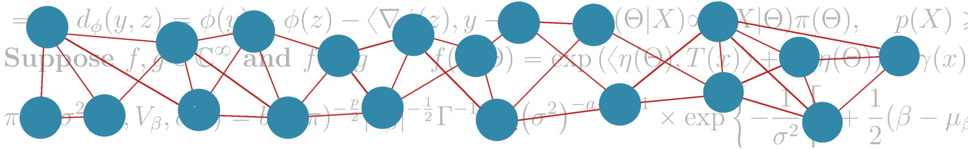


Distributed Bayesian Modelling

Kamil Dedecius

ÚTIA AV ČR, v.v.i. & FIT ČVUT v Praze

14.10.2021

Motto 1: Bayesianism is natural

Bayesian modelling is principally very similar to our thinking

prior opinion & *new observations* $\rightarrow$ *posterior opinion*

- opinion = knowledge = expectation = ...
- real life examples
  - learning to bike
  - learning to behave correctly
  - problem solving in maths
  - searching lost keys
  - even learning to cheat

Motto 2: Two heads are better than one

American Psychological Association: Two (or More) Heads Are Better Than One for Reasoning and Perceptual Decision-Making. (Dec. 18, 2014)

> However, because accuracy is often correlated with confidence, it may be that the most confident group member exerts the strongest influence regardless of whether their answer is right or wrong, and it just so happens that the most confident person is usually right.

…but we're not that far (yet)

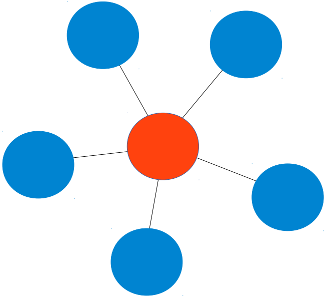
Communication strategies

Fusion center

- oriented graph (tree)
- nodes gather observations
- FC responsible for info processing
- ...and may send results back

+ effective information processing
+ node malfunction can be detected
+ relatively flexible (node addition/removal)
+ simple COMM protocol

- SPoF at FC
- high COMM requirements at FC



Communication strategies

Incremental strategy

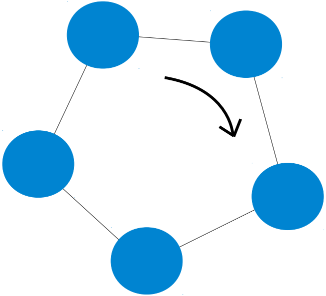
- oriented graph – Hamiltonian cycle
- nodes are equivalent

\+ easy information processing

\+ simple COMM protocol

\- SPoF at each node (recovery is NP hard)

\- info poisoning propagates further



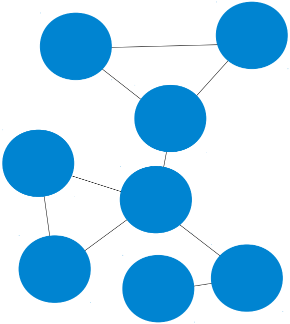
Communication strategies

Consensus and diffusion

- more complex graph with higher node degrees
- no dedicated nodes, no FC
- COMM among neighbors within 1 edge distance

+ COMM requirements relatively OK
+ excellent redundancy (no SPoF)
+ excellent flexibility (node addition/removal)
+ node malfunctions (poisoning) detectable

\- more complicated info processing



Communication strategies

Consensus

- multiple time scales
- → sensing step
- → information processing step
- → consensus iterations

- \- higher COMM reqs.
- \- higher processing reqs.
- \+ consensus reached

Diffusion

- single time scale
- → adaptation step
- → combination step

- \+ lower COMM reqs.
- \+ lower processing reqs.
- \- no consensus

Diffusion strategy

1. *Adaptation step:* observations are shared and incorporated into local knowledge.

   Bayesian update @ node

   prior opinion & **new observations** → *posterior opinion*

2. *Combination step:* posterior opinions are shared among neighbors.

   Information fusion @ node

   several **posterior opinions** → *one opinion*

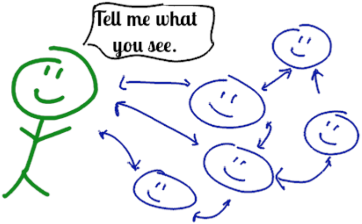
Diffusion strategy

- *Adaptation step:* observations are shared and incorporated into local knowledge.

Bayesian update @ node

prior opinion & **new observations** → *posterior opinion*

$$p(\theta|\text{observations}) \propto \mathcal{M}(\text{observations}|\theta) \times p(\theta)$$



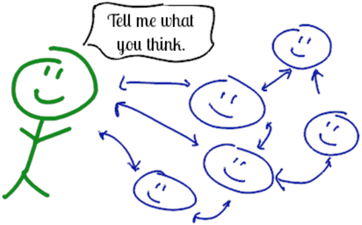
Diffusion strategy

- *Combination step:* posterior opinions are shared among neighbors.

Information fusion @ node

several **posterior opinions** → *one opinion*

$$\tilde{p}(\theta|\cdot) = \bigoplus_i p_i(\theta|\text{observations})$$



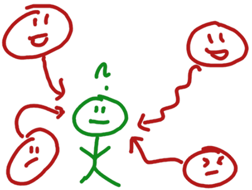
Diffusion strategy – possible settings

Two possible protocols:

- ATC – Adapt–then–Combine
- CTA – Combine–then–Adapt
- + isolated A and C

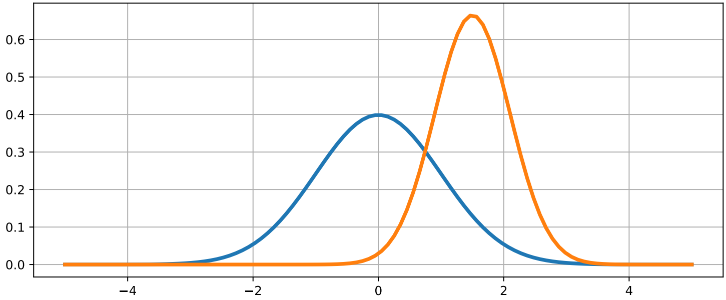
Information weighting

- Equality of neighboring nodes
- Discrimination and preferences of nodes



Combination of opinions (knowledge)

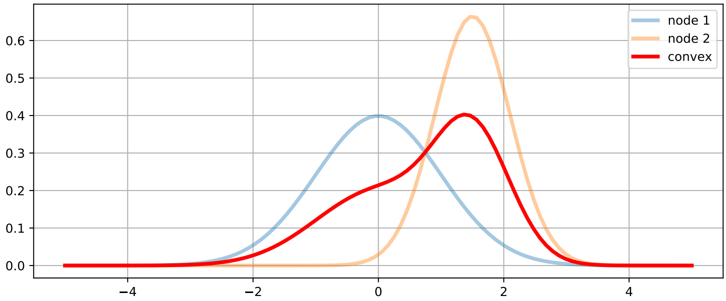
Assume that we have two (prior, posterior...) opinions:



How to merge them correctly???

Combination of opinions (knowledge)

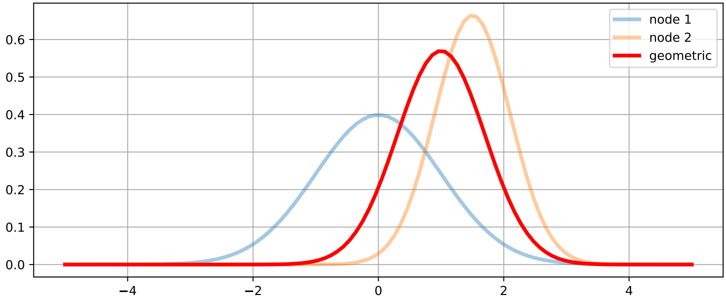
Assume that we have two (prior, posterior. . .) opinions:



How to merge them correctly???

Combination of opinions (knowledge)

Assume that we have two (prior, posterior...) opinions:



How to merge them correctly???

Combination of opinions (knowledge)

Kullback-Leibler divergence

Assume 2 pdfs p, q s.t. $q(x) \ll p(x)$. Then

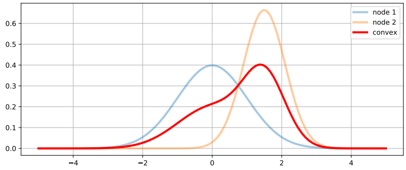
$$\mathcal{D}(p||q) = \mathbb{E}_{p(x)}\left[\log\frac{p(x)}{q(x)}\right]dx = \int p(x)\log\frac{p(x)}{q(x)}dx$$

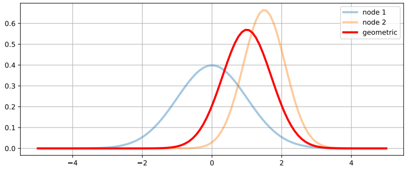
$\mathcal{D}$ is a premetric (nonnegative, asymmetric, does not fulfill $\triangle$ inequality).

Criterion to find the best approximating $\tilde{p}$:

$$\min\frac{1}{n}\sum_i \mathcal{D}(p_i||\tilde{p}) \Rightarrow \tilde{p} = \frac{1}{n}\sum_i p_i$$

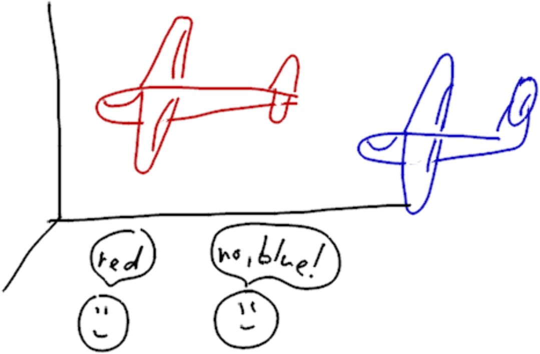
$$\min\frac{1}{n}\sum_i \mathcal{D}(\tilde{p}||p_i) \Rightarrow \tilde{p} \propto \prod_i p_i^{\frac{1}{n}}$$





Partially compatible knowledge

How to detect that our nodes refer about the same process?



Partially compatible knowledge

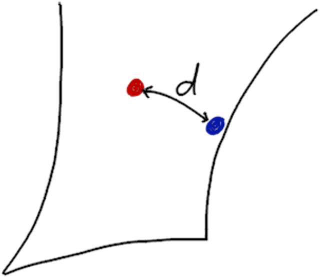
How to detect that our nodes refer about the same process?

Measure dissimilarity

- distance (Euclidean)
- divergence (prob. manifold)

Common parameters?

- factorize posteriors (VB)
- marginalize (submodels in KF)



Features of Bayesian setting

- \+ Excellent explainability
- \+ Natural interpretation, works similarly to human thinking
- \+ “Absolute” generality: no special assumptions or concrete models
- \+ Many existing solutions are only special cases
- \- Perceived as “too complex/complicated”
- \- Single-problem-oriented solutions may be superior

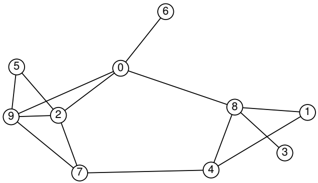
Open problems

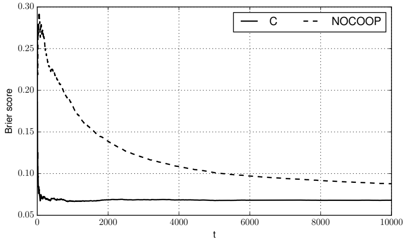
- Combination of predictions
- Partially compatible models
- . . . and submodels
- . . . and models of (generally) correlated processes
- Information weighting
- etc.

Example: Logistic regression

Bhatt and Dhall's skin-nonskin dataset:

- classes: skin, non-skin
- regressors: [1, B, G, R]
- 10,000 observations
- sequential classification & learning

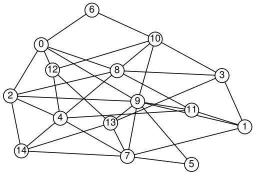


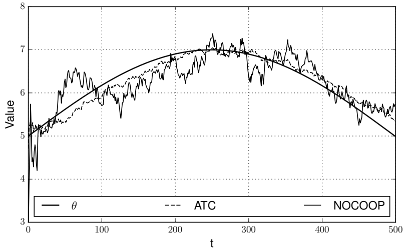


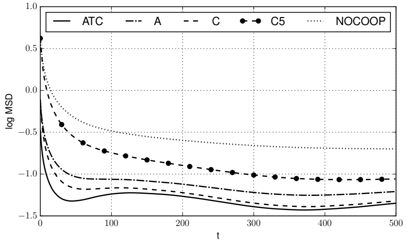
Example: Poisson process rate estimation

Simulated data, TV parameter:

- observations: Poisson variable
- 500 observations
- sequential estimation of time-varying rate



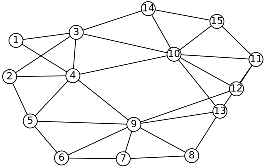


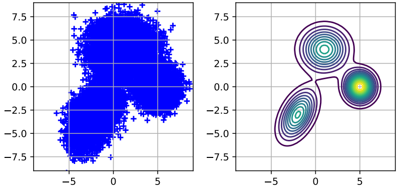


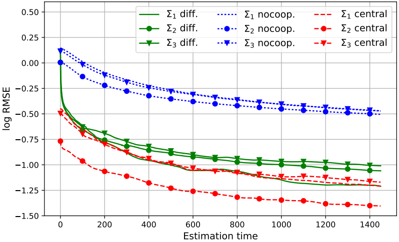
Example: Mixture estimation

Simulated data:

- observations from mix
- 1500 observations
- sequential estimation
- floating window: 50 obs.



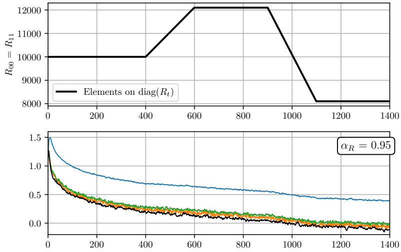




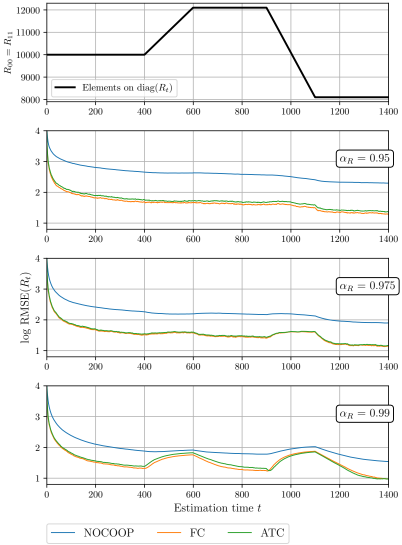
Example: VBKF w/ unknown noise covs.

Simulated data:

- 2D trajectories, CVM
- sequential estimation
- est. of states and TV MNCM



log RMSE($[x_{1,t}, x_{2,t}]$)



Further reading

- Collaborative sequential state estimation under unknown heterogeneous noise covariance matrices, IEEE Trans. Signal Process. 68(10), 2020.
- Sequential Poisson Regression in Diffusion Networks, IEEE Signal Process. Lett., 27(1), 2020.
- Factorized Estimation of Partially Shared Parameters in Diffusion Networks, IEEE Trans. Signal Process., 65(19), 2017.
- Sequential estimation and diffusion of information over networks: A Bayesian approach with exponential family of distributions, IEEE Trans. Signal Process., 65(7), 2017.

Collaboration is not a sin! :)