basic emotions families. According to his work [Ekm92], there are nine characteristics that a basic emotion fulfills, that enables its identification. These characteristics are displayed in Table 2.1. Inspired by the ideas of Charles

Characteristics of basic emotions
1. Distinctive universal signals
2. Presence in other primates
3. Distinctive physiology
4. Distinctive universals in antecedent events
5. Coherence among emotional response
6. Quick onset
7. Brief duration
8. Automatic appraisal
9. Unbidden occurrence

Table 2.1: Characteristics for the identification and differentiation of basic emotions.

Darwin, Ekman set out to determine, whether expressions of emotions are dependent on culture and environment. He conducted studies in numerous cultures, of which some were even socially isolated, and his findings show, that the mentioned basic emotions are recognized and displayed in the same way by humans in all of these places. Figure 2.1 displays a reenactment of examples of the pictures Ekman used in these studies. The pictures show different emotional expressions and the people Ekman surveyed recognized the corresponding emotions. The misconduct of other researchers that the expression of emotions differs between cultures could be traced back to a masking of the felt emotions according to cultural rules by the subject. His model of the six universal emotions gained immense acceptance in experimental psychology as well as anthropology. [Ekm98] Yet, there are also researchers, who do not deem six emotion categories as sufficient.

Plutchik, for example, proposed a model that can be seen as a hybrid between a categorical and a dimensional model. According to his paper [Plu01], there are eight main emotions, that can be arranged in pair-wise opposites. He draws a parallel to the main colors, which is visible in Figure 2.2,

Figure 2.1: Depiction of emotional expressions oriented by the pictures that were used by Ekman to investigate the universality of emotions.

where the eight main emotions are organized in the center like a color-wheel with complementary colors on opposite sides. Carrying on with the color analogy, he states that other emotions are mixtures of two main emotions or lighter/less intense "shades" of a basic emotion.

A fully dimensional model is advocated by Russel and Mehrabian. Their emotion model is analogous to a 3-dimensional vector space, where the first axis is for pleasure-displeasure, the second for the degree of arousal and the third for dominance-submissiveness. With three dimensions, their idea is more fine-grained than most considerations about dimensional models before, that neglected the dominance dimension. Every point in this vector space can correspond to an emotion and for any given point in time, a person is at some point in the vector space [RM77]. Nonetheless, Ekman's model is used predominantly in emotion classification tasks [PG17] and due to the novelty of our approach, it seems reasonable to start with an uncomplicated model, thus we utilize Ekman's emotions categories in this work.

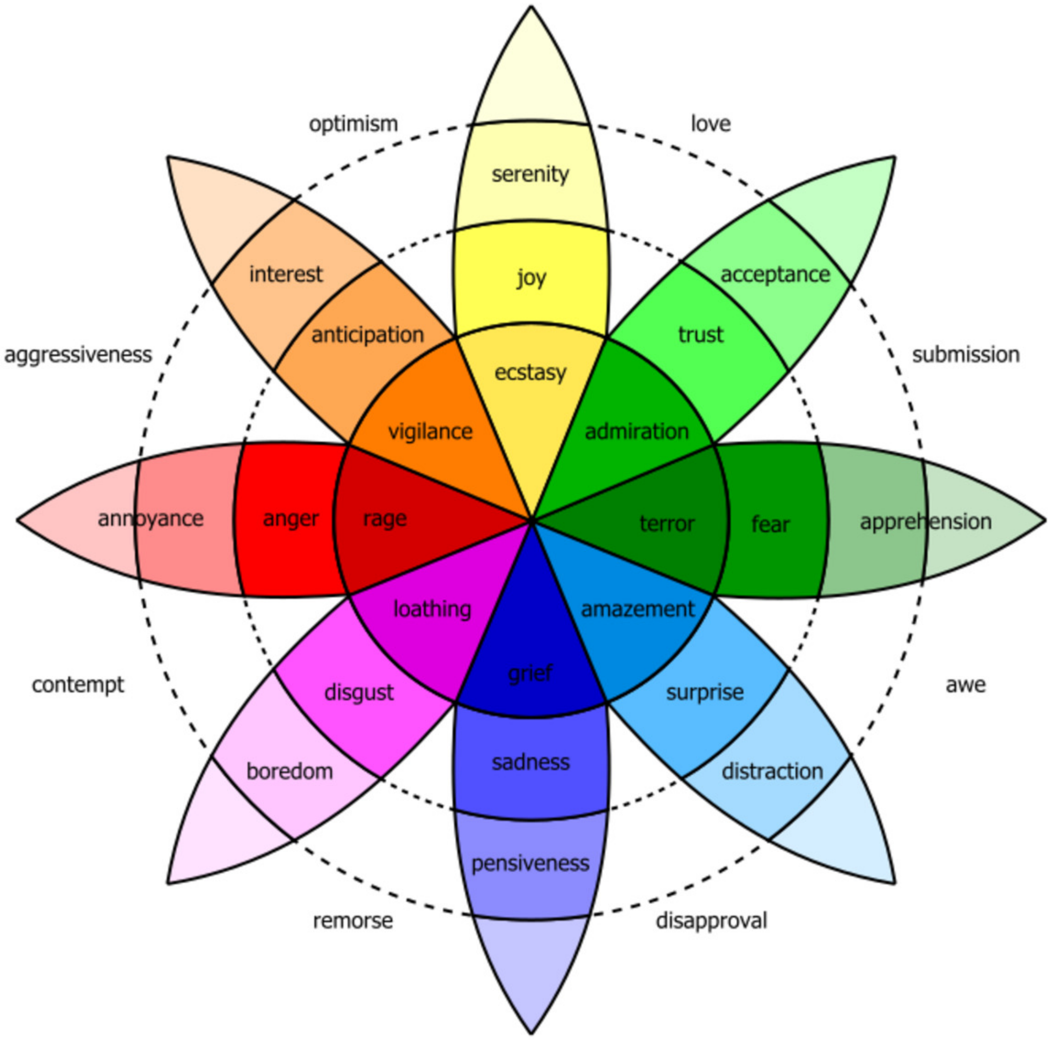

Figure 2.2: Emotion Model by Plutchik.[1]

2.1.2 Appraisal Models

There are also researchers, that approach the categorization of emotion from the way in which an emotion is developed through cognitive evaluation. These models divide the possible emotions by the kind of appraisal process that accompanies it.

The model of Smith and Ellsworth combines the dimensional representation with the appraisal approach. In [SE85], they introduce their idea of a model with several orthogonal dimensions, where each axis is one type of cognitive appraisal. The authors identify eight different types of appraisal, namely pleasantness, attention, control, certainty, perceived obstacle, legitimacy, responsibility, and

[1] https://en.wikipedia.org/wiki/Robert_Plutchik

anticipated effort. After investigating their model with experiments where participants describe their experiences of emotions, they find that obstacle and legitimacy do not qualify as distinct dimensions and thus they result with a six dimensional model. Despite the appraisal dimensions being adequate for categorizing emotions, they concede that emotions do not wholly consist of appraisal, but some additional components are needed for an emotion.

One of the most popular appraisal models is the one by Klaus Scherer, which is also the central focus of this work. Scherer states that emotions arise through a process of cognitive appraisals with impact on diverse parts of a person's senses and has published several works on this emotion model, which he calls the CPM. To further explain the CPM, we must first define its five components, which are cognitive appraisal, physiological symptoms, motor expression, motivational tendencies, and subjective feelings, in more detail. The following descriptions are derived from [Sch05] and [BS08]. The cognitive appraisal component is concerned with the evaluation of an experienced event. The event is assessed regarding first its relevance to the individual, secondly the implications and consequences it might lead to, thirdly the possible ways to cope with it and control it, and fourthly its significance according to personal values and social norms. The second component, physiological symptoms, is the component that regards automatically activated reactions and symptoms of the body, like changes in the heartbeat or breathing pattern. The motor expression component contains all movements, facial expressions, changes concerning the speech, and similar patterns. Actions like attention shifts and movement with respect to the position of the event, are part of the motivational tendencies component. The last component is about subjective feelings and takes into account, how strong, important, and persisting the felt sensations are. As the first component consists of several evaluation steps, the other components are influenced by it and might undergo several changes. With these explanations, he further argues, that it is possible to infer the emotion a person is experiencing by analyzing the set of changes in the five components [Sch05].

The OCC model, which is named after its creators Ortony, Clore, and Collins, falls into the category of appraisal theories, too, and pursues the idea that emotions represent situations. In [CO13] it is stated, that emotions arise as a result of an appraisal of an experienced situation and can be distinguished by whether they focus on events, actions, or objects. Those three categories can be further subdivided, by factors like the involved persons or the expectedness of the situation, resulting in 22 different types of emotions. The 22 emotion types, which include, for example, grief, pride, and love, can be regarded as equivalence classes containing a range of similar states that vary in intensity and symptoms.

2.2 Text and Emotion Classification

Text classification is the task of automatic assignment of one element out of a finite set of classes to a text instance. One of the most mentioned examples is spam filtering for e-mails. In this use-case, the classification program receives an e-mail as input and then decides to which of the two classes spam and no-spam it probably belongs. As programs can not know the classes of unseen instances, they have to work with probabilistic methods. One special case of text classification is emotion classification. In this case, the set of classes is composed of different emotion categories. There already exist several works on emotion retrieval in text. Here, we present some of those with a focus on works that use a similar text-domain as we do, and in the following subsections we will also mention works that use similar classification methods.

Yang, Lin, and Chen pursue the goal of classifying text by emotions with the help of machine learning in [YLC07]. They use a Support Vector Machine and a Conditional Random Field as two machine learning mechanisms and compare the results of both. Concerning a domain, they decided on web blogs, which is similar to social media posts in content and style. They conclude, that a context-aware classifier like the Conditional Random Field is the better choice at a sentence level and that choosing the emotion of the last sentence is the best choice at a document level. In [WW14] the emotions in social media microblogs are inspected. In this work, a lexicon-based and a Support-Vector-Machine approach are combined with class-sequential rules. Taking the order and relation of sentences into account instead of considering the blogs as a bag of words leads to better results. Their approach is able to improve the F-measure in comparison to all baseline methods. Go et al. [GBH09] analyze the expressed sentiment of social media posts from the platform Twitter. They do not consider a set of emotion categories, instead, they only differentiate between positive and negative moods, which is also called sentiment classification. Basing their work on [PLV02] who use the same approach for the text domain of movie reviews, they use Naive Bayes, Maximum Entropy, and Support Vector Machines, being able to achieve similar results with all of them. For testing their mechanisms, they used emoticons, which is a comparable approach as using hashtags like in our work.

2.2.1 Maximum Entropy

Maximum Entropy, which is also called MaxEnt or logistic regression is a statistical method to compute a distribution for an incomplete data set so that the data has maximal probability of occurrence [Ski88]. For discrete data, the entropy H is computed as

$$H(p_1, p_2, ..., p_n) = -K \sum_i p_i \log p_i \,,$$

with K being a positive constant and $(p_1, p_2, ..., p_n)$ being the probabilities for the values $(x_1, x_2, ..., x_n)$ a variable x can assume. With incomplete data, the probability distribution is to be chosen, that maximises this entropy, hence the name of this method [Jay57].

Nigam et al. investigate in [NLM99] the applicability of MaxEnt to text classification with promising results. The following explanations are based on their paper. In text classification the data points are text instances or documents that consist of a sequence of words and belong to one class out of a finite set of possible classes (multi-label classification tasks are neglected here). The task is then to find a probability distribution for all classes given a document. To aid this, the classifier can draw on knowledge from test documents for which the correct classes are known. This works by creating constraints on the distribution with the information of the labeled training data and then finding the most uniform distribution that satisfies all constraints, The constraints make use of features of the training data. A feature can generally be any function that depends on the document and/or the class, for instance *The word happy appears and the class is joy* could be a possible feature and can have the values 0 and 1. Let the features be denoted by $f_i(d, c)$ and let λ_i be the notation for real-valued factors. After we have defined features, the probability of a class c for a given unlabeled document d can then be computed by

$$P(c|d) = \frac{\exp \sum_i \lambda_i f_i(d,c)}{\sum_{c'} \exp \sum_i \lambda_i f_i(d,c')} \,.$$

This probability is computed for every class given the same document and then the class with the highest possibility is chosen for this document. Now it remains to define the factors λ_i. Those factors are found through an iterative scaling process, that starts by randomly initializing all λ_i and then finding the class distribution for the labeled documents with this model. Afterwards the λ_i are updated and the process is repeated until the maximum is reached. This maximum exists due to the convexity of the likelihood function l of a model Λ given the set of training documents D

$$l(\Lambda|D) = \log \prod_{d \in D} P_\lambda(c(d)|d) \, .$$

As [NLM99] proposes, MaxEnt is suited for text classification. For example, Zhu et al. confirm that with their work [ZJXG05]. They use MaxEnt for multi-labeled data from the Reuters-21578 corpus and achieve acceptable results. The MaxEnt method has also already been used for emotion classification, for instance in [WS12] that investigates suicide notes. The authors find a relation between performance and frequency of the emotion and reach an F1-value of 0.534.

2.2.2 CNN

CNNs are methods that obtain grid-like data as input and compute an output, with at least one of the operations they perform in the process being a convolution [GBC16]. Neural networks, in general, consist of several layers of nodes that are inspired by human neurons. Nodes on every layer are connected to some nodes of the next layer and every node represents a function that combines the inputs it gets from its connections to the former layer with weights and computes, usually with incorporation of an activation function, outputs which it passes on to the connections to the next layer. The input vector or matrix is passed to the first layer and the last layer produces the output vector or matrix. The exact structure of the network is generally not known and not of interest and it is learned by providing training data with the corresponding expected output to the network [Roj13]. In CNNs a convolution layer consists of a convolution stage, a detector stage and a pooling stage. Unsurprisingly, in the convolution stage, convolutions are applied on the input, afterwards in the detection stage, the results are combined with a non-linear activation function and in the pooling stage a function is applied that summarizes several neighboring values [GBC16].

As text instances are a list of words, they can be treated as a vector and can serve as input to such a CNN. For text classification, CNNs have been used by Kim [Kim14] in combination with word embeddings with remarkable results and in [HLKS17] where it is shown that CNNs outperform other methods in classification of medical texts. Hofmann et al. [HTSK20] use CNNs for appraisal prediction using appraisal categories proposed by Scherer and for emotion classification. They are able to improve emotion classification with appraisal annotations but not with appraisal predictions. The classifiers developed by them are used in this work.

2.2.3 Performance Metrics

In order to evaluate the performance of a text classification, fixed metrics are required. The metrics that are used for virtually all classification tasks depend on the amount of correctly and incorrectly classified instances. After the classification of a given set of texts for which the true classes are known, it is possible to determine for every text if the right class is predicted. For a binary classification, which means there is only one class, this determination has four possible cases. First, the text belongs to the class and the class is predicted. Second, the text belongs to the class, but the class is not predicted. Third, the text does not belong to the class and the class is not predicted. Fourth, the text does not belong to the class, but the class is predicted. Then, one can count the number of instances for all of these cases, which leads to the four numbers True Positives (TP) for the first case, False Negatives (FN) for the second case, True Negatives (TN) for the third case, and False Positives (FP) for the fourth case. In Table 2.2 the standard metrics Accuracy (A), Precision (P), Recall (R) and F1, that are also used in this work, are explained [MSR08].

Name	**Formula**	**Description**
Accuracy	$A = (TP + TN)/(TP + FP + TN + FN)$	Accuracy measures the percentage of instances for which the predicted class matches its real class out of all instances.
Precision	$P = TP/(TP + FP)$	Precision measures the percentage of instances that belong to the class out of all instances that were predicted to belong to the class.
Recall	$R = TP/(TP + FN)$	Recall measures the percentage of instances that were predicted to belong to the class out of all instances that really belong to the class.
F1	$F1 = (2 \cdot P \cdot R)/(P + R)$	F1 is the harmonic mean of precision and recall.

Table 2.2: Performance metrics for binary classification.

2.3 Twitter and TEC

Twitter is an online service with the main functionality of enabling its users to upload messages with arbitrary content, which are also called tweets, and subscribe other users and topics. As soon as something is uploaded, anybody whose subscriptions it matches, will see it automatically in their account's homepage and everybody else can access it by explicitly searching for it. Twitter is both used as a source for news as well as to communicate with acquaintances and strangers [Enc20]. There is an upper limit of 140 characters for all tweets, that traces back the beginnings of the platform when it used SMS protocol standards. Since the first tweet in 2006, the platform recorded immense growth and nowadays up to 60000 tweets per second are posted [Mac20]. One special feature of Twitter is the hashtag. A hashtag is any string beginning with the pound symbol (#) that can be included in tweets. These hashtags are mostly used to state the core ideas of a tweet or give additional information and also function as links that lead to an overview of all tweets with the same hashtag [Nat20]. Thus, the hashtags provide uncomplicated access to categorized text instances, supporting the utilization of tweets for NLP tasks.

In [Moh12] the validity of hashtags as classification labels is investigated. The resolving of this question starts with retrieving approximately 21000 tweets which end with one of the six basic emotions proposed by Ekman (see Section 2.1) as a hashtag. Examples of the tweets that Mohammad gathered are displayed in Table 2.3. Succeeding this, the emotion-hashtags are removed from the tweet and used as an annotation label and a binary Support Vector Machine (SVM) classifier with 10 fold-cross validation is applied to this corpus for each of the six emotions. Difficulties like the fact that these tweets are from about 19000 people are considered to possibly diminish the usefulness of Twitter for text classification, but the authors reach an F1-score of 49.9, which is commensurable to an inter-annotator agreement, calculated as Pearson's correlation, of 0.54, they measure for another emotion-labeled corpus. With these results, it can be concluded that hashtags are consistent and can function as adequate labels for emotion classification tasks. The corpus with the emotion-labels is published under the name TEC and is used in our work.

Emotional tweets
1. Feeling left out... #sadness
2. My amazing memory saves the day again! #joy
3. Some jerk stole my photo on tumblr. #anger
4. Mika used my photo on tumblr. #anger
5. School is very boring today :/ #joy
6. to me.... YOU are ur only #fear

Table 2.3: Example tweets with emotion-word hashtags.

3 Methodology

This chapter outlines the work process and the methods we apply in order to answer the problem statements. The work can be structured coarsely into the three parts corpus creation, CPM detection and emotion classification. Figure 3.1 demonstrates the workflow and combination of these parts, where circles represent tasks and rectangles represent artifacts. It can be seen, that the first step is to establish a text corpus that has annotations for the CPM components as well as for emotions. This step consists of the formulation of annotation guidelines and the actual annotation. The sequence goes on with implementing a classifier that detects CPM components in text. With the output of this classifier as an input, two different emotion classifiers are created. In the following, the methods utilized in the individual parts of the work sequence are described in more detail.

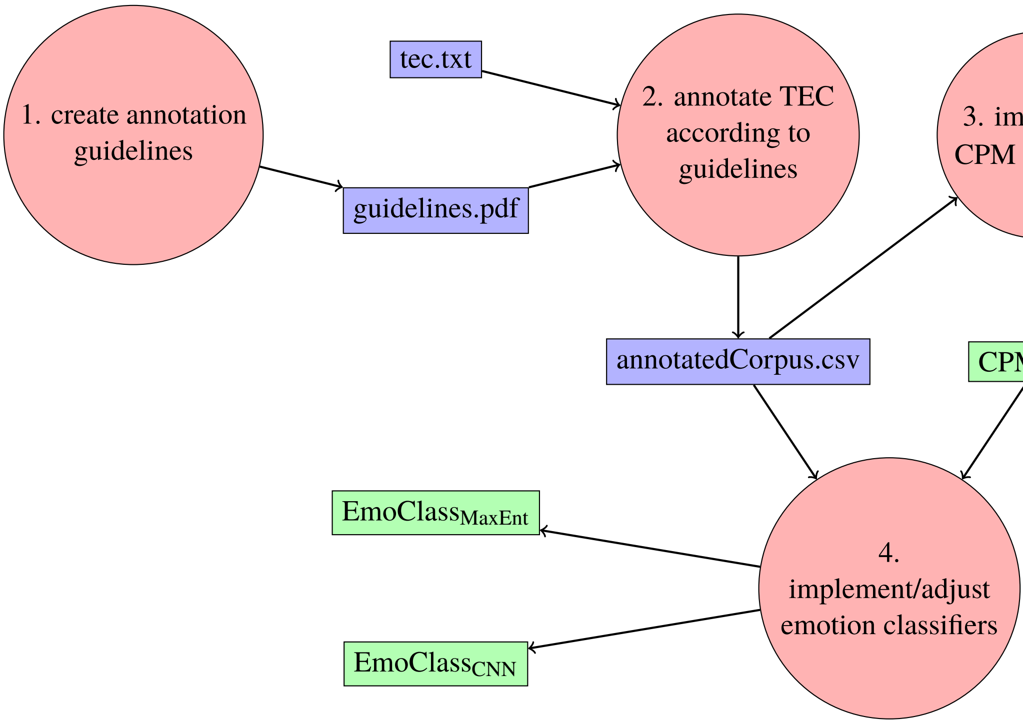

Figure 3.1: Workflow and depiction of the interaction of this work's steps.

3.1 Methods for Corpus Creation

With the term corpus creation, we summarize steps 1 and 2 of Figure 3.1. To answer the first problem statement posed in Chapter 1 and find out, whether a program can detect instances of the components of the CPM in text, a text database with annotations for the CPM is crucial. As a result of the novelty of this approach, no such database exists, and therefore, one is constructed for this work. We want to use the result of the classification according to the CPM for emotion detection afterwards, which motivates the decision to use TEC, that is described in Section 2.3, with its existing emotion annotations and extend it with CPM annotations. The fact that thiss annotation is conducted by a single person necessitates fine-grained guidelines. In order to assert the objectiveness of the annotation guidelines, these are established by two persons in an iterative process of annotating sample instances, computing the inter-annotator agreement, and refining the guidelines. The two persons are both students at the University of Stuttgart, one computational linguistics master student and one computer science bachelor student, that both work on their thesis about CPM at the Institut für Maschinelle Sprachverarbeitung (IMS). Due to the nature of the CPM, the annotation of an instance is defined by a set of five binary decisions. For every component of the CPM it has to be decided, to either annotate the instance with the value 1, meaning that a reference to this component is existent in this instance, or with the value 0 otherwise. Therefore, the guidelines consist of a list for every CPM component with rules that state in which cases to annotate the value 1 for this component. After the annotators agree on these lists, the annotation process goes on by randomly selecting 20 instances from the TEC and 20 instances from the Reman literary corpus [KK18]. Those instances are annotated by the two annotators independently. The final step in the iteration is then to discuss instances where the annotations do not match and refine the guidelines, so that the occurring disagreements are resolved. Afterwards, the steps are repeated with the new guidelines until the mismatches between the annotators are sufficiently scarce. The first version of guidelines is derived from a table that Scherer himself presented in [Sch05] to give examples for the components. Table 3.1 displays those examples.

Cognitive appraisal	**Physiological symptoms**	**Action tendencies**	**Motor expressions**	**Subjective feelings**
How suddenly and and abruptly did the event occur?	Moving attention towards the event	Feeling cold shivers (neck, chest)	Smiling	Intensity
How familiar was the person with the event?	Moving attention away from the event	Weak limbs	Mouth opening	Duration
How probable is the occurence of the event in general?	Information search	Getting pale	Mouth closing	Valence
How pleasant is the event in general, independently of the current situation?	Attention self-centered	Lump in throat	Mouth tensing	Arousal
How unpleasant is the event in general, independently of the current situation?	Attention directed towards others	Stomach troubles	Frown	Tension
How important/relevant is the event to th person's goals or needs?	Physically moving towards the event	Heart beat slowing down	Eyes closing	
How likely is it that the event was mostly caused by change or natural causes?	Physically moving away from the event	Heart beat getting faster	Eyes opening	
How likely is it that the event was mostly caused by the person's own behavior?		Muscles relaxing, restful (whole body)	Tears	
How likely is it that the event was mostly caused by someone else's behavior?		Muscles tensing, trembling (whole body)	Other changes in face	
If the event is caused by a behavior, how likely is it that it was caused intentionally?		Breathing slowing down	Voice volume increasing	
Are the potential consequences of the event clearly envisaged and may they occur in the near future?		Breathing getting faster	Voice volume decreasing	
		Feeling warm, pleasant	Voice trembling	
		Perspiring, moist hands	Voice being assertive	
		Sweating (whole body)	Other changes in voice	
		Feeling hot, puff of heat (cheeks, chest)	Abrupt bodily movements	

Table 3.1: Examples for instantiations of the CPM components.

To operationalize the inter-annotator agreement, we use Cohen's κ, which is defined in [Coh60] as

$$\kappa = \frac{p_0 - p_c}{1 - p_c} ,$$

with p_0 being the fraction of units where the two annotators or judges (this measure is generally defined for all use cases with nominal scales where validity is measured with agreement between independent persons, and is not restricted to annotation processes) agree and p_c being the fraction of matching judgments that is expected by chance. p_0 represents the actual agreement and is calculated by summing up the amount of units for which the judges chose the same class and dividing this by the total number of units. p_c represents the agreement expected by chance and is calculated by a class-wise multiplication of the fraction of units that were assigned to this class by the judges separately and then adding up the products. It can be seen, that the numerator in the formula for κ measures the difference between the actual agreed-upon units and those expected by chance. The denominator of this formula corresponds to the highest possible value of the numerator, which would occur in a scenario where the annotators agree on every instance. It is used for normalization purposes, leading to 1.0 being the highest possible value for κ. A κ-value of 0 appears, if p_0 equals p_c and shows that the agreement is the same as the one that would be assumed from random annotations with the same marginal distributions. A negative value means that there is a trend of the two annotators choosing different classes fo the same instance that is higher than it would be expected by chance.

3.2 Methods for CPM Detection

In order to answer the first problem statement, we build a program that tries to predict, which components of the CPM are likely to be contained in a piece of text. This work step, whose structure is now explained, corresponds to the third step in Figure 3.1. Section 2.2 approves MaxEnt as a favorable method for the task of text classification that also facilitates the use of different kinds of features and therefore, MaxEnt seems to be a fitting choice for our experiments. The five CPM components can appear in any combination in a piece of text and thus we need to predict each of them separately which leads to five binary classifiers. The set of these five classifiers together is referred to as the CPM classifier in this work. We first build a baseline version of this classifier with only plain word occurrences as features and then extend it with several further types of features, which are explained below, to have an advanced version. These two versions of the clasifier are referred to as $\text{CPMClass}_{\text{Basic}}$ and $\text{CPMClass}_{\text{Adv}}$ in this work. To prevent the classifier from undesirable behavior, a preprocessing method is implemented, that removes some punctuation, converts everything to lowercase and stems all words.

3.2.1 Settings for $\text{CPMClass}_{\text{Basic}}$

For the basic version of the CPM classifier, we use word occurrences as features. After preprocessing, the text instances are converted into vectors of the Term Frequency-Inverse Document Frequency (TFIDF) format. More precisely, that means with $C = c_1, c_2, ..., c_n$ being the list of all words appearing in the corpus and X being the set of all text instances, the text instance $x = x_1 x_2 \ldots x_m$ is converted into the vector $v = (\text{tfidf}(c_1, x, X), \text{tfidf}(c_2, x, X), \ldots, \text{tfidf}(c_n, x, X))$. A single TFIDF-value $\text{tfidf}(c_k, x, X)$ is calculated by

$$\text{tfidf}(c_k, x, X) = \text{tf}(c_k, x) \cdot \text{idf}(c_k, X) = \frac{|\{x_l \in x | x_l == c_k\}|}{|x|} \cdot \log \frac{|X|}{|\{x' \in X | c_k \in x'\}|}$$

[MSR08]. We utilitze the tfidfVectorizer of the scikit-learn module[1] for python and set the parameters according to Table 3.2.

Name	**Description**	**Value**
max_features	Only the top max_features features are selcted and returned.	10000
min_df	Terms that appear in less than min_df instances are ignored because they are expected to be of no importance.	3
max_df	Terms that appear in more than max_df of all instances are ignored because they are expected to be of no importance.	0.8
ngram_range	Ngram of a size smaller than the first argument of ngram_range or larger than the second argument of ngram_range are ignored. In our case only unigrams and bigrams are used.	(1,2)
stop_words	Every word that occurs in the stop_words list is ignored because they are expected to carry no information.	stopwords.words('english') from nltk[2]

Table 3.2: Parameter settings for the tfidfVectorizer.

3.2.2 Settings for CPMClass$_{\text{Adv}}$

The basic version of the classifier uses only the most elementary components needed for classification, hence motivating us to build in more specialized additions to form an advanced version. As stated in Section 2.2, features for a MaxEnt classifiers are not limited to word occurences, therefore, we want to include a wider variety of features that are explained in the following.

Dictionaries

Dictionaries are word lists that contain words expected to appear in instances of a class. For each component we create such a dictionary with corresponding words. For example, the dictionary for motor expressions contains the word "face". In the dictionary for subjective feelings, we furthermore

[1] https://scikit-learn.org/stable/
[2] https://www.nltk.org/

include smileys like “:)”. The issue of how to combine the dictionary features with the existing TFIDF features is resolved by integrating the units in those dictionaries like additional text instances. Thus, it equals having an additional sentence in our training set for every element in the dictionary of the respective component.

Embeddings

Another idea is that the TFIDF vectors can be substituted by word embeddings. Unlike TFIDF, embeddings are not derived from the train and test sets. Instead, they represent the words occuring in another corpus. We choose to incorporate twitter embeddings from GloVe[3], as they are from the same text domain as our instances. The incorporation is realized by loading an already existing corpus and transforming our text instances from the train and test sets into vectors accordingly. On top of substituting the TFIDF vectors with the embedding vectors, we also try to concatenate both.

Part-Of-Speech (POS)-Tags

Motivated by the observation that the action tendency component is often found in sentences formulated in future tense, POS-tags seem to be a possibility for improvement. We utilize the tag_-option from spaCy[4], that includes the word class, the tense, the person and more. We extend every instance with the POS-tag of every word in it before transforming them into vectors. Thus the tags are treated like words occuring in the instances.

Appraisal Classifier by Hofman et al.

Lastly, we make use of the method for appraisal detection in text that was recently developed at the IMS and is mentioned in Section 2.2 [HTSK20]. This project includes a script that predicts six types of appraisals for text instances, which we use on our corpus. The six types of appraisal evaluate (1) pleasantness (2) expected effort (3) certainty (4) attention (5) responsibility and (6) control [HTSK20]. Following this, we adjust our classifier for cognitive appraisal to predict true if at least one of the six appraisal types is predicted by the IMS classifier and our previous method says true. This feature can not be used for any of the other components except cognitive appraisal.

3.3 Methods for Emotion Classification

The second problem statement is concerned with the usefulness of the CPM for the task of emotion classification. To find an answer, we take two emotion classifiers, one we build and one that already exists, and evaluate whether their performance changes after including CPM data. For optimal comparability, we create three versions of both emotion classifiers, one baseline version without CPM information, one that incorporates the CPM information as annotations, and one that

[3] https://nlp.stanford.edu/projects/glove/
[4]https://spacy.io/usage/linguistic-features#pos-tagging

incorporates it as predictions from the classifier described in Section 3.2. The detailed structure of those classifiers is explained below. For reasons of clarity, each of these classifiers has a name, that is listed in Table 3.3. In the TEC every instance has exactly one emotion, therefore, we use one multiclass classifier, that is, the emotions are not predicted separately but exactly one is chosen for each text instance. This work step corresponds to step 3 in Figure 3.1.

Classifier name	**Classifier description**
$\text{EmoClass}_{\text{MaxEnt-BL}}$	MaxEnt classifier without CPM information
$\text{EmoClass}_{\text{MaxEnt-Anno}}$	MaxEnt classifier with CPM annotations as additional features
$\text{EmoClass}_{\text{MaxEnt-Pred}}$	MaxEnt classifier with CPM predictions as additional features
$\text{EmoClass}_{\text{CNN-BL}}$	CNN classifier without CPM information
$\text{EmoClass}_{\text{CNN-Anno}}$	CNN classifier with CPM annotations as additional features
$\text{EmoClass}_{\text{CNN-Pred}}$	CNN classifier with CPM predictions as additional features

Table 3.3: Overview of the names for the emotion classifiers.

3.3.1 Settings for $\text{EmoClass}_{\text{MaxEnt}}$

With the new classifier we create, the focus is on seeing the alteration through CPM information and not on building a high-performance emotion classifier and thus, only a plain MaxEnt version with word occurrences as features is developed. To evaluate the helpfulness of the CPM information, we need a baseline version to compare the results to. The $\text{EmoClass}_{\text{MaxEnt-BL}}$ classifier works with MaxEnt and uses TFIDF vectors as features. The same preprocessing method as in Section 3.2 is used. Furthermore, we build two copies of this classifier with five additional features that correspond to the CPM components. One of these copies is $\text{EmoClass}_{\text{MaxEnt-Anno}}$ that knows the annotated values of the CPM components, which are assumed to be largely correct, because it is of interest how much increase in correctness could be possible with perfect CPM information. The second copy is $\text{EmoClass}_{\text{MaxEnt-Pred}}$ that knows the CPM predictions. This version is of interest due to the fact that in most use-cases CPM annotations will likely not exist. To integrate the CPM data in $\text{EmoClass}_{\text{MaxEnt-Anno}}$ and $\text{EmoClass}_{\text{MaxEnt-Pred}}$, the vector of every text instance is appended with values for the five CPM components, with the value being 1.0 for components, this instance was annotated/predicted with and 0.0 for the other components. For $\text{EmoClass}_{\text{MaxEnt-Anno}}$ this is possible by simply giving the text file obtained by the annotation as an input to the classifier, which can then read in the annotations for the features. For the prediction, we enable $\text{EmoClass}_{\text{MaxEnt-Pred}}$ to access $\text{CPMClass}_{\text{Adv}}$. Upon the input of a text file $\text{EmoClass}_{\text{MaxEnt-Pred}}$ can then query $\text{CPMClass}_{\text{Adv}}$ for CPM labels and afterwards treat these labels in the same way as $\text{EmoClass}_{\text{MaxEnt-Anno}}$ treats the annotated labels.

3.3.2 Settings for $\text{EmoClass}_{\text{CNN}}$

It is imaginable that a plain classifier as the one introduced above can easily be improved. To ascertain whether CPM information can also be beneficial for a classifier with state-of-the-art performance, we replicate our experiments with the emotion classifier from [HTSK20]. This classifier utilises a

CNN and word embeddings, which we set to a 200 dimensional Twitter embedding from GloVe[3]. For $\text{EmoClass}_{\text{CNN-BL}}$ no further configuration of the classifier is performed. To incorporate the CPM information in $\text{EmoClass}_{\text{CNN-Anno}}$ and $\text{EmoClass}_{\text{CNN-Pred}}$, we append the names of every component an instance is annotated with to this instance's text. In doing so, the CPM information is treated as words of the vocabulary that occur in every instance that is annotated with this component. The CPM classifier and the emotion classifier by Hofmann were developed independently and therefore, differ in their environment and requirements, making an assembly cumbersome. Hence, for the $\text{EmoClass}_{\text{CNN-Pred}}$, we decide to first get the predictions from $\text{CPMClass}_{\text{Adv}}$ in a separate work step and then provide a file with the results to $\text{EmoClass}_{\text{CNN-Pred}}$, which then is able to treat this in the same way as $\text{EmoClass}_{\text{CNN-Anno}}$.

4 Experiments and Results

In this chapter, we demonstrate, how we conduct experiments with the methods proposed in Chapter 3 and present the thereby obtained results. As mentioned above the worklow is parted into three main tasks and this chapter is also structured accordingly. Consequently, the annotation process and the properties of the annotated corpus are explained first, followed by a description of the CPM classifier's development and outcome. The chapter closes with a presentation of the developed versions of emotion classifiers and their comparison.

4.1 Corpus Creation

The first step of the corpus creation is the establishment of guidelines according to the process introduced in Section 3.1. To illustrate the guidelines, Table 4.1 gives an example rule for each component. Two annotation iterations are performed and Table 4.2 displays the achieved inter-

CPM component	Example rule
Cognitive appraisal	The instance describes an individual evaluation of the expectedness of an event.
Physiological symptoms	The instance describes a change in the body temperature.
Action tendencies	The instance describes the urge to avoid an event.
Motor expressions	The instance contains a description of the voice volume.
Subjective feelings	The instance describes the intensity of an individual's feelings.

Table 4.1: Excerpt from the final annotation guidelines.

annotator agreement. In the first round of sample annotations with the guidelines that were initially agreed upon, there are several mismatches between the decisions of the two annotators. Hence, the instances with differing labels are discussed and the guidelines are modified to preempt further insecurities in the annotation decision. This leads to the following adjustments. We extend the cognitive appraisal guideline list with rules that ensure the annotation of instances, where there is no explicit evaluation of an event but an event is described that necessarily invokes a cognitive evaluation and instances with an evaluation of a state. Before these additions, this list only contained rules concerning the evaluation of different aspects of an event. Regarding the guideline list for physiological symptoms, we add a rule to annotate instances about the strength or tiredness of an individual. The list for action tendencies obtains two new rules for the inclusion of instances

[1] In this case all instances were annotated with the value 0 by both annotators, leading to an undefined value due to division by zero.

CPM Component	**Cohen's κ value after first annotation round**	**Cohen's κ value after second annotation round**
Cognitive appraisal	0.288	0.777
Physiological symptoms	0.459	undefined[1]
Action tendencies	0.444	0.732
Motor expressions	0.643	0.617
Subjective feelings	0.733	0.793

Table 4.2: Inter-annotator agreement during the annotation guideline creation measured with Cohen's κ.

that describe the urge to refrain from an action and instances with descriptions of situations that evidently lead to an action tendency. In the sample annotations are no instances that give rise to an additional rule for motor expressions, thus this guideline list remains unchanged. For the subjective feelings list, we add rules declaring that instances with smileys or emojis and instances that explicitly mention a feeling are to be annotated. After the second round of sample instance annotations, the values for Cohen's κ are satisfying. The rather low value in motor expressions traces back to a different understanding of the annotators of the expression ”verbal communication about an emotional reaction to an event”, which is part of one of the guidelines. The cases that are affected by this misunderstanding are resolved by the addition of a rule to the motor expression guideline list, that states, that instances containing interjections are to be annotated with a value of 1 for this component. After this adjustment, the guidelines are complete and the final annotation of 2100 instances out of TEC is conducted. For the sake of the available time, the number of instances is limited to 2100 which are randomly selected out of the 21051 original TEC instances. The complete guidelines including examples, that are also used for the final annotation are given in Appendix A.

After the final annotation, we are in possession of a corpus that has one categorical emotion label and five binary CPM labels for each text instance, and can thus be used for emotion as well as CPM detection. For the resolution of our problem statements, we are interested in the distribution of the CPM components over the different emotions. It is expected, that a large difference in the distributions is advantageous. Table 4.3 presents the results obtained from the annotation. In this figure the rows represent the emotions and the columns represent the CPM component. Each cell displays how many instances are annotated with the respecting emotion and component in black and the percentage of instances annotated with this component out of all instances annotated with this emotion in blue. We can see divergences of up to 23 percent between the emotions for a CPM component. The table shows, for example, that in our corpus physiological symptoms are noticeably more likely to appear in a tweet of the disgust class that in a tweet of any other emotion class and that there is a high probability of joy and subjective feelings occurring together.

	Cognitive appraisal	Physio-logical symptoms	Action ten-dencies	Motor ex-pressions	Subjective feelings	Total
Anger	130 76%	8 5%	30 17%	20 12%	50 29%	172
Disgust	66 81%	12 15%	7 9%	17 21%	21 26%	81
Fear	198 73%	9 3%	37 14%	28 10%	131 49%	270
Joy	626 71%	61 7%	186 21%	98 11%	243 28%	876
Sadness	330 86%	14 4%	61 16%	56 15%	147 38%	385
Surprise	222 70%	2 1%	57 18%	57 18%	85 27%	316
Total	1572 75%	106 5%	378 18%	276 13%	677 32%	

Table 4.3: Relation between CPM components and emotion categories.

4.2 CPM Detection

To resolve, whether it is possible to detect CPM components in text, we utilize $\text{CPMClass}_{\text{Basic}}$ and $\text{CPMClass}_{\text{Adv}}$ with the corpus annotated according to Section 4.1 and evaluate their performance. Of the 2100 annotated instances, 200 are reserved to have an independent set for testing the emotion classification that makes use of the CPM classifier, leaving 1900 instances available for the evaluation of the CPM classifier. For $\text{CPMClass}_{\text{Basic}}$ we use a train-test-split of the corpus, with the training set being three times the size of the test set. For each of the five components, a MaxEnt classifier is fit on the vectors of the training instances and thereupon queried for predictions for the test set. For $\text{CPMClass}_{\text{Adv}}$, we employ an inclusion of 10-fold cross-validation. That means the traning set is split into ten parts and then, in ten iterations a MaxEnt model is fit on nine of these parts and validated on the remaining part, in a way that every part acts as the validation set exactly once. Following this, the best model is chosen with accuracy as the ranking metric. On top, we still use 30% of the instances as a test set on which we measure the performance of the best model that was selected in the cross-validation process. In $\text{CPMClass}_{\text{Adv}}$ we only gradually include the additional features, as we want to see the improvement an individual feature category leads to. To measure the performance, we use the metrics introduced in Section 2.2. For visualization, Figure 4.1 compares the F1-values of $\text{CPMClass}_{\text{Basic}}$ to the final version of $\text{CPMClass}_{\text{Adv}}$ that incorporates the best performing features for every component. The figure displays the F1-scores for every component as well as a simple average over all components. The features which leads to the best result for each component is listed in Table 4.4 and thus, according to this table, we set the final version of $\text{CPMClass}_{\text{Adv}}$. A thorough presentation of the CPM classifiers performance results, including the performance of $\text{CPMClass}_{\text{Adv}}$ with only individual features incorporated is given in Table 4.5. In the table, the columns correspond to the CPM components and the rows state the classifier version and which additional features are incorporated for the specific run. The other cells then display tuples of A, P, R, and F1 for the respective component and classifier version.

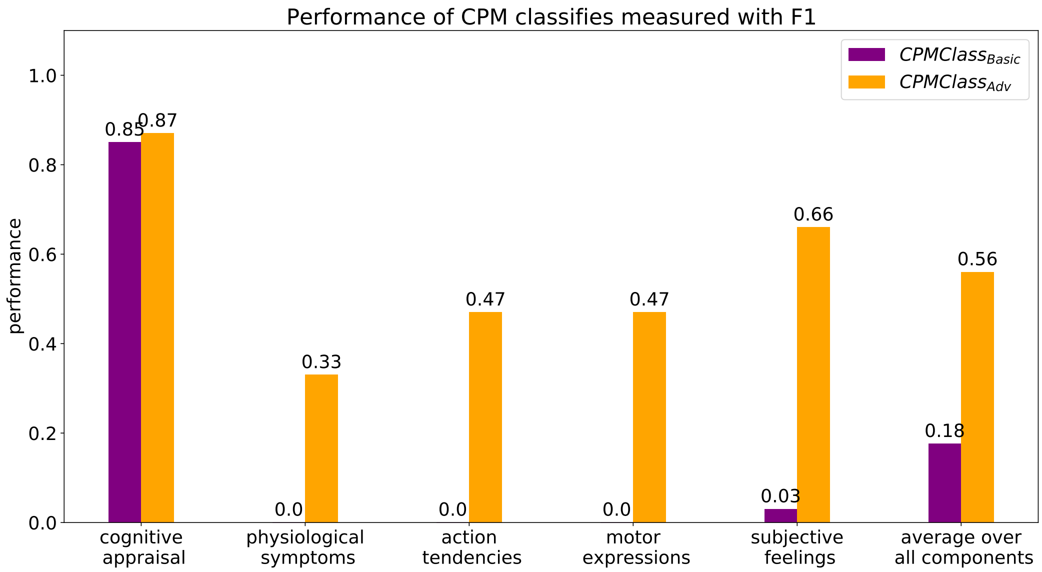

Figure 4.1: Comparison of the F1-scores of $\text{CPMClass}_{\text{Basic}}$ and $\text{CPMClass}_{\text{Adv}}$ with the best feature settings.

Component	**Included features**
Cognitive appraisal	dictionary; embeddings combined with TFIDF; Appraisal-Detection
Physiological symptoms	dictionary
Action tendencies	dictionary
Motor expressions	dictionary; embeddings instead of TFIDF
Subjective feelings	dictionary; embeddings combined with TFIDF

Table 4.4: Feature settings for the $\text{CPMClass}_{\text{Adv}}$ that lead to the best results for the individual CPM components.

Classifier configuration	Component				
	Cognitive appraisal (A, P, R, F1)	**Physiological symptoms** (A, P, R, F1)	**Action tendencies** (A, P, R, F1)	**Motor expressions** (A, P, R, F1)	**Subjective feelings** (A, P, R, F1)
$\text{CPMClass}_{\text{Basic}}$	(0.74, 0.74, 1.0, 0.85)	(0.95, 0.0, 0.0, 0.0)	(0.81, 0.0, 0.0, 0.0)	(0.87, 0.0, 0.0, 0.0)	(0.71, 0.5, 0.01, 0.03)
$\text{CPMClass}_{\text{Adv}}$ + dictionary features	(0.76, 0.76, 1.0, 0.86)	(0.94, 0.33, 0.32, 0.33)	(0.78, 0.42, 0.53, 0.47)	(0.83, 0.36, 0.4, 0.38)	(0.81, 0.69, 0.7, 0.7)
$\text{CPMClass}_{\text{Adv}}$ + embeddings instead of tfifd	(0.77, 0.78, 0.96, 0.86)	(0.87, 0.15, 0.36, 0.21)	(0.72, 0.25, 0.56, 0.43)	(0.82, 0.37, 0.63, 0.47)	(0.72, 0.54, 0.63, 0.58)
$\text{CPMClass}_{\text{Adv}}$ + embeddings and tfifd	(0.78, 0.79, 0.96, 0.87)	(0.9, 0.23, 0.43, 0.3)	(0.75, 0.38, 0.57, 0.46)	(0.82, 0.36, 0.58, 0.44)	(0.79, 0.65, 0.69, 0.67)
$\text{CPMClass}_{\text{Adv}}$ + POS-tags	(0.75, 0.75, 0.98, 0.85)	(0.9, 0.14, 0.21, 0.17)	(0.75, 0.37, 0.5, 0.43)	(0.75, 0.2, 0.34, 0.26)	(0.75, 0.58, 0.63, 0.6)
$\text{CPMClass}_{\text{Adv}}$ + appraisal detection	(0.76, 0.76, 1.0, 0.86)	(0.93, 0.25, 0.18, 0.21)	(0.77, 0.41, 0.54, 0.47)	(0.81, 0.3, 0.33, 0.31)	(0.79, 0.64, 0.68, 0.66)
$\text{CPMClass}_{\text{Adv}}$ + all features	(0.76, 0.78, 0.94, 0.85)	(0.89, 0.2, 0.43, 0.27)	(0.77, 0.41, 0.59, 0.49)	(0.8, 0.33, 0.55, 0.41)	(0.77, 0.62, 0.64, 0.63)
$\text{CPMClass}_{\text{Adv}}$ + best combination of features	(0.79, 0.79, 0.97, 0.87)	(0.94, 0.33, 0.32, 0.33)	(0.78, 0.42, 0.53, 0.47)	(0.82, 0.37, 0.63, 0.47)	(0.79, 0.64, 0.67, 0.66)

Table 4.5: Performance of the different versions of the CPM classifier measured with Accuracy, Precision, Recall and F1.

The first row is for $\text{CPMClass}_{\text{Basic}}$ and the following rows are for $\text{CPMClass}_{\text{Adv}}$. In every row different additional features on top of the TFIDF features are incorporated. The last row shows the best result possible for every component, here the components do not all get the same features. It is visible, that the predictions of the baseline version are erroneous in many cases. The difference between cognitive appraisal and the other components suggests, that the detection of Appraisal performs superior. This difference can be traced back to the frequency with which the components were annotated. In Section 4.1 it can be seen, that 75% of instances were annotated with Appraisal whereas the other components were annotated in at most 32% of instances. If we take a closer look at the predictions of the classifier, it comes clear that the combination (Appraisal-yes, physiological symptoms-no, action tendencies-no, Expressions-no, Feelings-no) is chosen in 471 out of 475 cases, which means that the classifier almost always predicts the most frequent annotation in the training set for every component. If we look at the results of the final classifier in the last row, it can be seen that, while there is only a minor improvement for cognitive appraisal, the precision and recall of all the other components could be increased remarkably. This table shows that dictionaries are helpful for every component, while POS-tags are helpful fo none. For the components physiological symptoms and action tendencies TFIDF-vectors perform superior to embeddings, for the component motor expressions this is inverse, and for the components cognitive appraisal and subjective feelings a concatenation of both is the best solution. Due to the novelty of our methods, there are no other CPM predictions to compare our results to.

4.3 Emotion Classification

We are interested in the role CPM detection can play in an emotion classification task. To analyse this, we run every classifier version explained in Section 3.3. As test set, the 200 instances that were reserved after the annotation are used and the remaining 1900 instances serve as the training set. For the versions of $\text{EmoClass}_{\text{MaxEnt}}$, the best model is chosen through 10-fold cross-validation and validated on the test set, for the versions of $\text{EmoClass}_{\text{CNN}}$, no cross-validation is used. The amount of available annotated instances are limited, therefore, we chose to refrain from using a training set for $\text{EmoClass}_{\text{MaxEnt-Pred}}$ and $\text{EmoClass}_{\text{CNN-Pred}}$ that gets its CPM information predicted, and instead use annotations for the training set and only predict the labels for the test set. The performance of the runs of all classifiers are displayed in Table 4.6 in tuples of P, R, and F1, with the values for the CNN classifiers being averages over 10 runs. The emotion classification is a multi-class problem, which enables us to furthermore compute its overall performance for all classes together. Those computations are displayed in the last two rows of the table with micro average and macro average. Those averages differ in that the micro average adds up TP, FP and FN for all classes and then computes a global F1-value with these, whereas macro average simply averages the F1-values of the individual classes. Furthermore, we summarize the F1-score for the separate emotion classes in Figure 4.2 for better visualization. According to this figure, the CPM information can bring an improvement of up to 0.03 (disgust), but can also decrease the performance up to 0.02 (surprise) compared to the baseline classifiers. Overall it is noticeable that the performance of the classifiers with CPM data is in a very close range of the baseline performance in all cases. The MaxEnt classifier and the CNN classifier react in a similar way to the additional features.

Class	Classifier version					
	EmoClass$_{\text{MaxEnt-BL}}$ (P, R, F1)	EmoClass$_{\text{MaxEnt-Anno}}$ (P, R, F1)	EmoClass$_{\text{MaxEnt-Pred}}$ (P, R, F1)	EmoClass$_{\text{CNN-BL}}$ (P, R, F1)	EmoClass$_{\text{CNN-Anno}}$ (P, R, F1)	EmoClass$_{\text{CNN-Pred}}$ (P, R, F1)
Anger	(0.54, 0.61, 0.58)	(0.55, 0.62, 0.59)	(0.54, 0.61, 0.58)	(0.72 , 0.35 , 0.46)	(0.66 , 0.38 , 0.48)	(0.69, 0.35 , 0.46)
Disgust	(0.56, 0.56, 0.56)	(0.56, 0.56, 0.56)	(0.56, 0.56, 0.56)	(0.65 , 0.25 , 0.35)	(0.42 , 0.25 , 0.3)	(0.7 , 0.28 , 0.38)
Fear	(0.5, 0.75, 0.6)	(0.52, 0.75, 0.61)	(0.53, 0.74, 0.62)	(0.64 , 0.61 , 0.62)	(0.63 , 0.62, 0.62)	(0.61 , 0.62 , 0.61)
Joy	(0.53, 0.69, 0.6)	(0.55, 0.69, 0.61)	(0.55, 0.69, 0.61)	(0.59 , 0.82 , 0.69)	(0.6 , 0.81 , 0.69)	(0.62 , 0.79 , 0.69)
Sadness	(0.56, 0.56, 0.56)	(0.56, 0.56, 0.56)	(0.56, 0.56, 0.56)	(0.57 , 0.42 , 0.49)	(0.59 , 0.44 , 0.50)	(0.56 , 0.45 , 0.5)
Surprise	(0.56, 0.56, 0.56)	(0.56, 0.56, 0.56)	(0.56, 0.56, 0.56)	(0.55 , 0.45 , 0.49)	(0.55 , 0.45 , 0.49)	(0.51 , 0.44 , 0.47)
All classes micro-averaged	(0.56, 0.56, 0.56)	(0.56, 0.56, 0.56)	(0.56, 0.56, 0.56)	(0.59 , 0.59 , 0.59)	(0.59 , 0.59 , 0.59)	(0.59 , 0.59 , 0.59)
All classes macro-averaged	(0.63, 0.35, 0.37)	(0.62, 0.36, 0.38)	(0.53, 0.36, 0.38)	(0.62 , 0.48 , 0.52)	(0.58 , 0.49 , 0.51)	(0.61 , 0.49 , 0.52)

Table 4.6: Results of the emotion detection of the different classifier versions measred with Precision, Recall and F1.

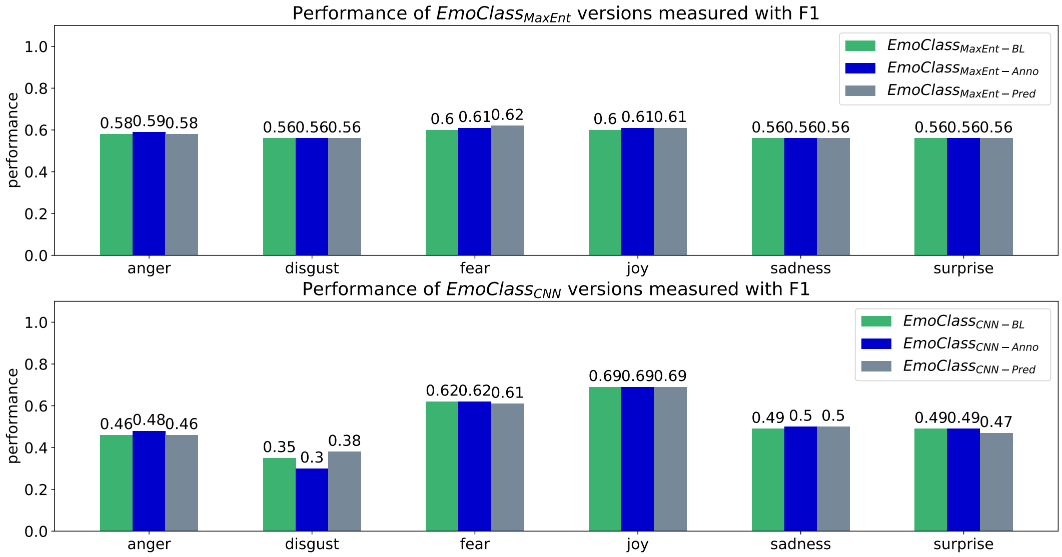

Figure 4.2: Comparison of the emotion classifier versions. The upper plot shows the F1-scores of $\text{EmoClass}_{\text{MaxEnt-BL}}$, $\text{EmoClass}_{\text{MaxEnt-Anno}}$ and $\text{EmoClass}_{\text{MaxEnt-Pred}}$, and the lower plot shows the F1-scores of $\text{EmoClass}_{\text{CNN-BL}}$, $\text{EmoClass}_{\text{CNN-Anno}}$ and $\text{EmoClass}_{\text{CNN-Pred}}$.

Nevertheless, there are some cases in which the classifiers benefit from the CPM data. Table 4.7 gives an overview of all of the text instances with their annotated emotion class, that are not correctly classified by $\text{EmoClass}_{\text{MaxEnt-BL}}$, but are correctly classified by at least one of $\text{EmoClass}_{\text{MaxEnt-Anno}}$ and $\text{EmoClass}_{\text{MaxEnt-Pred}}$. When looking for instance at the sentence in the fourth row, the word "wow", which indicates surprise might not have been detected by $\text{EmoClass}_{\text{MaxEnt-BL}}$, that predicted the class joy, but because it is annotated as a motor expression (interjection as an emotional reaction, more precisely) and surprise has a high co-occurence with motor expression, the classifiers with this information are able to make the correct assignment to surprise. The same argumentation holds true for the sentence in the last row, where we have cognitive appraisal which occurs most often with sadness. $\text{EmoClass}_{\text{MaxEnt-BL}}$ predicts joy for this instance.

Text instance	**Correct emotion class**
F-false E-evidence A-appearing R-real	fear
Who else is sad that Jim Gleason is currently out of the office? #depression #loneliness #Voldemort #WheresTheEDIT	sadness
Ohh! A horsey hore! In the middle of no where hahah hills! Sand! Im on Saltillo Coahuila yippiie! And just got my laser cirgury	sadness
Wow that's a long ass time!RT @murphylee: Y? i been one since 98 RT @CrazeLegs07: @murphylee You're a vegan??	surprise
Almosy made the deans list I just missed it by 5 points	sadness

Table 4.7: Text instances that are correctly classified by the MaxEnt emotion classifier only if CPM data is incorporated.

5 Conclusion

This chapter is devoted to giving a concluding overview of the process, gained insights and contributions of this work. After discussing the obtained results and their relevance, we give an outlook on possibilities for further research interests related to this work.

5.1 Discussion

The introduction already illustrated the usefulness of emotion classification and the interest in emotion detection programs. One complex psychological model of emotions that is accepted by researchers but so far has not been deployed for emotion classification is the CPM by Scherer. The five components of the CPM are concerned with several phenomena related with emotional experiences of humans. This motivated our two research questions, whether it is possible to automatically detect the CPM components in text and, whether this information about the presence of the components can be used to ameliorate emotion classification. To resolve these questions, we first create a corpus with annotations for the five CPM components and for emotions. This corpus is made up of 2100 Twitter posts and is used to train and test the classifiers we build.

For the CPM detection we implement a MaxEnt method. After investigating various setting, we find that a combination of TFIDF-vectors, specialized dictionaries, word embeddings and the incorporation of an existing appraisal classifier yields the best results. With an F1-value of 0.87, the cognitive appraisal component classification performs superior to the classification of the other components. Particular difficulties shows the detection of physiological symptoms, which despite 94% accuracy only reaches F1-values of 0.33. Physiological symptoms is also the least frequently annotated component with only 5% of instances being labeled with this component. For all components we observe equivalent trends of frequency of annotation and height of F1. Overall, we obtain an average F1-value of 0.56, which is acceptable for a new classification task. To conclude the first part of this work, we can affirmatively answer the first research question.

For the emotion detection we decide on using two distinctive classifiers for testing the change in performance through CPM information, in order to receive a more universal insight. We build a new MaxEnt classifier and adjust an existing CNN classifier and then compare their performance before and after enclosing the CPM information for the text instances that are to be classified. This CPM data is incorporated once as the actual annotations and once as the predictions of the created CPM classifier. Upon investigation of the predictions of these six classifier versions, it can be seen that the extended classifiers lead to F1-values in a range of 0.03 of those from their corresponding baseline versions. It was expected that the classifier version with acces to the CPM annotation would outperform the one with the CPM predictions, as we know that these predictions are erroneous in some cases, but this expectation could not be asserted. What we find is a superior performance of the MaxEnt classifiers over the CNN classifiers in the emotion classes of anger, disgust, sadness, and surprise. Although all versions of one classification method have almost identical performance, we recognized several instances which are only assigned to the correct emotion class, if CPM information is available. Therefore, concerning the second research question, we can state that indeed, CPM information can help in classifying text according to emotions, but we have to acknowledge that the improvement in performance we gained was but a rather minor one.

5.2 Outlook and Future Work

With this work we were able to take the first step towards CPM classification. Our classifier has no extraordinary performance, but the results are encouraging to further elaborate the detection of CPM components in text. Apart from social media posts as in this work, an investigation on other text domains like personal messages, product reviews or newspapers could also yield interesting results. In Appendix B we present results from the adaption of our methods on a literature corpus. Another idea is to refrain from using Ekeman's emotions. For instance, one could investigate the applicability of CPM information for sentiment detection. There is also the possibility to further focus on only individually chosen components of the CPM. For Human-Computer-Interaction-Programs it might be useful information to know that there are hints to action tendencies or cognitive appraisals regardless of which emotion they belong to.

Bibliography

[AM20] N. Alswaidan, M. Menai. "A survey of state-of-the-art approaches for emotion recognition in text". In: *Knowledge and Information Systems* (Mar. 2020). DOI: 10.1007/s10115-020-01449-0 (cit. on p. 14).

[AU17] M. Abdul-Mageed, L. Ungar. "Emonet: Fine-grained emotion detection with gated recurrent neural networks". In: *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)*. 2017, pp. 718–728 (cit. on p. 13).

[BS08] T. Brosch, K. R. Scherer. "Plädoyer für das Komponenten-Prozess-Modell als theoretische Grundlage der experimentellen Emotionsforschung". ger. In: Experimentelle Emotionspsychologie: Methodische Ansätze, Probleme und Ergebnisse. ID: unige:96545. Lengerich: Pabst Science Publ., 2008, p. 993. ISBN: 978-3-89967-450-7. URL: https://archive-ouverte.unige.ch/unige:96545 (cit. on pp. 13, 19).

[Can27] W. B. Cannon. "The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory". In: *The American Journal of Psychology* 39.1/4 (1927), pp. 106–124. ISSN: 00029556. URL: http://www.jstor.org/stable/1415404 (cit. on p. 13).

[CO13] G. L. Clore, A. Ortony. "Psychological Construction in the OCC Model of Emotion". In: *Emotion Review* 5.4 (2013). PMID: 25431620, pp. 335–343. DOI: 10.1177/1754073913489751. URL: https://doi.org/10.1177/1754073913489751 (cit. on p. 19).

[Coh60] J. Cohen. "A coefficient of agreement for nominal scales". In: *Educational and psychological measurement* 20.1 (1960), pp. 37–46 (cit. on p. 28).

[Ekm92] P. Ekman. "An argument for basic emotions". In: *Cognition & emotion* 6.3-4 (1992), pp. 169–200 (cit. on p. 16).

[Ekm98] P. Ekman. "Universality of emotional expression? A personal history of the dispute". In: *Darwin, C.(1998) The Expression of the Emotions in Man and Animals, 3rd editon. Oxford University Press: New York* (1998), pp. 363–393 (cit. on p. 16).

[Enc20] T. E. of Encyclopaedia Britannica, ed. *Twitter. microblogging service*. Encyclopædia Britannica, inc. Mar. 2020. URL: https://www.britannica.com/topic/Twitter (cit. on p. 22).

[GBC16] I. Goodfellow, Y. Bengio, A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016 (cit. on p. 21).

[GBH09] A. Go, R. Bhayani, L. Huang. "Twitter sentiment classification using distant supervision". In: *CS224N project report, Stanford* 1.12 (2009), p. 2009 (cit. on p. 20).

[GGF13] N. Gupta, M. Gilbert, G. D. Fabbrizio. "Emotion detection in email customer care". In: *Computational Intelligence* 29.3 (2013), pp. 489–505 (cit. on p. 13).

[HLKS17] M. Hughes, I. Li, S. Kotoulas, T. Suzumura. "Medical text classification using convolutional neural networks". In: *Stud Health Technol Inform* 235 (2017), pp. 246–50 (cit. on p. 21).

[HTSK20] J. Hofmann, E. Troiano, K. Sassenberg, R. Klinger. *Appraisal Theories for Emotion Classification in Text*. 2020. arXiv: 2003.14155 [cs.CL] (cit. on pp. 21, 30, 31).

[Jay57] E. T. Jaynes. "Information theory and statistical mechanics". In: *Physical review* 106.4 (1957), p. 620 (cit. on p. 20).

[Kim14] Y. Kim. "Convolutional Neural Networks for Sentence Classification". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1746–1751. DOI: 10.3115/v1/D14-1181. URL: https://www.aclweb.org/anthology/D14-1181 (cit. on p. 21).

[KK18] E. Kim, R. Klinger. "Who Feels What and Why? Annotation of a Literature Corpus with Semantic Roles of Emotions". In: *Proceedings of COLING 2018, the 27th International Conference on Computational Linguistics*. Santa Fe, USA, Aug. 2018 (cit. on pp. 26, 53).

[Köv12] Z. Kövecses. *Emotion concepts*. Springer Science & Business Media, 2012 (cit. on p. 15).

[Mac20] A. MacArthur. *The Real History of Twitter, in Brief. How the micro-messaging wars were won*. Sept. 2020. URL: https://www.lifewire.com/history-of-twitter-3288854 (cit. on p. 22).

[McI13] C. McIntosh. "emotion". In: *Cambridge Advanced Learner's Dictionary & Thesaurus*. 4th ed. Cambridge University Press, 2013. URL: https://dictionary.cambridge.org/dictionary/english/emotion#dataset_cacd (cit. on p. 15).

[Moh12] S. Mohammad. "#Emotional Tweets". In: *Proceedings of the First Joint Conference on Lexical and Computational Semantics, *SEM 2012, June 7-8, 2012, Montréal, Canada*. Ed. by E. Agirre, J. Bos, M. T. Diab. Association for Computational Linguistics, 2012, pp. 246–255. URL: https://www.aclweb.org/anthology/S12-1033/ (cit. on p. 23).

[MS12] K. Mulligan, K. R. Scherer. "Toward a Working Definition of Emotion". In: *Emotion Review* 4.4 (2012), pp. 345–357. DOI: 10.1177/1754073912445818. eprint: https://doi.org/10.1177/1754073912445818. URL: https://doi.org/10.1177/1754073912445818 (cit. on p. 15).

[MSR08] C. D. Manning, H. Schütze, P. Raghavan. *Introduction to information retrieval*. Cambridge university press, 2008 (cit. on pp. 22, 29).

[Nat20] D. Nations. *What Is a Hashtag on Twitter? Everything you need to know about using Twitter hashtags*. Sept. 2020. URL: https://www.lifewire.com/what-is-a-hashtag-on-twitter-3486592 (cit. on p. 22).

[NLM99] K. Nigam, J. Lafferty, A. McCallum. "Using maximum entropy for text classification". In: *IJCAI-99 workshop on machine learning for information filtering*. Vol. 1. 1. Stockholom, Sweden. 1999, pp. 61–67 (cit. on pp. 20, 21).

[PG17] S. P S, M. G S. "Emotion Models: A Review". In: *International Journal of Control Theory and Applications* 10 (Jan. 2017), pp. 651–657 (cit. on pp. 16, 17).

[Plu01] R. Plutchik. “The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice”. In: *American scientist* 89.4 (2001), pp. 344–350 (cit. on p. 16).

[PLV02] B. Pang, L. Lee, S. Vaithyanathan. “Thumbs up? Sentiment Classification using Machine Learning Techniques”. In: *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*. Association for Computational Linguistics, July 2002, pp. 79–86. DOI: 10.3115/1118693.1118704. URL: https://www.aclweb.org/anthology/W02-1011 (cit. on p. 20).

[RM77] J. A. Russell, A. Mehrabian. “Evidence for a three-factor theory of emotions”. In: *Journal of research in Personality* 11.3 (1977), pp. 273–294 (cit. on p. 17).

[Roj13] R. Rojas. *Theorie der neuronalen Netze: eine systematische Einführung*. Springer-Verlag, 2013 (cit. on p. 21).

[Sch05] K. R. Scherer. “What are emotions? And how can they be measured?” In: *Social Science Information* 44.4 (2005), pp. 695–729. DOI: 10.1177/0539018405058216. eprint: https://doi.org/10.1177/0539018405058216. URL: https://doi.org/10.1177/0539018405058216 (cit. on pp. 13, 15, 19, 26).

[Sch82] K. R. Scherer. “Emotion as a process: Function, origin and regulation”. In: *Social Science Information* 21.4-5 (1982), pp. 555–570. DOI: 10.1177/053901882021004004. eprint: https://doi.org/10.1177/053901882021004004. URL: https://doi.org/10.1177/053901882021004004 (cit. on p. 13).

[SE85] C. A. Smith, P. C. Ellsworth. “Patterns of cognitive appraisal in emotion.” In: *Journal of personality and social psychology* 48.4 (1985), p. 813 (cit. on p. 18).

[Sho05] E. Shouse. “Feeling, Emotion, Affect”. In: *M/C Journal* 8.6 (Dec. 2005). DOI: 10.5204/mcj.2443. URL: https://journal.media-culture.org.au/index.php/mcjournal/article/view/2443 (cit. on p. 15).

[Ski88] J. Skilling. “The axioms of maximum entropy”. In: *Maximum-entropy and Bayesian methods in science and engineering*. Springer, 1988, pp. 173–187 (cit. on p. 20).

[SS62] S. Schachter, J. E. Singer. “Cognitive, social, and physiological determinants of emotional state.” In: *Psychological review* 69 (1962), pp. 379–99 (cit. on p. 13).

[WS12] R. Wicentowski, M. R. Sydes. “Emotion Detection in Suicide Notes using Maximum Entropy Classification”. In: *Biomedical Informatics Insights* 5s1 (2012). PMID: 22879760, BII.S8972. DOI: 10.4137/BII.S8972. eprint: https://doi.org/10.4137/BII.S8972. URL: https://doi.org/10.4137/BII.S8972 (cit. on p. 21).

[WW14] S. Wen, X. Wan. “Emotion Classification in Microblog Texts Using Class Sequential Rules”. In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*. 2014, pp. 187–193. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8209 (cit. on p. 20).

[YLC07] C. Yang, K. H.-Y. Lin, H.-H. Chen. “Emotion classification using web blog corpora”. In: *IEEE/WIC/ACM International Conference on Web Intelligence (WI’07)*. IEEE. 2007, pp. 275–278 (cit. on p. 20).

[ZJXG05] S. Zhu, X. Ji, W. Xu, Y. Gong. "Multi-Labelled Classification Using Maximum Entropy Method". In: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '05. Salvador, Brazil: Association for Computing Machinery, 2005, pp. 274–281. ISBN: 1595930345. DOI: 10.1145/1076034.1076082. URL: https://doi.org/10.1145/1076034.1076082 (cit. on p. 21).

A Annotation Guidelines

Table A.1 illustrates the annotation guidelines used in this work. Every instance of text is assigned one value out of 0, 1 for each of the five components. The illustrated examples are extracted from TEC and serve to clarify the use of these guidelines.

CPM Component	Rules	Examples
Cognitive Appraisal	•The instance describes an individual evaluation of the **novelty** of an event. •The instance describes an individual evaluation of the **expectedness** of an event. •The instance describes an individual evaluation of the **expectedness** of an event. •The instance describes an individual evaluation of the **pleasantness** of an event. •The instance describes an individual evaluation of the **responsibility** of someone for an event. •The instance describes an individual evaluation of the **consequences** of an event. •The instance describes an individual evaluation of someones **coping potential** regarding an event. •The instance describes an individual evaluation of the **violation of social norms** by an event. •The instance describes an individual evaluation of **some other cognitively evaluateable feature** of an event. •The instance describes an event that **necessarily evokes a cognitive appraisal** process in an individual. •The instance describes an individual's **evaluation of a state**.	•The instance *'Thinks that @melbahughes had a great 50th birthday party'* would be annotated with a value of 1, because it describes an individual's evaluation of the pleasantness of the event *melbahughes' 50th birthday party*. •The instance *'too tired for black friday fun.'* would be annotated with a value of 0, because, despite containing an evaluation of a personal state, it does not contain an evaluation of an event.

Physio-logical Arousal	•The instance describes a change in the **heart beat**. •The instance describes a change in the **breath pattern**. •The instance describes a change in the **muscle tension**. •The instance describes a change in the **body temperature**. •The instance describes a change in the **perspiration rate**. •The instance describes a change in the **blood flow**. •The instance describes a **gastrointestinal** reflex. •The instance describes a case of **piloerection**. •The instance describes the state of the **strength or tiredness** of an individual. •The instance describes **any other physiological symptoms**.	•The instance *'Loves when a song makes your heart race practically forcing you to dance around the living room'* would be annotated with a value of 1, because it describes the behavior of an individual's heart beat. •The instance *'its almost the end of november and i still dont feel cold.'* would be annotated with a value of 0, because, despite containing a statement about an individual's body temperature, that body temperature is not a physiological reaction to an event.
Action Tenden-cies	•The instance describes the urge to **approach an event**. •The instance describes the urge to **avoid an event**. •The instance describes the urge to **observe an event**. •The instance describes the urge to **shift attention away** from an event. •The instance describes the urge to **submit** to an individual or event. •The instance describes the urge to **control** an individual or event. •The instance describes the urge to **attack** an individual or object. •The instance describes the urge to **initiate** any other kind of action. •The instance describes the urge to **refrain from** a action. •The instance describes a **situation that evidently leads to** one of the mentioned action tendencies.	•The instance *'sometimes when i think bout you i want to beat the shit out of your face so everyone can see how ugly you are inside and out'* would be annotated with a value of 1, because it describes the urge to approach and attack an individual. •The instance *'Looks like I'm watching an Indian movie tonight'* would be annotated with a value of 0, because, despite containing a statement about an upcoming action of an individual, it does not contain information about what the individual wants to do.

Motor Expression	•The instance contains a description of a **facial expression**. •The instance contains a description of the **state of any facial feature**. •The instance contains a description of **tears**. •The instance contains a description of the **vocal quality**. •The instance contains a description of the **voice volume**. •The instance contains a description of the **body posture**. •The instance contains a description of a **movement towards** something or somebody. •The instance contains a description of a **movement away** from something or somebody. •The instance contains a description of **silence**. •The instance contains a description of a **gesture**. •The instance describes a **interjection** as an emotional reaction to an event. •The instance describes a **non-verbal communication** about an emotional reaction to an event.	•The instance *'@TheBodyShopUK when I walk in the room and my 9month old nephew recognises me and his face lights up with the biggest smile thats 100%'* would be annotated with a value of 1, because it describes the state and movement of an individual's face and mouth. •The instance *'My parents are talking about my dogs belly button... '* would be annotated with a value of 0, because, despite containing a statement about a communication, that communication has no emotional content or reason.
Subjective Feeling	•The instance describes the **duration** of an individuals feelings. •The instance describes the **valence** of an individuals feelings. •The instance describes the **intensity** of an individuals feelings. •The instance describes **any other property** of an individuals feelings. •The instance explicitly mentions the **feeling** of an individual. •The instance contains an emoji that represents a feeling.	•The instance *'Feelin a bit sad tonight'* would be annotated with a value of 1, because it describes that an individual experiences the feeling of sadness. •The instance *'Listening to Wiz Khalifa makes it feel like summer again'* would be annotated with a value of 0, because, despite containing a statement about how something feels, it does not contain information about a property of an individual's feelings.

Table A.1: Annotation guidelines.

B CPM and Emotion Classification in Literature

For further insights, we test $\text{CPMClass}_{\text{Adv}}$, $\text{EmoClass}_{\text{MaxEnt-BL}}$, $\text{EmoClass}_{\text{MaxEnt-Anno}}$ and $\text{EmoClass}_{\text{MaxEnt-Pred}}$ on a literary corpus. This corpus [KK18] was annotated by a master computational linguistics student of the university of Stuttgart with labels for the CPM components according to the guidelines presented in Appendix A. We have 995 text instances at our disposal, of which we reserve 200 for the validation of the emotion classifier. $\text{CPMClass}_{\text{Adv}}$ is run on the 795 remaining instances with its settings remaining unchanged. Figure B.1 summarizes the result of this experiment and is comparable to Table 4.5. Averaged $\text{CPMClass}_{\text{Adv}}$ achieved an Accuracy-score of

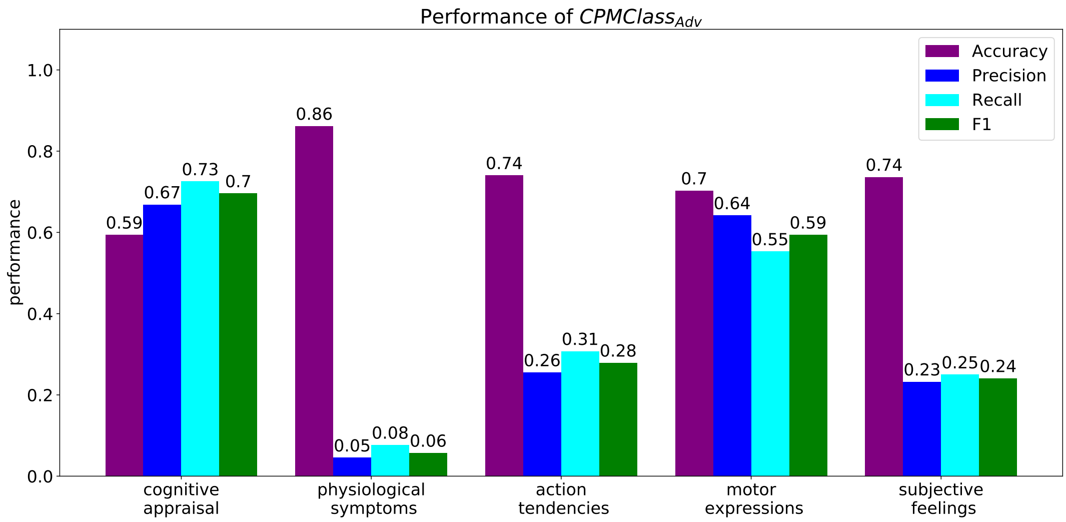

Figure B.1: Resulting performance of $\text{CPMClass}_{\text{Adv}}$'s prediction of the separate CPM components on a literature corpus measured with Accuracy, Precision, Recall and F1.

0.73, a Precision-score of 0.37, a Recall-score of 0.38, and an F1-score of 0.37, which are lower values than those of the Twitter corpus. Upon comparison of the F1-scores of $\text{CPMClass}_{\text{Adv}}$ for both corpora, it is visible that the distribution between the components is similar, again cognitive appraisal scores best and physiological symptoms is the hardest to detect.

The set of possible emotions for the literature corpus is {anger, anticipation, disgust, fear, joy, neutral, other, sadness, surprise, trust} and a text instance can belong to several emotion classes. Our emotion classifiers only predict one emotion label for each instance. It is decided to resolve this discrepancy by counting an emotion prediction as correct, if the predicted label is in the set of annotated labels of the text instance. Figure B.2 demonstrates the resulting performance of $\text{EmoClass}_{\text{MaxEnt-BL}}$, $\text{EmoClass}_{\text{MaxEnt-Anno}}$ and $\text{EmoClass}_{\text{MaxEnt-Pred}}$ upon training on 795 instances and testing on the reserved 200 instances in the same way as described in Section 4.3. The figure is displayed in a format that facilitates comparison to Figure 4.2. Unlike the observed behavior on the

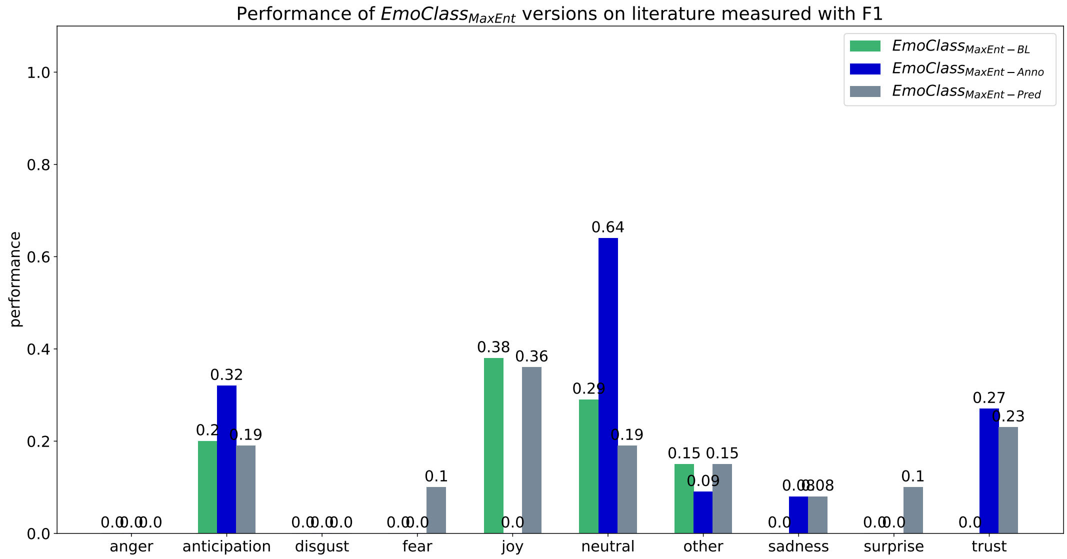

Figure B.2: Comparison of the performance of $\text{EmoClass}_{\text{MaxEnt-BL}}$, $\text{EmoClass}_{\text{MaxEnt-Anno}}$ and $\text{EmoClass}_{\text{MaxEnt-Pred}}$ when applied to the literature corpus.

Twitter text instances, the three classifier versions perform remarkably different on the literature corpus. Overall, the CPM seems to be helpful for the emotion classification.

C Zusammenfassung in Deutscher Sprache

Der Begriff Emotion ist, trotz seiner häufigen Verwendung, noch immer nicht in seiner Bedeutung geklärt. Dieser Sachverhalt führt zu Schwierigkeiten in der automatischen Emotionserkennung in Text. Parallel dazu, nimmt die Anzahl der Anwendungen für Emotionsklassifizierung in der heutigen digitalen Gesellschaft, in der Menschen konstant mit Maschinen interagieren, stetig zu. Folglich wächst die Notwendigkeit einer Verbesserung der Emotionsklassifikatoren auf dem aktuellen Stand der Technik. Der schweizer Psychologe Klaus Scherer publizierte ein Emotionsmodell, welchem zufolge eine Emotion aus Änderungen in den fünf Komponenten kognitive Evaluierung, physiologische Symptome, Handlungstendenzen, motorische Ausdrücke und subjektive Gefühle besteht. Dieses Modell, welches er Komponenten-Prozess-Modell nennt, erreichte Zustimmung in den Bereichen Psychologie und Philosophie, wurde bisher jedoch nicht für Anwendungen aus der maschinellen Sprachverarbeitung benutzt. Mit dieser Arbeit untersuchen wir, ob es möglich ist, die Komponenten des Komponenten-Prozess-Modells automatisiert in Nachrichten aus einem sozialen Netzwerk zu erkennen und, ob die Information über das Vorhandensein der Komponenten hilfreich für Emotionserkennung sein kann. Wir erstellen einen Textkorpus bestehend aus 2100 Twitter Nachrichten, in dem jede Instanz mit exakt einer Emotion und einem binären Label für jede Komponenten-Prozess-Modell Komponente annotiert ist. Mit einem Maximum Entropy Klassifikator erreichen wir eine Erkennung der Komponenten mit einem durchschnittlichen F1-Wert von 0.56 und einem durchschnittlichen Accuracy-Wert von 0.82 auf diesem Korpus. Des Weiteren, vergleichen wir baseline Versionen eines Maximum Entropy und eines CNN Emotions Klassifikators mit Erweiterungen dieser Klassifikatoren mit den Annotationen oder Vorhersagen der Komponenten-Prozess-Modell Komponenten als zusätzliche Eingabe. Wir stellen eine leichte Erhöhung der Performanz, um bis zu 0.03 bei den F1-Werten durch Einbinden der Komponenten-Prozess-Modell Information fest.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

place, date, signature