1 Introduction

Data Revolution

- Development of machine = better information system = faster human knowledge

Medicine

- Gene expression, bacteria, diet, antibiotic impact = our epoch is a dark age

Data Analysis

- Methodology, Statistic, Computer System, Big Data = new mathematical field

1.1 Psychological Profiling

Psychological Screening

- Put someone in an environment and look at reaction (revolution: bravery or ingenuity, depending on your political conviction)
- Stimulate someone and look at response
- Autism Screening: 1 hour play, behavior observation, ADOS procedure
- ADOS = Autism Diagnosis Observation Schedule

Need for an App

- Casting the schedule on a Human-Machine Interaction framework
- Hard to go to the psychologist: screening at 4 instead of 18 months.
- Subjective inventories.
- An app, cheap, easy to access and use.
- Able to catch and better analyze differences, speed up understanding of autism

1.2 Face Analysis

The Data

- How to design an app to screen for autism
- Videos trigger specific autistic behavior
- Tower fall, social referencing
- While playing the video on a tablet, we recorded the child reaction with the front camera,
- 20 autistics, 80 control in the control group

- Data preprocessed into landmarks
- Why forgetting movement like ”pointing finger”?
- Homogenous Data
- Beforehand on face expression = easy to reuse algorithm Smiling = Joy
- Smiling = Joy

Ideal Analysis Pipeline

- Muscle contraction analysis to recognize instantaneous feelings
- Analysis of those feelings to recognize behaviors
- Analysis of those behaviors to screen for autism
- Learning or Psychologist

1.3 Table of Content

Features Engineering

- Parameterize the face with muscle activity
- Decompose the signal into simple component and Deal with missing frame
- Dynamic Characterization based on random process classification ideas
- Draw deeper ideas about pattern recognition, linked with psychological ideas of observational behaviors screening

Statistical Learning

- Blind Machine Learning: Explore Technics and Guarantees
- Understandable Learning: Few Features to diagnose
- Modeling of Normal Behaviors: Ask psychologists for descriptors and generative modeling

2 Features Engineering

- Raw Data: High-dimensional, no clear space structure
- Features Engineering: Transform this space to give him structure
- Invariance, Symmetry, this at different scale, and other ideas

2.1 Face Parameterization

Head Pose

- Autistic: back and forth motion, orient their head randomly
- Retrieve a 3D model from the 2D
- Inverse Problem: Easy to go from 3D to 2D
- Not convex cause of the cinematographic projection
- Look at dilatation and ratio between distance
- No need for a perfect solving: Moreover Algebraic formulae could be unstable

Facial Muscles

- Muscular Action
- Absence of subtile of supervision or huge quantity of data = overfitting rapidly threaten
- Keep simple descriptors
- Normalization: remove unwanted variability
- Distance and Angle = 2D translation and rotation invariance
- Normalization given head pose, but rotation = risk of unstability
- Remove morphology variation by describing comparing to distance at rest of muscle computed as the median

2.2 Signal Processing

Sparse Representation

- Understand a phenomemon: separate its differences causes, its sources
- Signals are high-dimensional, noisy and frames missing
- Five momentary head-turn and one slight one all along the video
- Mathematical Formulation, how to solve this problem
- Trade-off between basis = uniqueness, easy to compute and to compare two signals & Dictionary containing more simple component, for more accurate decomposition
- atom should be translation invariant and eventually scaling

Wavelet Solution

- A orthogonal dictionary of localized, translated and scaled atom = wavelet
- Easy to compute, with cascading to faster the transform
- Simple Algorithm: Compute wavelet coefficients on the signal and smooth them in this dictionary to extrapolate on the missing frame
- Not satisfying, atoms pics aren't well positioned
- One can see the Daubechies wavelets form

Dictionary Solution

- Take a simple atom, translate and scale it
- Translation invariance = convolution = fast dictionary transform
- Two types of methods: First = Convex Relaxation
- Close reconstruction = euclidean norm, Sparsity = norm L1
- Second = Heuristic Method
- Look at the biggest response and find this as the biggest component
- Remove this component and look at the second biggest component as the new biggest component, and so on

Final Solution

- Convex Relaxation: too long to optimize or too manage well a fast optimization (we have tried simple proximal splitting method, looking like alternative projection)
- Heuristic method, Improvement with scaling response
- Why? Look at scale-space
- New Idea: Look at maxima in scale-space. It works best for reconstruction

2.3 Characterization of Dynamics

- Further analysis: extract descriptors on the signals
- video = random process with a simple law to discover
- Characterization with moment or density

Independence Hypothesis

- Strong Hypothesis: frame independent
- Describe the distribution : explain the formula L, X, x_i
- moment descriptors. It could be meaningful as variance = kid agitation
- or density descriptors, simpler = histogram = cluster regarding muscle activity = highly or slightly, contracted or extended

Frame Correlation

- Correlate channels = Muscle Correlation are meaningful and we want to extract this info
- Example: smile and wringles
- Distribution Descriptor: Moment (correlation), Density (joint-histogram, or joint-density estimation)
- Example of organized motion in normal kids = turn to turn at their mother

Time Correlation

- Described correlation in time: access speed or acceleration
- Mathematically a lot of random process are defined by their increment properties (Levy processes among which is Browian motion)
- So described increment and at different scaled
- Increment = filtering against a filter [-1, 0..., 1] : smoothing = Derivative of Gaussian
- New Ideas = A first layer retaking increment description and signal processing idea = two filter at 7 scaled

Non-Stationarity

- Simple idea = take density descriptor on first layer
- Supposed stationarity of the process or it is false since response to event
- So localized density descriptor. We suggests a simple framework: explanation
- Second Layer: Explanation. T2 = Pooling to remove local redundancy (removed features known as redundant allow to fight against overfitting)

2.4 Pattern Recognition

- Size and Supervision of the dataset = hard to do comprehensive deep learning
- Yet we will suggest some ideas for the learning

Convolution Network

- Remember the ideal analysis pipeline
- Raw data: extract facial muscular action
- Parameterization and filter to recognize the action eventually learn the filter to learn the type of meaningful muscular activity
- Combine Action to recognize Facial Expression
- Combine it to recognize Behaviors

- Perform statistics to screen for autism
- Let the network learnt the filter with back-propagation

Clustering Network

- If you don't like the filtering framework, you can think with a word presence framework
- Remember the work of Yaniv Romano and Michael Elad

3 Statistical Learning

3.1 Machine Learning

Blind Learning

- Let first look at blind machine learning with a simple example: how to classify the red dot?
- Any child would say: well it should be yellow it is in a yellow see
- In other term look at neighborhood
- But how to define the neighborhood mathematically?
- Which metric to take: euclidean for facility
- Lean a metric that group some clusters
- And separate between clusters (margin condition)
- See on a real example (learning hard to manipulate)
- Linear Separation: Svm
- Power of generalization: A separation is often a line in a new space (see kernel space for Svm).
- If you rely on Euclidean info you only need to describe the span of the data and not the whole space.
- Historically description through kernel space, for us simple Gram-Schmidt.

Scoring

- We have designed blind method how to know if they work?
- See if they predict autism on new data?
- More precise guarantee given by statistics. Law on the data
- A measure of dissatisfaction to rank the methods, usually reading as mean over each sample since it is easy to manipulate statistically
- Approach this measure empirically
- If you try a lot of method, more likely one of them will randomly perform well on your scoring method, so you have to be careful with this method
- Finding guarantees is still an open problem
- Not asymptotical statistics, and unknown statistics
- Concentration inequality. Make sure that randomness in empirical quantity stay reasonable.
- What are the good score?
- Recall: how much alert your method doesn't give
- Precision: on the alert: how much are in fact meaningful

- For our application we want to raise a warning flag for the child to go at the psychologist get a screening. So we want to maximize the precision while keeping a recall close to one (don't miss a case).
- Area Under the Curve: the method give an index of warning, regarding how to threshold this index we obtain different (recall-precision)-couple, or other trade-of couple, describing this curve is useful. A good curve often maximize area under it.
- The hard thing is to distinguish between training and testing. To do so, a good way is to let one autistic and one normal child away and look at which one as the higher warning index. In other term, is our method able to distinguish an autistic from a normal child.

3.2 Understandable Learning

- For medical purpose we want our method to be easily understandable putative discriminative features of our descriptors.

Sparse Regression

- One way to do it, is to force the classifier to rely on few of the
- Again the sparsity problem
- New heuristic idea: Recursive Features Elimination. We learn the linear model than we removed the features that aren't really used by the classifier. And we refine iteratively.
- We can also used convex relaxation like this is the well-known "lasso"
- Those convex relaxation doesn't generalized well yet, how would we do it?
- mask the data in order to remove features with a matrix W
- Then cast Wx in an optimization framework adding penalty on W to mask the data
- mask the data = used few columns of x = measure with operator norm = norm p on the column then 0 relax in 1 on the lines
- which p to choose: p = 1: column = line
- Thinks about the L1 relaxation why it work?
- if allowing L1 norm to be at most equal to x, because of its form it is more likely to meet an other space on an edge corresponding to a lot of null coordinate
- A simple understanding of the ℓ^1 penalty is the following, let's consider the least square problem, the solution is given by

  $$w \in X^\dagger y + \ker X$$

  If we want to minimize $\|w\|$ keeping this constraint, a graphical way to obtain the solution can be seen on the figure. It consists in blowing the ℓ^1

ball as a balloon, and wait for the ballon to hit the hyperplane $X^\dagger y + \ker X$. Because this balloon is like a diamond (ℓ^1 ball), it is more likely to hit the plane on a corner or a sharp edge rather than a plane surface and this sharp edge correspond to some coordinate being null.

- So we want to design a ball on the matrices such that it is wider closer to the space corresponding to some columns being null.
- Because of parameterization symmetry of the columns, Intuition: p = 2
- however, can be hard to cast this in a nice framework without loosing convexity

Features Reduction

- Reuse the Euclidean Info summary ideas
- Find a span basis of the features
- Order the features than iteratively add them if not correlated and linear independent
- For the order, look at the distribution tails? Does it contains more autistics than a random features? Really easy to compute.
- An other method, Decision tree, like MCHAT, allow to easily combine simple test based on a unique features to choose

3.3 Modeling of Normal Behaviors

- Need to work closer to psychologist knowledge
- Autism = a condition as a sum of weird behaviors
- Try to spot those strange behaviors
- Ask psychologist to help us, learn on the dataset differences

Long-Term Behaviors

- blink rate: more or less a technical problem
- asymmetry: compute difference between left and right part of the face
- micro-movement: look at noise (comparison with smooth signal)
- non-smooth movement: correlation in time
- cumulative motion: l1 distance on speed
- motion variability: variance
- motion organization: correlation between channels/muscles

Event Responses

3.4 Long-Term Behaviors

Generative Modeling

- Some features seems to show differences of distribution between autism and normals
- Generative model: model distribution of $(X|y)$
- Use Bayes to deduce $(y|X)$
- To design a test statistical guarentee: such as Neyman-Pearson preaching for likelihood ratio

Detection Framework

- Some features seems to extract few response as unusual
- How to interpret them: a normal child responding, or a weird autistic response
- So singularity = detect low-likelihood
- then combine singularity to obtain a diagnosis
- weight each singularity: some preach for normality, some for autism
- To interpret a singularity: look at which class present the most of low likelihood
- These to weight a singularity before looking at at each of them and suming
- Method that I design a bit from out of nowhere and it works quite well

3.5 Event Response

- harder to characterize cause a lot of variability
- so try our clustering framework
- cluster the different face appearing (10 cluster) (on face parameterization)
- and some do an histogram for each of the kids
- Those cluster could be interpret as pattern and used to filter the signal
- They didn't show impressive results beside the fact that once against distribution were a bit different for each class, for example, percentage of people showing a low response against one filter.

4 Conclusion

- AUC is a bit meaningless because compute on small fold on then average
- results are good if we preprocess the data with the ordered gram-schmidt process before but if we do it on training it is a bit more random.
- Improvement: technical ideas could be better implemented such as extracted the facial muscle
- too much variability due to meaningless features, more education or more data to remove those