**Transfer Learning**

Transfer learning in machine learning: goals, formulations, challenges, research, approaches:

This section of the document focuses on the concept of transfer learning in machine learning. It explains that transfer learning refers to improving learning in a new task by transferring knowledge from a related task that has already been learned. The section discusses the goals, formulations, and challenges of transfer learning, as well as current research in the field. It also explores transfer learning in both inductive learning and reinforcement learning, and discusses the issues of negative transfer and task mapping. The section distinguishes transfer learning from multi-task learning, where several tasks are learned simultaneously. It also delves into different approaches to transfer learning, including inductive transfer, Bayesian transfer, and hierarchical transfer.

Transfer learning approaches:

This section of the document discusses various methods and approaches for transfer learning in different learning scenarios, such as supervised learning, reinforcement learning, and unsupervised learning. Some important details include:

- Stracuzzi's work focuses on selecting relevant source-task Boolean concepts from a knowledge base to assist in learning more complex concepts.

- Taylor et al. propose a transfer hierarchy that orders tasks by difficulty to improve the effectiveness of transfer learning.

- Transfer with missing data or class labels can be beneficial in situations where there are limited amounts of data or class labels for a target task. Bayesian transfer methods and boosting algorithms are mentioned as approaches to address this problem.

- Shi et al. discuss transfer learning in unsupervised and semi-supervised settings, where a reasonably-sized dataset exists in the target task but is mostly unlabeled. They propose an active learning approach to label examples only when necessary.

- The section also discusses transfer learning in reinforcement learning (RL), highlighting the RL agent's operation in a Markov decision process (MDP). Different RL algorithms and methods, such as temporal-difference methods, policy-search methods, and imitation methods, are explained.

- Starting-point methods and imitation methods are two types of RL transfer methods that aim to speed up the learning process by leveraging knowledge from source tasks.

- The section further discusses the importance of selecting a suitable source task for transfer learning and provides examples of approaches that order tasks by difficulty or similarity to aid in this process.

Overall, this section provides an overview of various transfer learning approaches and their applications in different learning scenarios.

Task mapping challenges:

This section of the document discusses the challenges and methods of automatically mapping tasks in transfer learning. It highlights the importance of recognizing the correspondences between tasks in order to apply knowledge from one task to another. The section discusses different approaches to address the mapping problem, such as equalizing task representations, trying multiple mappings, and mapping by analogy. It also mentions the challenges and future directions of transfer learning, including the avoidance of negative transfer and the automation of task mapping. The section concludes by emphasizing the benefits of transfer learning and its increasing relevance in machine learning.

**Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning**

Facial expression recognition:

This section of the document discusses a paper on the techniques used in a team's submission to the Emotion Recognition in the Wild contest. The team's submission focused on the sub-challenge of static facial expression recognition in the wild. The paper outlines their use of transfer learning and fine-tuning on deep Convolutional Neural Network (CNN) architectures. They started with a network pretrained on the ImageNet dataset and then performed supervised fine-tuning on datasets relevant to facial expressions and the contest's dataset. The experimental results showed that their cascading fine-tuning approach achieved better results compared to a single-stage fine-tuning. The paper also discusses related works in the field and provides details on the data preparation, architectures, and training process used in their experiments.

Fine-tuning improves accuracy:

This section of the document provides a comparison between the FER-2013 dataset and the EmotiW dataset, which are both used for training and validating fine-tuned models. The section discusses the different schemes used for fine-tuning the pre-trained CNN model with these datasets. The training procedure for the CNNs is also described.

The results and discussion section highlights the effects of supervised fine-tuning and the availability of more labeled data on the performance of the models. It is noted that fine-tuning with the auxiliary FER-2013 dataset generally increases accuracy on the test set by 10% compared to the baseline method. Further fine-tuning with the EmotiW dataset improves accuracy by a few more percentage points. Directly fine-tuning with only the EmotiW training data results in lower accuracy, suggesting the importance of using auxiliary data.

Additionally, it is observed that even models fine-tuned only on the FER-2013 dataset, without any data from the EmotiW dataset, can achieve competitive results on the test set. The quantity of data appears to be more important than the quality in improving the performance of CNN-based face expression classifiers at the early stages. The section also mentions the use of low resolution, larger datasets in fine-tuning, which aligns with neuroscientific findings that facial expression recognition in humans is holistic rather than feature-based.

Important details include the different schemes for fine-tuning, the increase in accuracy with supervised fine-tuning, the lower accuracy when directly fine-tuning with only the EmotiW dataset, and the surprising performance of models trained only on the FER-2013 dataset.

Training difficulties, imbalanced data:

This section of the document discusses the difficulties faced in training models for certain expressions in the EmotiW dataset. The classes "disgust", "fear", "surprise", and "sad" were found to be harder to train, likely due to the fact that they have fewer training samples compared to other classes. These expressions are also more nuanced, making it difficult for both humans and models to accurately label them. It is suggested that the difficulty in assigning correct labels to these expressions may have affected the performance of the models. The section also mentions the imbalance in the number of training samples for the "disgust" class, which could explain why the models consistently struggle to predict this expression correctly. The section concludes by emphasizing the need for larger datasets in training deep neural networks for face expression recognition.

References for facial emotion recognition studies:

This section of the document provides references for various studies and papers related to facial emotion recognition. Some of the important details mentioned include the distribution of classes on different training and validation sets, the challenges in representation learning, and the combination of modality-specific deep neural networks for emotion recognition in videos.

**MULTI-MODAL EMOTION DETECTION WITH TRANSFER LEARNING**

Multi-modal emotion detection:

This section of the document introduces a multi-modal approach for automated emotion detection in speech. The authors address the challenges of small datasets and incompatible labeling idiosyncrasies by using transfer learning and pLDA classification. They train a multilayer TDNN on the task of speaker identification and then fine-tune it on the task of emotion identification. Speech embeddings are extracted from each layer of the network, and text embeddings are generated using a fine-tuned BERT model. These embeddings are then used to train an LDA-pLDA classifier. The authors evaluate the predictive power of each component and achieve an EER of 38.05% on IEMOCAP with the best variant. They also discuss the limitations of existing emotion corpora and the need for models that can adapt to unseen emotions and domains.

Data and methodology:

This section of the document discusses the data and methodology used in the emotion detection pipeline.

In terms of data, it mentions that different corpora have different sets of labeled emotions. While classic models of emotions identify six fundamental emotions (happiness, sadness, anger, disgust, surprise, and fear), more recent research suggests clustering emotions into four basic categories. To balance the classes in their training corpus, the authors adopt the suggested grouping. They use the Crema-D dataset for emotion fine-tuning and the DailyDialog dataset for text-based emotion recognition. They also evaluate their model on the IEMOCAP dataset.

The methodology section describes a three-stage emotion detection pipeline. The first stage involves extracting unimodal embeddings using a time-delay neural network (TDNN) for speech and BERT for text. These embeddings are then concatenated to create a multimodal embedding vector. The second stage involves using an LDA/pLDA classifier to infer emotions from the multimodal embedding vector.

The section also provides details on the experimental setup, including data preprocessing, pre-training on speaker recognition, and fine-tuning on emotion detection. Finally, it presents results of an evaluation of the pipeline, showing the performance of the fine-tuned speech TDNN models and discussing the implications of the results.

Emotion identification results:

This section of the document discusses the results of emotion identification using different variants of a fine-tuned speech TDNN model. The results are presented in tables, showing the accuracy and F1 scores for different configurations of the model, trained with or without the IEMOCAP dataset.

The section also discusses emotion identification using LDA/pLDA classifiers trained on speech and text embeddings. Results for different variants of the TDNN model and the fine-tuned BERT model are presented in tables. It is noted that including the IEMOCAP dataset during training improves the accuracy of the models.

Additionally, the section discusses emotion confirmation results on the IEMOCAP dataset using pairwise testing. The results show that combining speech and text embeddings improves the performance of the pLDA classifier.

The conclusion of the section highlights the effectiveness of the multi-modal approach in emotion detection, demonstrating the ability of the models to adapt to unseen emotions and domains.

**Emotion Detection in Text- a Review**

Emotion detection challenges:

This section of the document is titled "Emotion Detection in Text: a Review" and it provides an introduction to the topic of emotion detection in text. The authors explain that the interest in emotion detection has grown in recent years due to its potential applications in various fields such as marketing, political science, psychology, artificial intelligence, etc. They argue that although many techniques and models have been created to detect emotion in text, there are still challenges that make these methods insufficient, including the complexity of human emotions and the use of implicit and metaphorical language in expressing emotions. The section also discusses the relationship between emotion detection and sentiment analysis, highlighting the need for more precise identification of discrete emotions. The authors emphasize the importance and potential uses of emotion detection in various fields, such as understanding consumer reactions, improving human-computer interaction, and profiling authors based on their emotional expressions. The section concludes by stating that the main contribution of the paper is to provide a review of the literature in the field and to identify the shortcomings and future directions for emotion detection in text.

Dimensional model emotions compared with basic emotions, complexity of emotion expression, implicit expressions, scarcity of datasets, emotion detection methods using supervision and microblogs:

This section of the document discusses the dimensional model of emotions, which defines emotions based on two or three dimensions. It contrasts this model with the basic emotions theory, which states that different emotions correspond to different neurological subsystems in the brain. The section also highlights the complexity of expressing emotions in language, noting that emotion expression is context-sensitive and can consist of a cluster of emotions rather than just one. The importance of understanding implicit expressions of emotions and the role of context in identifying emotions is emphasized. The section further mentions the scarcity of annotated datasets for emotion detection in text and discusses various resources and methods for detecting emotions, including emotion lexicons and word embedding models. Finally, it mentions the use of supervised approaches for emotion detection, particularly on data gathered from microblogs like Twitter using hashtags or emoticons as emotional labels.

Methods for emotion classification using hashtags, emoticons, and emoji as labels. Machine learning algorithms used. Imbalanced training data challenge. Importance of commonsense knowledge and deeper analysis. Brief mention of unsupervised methods:

This section of the document discusses various approaches and methods used for emotion classification in text. Several studies are mentioned, highlighting the use of hashtags, emoticons, and emoji as labels for training data. The accuracy of the classification models varies depending on the emotion being classified, but overall, using these labels is considered a promising strategy. Different machine learning algorithms such as SVM, Naive Bayes, Decision Tree, KNN, and Support Vector Regression have been used in these studies. The section also mentions the challenge of imbalanced training data and proposes different methods to overcome this issue. Additionally, the importance of incorporating commonsense knowledge and deeper analysis in emotion detection systems is emphasized. The use of unsupervised methods for emotion detection is also briefly mentioned, with one study using categorical and dimensional models of emotions.

Emotion detection challenges:

This section of the document discusses the challenges and open problems in detecting expressed emotions in text. It highlights the complex nature of emotion expression in text, including the use of metaphorical language and the context-dependent nature of emotion expression. The section also emphasizes the shortage of high-quality annotated data, which is crucial for training emotion detection models. Additionally, it mentions the inefficiency of current emotion detection models and suggests the need for new approaches, such as multi-class classification methodologies and the consideration of language flow and composition. The section concludes by stating that while emotion detection has made progress, there is still a long way to go before it becomes a reliable and applicable tool in natural language processing.