Interactive Machine Learning for Information Visualization

Viet Minh Vu

under the supervision of Prof. Benoît Frénay

17/05/2019

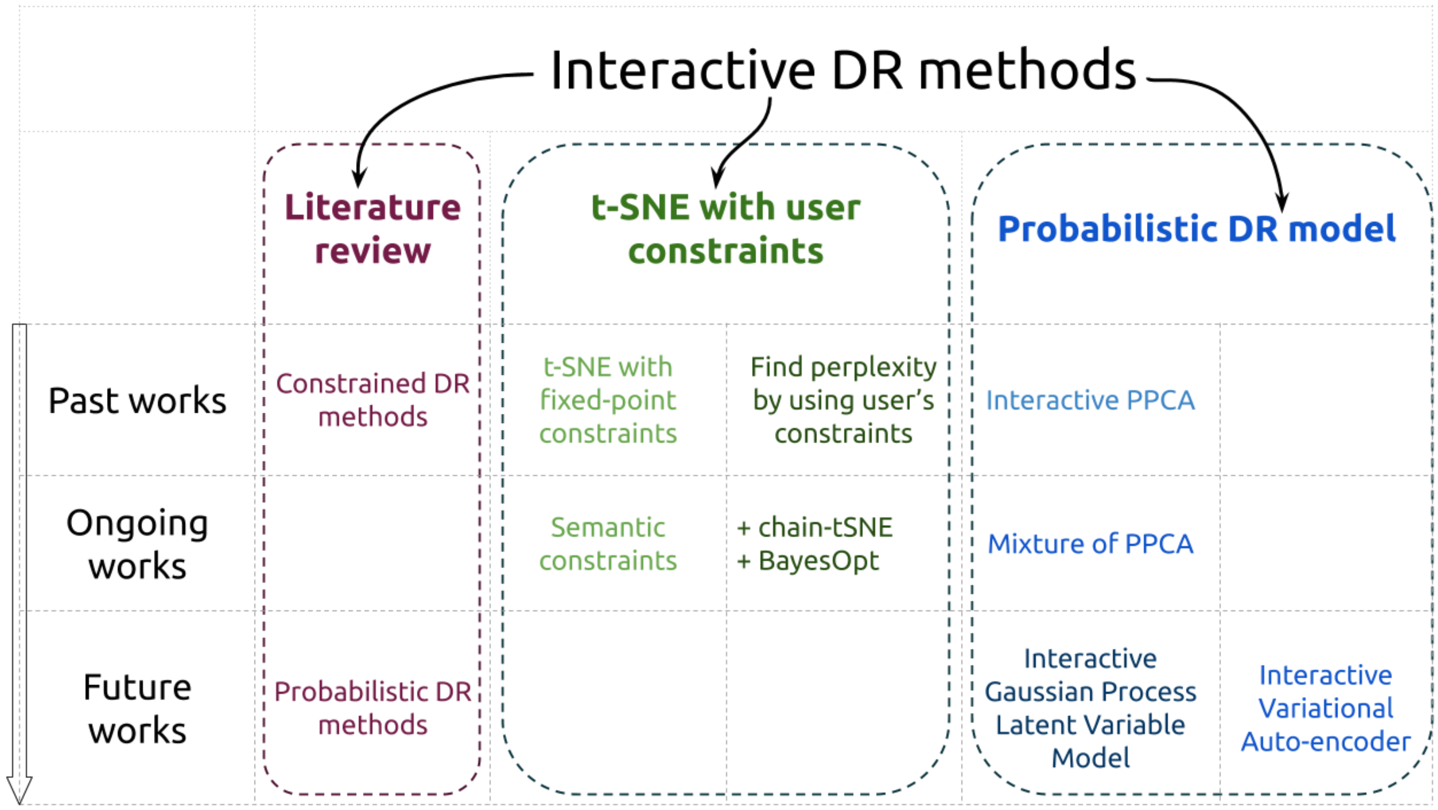
Scope

+ Machine Learning for Information Visualization $\longrightarrow$ Dimensionality Reduction (DR) methods

+ Users interact with the visualization to give their feedbacks

+ Research goal: transform the cognitive feedbacks of the users into constraints for the DR methods

Work plan Overview



What we’ve done so far

+ Literature review of DR methods with constraints

+ Two works with t-SNE:

- Incorporate fixed-point constraints to t-SNE
- Use the user's pairwise constraints to find the best t-SNE visualization

+Incorporate fixed-point constraints to the probabilistic PCA model

Review of DR methods with constraints

+ Consider the user's feedback as constraints to the system

+ Categorization

- Instance-level
- Group-level
- Feature-level
- Dataset-level

→ The review scope is too large, the categorization is not sound.

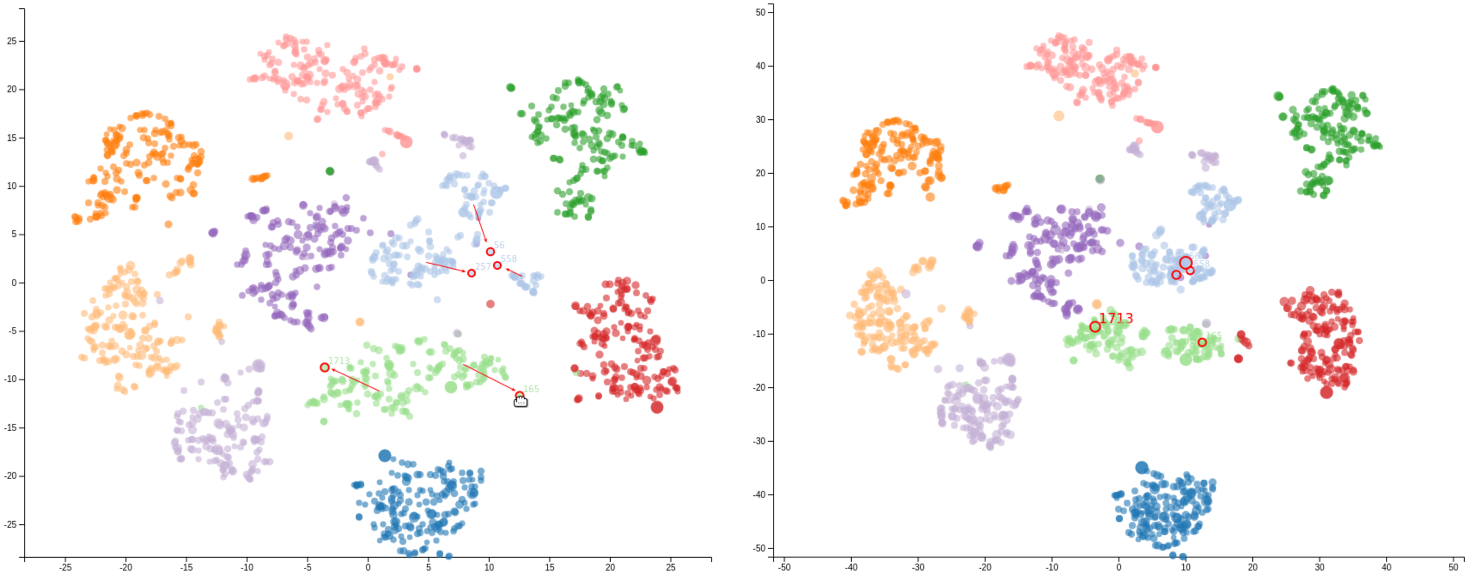
Fixed point constraints with t-SNE

+ Fix the position of some anchor points

+ Attract their neighbors to move towards the fixed points

+ Goal:

- Merge the small clusters
- Split the large cluster



6.2

Tuning perplexity with user constraints

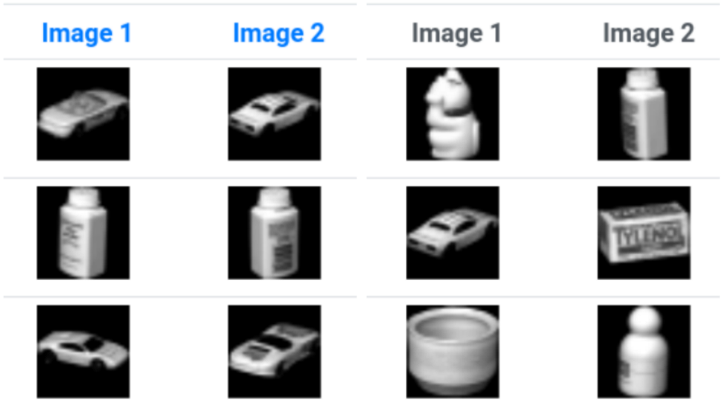
+ t-SNE's perplexity is not easy to tune and to find the best one

+ Let the users define their requirements about the visualization in form of pairwise constraints:

- Similar links to connect similar objects
- Dissimilar links to connect dissimilar objects

+ Transform the user defined constraints into a *constraint-preserving score*

+ Use this score as a criteria to select the best perplexity



(a) Must-link constraints. (b) Cannot-link constraints.

We have:

- the user's feedback in form of constraints,
- the *relationship* between points:

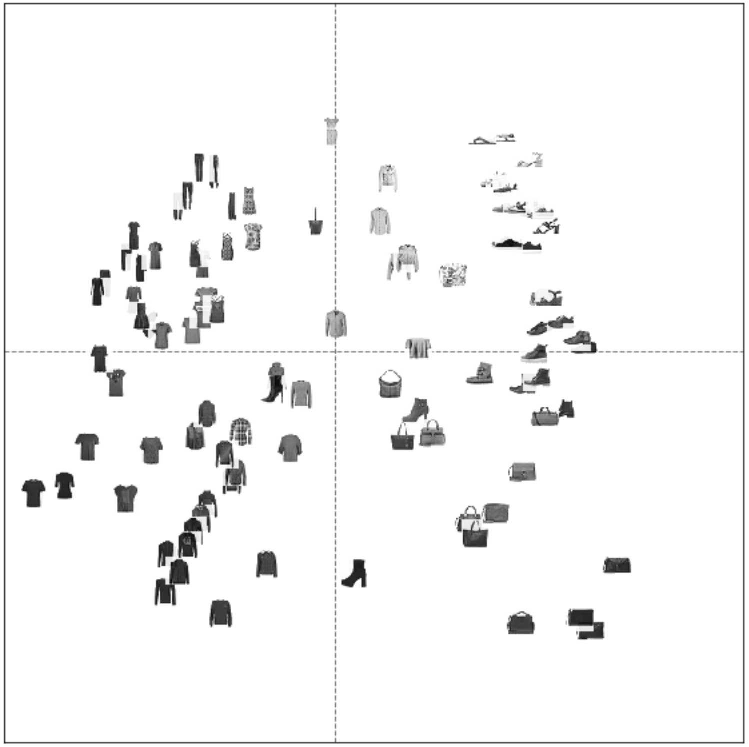
$$q_{ij} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_{k \neq l} (1 + ||y_k - y_l||^2)^{-1}}.$$

So we could transform the constraints into the scores:

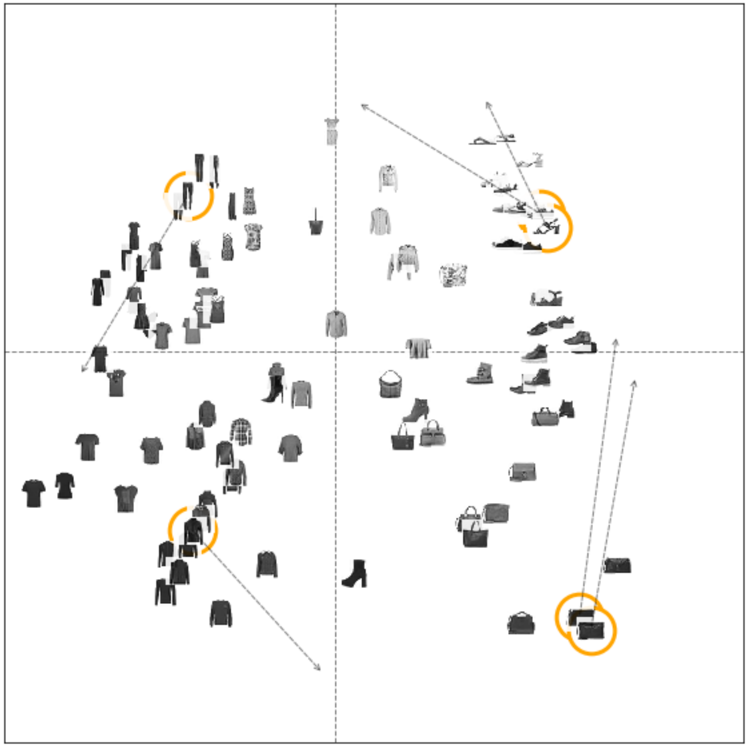
$$S_{\mathcal{M}} = \frac{1}{|\mathcal{M}|} \sum_{(i,j) \in \mathcal{M}} \log q_{ij}. \qquad S_{\mathcal{C}} = -\frac{1}{|\mathcal{C}|} \sum_{(i,j) \in \mathcal{C}} \log q_{ij}.$$

Interactive PPCA

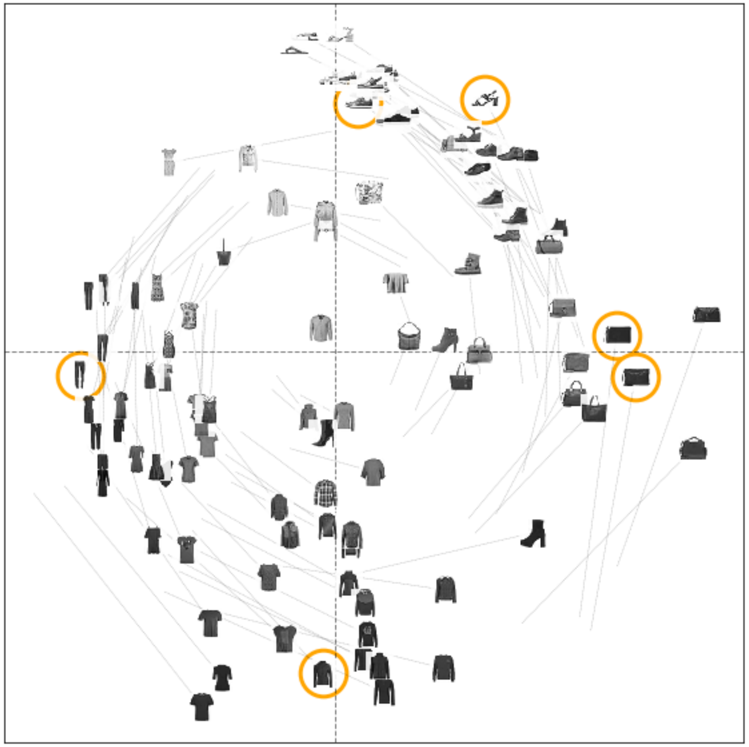
Initial visualization with the PPCA model.



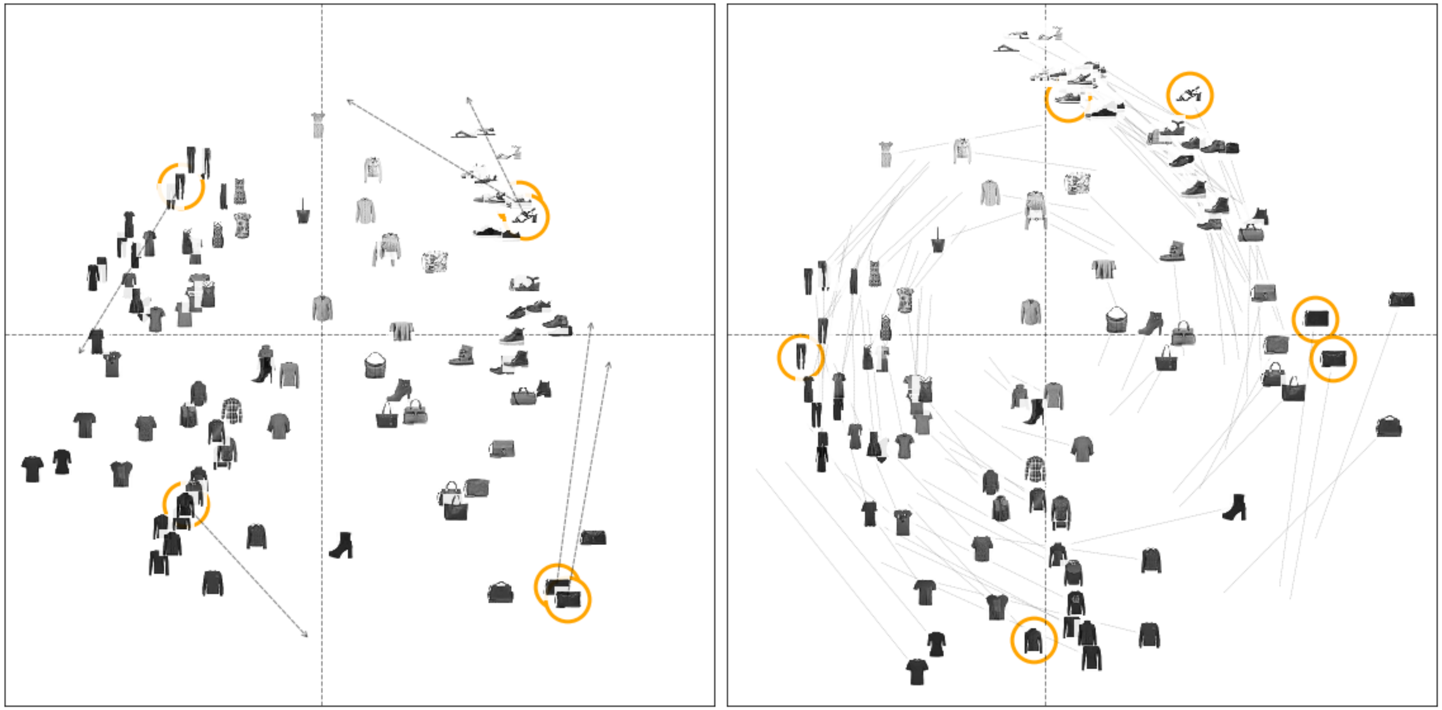
The user can manipulate the visualization by moving some points.



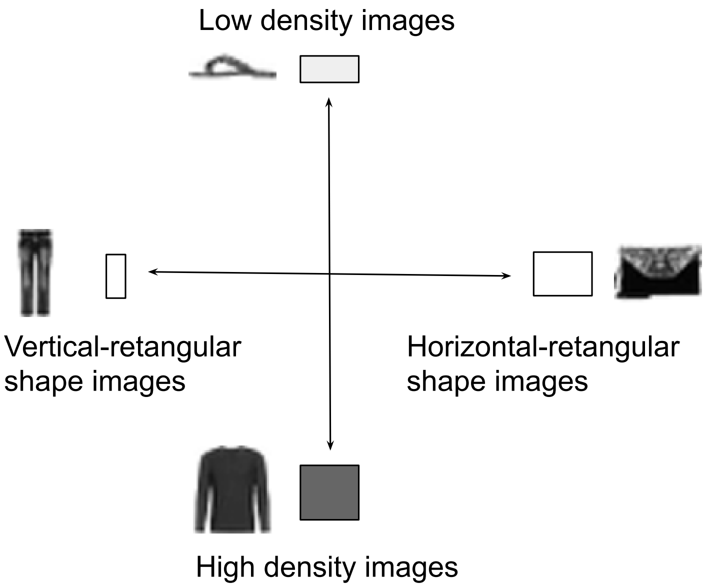
The result of the interactive model is explainable to the users.

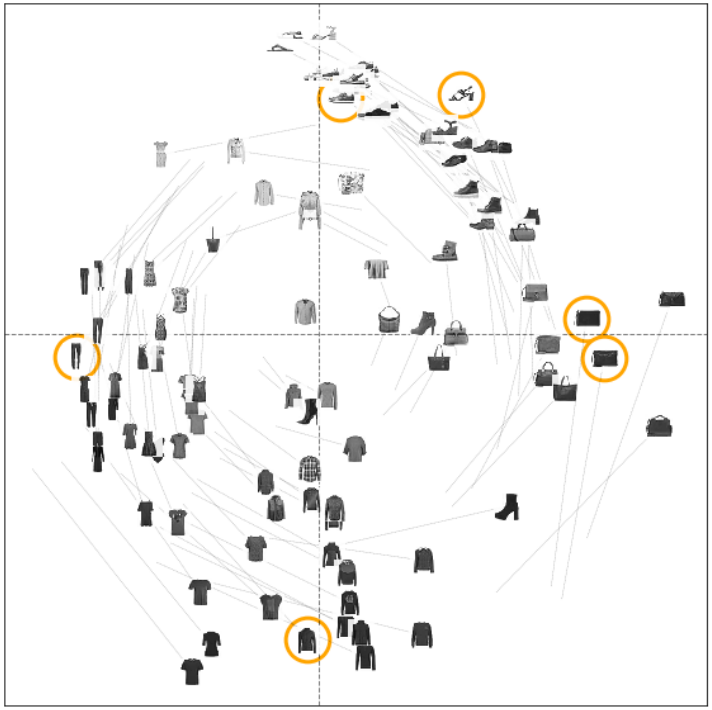


The result of the interactive model is explainable to the users.



Goal: reveal the pattern that is not clear.





+ The embedding in 2D space $\approx$ the latent variables

+ The position of the selected points is modeled directly into the prior distribution of the latent variables

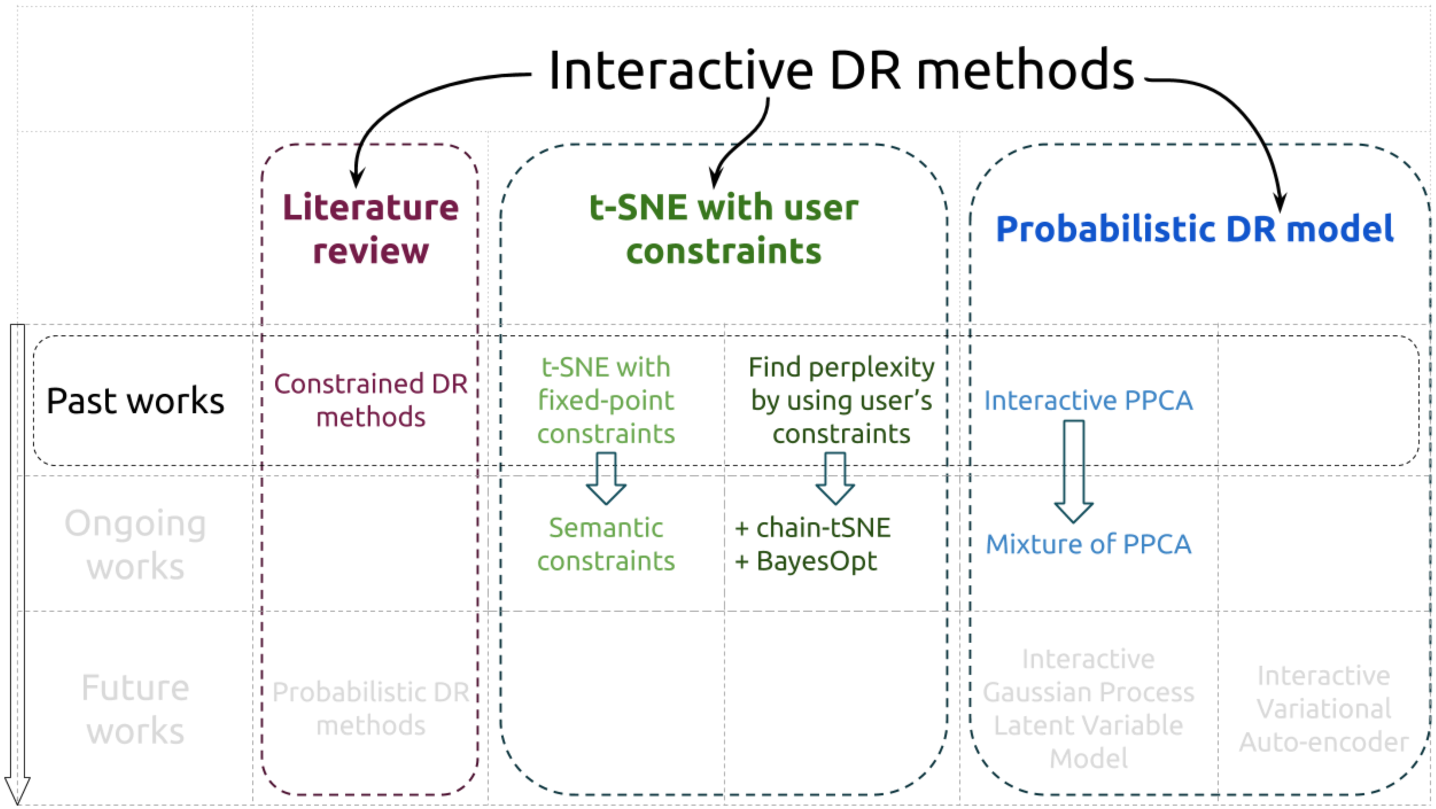
+ Can use the black-box inference toolbox to infer the latent variables

+ Usage: Communicate the result, make some patterns in the visualization easier to understand

---

`pyro, Tensorflow probability`

Summary the past works



Ongoing works with t-SNE

+ Integrate pairwise constraints instead of fixed-point constraints

+ Find best perplexity for t-SNE by using the user defined pairwise constraints

Two drawbacks of t-SNE

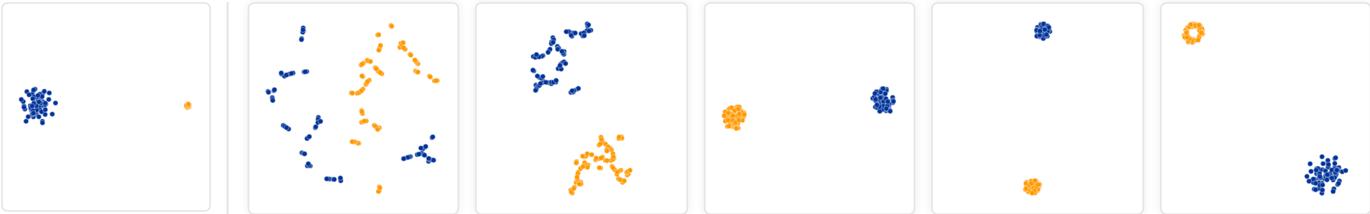
- Within-cluster distances
- Between-clusters distances

are not always well preserved in t-SNE

---

*Do we need a t-SNE embedding in which the **cluster sizes** and the **distances between clusters** are well preserved?*

Cluster sizes in a t-SNE plot mean nothing



Original

Perplexity: 2
Step: 5,000

Perplexity: 5
Step: 5,000

Perplexity: 30
Step: 5,000

Perplexity: 50
Step: 5,000

Perplexity: 100
Step: 5,000

Distances between clusters might not mean anything

Original

Perplexity: 2
Step: 5,000

Perplexity: 5
Step: 5,000

Perplexity: 30
Step: 5,000

Perplexity: 50
Step: 5,000

Perplexity: 100
Step: 5,000

Integrating pairwise constraints

Semantic Similarity

$$\sum_{x,x^+,x^-} \left[-\log\left(\frac{e^{f(x)^T f(x^+)}}{e^{f(x)^T f(x^+)} + e^{f(x)^T f(x^-)}} \right) \right]$$

$f(x)$ is a representation of a sample x,

x^+ and x^- are the corresponding *positive* and *negative* samples of x

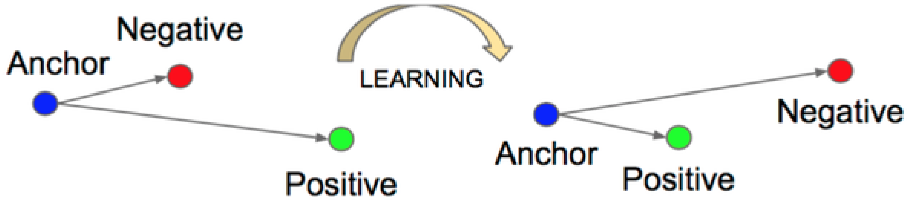
---

`An efficient framework for learning sentence representations [Logeswaran2018]`

`A Theoretical Analysis of Contrastive Unsupervised Representation Learning [Arora2019]`

Triplet constraints (*triplet loss* as a regularization term)

$$\sum_{x,x^+,x^-} \left(\| f(x) - f(x^+) \|^2 - \| f(x) - f(x^-) \|^2 + marg \right.$$



Facenet: A unified embedding for face recognition and clustering [Schroff2015]

Finding the best perplexity for t-SNE

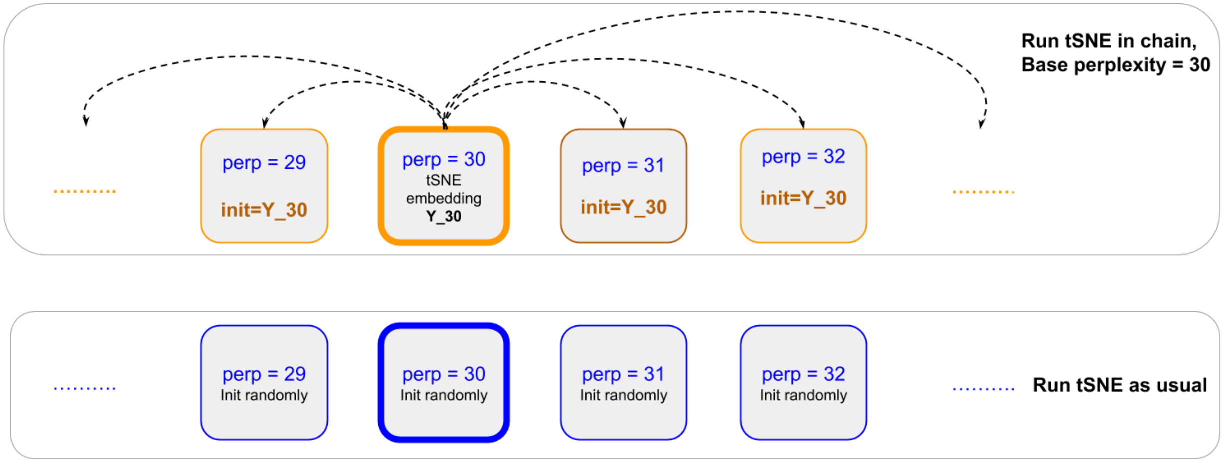
+ Proposed to use the constraint-preserving score to find the best perplexity

+ Should pre-calculate many t-SNE embeddings corresponding to different perplexity

+ Calculate the constraint scores for all embeddings and search for the perplexity that makes the constraint score maximum

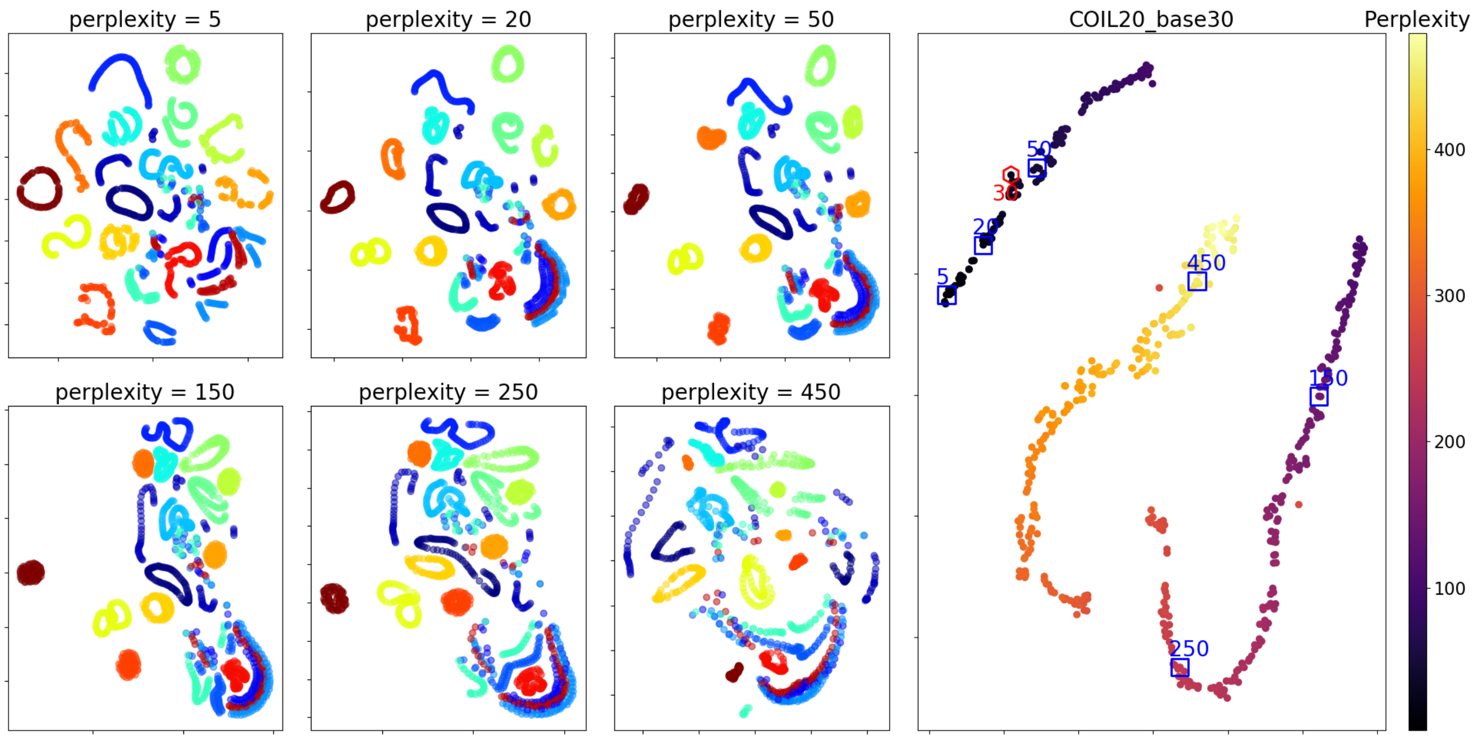
→ scalability problem

Proposed solution with *chain-tSNE*



~~Do not initialize randomly~~, Do initialize with a precalculated solution (of a base-perplexity)

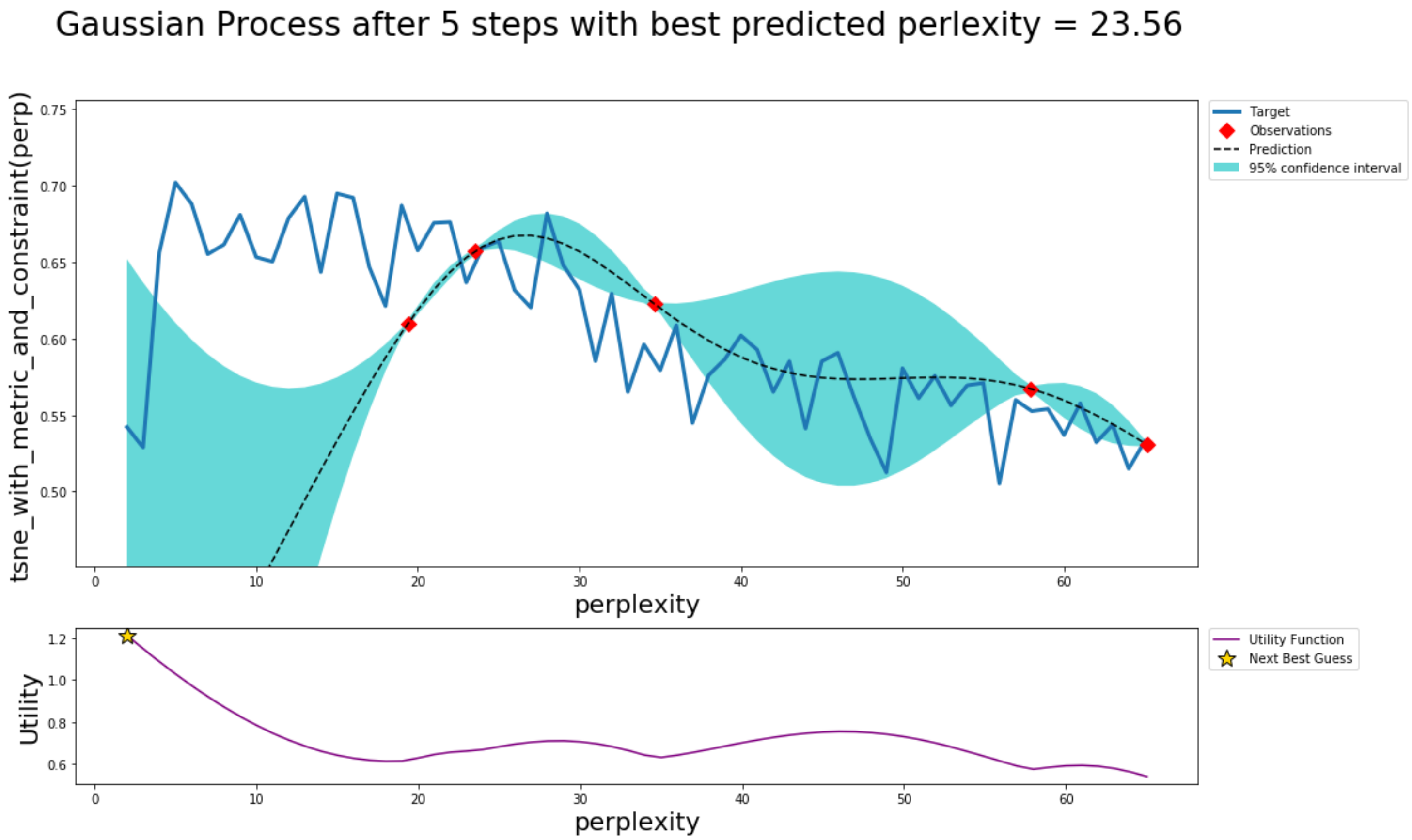
Proposed solution with *chain-tSNE*



Proposed solution with *Bayesian Optimization*

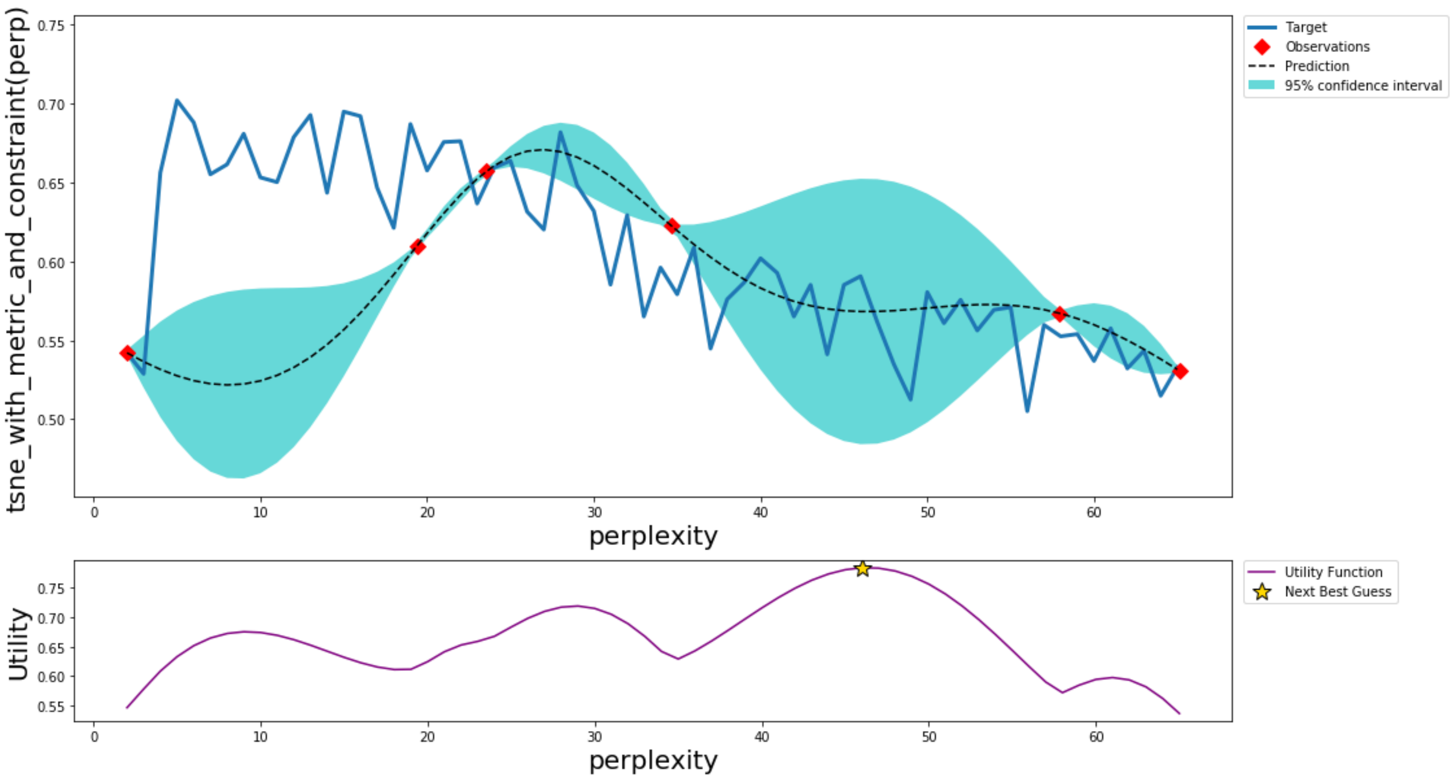
- Have calculated some initial embeddings
- What is the next perplexity we should try, ...
- ... in order to maximize the chance to approach the best perplexity?

Always using the *constraint preserving scores* as a criteria to find the next perplexity


Gaussian Process after 5 steps with best predicted perlexity = 23.56
tsne_with_metric_and_constraint(perp)
Target
Observations
Prediction
95% confidence interval
perplexity
Utility
Utility Function
Next Best Guess
perplexity

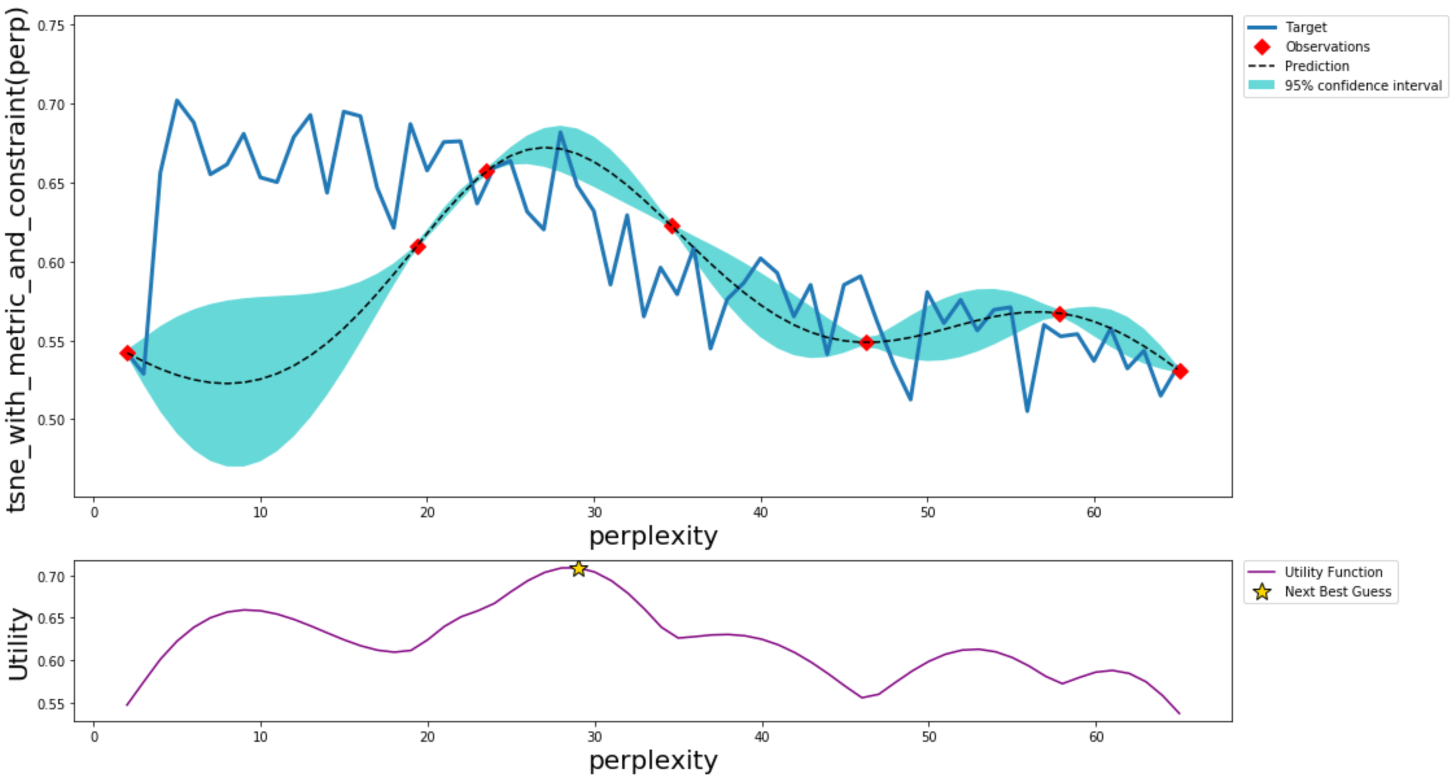
13.5

Gaussian Process after 6 steps with best predicted perlexity = 23.56



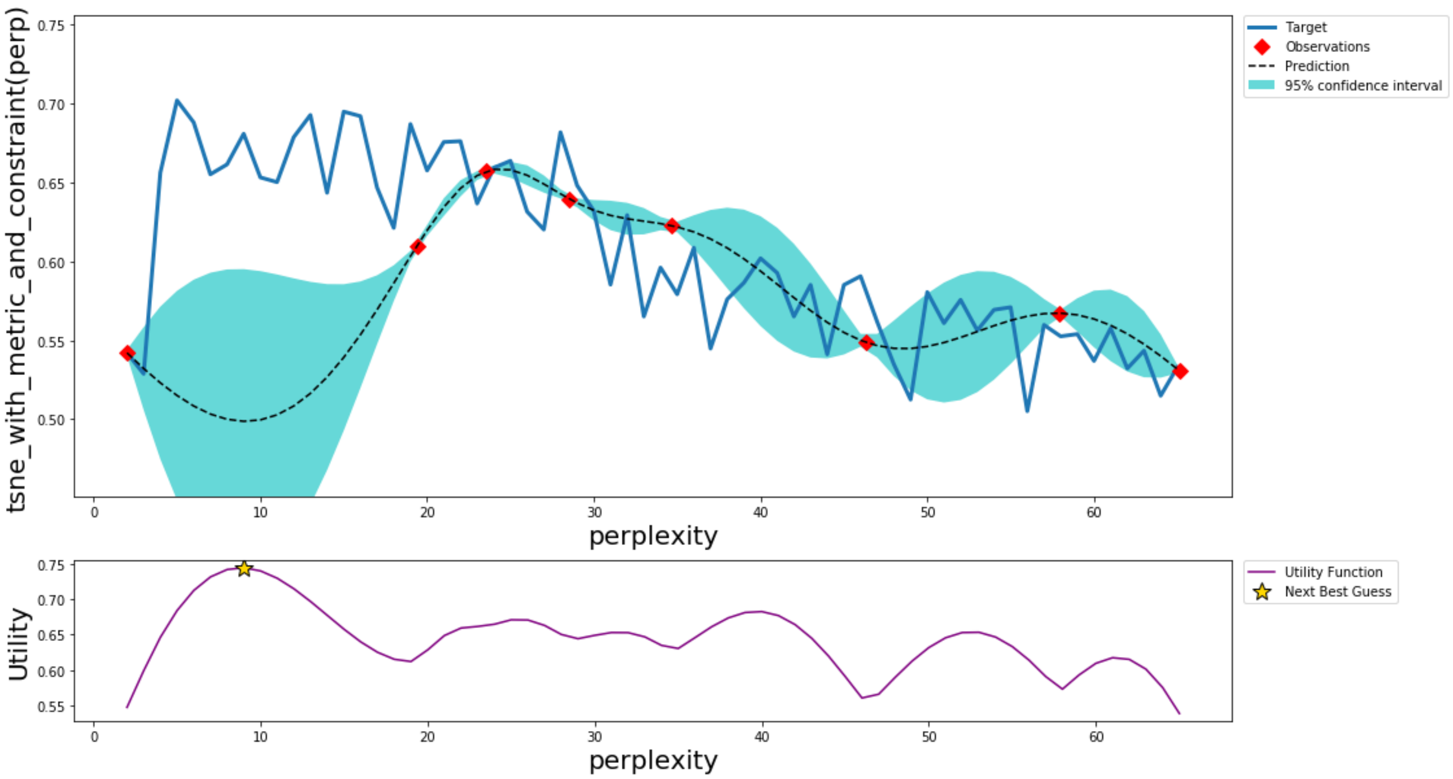
13.6

Gaussian Process after 7 steps with best predicted perlexity = 23.56

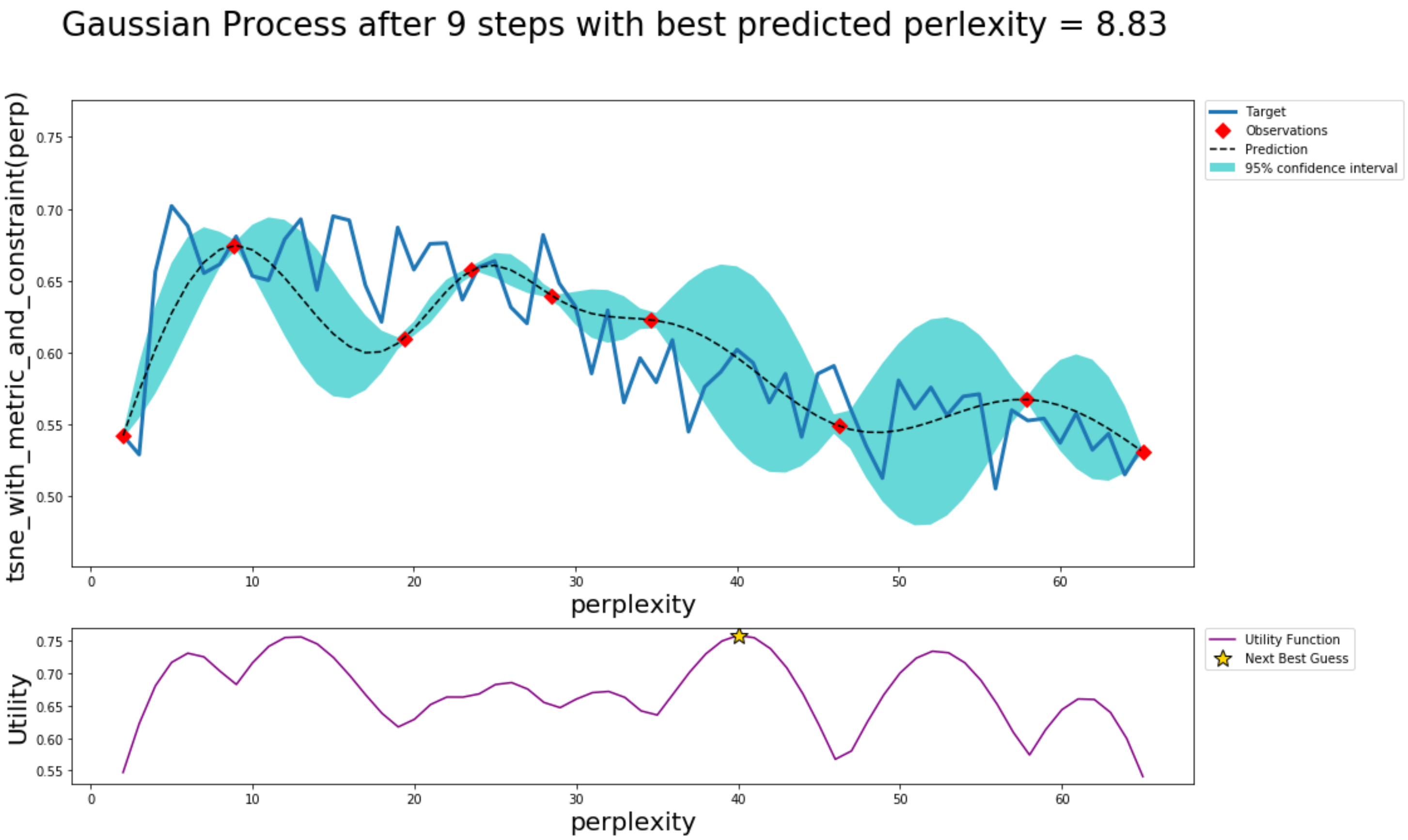


13.7

Gaussian Process after 8 steps with best predicted perlexity = 23.56



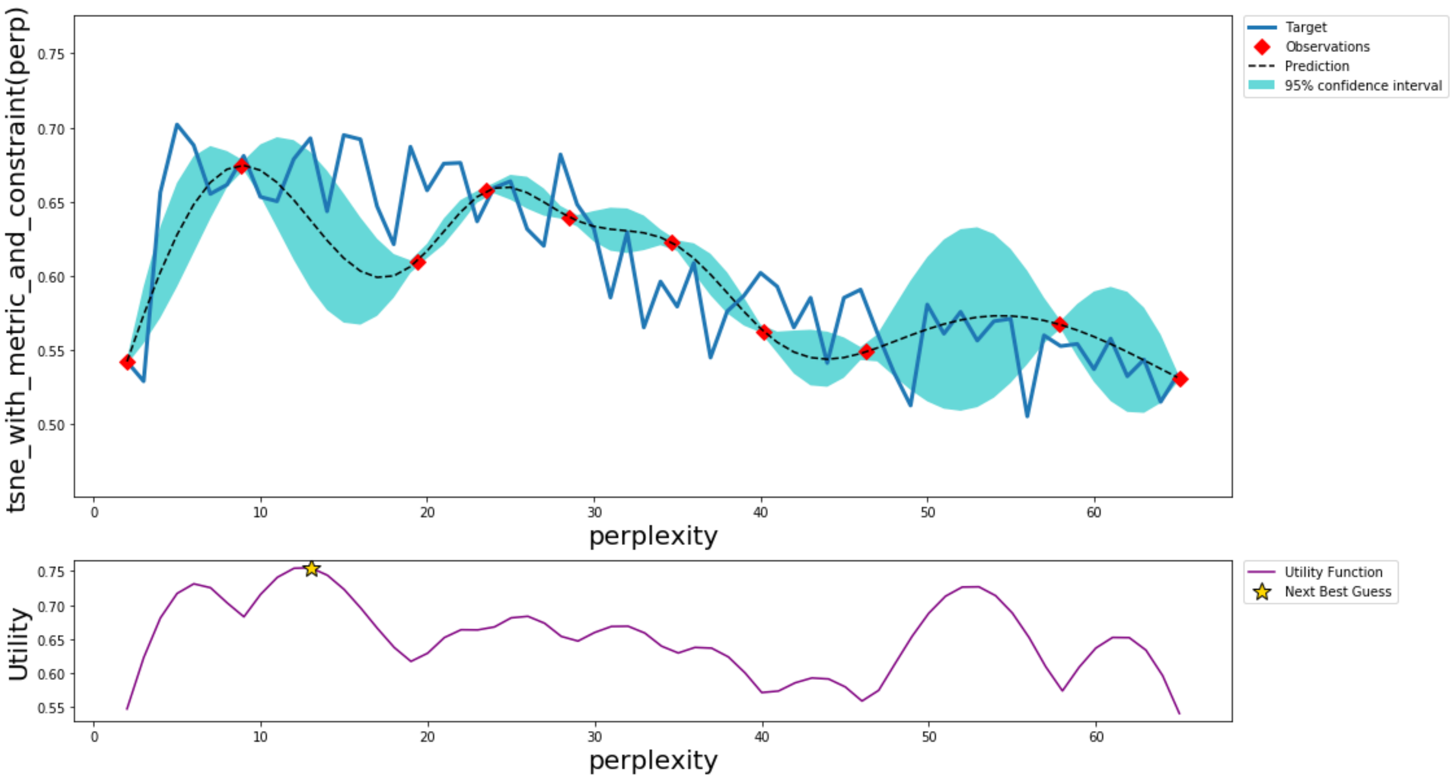
13.8



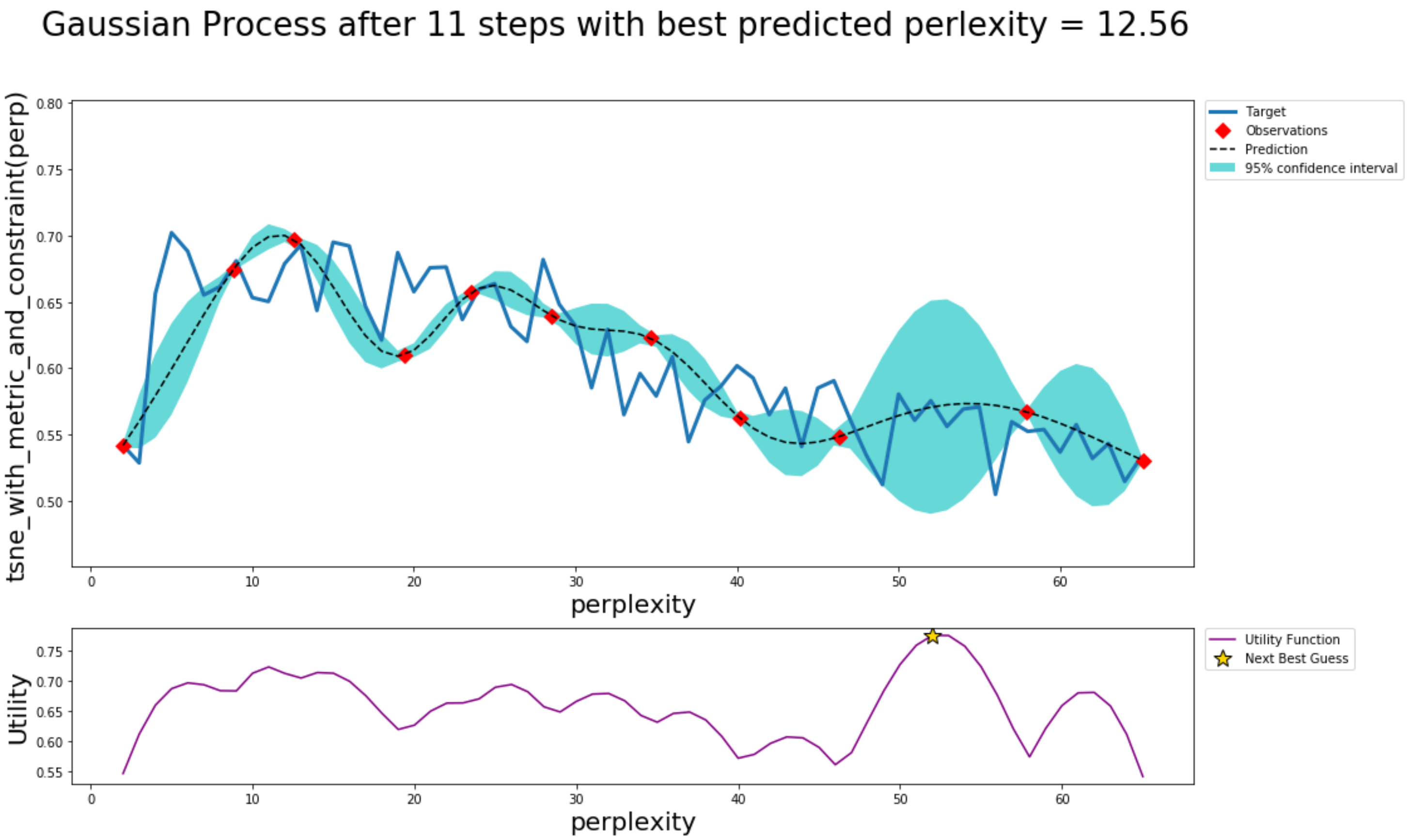
Gaussian Process after 9 steps with best predicted perlexity = 8.83
tsne_with_metric_and_constraint(perp)
Target
Observations
Prediction
95% confidence interval
perplexity
Utility
Utility Function
Next Best Guess
perplexity

13.9

Gaussian Process after 10 steps with best predicted perlexity = 8.83

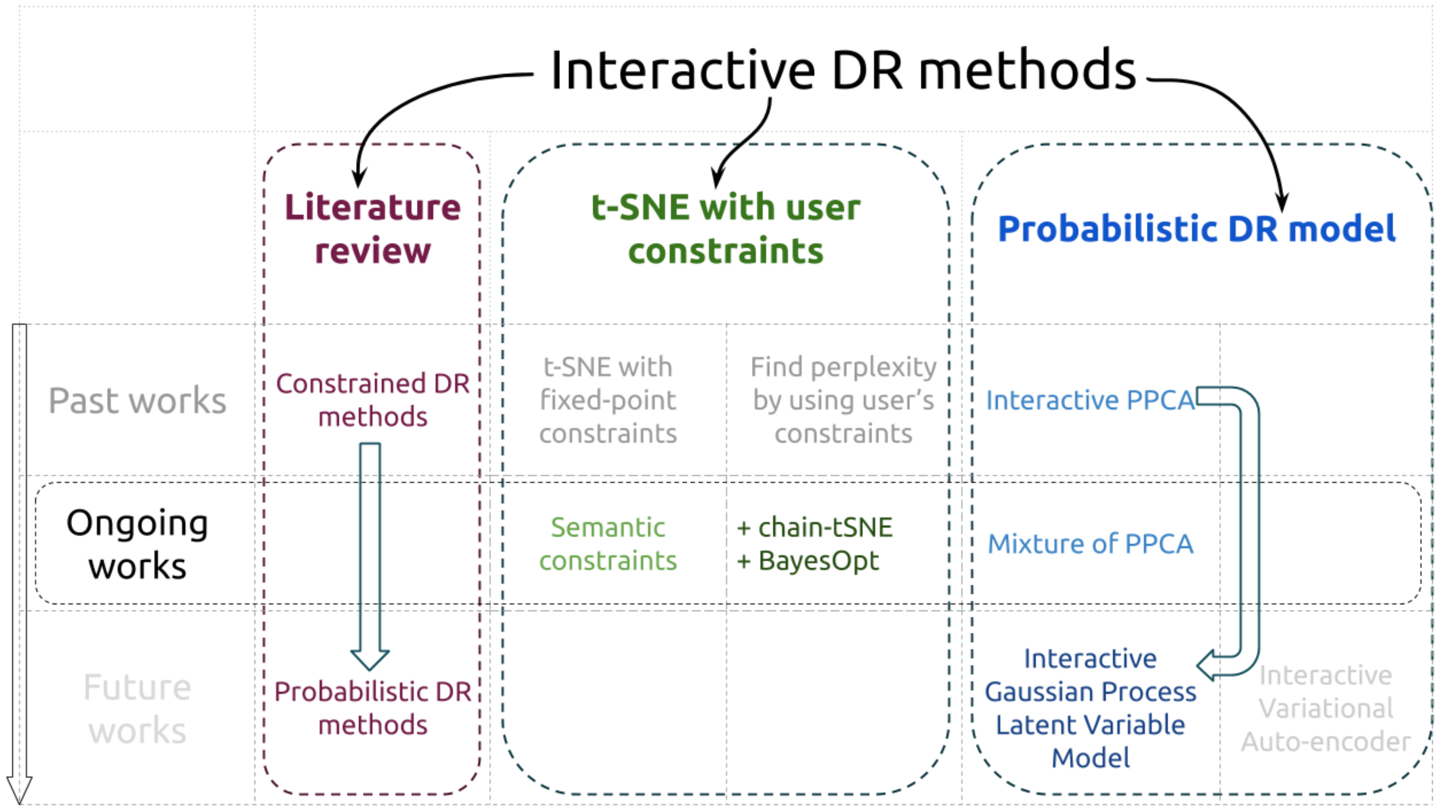


13.10

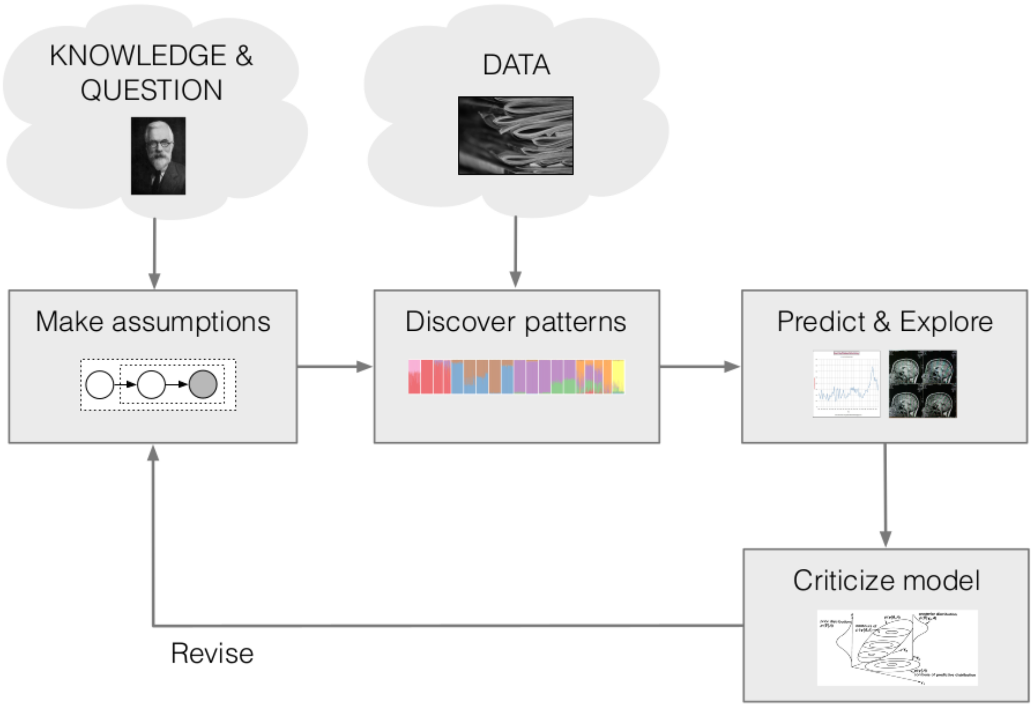


Gaussian Process after 11 steps with best predicted perlexity = 12.56
tsne_with_metric_and_constraint(perp)
perplexity
Target
Observations
Prediction
95% confidence interval
Utility
perplexity
Utility Function
Next Best Guess

Summary the ongoing works with t-SNE



Future works: probabilistic DR models



Variational Inference: Foundations and Modern Methods [Blei2016]

- Integrate user's constraints into Gaussian Process models
- Modify the structure of the latent space or the generated samples with Variational Auto-encoder (VAE)
- Understand the latent representation of the probabilistic generative model by visualizing their features in the network

Gaussian Process Latent Variable Model

PPCA $\leftrightarrow$ Dual PPCA $\rightarrow$ GP-LVM

- We would learn a mapping (a function) to map the data from HD to LD (or vice versa)
- There exist many mapping functions
- Gaussian Process allows us to express a distribution over these functions
- and help us to find the most promising function that represent well the data

---

Gaussian process latent variable models for visualisation of high dimensional data [Lawrence2004]

Probabilistic non-linear principal component analysis with Gaussian process latent variable models [Lawrence2005]

Why GP-LVM? An example

Data

- Real-time / quantitative polymerase chain reaction (Real-Time PCR) data
- 48 genes obtained from mice
- Data values are measurements of each gene ...
- ... at various stages of development: from the 1-cell stage to the 64-cell stage

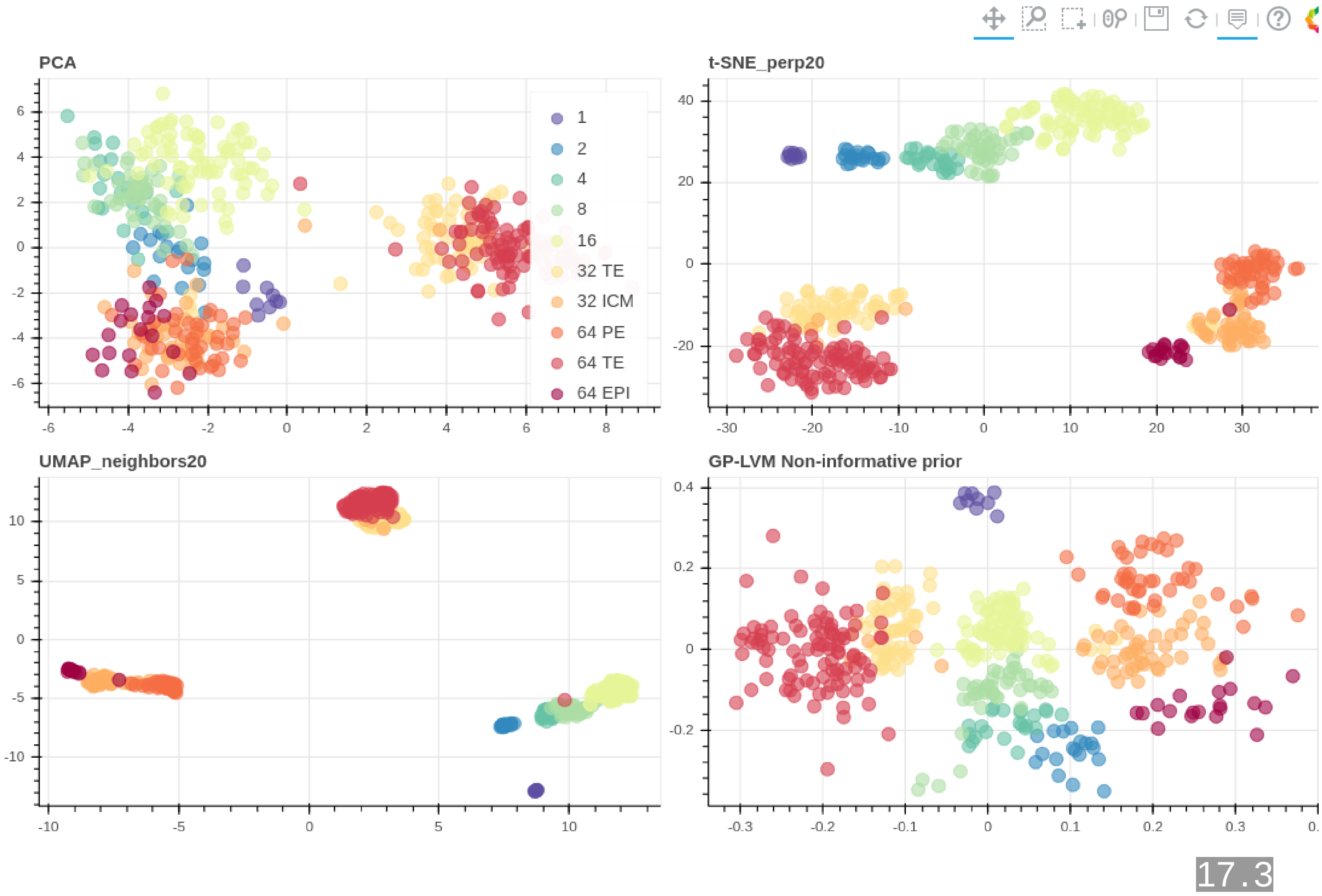
`Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst.` [Guo2010]

Data

```
Stage labels: ['1', '2', '4', '8', '16', '32 TE', '32 ICM', '64 PE', '64 TE', '64 EPI']

qPCR data:

             Actb      Ahcy      Aqp3    ...      Tcf23      Utf1    Tspan8
    '1'  0.541050 -1.203007  1.030746    ...   0.942981  1.348892 -1.051999
'32 TE'  0.680832 -1.355306  2.456375    ...   1.064399  1.469397 -0.996275
    '2'  1.056038 -1.280447  2.046133    ...   1.211529  1.615421 -0.651393
'64 PE'  0.732331 -1.326911  2.464234    ...   1.071541  1.476485 -0.699586
   '16'  0.629333 -1.244308  1.316815    ...   1.114394  1.519017 -0.798985
...
```


PCA
t-SNE_perp20
UMAP_neighbors20
GP-LVM Non-informative prior
1
2
4
8
16
32 TE
32 ICM
64 PE
64 TE
64 EPI
17.3

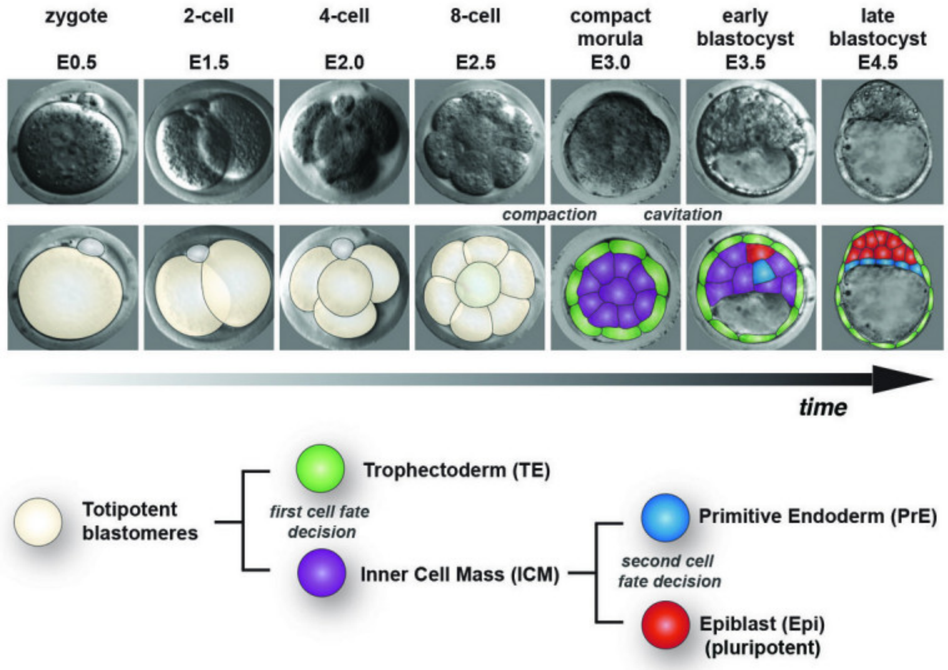
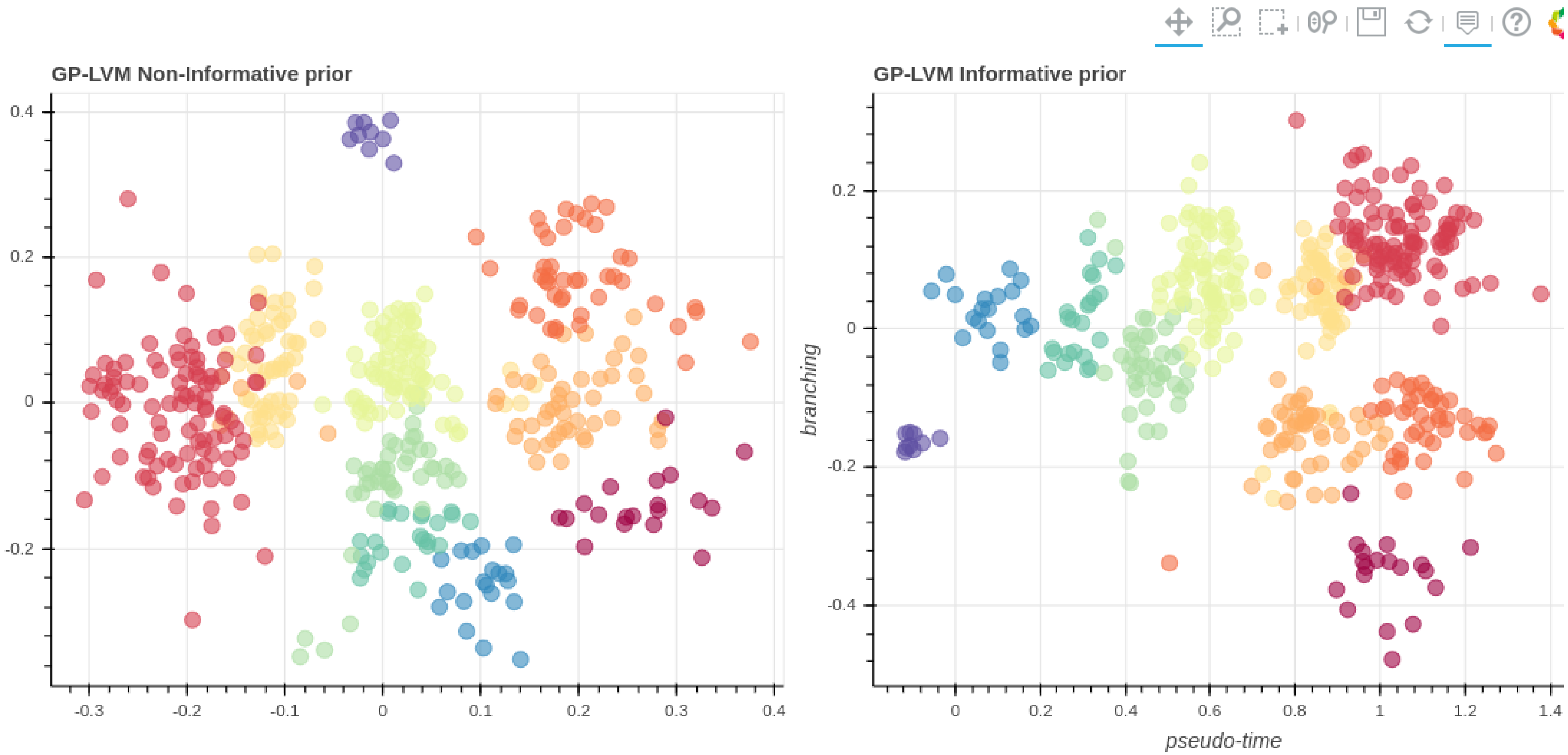


Image credit: Formation of the blastocyst during early mammalian embryonic development. [url]



Reproduce from: GrandPrix scaling up the Bayesian GPLVM for single-cell data [Ahmed2019], with pyro tutorial on GP-LVM [url]

Variational Auto-encoder (VAE)

- The encoder and decoder can be neural networks
- Learn the representation of the data in the latent space
- Has an explicit mapping for the new examples

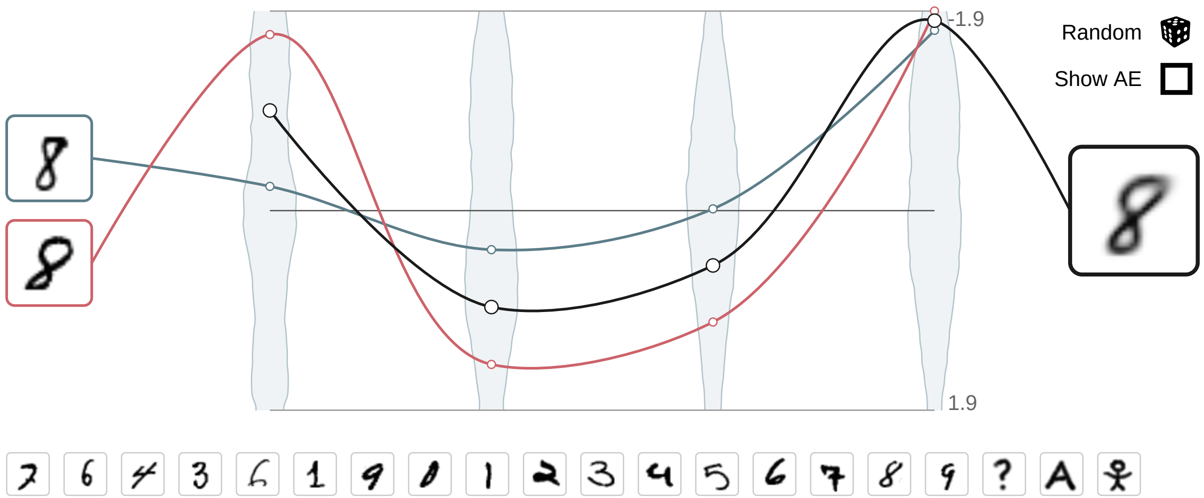
User's constraints to

- Modify the representation of the latent space
- Control how the new sample is generated

Semi-supervised learning with deep generative models [Kingma2014]

Learning structured output representation using deep conditional generative models [Sohn2015]

Credit: VISxAI 2018: Towards an Interpretable Latent Space, Thilo Spinner, Jonas Körner, Jochen Görtler and Oliver Deussen



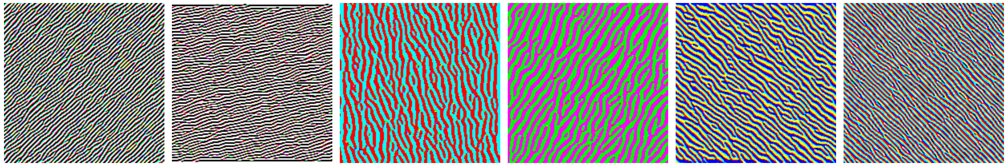
Feature viz. for a generative model

E.g., visualize the latent representation of VAE

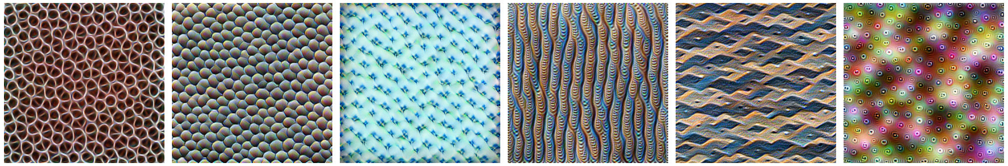
Goal: understand the latent structure, the contribution of the latent features to the output generated sample.

Feature Visualization

How neural networks build up their understanding of images



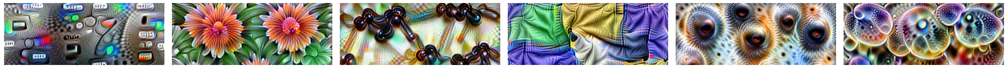
Edges (layer conv2d0)



Textures (layer mixed3a)

Patterns (layer mixed4a)

19.2



Summary the future works

- Integrate user's constraints into Gaussian Process and VAE models
- More general, build a unifying framework that can represent different kinds of user's constraints for the probabilistic DR models
- Visualizing the internal features of a generative model

Focus only on the modeling problem, not inference problem

Bayesian visual analytics: Bava [House2015]

Other discrete ideas

Uncertainty Visualization: Incorporating an indication of uncertainty into visual representation

---

References:

- *Visualization of uncertainty* vs *Uncertainty of the visualization*

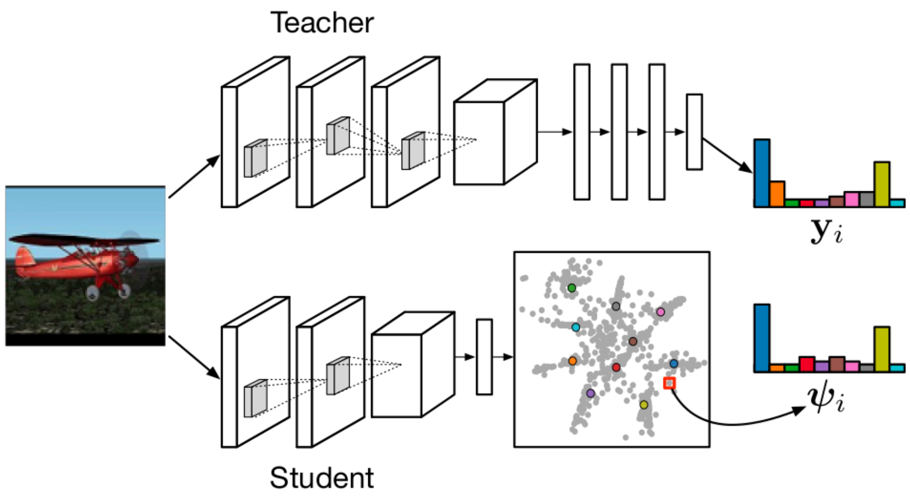
  A review of uncertainty in data visualization. Expanding the Frontiers of Visual Analytics and Visualization [Brodlie2012]

- Visualising uncertainty in dropout Bayesian neural networks, e.g., the dropout masks and the dropout probability

  Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning

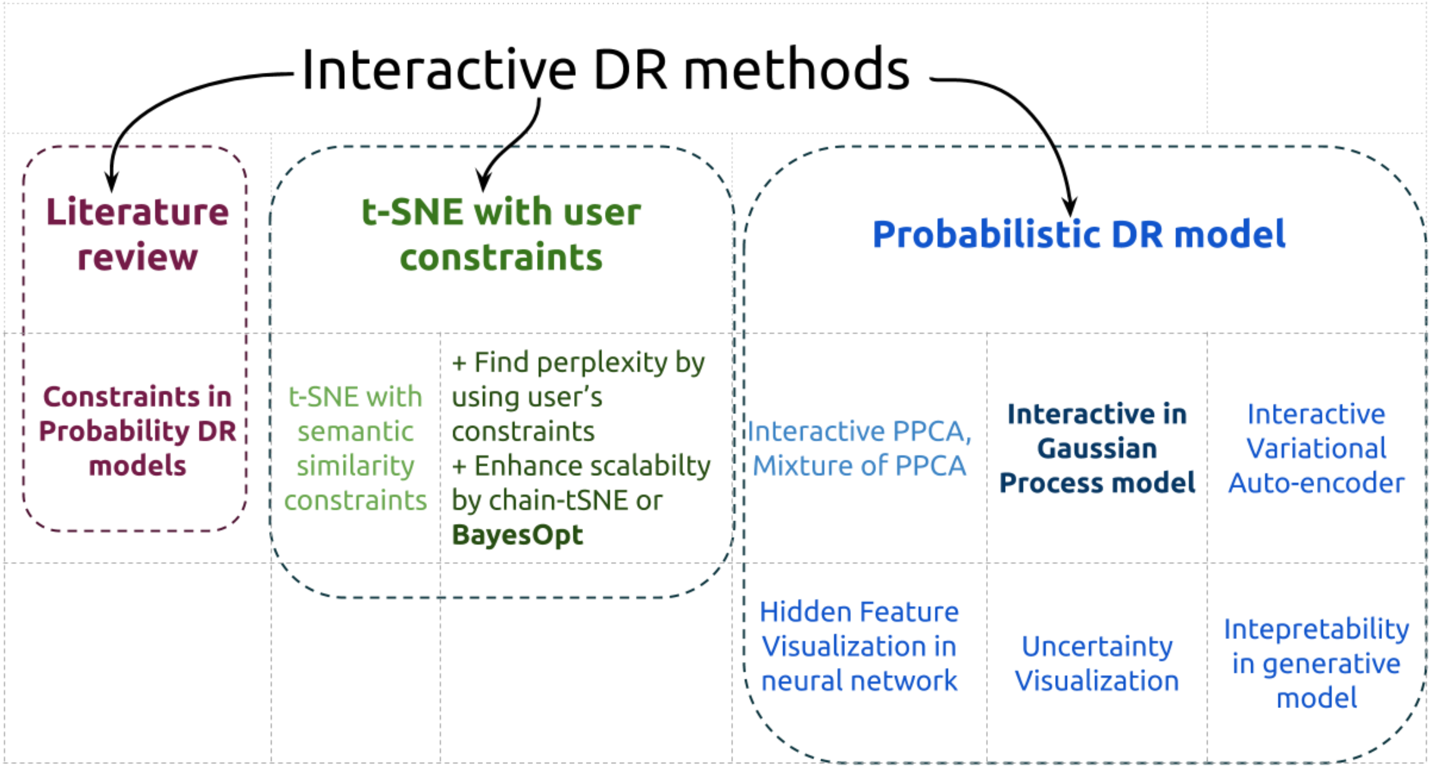
Other discrete ideas

Enhance the visualization by pre-trained models



Dimensionality Reduction for Representing the Knowledge of Probabilistic Models [Law2019]

Recap: Some ideas we can work on



Thank you!

Slide configs

Black (default) - White - League - Sky - Beige - Simple
Serif - Blood - Night - Moon - Solarized