Company: Corti

Project title: Generating Translated Speech from Transcripts

Motivation:
At Corti we are developing a digital assistant that utilizes artificial intelligence to enhance the decision-making of emergency dispatchers in real-time. We are looking for motivated, and self-driven students that are interested in solving a challenging machine learning project in the domain of automatic speech recognition. We expect the student to have prior machine learning experience and ideally, to have worked with a deep learning framework like TensorFlow or PyTorch. We offer the student a challenging real-life machine learning problem and supervision by machine learning experts.

Project description:
Acquiring training data is expensive, labeling it correctly even more so. Recent progress in generative modelling has made it possible to generate high-quality samples which appear realistic to the user (both in the image as in the audio domain). The aim of this project is combining generative models with machine translation to generate inexpensive training data. While there is a lot of training data available for languages like English or Spanish, languages with a fewer number of speakers typically offer less available data. We would like the student to train a generative model on a low-resource language to generate samples from transcripts which are translated from English. The student should evaluate both the intermediate steps of this pipeline, and the generalization error of a supervised learner trained on the synthetic data.

References:

- van den Oord, et al. “WaveNet: A Generative Model for Raw Audio”. arXiv:1609.03499, 2016.
- van den Oord, et al. “Parallel WaveNet: Fast High-Fidelity Speech Synthesis”.arXiv:1711.10433, 2017.

Company: FOSS

Project title: Image autoencoder as feature extractor for modelling grain kernel images

Motivation:
The EyeFossTM computer vision system replaces human visual inspection in grain quality control. The system uses machine learning, and thus requires large amounts of labeled image data. Currently, these images must be manually labeled by human experts, which is time-consuming and expensive.

Project description:
The purpose of this project is to develop and evaluate image autoencoders for image feature extraction, as part of a larger effort by Foss to evaluate methods for automating image labeling. The autoencoder will be be evaluated on the ability of its extracted features to be used as input to supervised (ANN, SVM etc.) or unsupervised (Kmeans, DBSCAN etc.) grain models. A dataset with ~115k well-labeled grain images is available for this work.

Company: WSAudiology

Project title: Optimising hearing aid settings with collaborative filtering in real-life situations

Motivation:
According the WHO, 466 million people live with disabling hearing loss today, and almost 700 million live with some level of hearing loss. Not only does a hearing loss influence quality of life and economic opportunity, but research show that the chances of mental decline go up, the worse your hearing problems are [Lin, 2011; Abrams, 2017].

WS Audiology is a global leader in the hearing industry, with a clear ambition to redefine it and be number one. One in three hearing aids worn by people around the world has been manufactured by us. WS Audiology continues to pioneer the use of technology to help make hearing diagnoses and care accessible to all, as well as improving the quality of life for our users.

For our Widex hearing aids we have launched “SoundSense Learn” – the first ever AI feature within the industry – as part of the new Widex Evoke™ platform. SoundSense Learn relies on machine learning together with reinforcement learning to perform “AI at the edge”. SoundSense Learn can learn the end-user's preferences in the current environment fast and reliably and adjust the hearing-aid sound processing accordingly.

Project description:
The objective of this project is to develop an alternative approach to suggest hearing aid settings on the mobile device. Currently it is possible to adjust a three-band equalizer for the hearing aid (HA) from the mobile app. This adjustment can happen either directly through a personal program or using the SoundSense Learn functionality to create programs. These adjustments have been collected from users for more than 18 months accumulating >100.000 unique HA settings.

We will investigate if the collected data can be used to build a recommender system to suggest new settings to users. Since the data we have are pairwise assessments of two particular settings at the time, we either need to customize the data to fit standard recommender system approaches or build a specialized recommender system to work on pairwise data. Recommender systems have often been built using collaborative filtering techniques.

Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). Collaborative filtering systems have many forms, but many common systems can be reduced to two steps:
- Look for users who share the same rating patterns with the active user (the user whom the prediction is for).
- Use the ratings from those like-minded users found in step 1 to calculate a prediction for the active user
This falls under the category of user-based collaborative filtering. A specific implementation of this is the user-based nearest neighbor algorithm [Su, 2009].

References:

- Abrams, H. (2017). Hearing loss and associated comorbidities: What do we know. Hearing Review, 24(12), 32-35.
- Lin, F. R., et al. (2011). Hearing loss and incident dementia. Archives of neurology.
- Su, X., Khoshgoftaar, T. M. (2009) A survey of collaborative filtering techniques, Advances in Artificial Intelligence archive.

Company: eCapacity

Project title: Digital jewelry recommendation system using image similarity

Motivation: It can be hard to find that special piece of jewelry that you have in mind – even if you already have a picture of it. A digital jewelry recommendation tool, that can generate jewelry recommendations based on an existing image, will greatly improve the user experience of customers visiting a website.

Most e-commerce recommendation systems are relying on a key word matching system. Adding an option for visual search instead of the user having to provide product description in terms of key words, will be a quicker an easier way, than figuring out the correct set of key words and filters to use.

The recommendation system will help the customers that are in search for a familiar piece of jewelry as well as making the search for the perfect product a simpler and easier affair.

Project description: The input to the algorithm should be an image of a product taken by the customer. Using this input, you should find a way to extract the features of the image and do a similarity calculation to either match the image to a product that is sold in the e-store or find the closest match for it. An image of a piece of jewelry not sold on the e-Store should be able to be used. The features from this new image should be extracted as well as the features from the images on the e-Store in order to create the recommendation system.

Ideally, the system would be able to give an output in real-time.

The feature extraction part of the system could possibly be done using Transfer Learning, Convolutional Neural Networks or something else entirely. It is up to you to find the best approach.

Example:

Input from user Output from recommendation system

References:

- https://towardsdatascience.com/zalando-dress-recomendation-and-tagging-f38e1cbfc4a9
- **https://medium.com/@akarshzingade/image-similarity-using-deep-ranking-c1bd83855978**

Supervisor: Tobias Andersen

Project title: EEG Functional connectivity changes during multisensory perception

Motivation: The brain processes different aspects of incoming sensory information in parallel. As an example, separate areas in the visual cortex process motion and color. The brain needs to integrate, or bind, this information together in order to make decisions. This is known as the binding problem and we still do not know exactly how the brain solves it. Studies of functional connectivity using measures such as phase coherence or envelope correlation has shown that the functional connectivity of the brain areas involved increases when the information they process is bound together (Keil & Senkowski, 2018).

The brain also integrates information across the sensory modalities. One example is in speech perception where seeing the face of the talker facilitates comprehension and influence activity in the auditory cortex. Tuomainen et al. (2005) studied this effect using a degraded speech signal that naive observers did not recognise as speech while trained observers did recognise it as speech. Only trained observers integrated auditory and visual information. This approach has been used to show that effects of audiovisual integration can be measured by comparing the EEG of trained and naïve observers (Baart et al., 2014; Lindborg et al., 2019). We wish to test whether these effects are also present in functional connectivity measure.

Project description:
In this project we wish to compare the functional connectivity from pre-recorded EEG data for trained and naïve subjects to investigate whether it reflects the difference in audiovisual integration between the two groups.

References:

Baart, M., Stekelenburg, J. J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia*, *53*(1), 115–121.

Keil, J., & Senkowski, D. (2018). Neural Oscillations Orchestrate Multisensory Processing. *Neuroscientist*, *24*(6), 609–626.

Lindborg, A., Baart, M., Stekelenburg, J. J., Vroomen, J., & Andersen, T. S. (2019). Speech-specific audiovisual integration modulates induced theta-band oscillations. *PLOS ONE*.

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. *Cognition*, *96*(1), B13-22.

Supervisor: Line H Clemmensen

Project title: Wearable to assist children with OCD

Motivation

In Denmark, the prevalence of psychiatric diagnoses among children has been steadily rising since 2010, while the use of private and public services has remained constant (Sundhedsdatastyrelsen). Many children who need mental health services do not receive them due to waiting times, cost, location, and quality of services (Reardon et al., 2017). The research project WristAngel mission is to develop a wearable which through the use of sensors and AI assists children with psychiatric diagnoses. The first aim is to investigate obsessive compulsive disorder (OCD) with children.

Project description:

The project concerns initial end-to-end proof of concept to use skin conductance and heart rate from wearable wrist sensors to detect anxiety, stress, and similar OCD-related symptoms. This includes setting up an experiment, collecting data, and analyzing the sensor data using deep learning methods. Previous research indicates the feasibility of this approach (McEuen et al., 2017).

References:

McEuen et al. (2017), &You: Design of a sensor-based wearable device for use in cognitive behavioral therapy. In advances in human factors and ergonomics in healthcare (pp. 251-260). Springer, Cham.

Reardon et al. (2017). What do parents perceive are the barriers and facilitators to accessing psychological treatment for mental health problems in children and adolescents? A systematic review. Eur Child Adolesc Psychiatry, 26(6), 623-647.

Sundhedsdatastyrelsen: https://sundhedsdatastyrelsen.dk/da/nyheder/2018/psykiske-lidelser-boern_unge_06012018

Supervisor: Lars Kai Hansen

Project title: EEG / epilepsy project

Motivation

WHO has pointed out that there is huge epilepsy treatment gap in low and middle income countries (LMIC). It is estimated that there are 30 million people with undiagnosed epilepsy in LMIC, and 2 million annual new undiagnosed cases. WHO estimates that 70% of these can be treated with inexpensive medication if we can help to provide a diagnosis! The Danish start-up Braincapture's mission is to create an EEG cloud solution for massive epilepsy diagnosis based on affordable hardware, automated quality control and AI supported diagnostics.

Project description:

The project concerns deep learning in EEG data with a focus on data augmentation, outlier detection, and denoising. Aims are: Assist with real-time quality control and on-site assistance to medical staff and classify EEG data in support of diagnostics. Braincapture co-founder Farrah Matteen has already collected data in Bhutan and Guinea which will be available for the project. Outlier detection can further be based on large international open source data bases. Data augmentation can be based on statistical methods or physical simulation.

References:

Braincapture - braincapture.dk

Chalapathy, R. and Chawla, S., 2019. Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.

Chen, S., Dobriban, E. and Lee, J.H., 2019. Invariance reduces variance: Understanding data augmentation in deep learning and beyond. arXiv preprint arXiv:1907.10905.

Supervisor: Mikkel N. Schmidt

Titel: Reverse curriculum generation for reinforcement learning

Beskrivelse

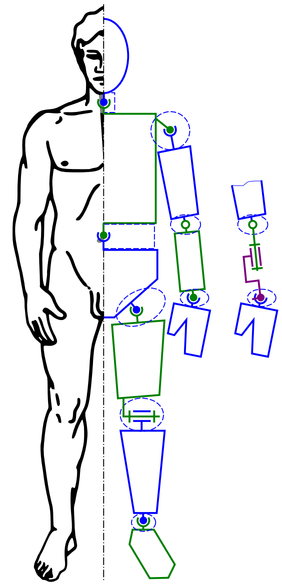
Mange reinforcement learning problemer er karakteriseret ved et meget stort tilstandsrum og kun én eller et snævert sæt af mål-tilstande. Eksempler kunne være systemer som skal løse kombinatoriske opgaver, såsom at finde vej gennem en labyrint, løse en professorterning, eller styre en robot til en specifik tilstand. Disse problemer er karakteriseret ved en sparse reward-funktion, og det kræver derfor store resurser at udforske tilstandsrummet hvis problemet løses med traditionel reinforcement learning. Der findes forskellige metoder som kan hjælpe med at løse disse problemer mere effektivt, heriblandt metoder som tilføjer en mere informativ designet reward-funktion, metoder som anvender mere effektive søgningsheuristikker, og metoder som benytter ekspert-eksempler. I dette projekt vil vi arbejde med metoden *reverse curriculum generation*[1], som genererer træningsdata ved at "regne baglæns" fra en kendt løsning. I projektet udvælges et passende reinforcement learning problem, og en reinforcement learning algoritme (fx. deep q-learning eller policy gradient) implementeres og afprøves.

Referencer

1. Carlos Florensa, David Held, Markus Wulfmeier, Michael Zhang, Pieter Abbeel, "Reverse Curriculum Generation for Reinforcement Learning", Proceedings of the 1st Annual Conference on Robot Learning, PMLR 78:482-495, 2017.

Maximum a posteriori inverse kinematics

Supervisor: Søren Hauberg, sohau@dtu.dk



Project description. Inverse kinematics (IK) is a cornerstone work-horse in animation and graphics. The objective of IK is to find an optimal set of joint angles, such that a stick figure can place a select joint at a specified position. For example, determine how to bend your elbow and shoulder joints such that your hand touches your nose. This problem is fundamentally ill-posed as there are many solutions (you can touch your nose in many equally good ways). The probabilistic approach to make things less ill-posed to assume a prior over possible poses (you then prefer to touch your nose in whichever way your prior dictates). In this project, you will use machine learning to construct a prior of human motion from motion capture data. From this prior, you will build a probabilistic inverse kinematics model.

Learning goals. Upon completion of the project you shoulder

- understand and be able to explain *forward kinematics*, i.e. the process of converting joint angles to joint positions.
- be able to construct a prior over the joint angles of human motion using machine learning.
- be able to apply the change of variable theorem to create a prior over joint positions.
- combine the prior with a simple likelihood and construct a probabilistic model of inverse kinematics.

The project should be implemented in `python` in combination with `pytorch`.

Learning probabilistic trees for causal reasoning

Tue Herlau

January 3, 2021

Causality is often mentioned as one of the most important challenges for machine learning in the comming decade. Causal dependencies is presently described using the language of causal Baysian networks (CBNs) and structural causal models. Both are instances of Baysian networks, i.e. the earthquake-firealarm example from 02450, and describe the independence relations. This allows one to define (and predict) the effect of causal manipulations.

A recent paper by Google Deepmind[GMD+20] has proposed discrete probababilistic trees (DPTs) as a new way to define the above causal terms (interventions and counterfactuals), and also contains an interesting discussion of how DPTs may overcome certain limitations of CBNs. While the paper provides code to build/compute causal effects of DPTs, it does not contain any information of how to construct DPTs in the first place. Judea Pearl, one of the founders of Causal analysis, has been critical of the method claiming it is unable to account for toy examples[1]. The paper may therefore be DOA or one of the most important developments in 2020. This is where you come in. The goal of the project is to extend deepminds work to:

- Build a proof of concept algorithm for fitting DPTs on mixed observational/interventional data
- Test method on problems from supplementary material of paper
- Show it can realiably recover causal structure when interventional data is available
- Assess if placing limitations on the class of available trees $\mathcal{T}$ (below) can overcome some of the issues Pearl brings up
- Assess the method on a real causal learning problem such as a (binarized) version of the COMPAS discrimination problem

Details To allow you to think about how difficult this project would be I imagine the following solution. Define:

- A dataset D consisting of N observations x_i (of M discrete variables), along with any causal manipulations m_i that was active during the observation
- Let $T \in \mathcal{T}$ be a causal tree structure belong to a class of trees $\mathcal{T}$ (as in paper)
- Let θ be the probabilities on each edge (as in paper)

We can now use Bayes rule to get:

$$P(T|D) = P(T|(x_i, m_i)_i) = \frac{\prod_{i=1}^N P(x_i|T, m_i)p(T)}{P(D)} = \frac{\int \prod_{i=1}^N P(x_i|\theta, T, m_i)p(\theta|T)p(T)d\theta}{P(D)} \tag{1}$$

The critical term is $p(x_i|T, \theta, m_i)$ – this is computed by applying the manipulation (i.e. a do-operation, see figure 9 of [GMD+20]) and then simply computing the likelihood of the data. The rest may be computed by ignoring the denominator and use a beta-prior (as in 02450) for θ (or maximum likelihood); this should give us a closed-form expression for the probability of each tree. Implementing this is therefore about as difficult as implementing the pseudo-code in the paper plus some additional details like generating the trees.

References

[GMD+20] Tim Genewein, Tom McGrath, Grégoire Déletang, Vladimir Mikulik, Miljan Martic, Shane Legg, and Pedro A Ortega. Algorithms for causal reasoning in probability trees. *arXiv preprint arXiv:2010.12237*, 2020.

[1]`https://twitter.com/yudapearl/status/1322900367101718528?lang=da`

Supervisor: Morten Mørup

Title: Learning and Visualizing Bipartite Network Embeddings

Motivation:
Bipartite networks are ubiquitous in science and society from social networks spanning recommender systems (consider for instance the celebrated Netflix problem recommending movies to users based on a partially observed user x movie rating matrix) to the life-sciences (i.e., data on drugs and their side-effects can be represented as a bipartite graph as can many datasets within bioinformatics). Unfortunately, understanding the structure of these bipartite networks and visualizing their organization is a major challenge and many approaches have been proposed to extract patterns ranging form simple low rank matrix decompositions (such as PCA) to various advanced latent class and latent feature models.

Project description:
This project will take a network embedding approach in which we will embed bipartite networks into a so-called latent space. Notably, the learned representation will be amenable to potentially interesting data visualizations and facilitate human understanding of complex data. I.e., for a user x movie rating matrix the embedding approach should ideally position users close to the movies they like and similar users and movies naturally be in close proximity to each other.

The project focus will both focus on developing machine learning tools to learn embeddings of bipartite networks efficiently as well as create tools visualizing the learned embeddings. The project will further investigate to what extent the embedded representations can provide state-of-the-art predictions as well as facilitate human understanding of large datasets. The project will depending on your interests potentially also include collaboration with a bioinformatics group with access to very large single cell transcriptomics dataset for which the developed methods can have important applications.

Literature:

Sewell, Daniel K., and Yuguo Chen. "Latent space models for dynamic networks." *Journal of the American Statistical Association* 110.512 (2015): 1646-1657.

Karrer, Brian, and Mark EJ Newman. "Stochastic blockmodels and community structure in networks." *Physical review E* 83.1 (2011): 016107.

Pavlopoulos, Georgios A., et al. "Bipartite graphs in systems biology and medicine: a survey of methods and applications." *GigaScience* 7.4 (2018): giy014.

Cui, Peng, et al. "A survey on network embedding." *IEEE Transactions on Knowledge and Data Engineering* 31.5 (2018): 833-852.

Supervisor: Morten Mørup

Title: Deep Speech Separation

Motivation:
Isolating the speech of a specific talker from a noisy recording of multiple talkers is called speech separation. We are increasingly relying on digital communication channels e.g. in connecting with friends and family, or in working remotely by participating in online meetings. Improved speech separation, especially in noisy, diverse environments, greatly improves how well we can communicate and connect.

Improved speech separation is especially beneficial for people with a hearing loss using hearing aid devices. An estimated 466 million people world-wide have a disabling hearing loss (a number expected to double over the next 30 years, with an increasing number of young people with hearing loss) and untreated hearing loss costs an estimated 750 billion US dollars annually [1].

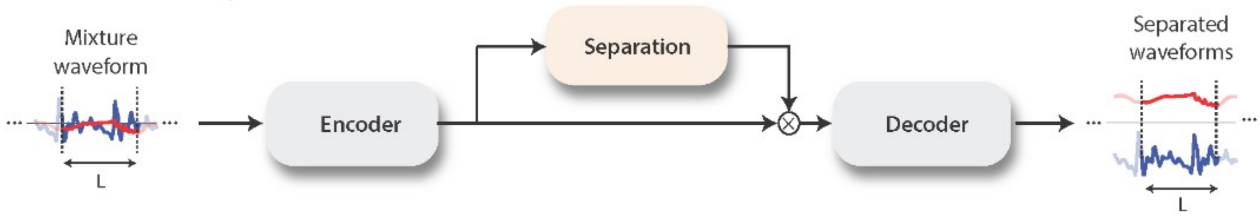


Figure 1: Visualization of example speech separation architechture. Figure is TasNet, adapted from [Fig. 1A, 4].

Project description:
The project will implement and work towards advancing state-of-the-art speech separation systems using deep learning. The implementation will be based on the frameworks available at [2]. The project will systematically investigate how well sound sources can be separated when contaminated with noise using the data set developed in [3].

For instance, building upon the convolutional time-domain audio separation network (Conv-TasNet, see Figure 1) [4], the project would implement an encoder-decoder structure, where the aim is to learn an encoding such that speakers are separable in the encoded space.

Literature:

[1] WHO: deafness and hearing loss, key facts (who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss)

[2] Papers with Code - task: speech separation (paperswithcode.com/task/speech-separation)

[3] Cosentino, Joris, et al. "LibriMix: An open-source dataset for generalizable speech separation." *arXiv preprint arXiv:2005.11262* (2020). (arxiv.org/abs/2005.11262)

[4] Luo, Yi, and Nima Mesgarani. "Conv-tasnet: Surpassing ideal time–frequency magnitude masking for speech separation." *IEEE/ACM transactions on audio, speech, and language processing* 27.8 (2019): 1256-1266. (arxiv.org/abs/1809.07454)