Machine Learning research

School of Information and Communication Technology
Hanoi University of Science and Technology

2021

Key members

Khoat Than
SOICT, HUST,
khoattq@soict.hust.edu.vn
https://users.soict.hust.edu.vn/khoattq/

Ngo Van Linh
SOICT, HUST,
linhnv@soict.hust.edu.vn
https://users.soict.hust.edu.vn/linhnv/

- **30+** research students

Recent research directions

- Continual learning
- Generative models
- Self-supervised learning
- Representation learning

Some recent sponsors

NAFOSTED
National Foundation
for Science & Technology Development

BỘ KHOA HỌC VÀ CÔNG NGHỆ

ASIAN OFFICE OF AEROSPACE R&D
BRIDGING SCIENCE & TECHNOLOGY

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
UNITED STATES AIR FORCE

Office of Naval Research
ONR
Science & Technology

VINGROUP
Innovation Foundation

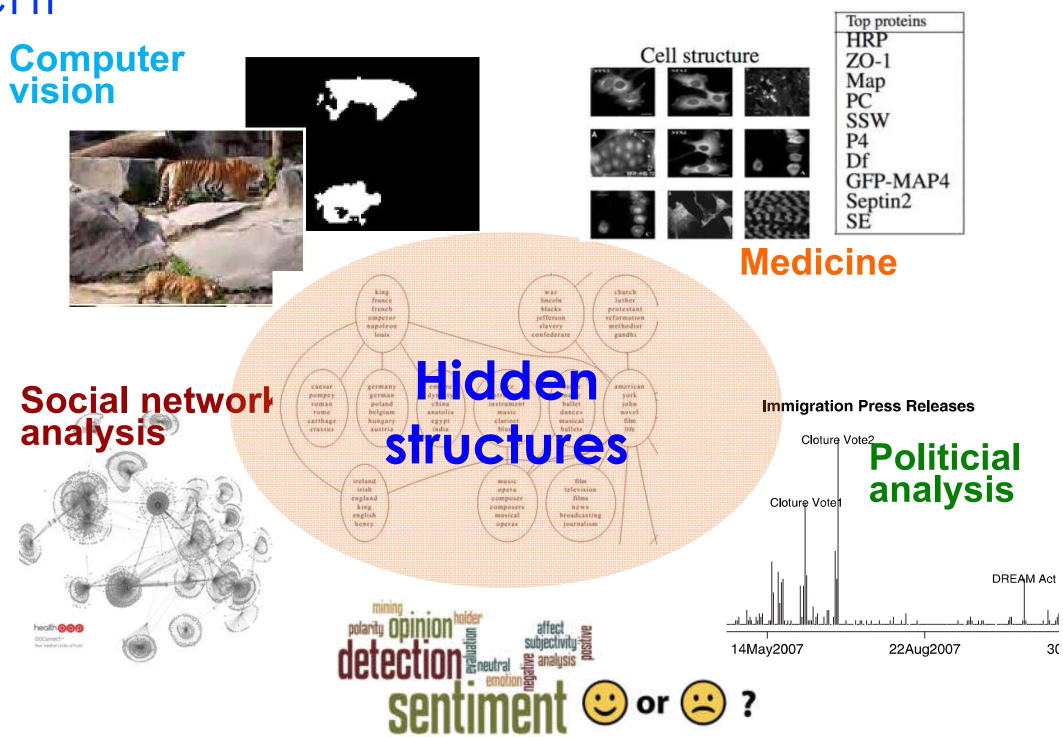
U.S. ARMY
RDECOM

Some projects

- VINIF (2019-2022, VN):
  - *Director:* Khoat Than
  - *Funding agency:* Vingroup Innovation Foundation
  - Area: Machine learning, Big data
- ONRG (2018-2020, USA):
  - *Director:* Khoat Than
  - *Funding agency:* US Navy & Air Force Office of Scientific Research
  - Area: Representation learning, Big data
- NAFOSTED (2015-2017, VN):
  - *Director:* Khoat Than
  - *Funding agency:* National Foundation for Science and Technology Development
  - Area: Machine learning, Big data
- AFOSR (2015-2017, USA):
  - *Director:* Khoat Than
  - *Funding agency:* Air Force Office of Scientific Research, and U.S. Army International Technology Center, Pacific
  - Area: Machine learning, Big data

ML: Inference & Learning

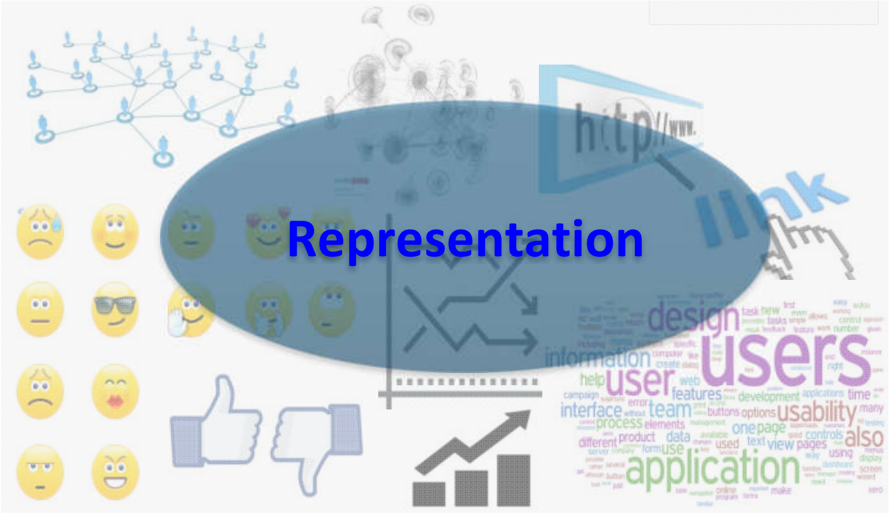
- Background: Big data are hard to interpret. Recovering **hidden semantics/structures** is beneficial, but very challenging.
- **Objective:** Develop efficient methods for analyzing the hidden structures from **big/streaming** text collections



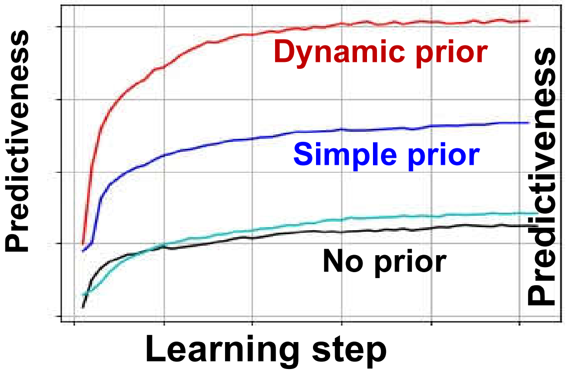
- Tools:
  - Topic models
  - Matrix factorization
  - Deep neural networks
- Research projects: NAFOSTED (VN), AFOSR (USA, 2015-2017)

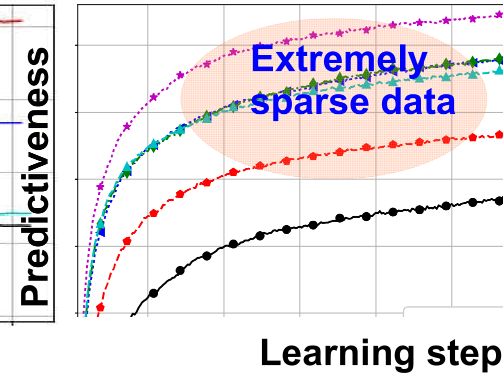
ML: Knowledge representation

- Background: **Semantic representation** is always challenging
- ***Objective:*** *Develop methods for learning efficient representations of the hidden semantics from big/streaming data*



- Tools:
  - Topic models + Neural nets (extracting hidden semantics)
  - Manifold learning (analyzing local structures)
  - Human knowledge as prior (providing supervision)



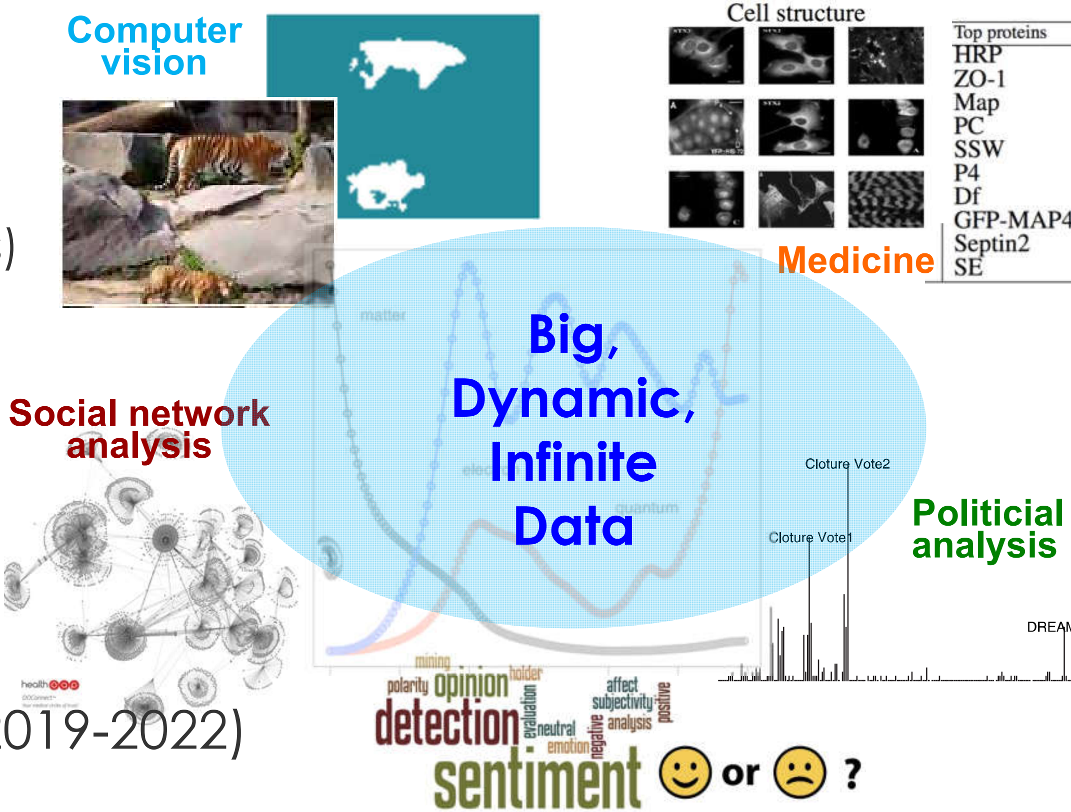


25% improvement in predictiveness

- Research projects: **ONRG** (USA, 2018-2020)

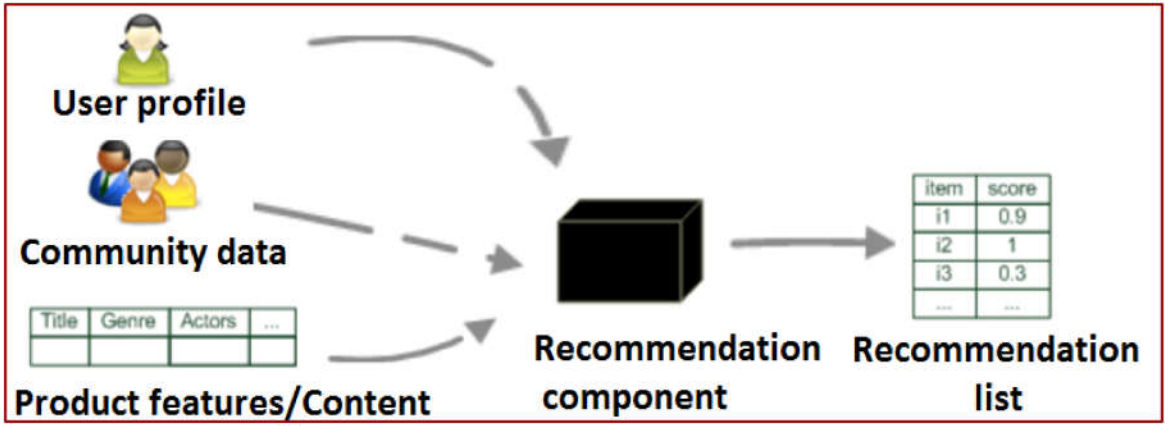
ML: Human knowledge + Learning

- Background: Integrating human knowledge into machine learning models is always beneficial but challenging.
- **Objective:** Develop efficient methods for learning from big or streaming data that can exploit human knowledge
- Tools:
  - **Topic models + Neural nets** (extracting hidden semantics)
  - **Human knowledge as prior** (providing supervision)

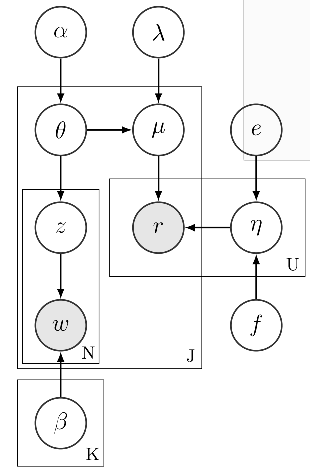


- Research projects: **VINIF** (2019-2022)

Application: Recommender systems



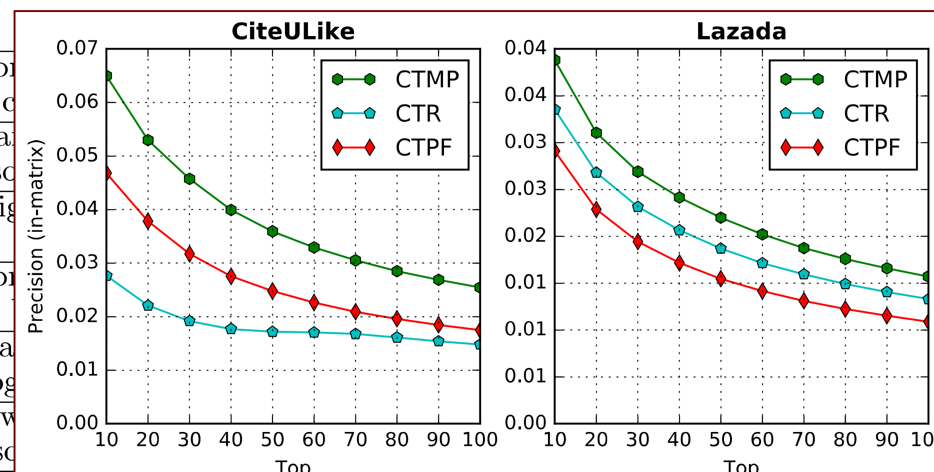
Hybrid contents and ratings/feedbacks



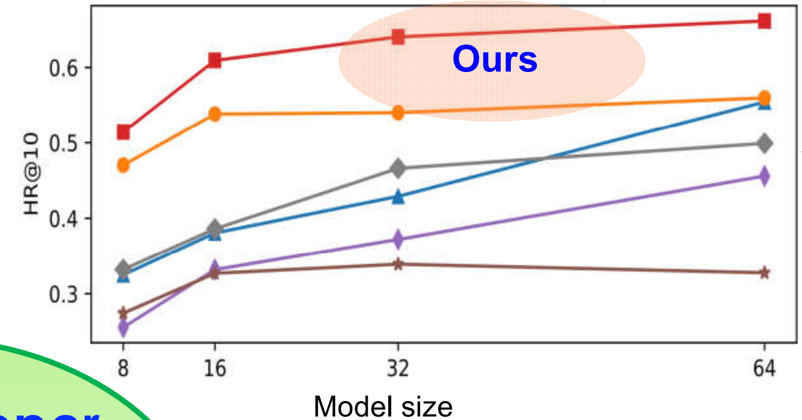
CTMP **(Our model)**

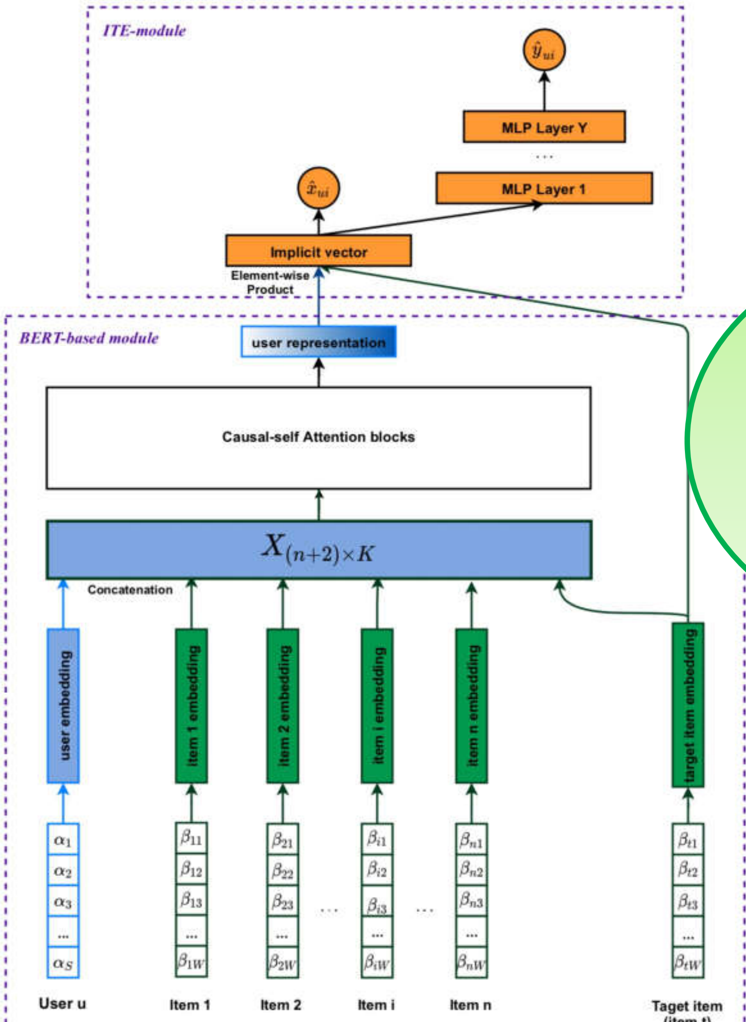
User	Profile
1	small, compact, design, charge, accessories, battery, device, pr black, sound, phone, camera, material, usage, china, laptop, c
	child, diaper, infant, tool, toy, children, hygiene, for, child, sa patch, absorptive, package, produce, pant, velcro_diaper, abso
2	wood, interior, houses, material, life, goods, room, chair, desig for, produce, vietnam, shelf, guest, sofa, bed, luxurious, size
	design, material, luxurious, fashionable, phone, usage, safe, pr goods, accessories, machine, leather, houses, black, feminine,
3	TV, LED, HD, sound, monitor, black, video, model, full, sma product, resolution, deliver ... free, inch, technolog
	machine, household_pro ... r, fridge, cold, w air_conditioner, blender ... g, produce, abso

Users' interests



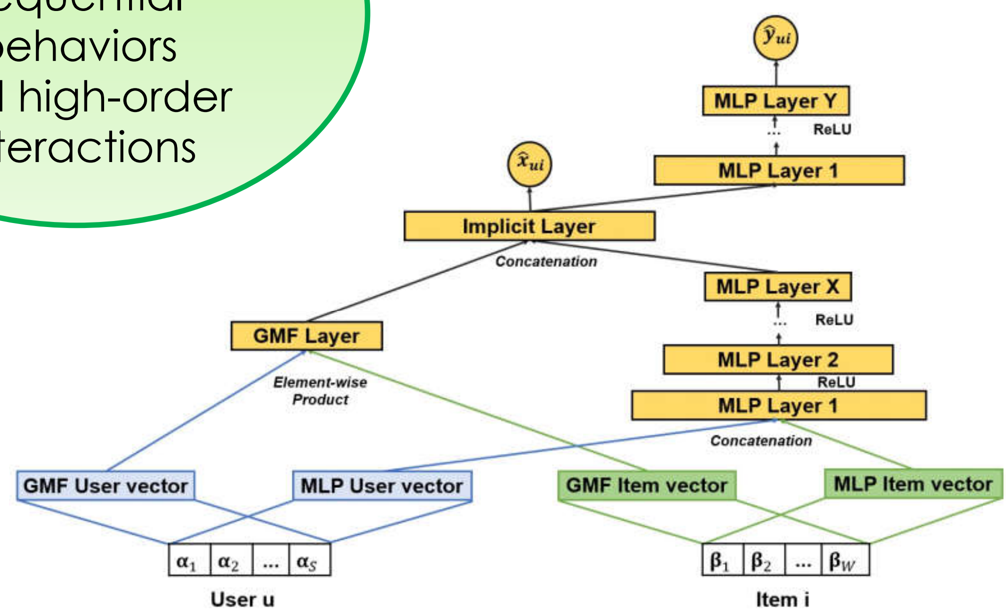
Application: Recommender systems





Going deeper

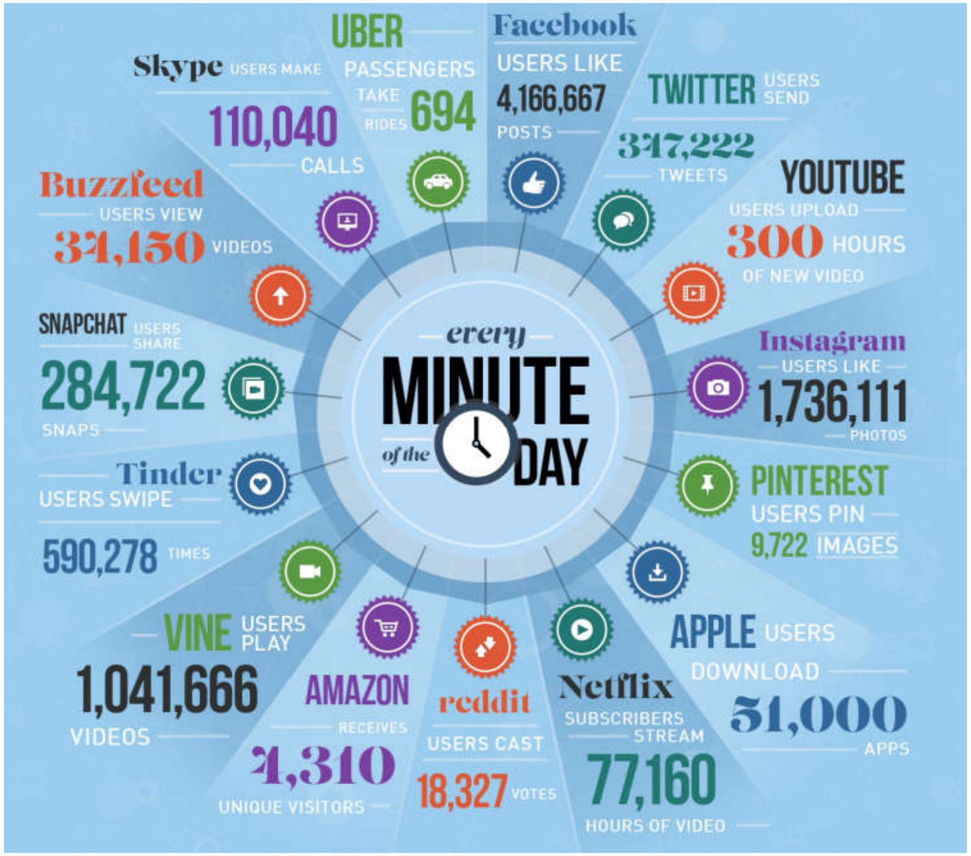
Modeling sequential behaviors and high-order interactions

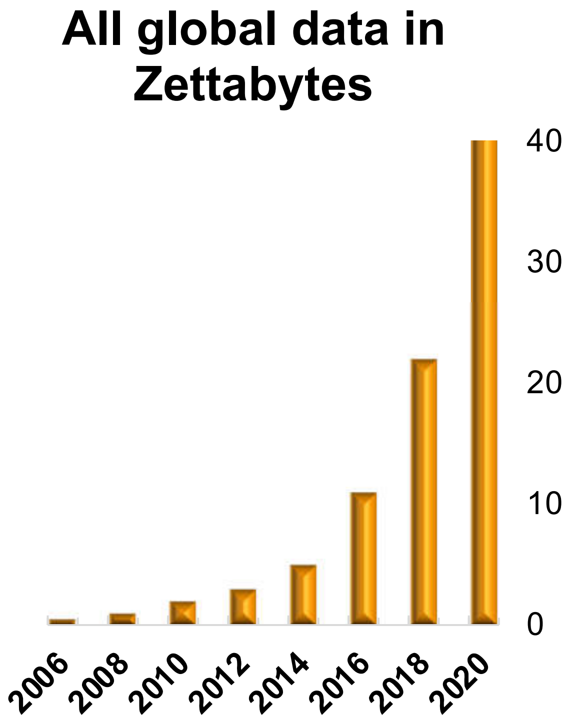


A recent research line

Question

- How to train a machine from an **infinite** sequence of data?
  - Learning from a data stream





https://www.umanis.com/lesechosdeladata/avis-dexpert/digital-listening-social-media/attachment/data-digital-social-media-3/?thc-month=201606

Applications

- Online services

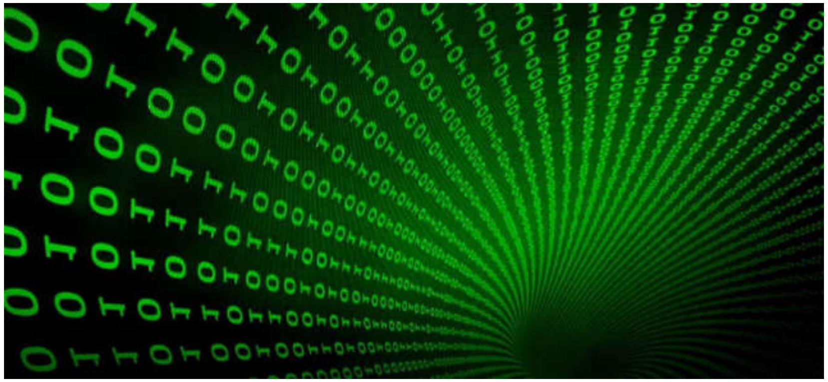
amazon

"The company reported a **29% sales increase** to \$12.83 billion during its second fiscal quarter, up from \$9.9 billion during the same time last year."
– Fortune, July 30, 2012

- Lifelong learning

Question: challenges

- Learning from infinitely many observations (data)



- The dynamic nature

- The NP-hardness of inference

- ...

Catastrophic forgetting

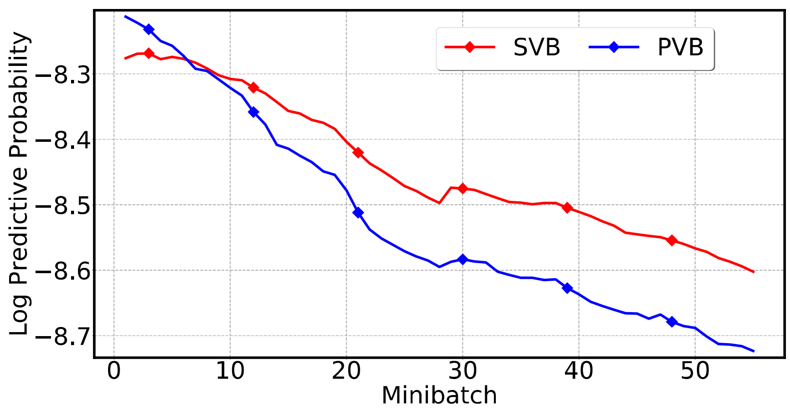
- Forget the tasks that have learned before
- This problem appears in most learning systems, but
  - Most existing focuses on neural networks [Parisi et al., 2019; French, 1999; Mermillod et al., 2013]
  - Lack of theory: forgetting rate, …
- Our finding: after training *b* minibatches

$$\|\xi_b - \xi_0\|_1 = \left\|\hat{\xi}_b + \cdots + \hat{\xi}_1\right\|_1 \geq b$$

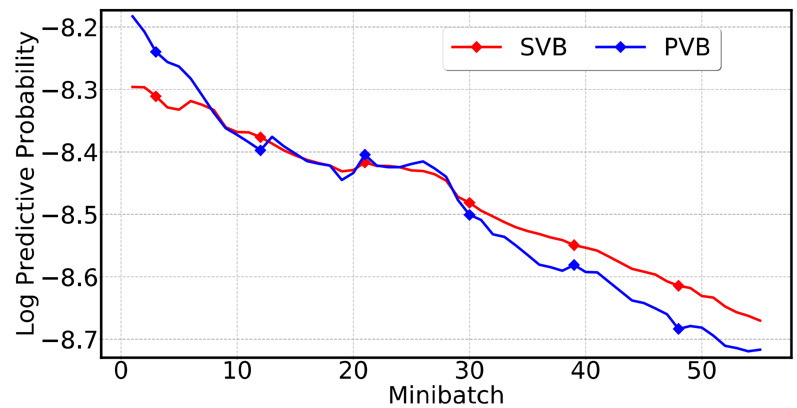
  - The knowledge ξ_0 will quickly become *negligible*
  - Forgetting rate: $\mathcal{O}(b^{-1})$ → Much faster than human: $\mathcal{O}(b^{-0.67})$

Bach, Tran Xuan, Nguyen Duc Anh, Linh Ngo Van, Khoat Than. "Dynamic transformation of prior knowledge into Bayesian models for data streams." Conditionally accepted, *IEEE Transactions on Knowledge and Data Engineering*, 2021.

Catastrophic forgetting: example



(a) Rec.autos



(b) Rec.sport.hockey

The accuracy of prediction about the first learned concept (*Rec.autos* for (a), *Rec.sport.hockey* for (b)) decreases quickly as learning from more new concepts. (Higher is better)

Stability-Plasticity dilemma

- Stability: should not forget what has been learned
- Plasticity: *quickly adapt with new changes* *[Mermillod et al., 2013]*
- **How to balance efficiently?**

Old knowledge <=> New knowledge

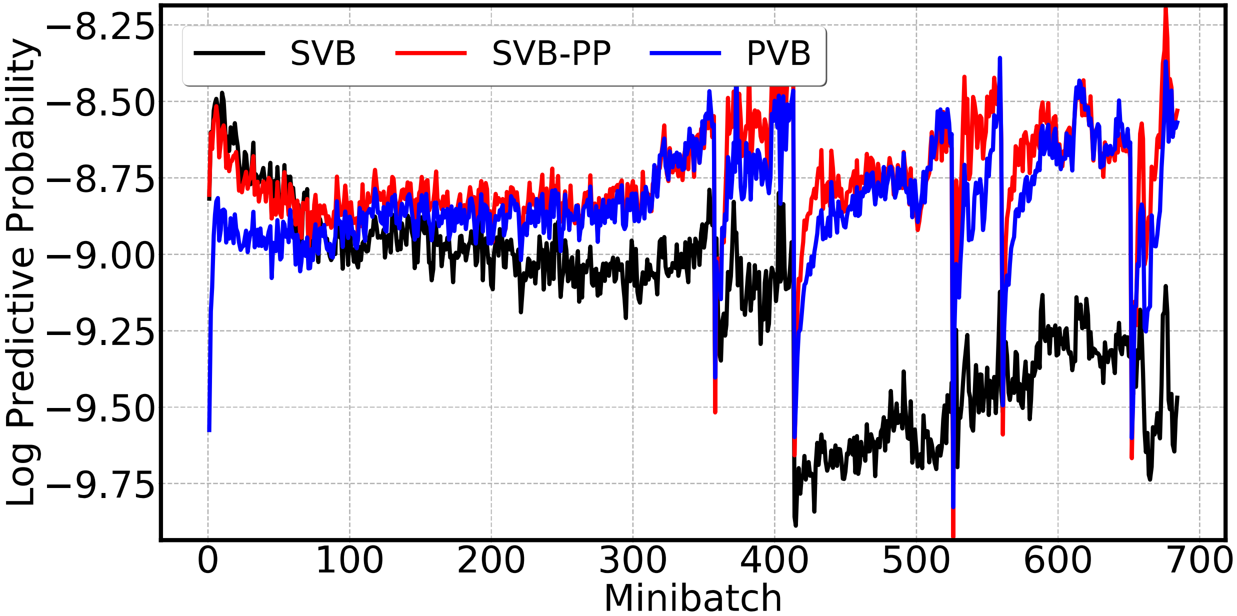
- Our finding for SVB:

$$\|\xi_b\|_1 \xrightarrow{b\to\infty} \infty \quad \xrightarrow{\text{yields}} \quad var(\xi_b) \xrightarrow{b\to\infty} 0$$

  - a model will evolve slowly (too stable)
  - could not deal well with sudden changes (Plasticity?)

Van-Son Nguyen, Duc-Tung Nguyen, Linh Ngo Van, Khoat Than, “Infinite Dropout for training Bayesian models from data streams”, IEEE Big Data, 2019.

Stability-Plasticity: concept drift



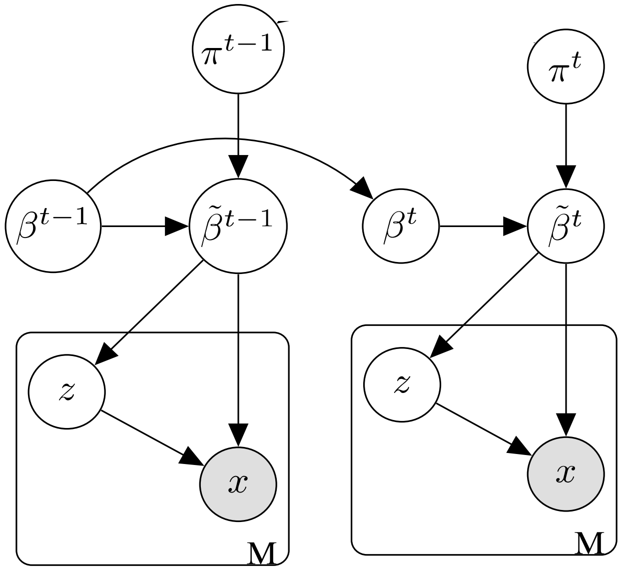
Generalization of a model when concept drifts sometimes appear over time.
(Higher is better)

Stability-Plasticity: our solution

- Infinite Dropout: ensemble of infinite learners in streaming conditions
  - Transition from time $t-1$ to t:
$$\beta^t \sim Normal(\beta^{t-1}, \sigma I)$$
  - Drop out some of the global variable:
$$\hat{\beta}^t = f(\beta^t \otimes \pi^t)$$
$$\pi^t \sim Bernoulli(dr)$$

Balance the old & new

Room for new knowledge

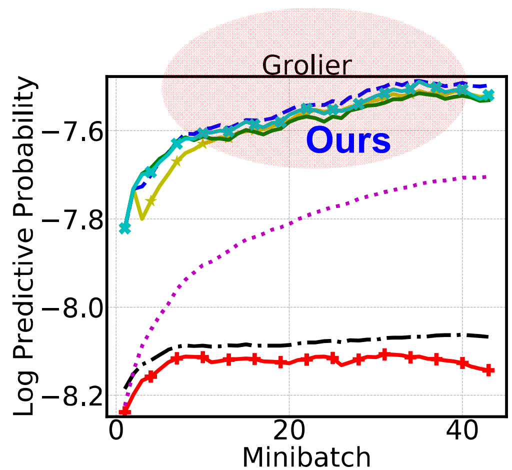


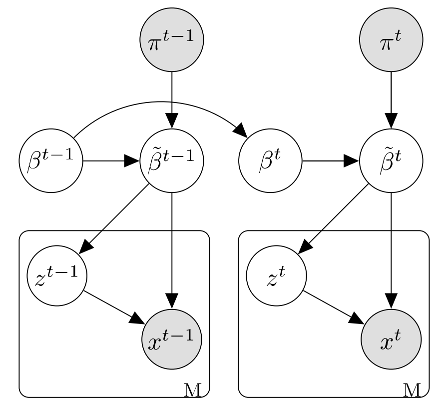
(a) iDropout for $B(\beta, z, x)$

Van-Son Nguyen, Duc-Tung Nguyen, Linh Ngo Van, Khoat Than, "Infinite Dropout for training Bayesian models from data streams", *IEEE Big Data*, 2019.

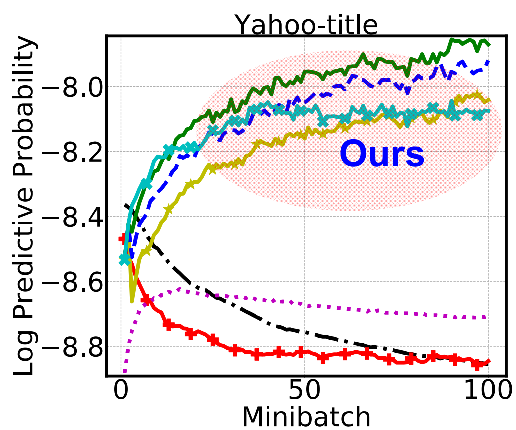
Stability-Plasticity: our solution

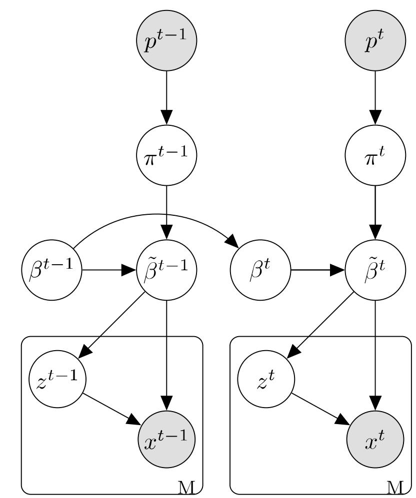
- iDropout & aiDropout
  - Adapt well with sudden changes
  - Deal well with noisy and sparse data



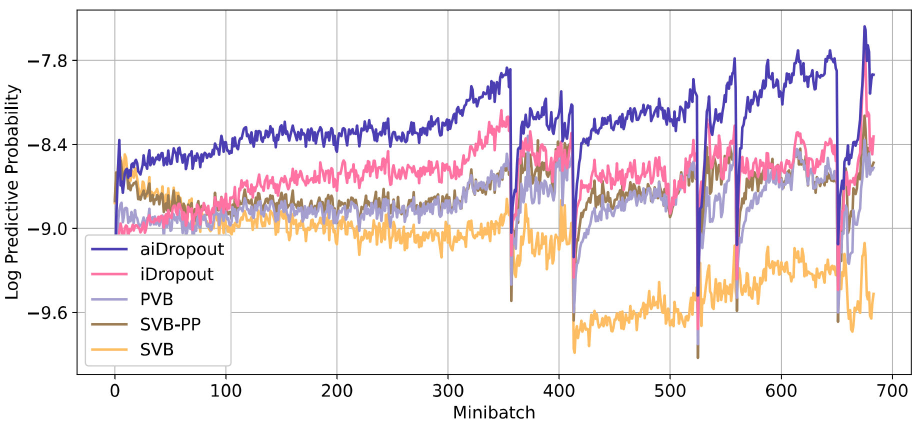


(a) iDropout for $B(\beta, z, x)$



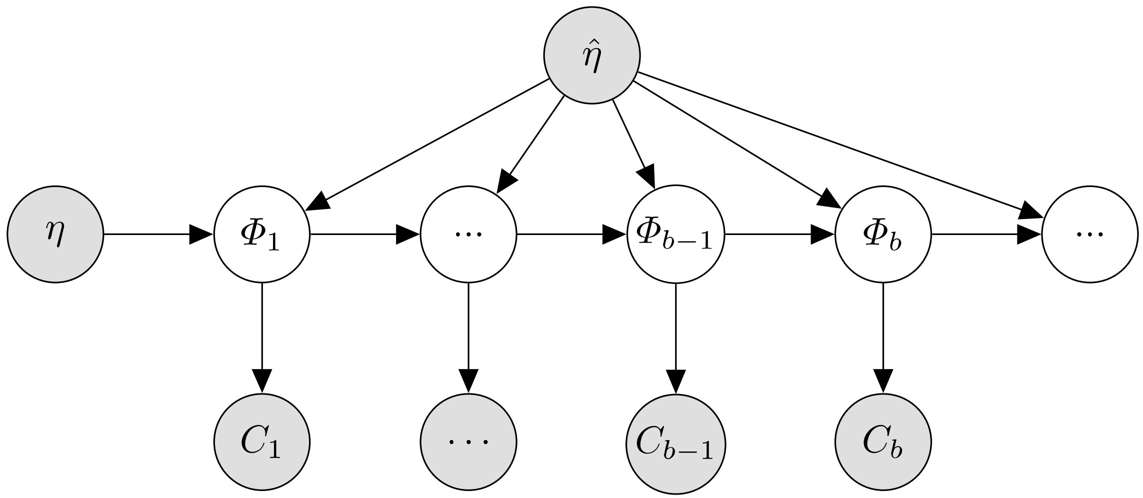


(b) aiDropout for $B(\beta, z, x)$



Ha Nguyen, Hoang Pham, Son Nguyen, Linh Ngo Van, Khoat Than. “Adaptive Infinite Dropout for Noisy and Sparse Data Streams,” Conditionally accepted, *Machine Learning journal*, 2021.

Catastrophic forgetting: our solution



- We propose a simple framework (**BPS**) to keep human knowledge at every step

$$p(\Phi|C_1, \dots, C_b, \eta) \propto p(\Phi|\hat{\eta}_b)p(C_b|\Phi, \eta)p(\Phi|C_1, \dots, C_{b-1}, \eta)$$

- *So, learning = boosted prior + (current + past) statistics*

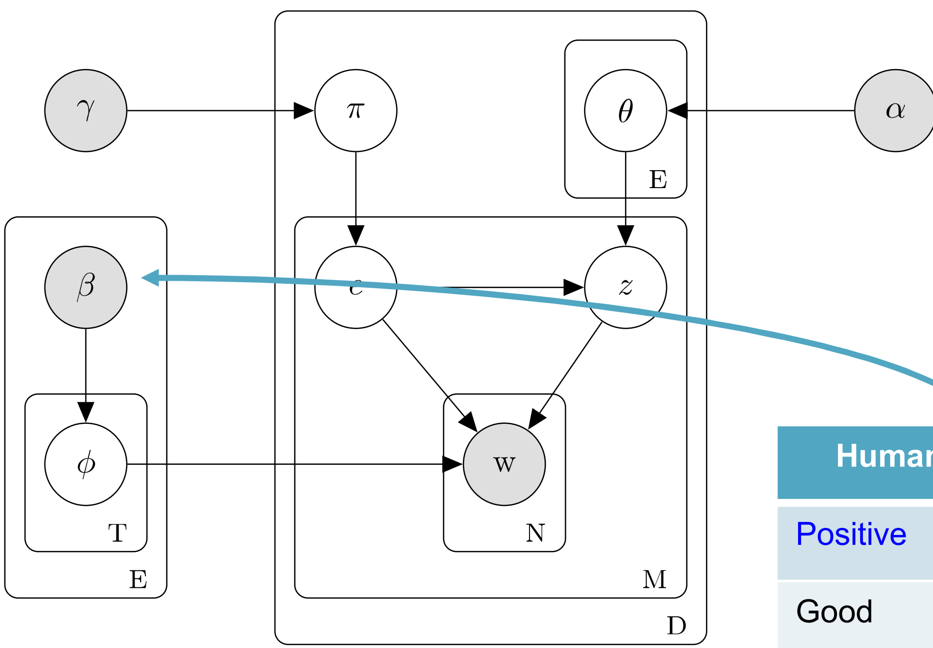
$$\boldsymbol{\xi_b = \xi_0^b + \hat{\xi}_b + \xi_{b-1}}$$

- Where ξ_0^b is the knowledge to be preserved

Anh Nguyen, Ngo Van Linh, Anh Nguyen Kim, Canh Hao Nguyen, and Khoat Than. "Boosting Priors in streaming Bayesian learning." *Neurocomputing*, vol. 424, 2021.

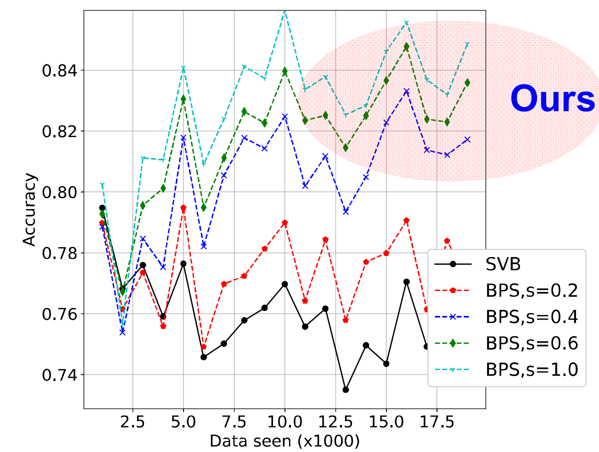
Our BPS: sentiment analysis

- Unsupervised Sentiment analysis
  - BPS mostly outperforms SVB [Nguyen et al., 2021]

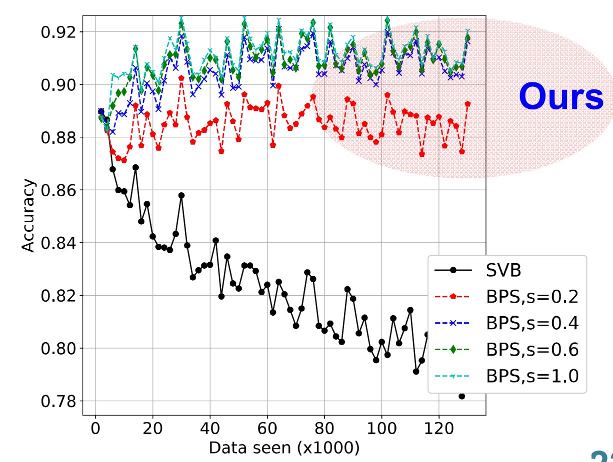


Aspect-sentiment unification model
[Jo & Oh, 2011]

Human knowledge	
Positive	Negative
Good	Bad
Awesome	Sad
Excellent	Boring



(b) Yelp.



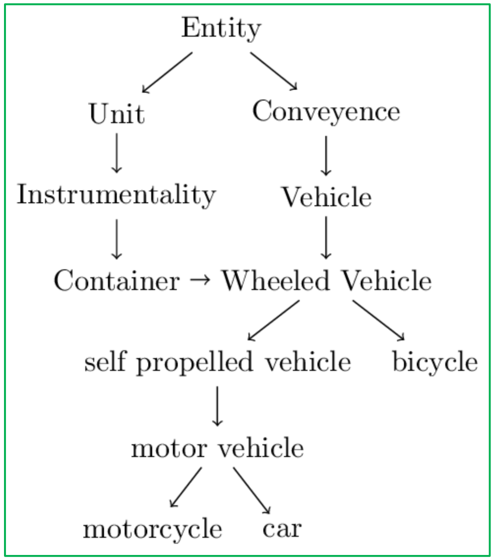
(d) Music.

Human knowledge: TPS

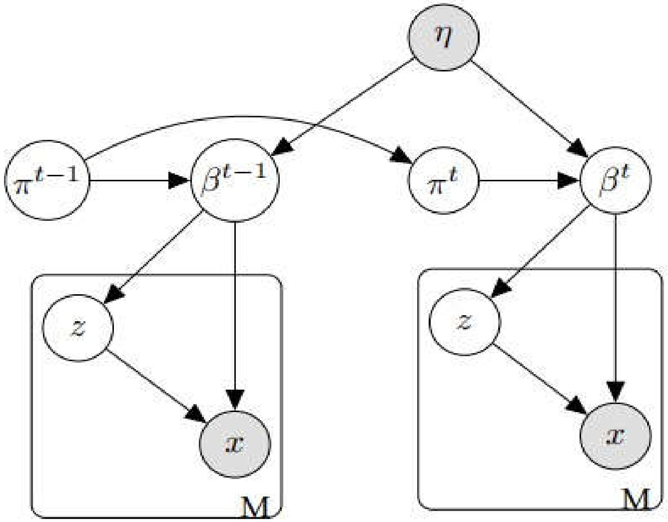
- Prior knowledge is exploited as learning from a stream
  - The knowledge is transformed into the global variable:

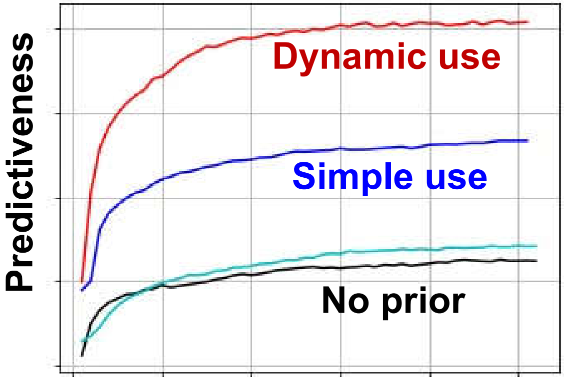
$$\beta^t = f(\pi^t \eta)$$

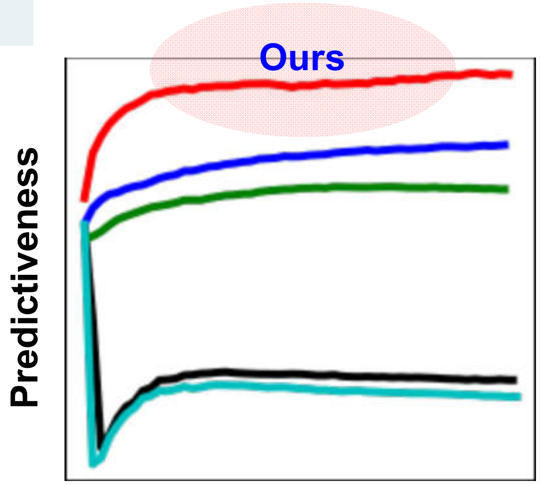
  - The transformation (π) is **dynamic**



Human knowledge	
Positive	Negative
Good	Bad
Awesome	Sad
Excellent	Boring

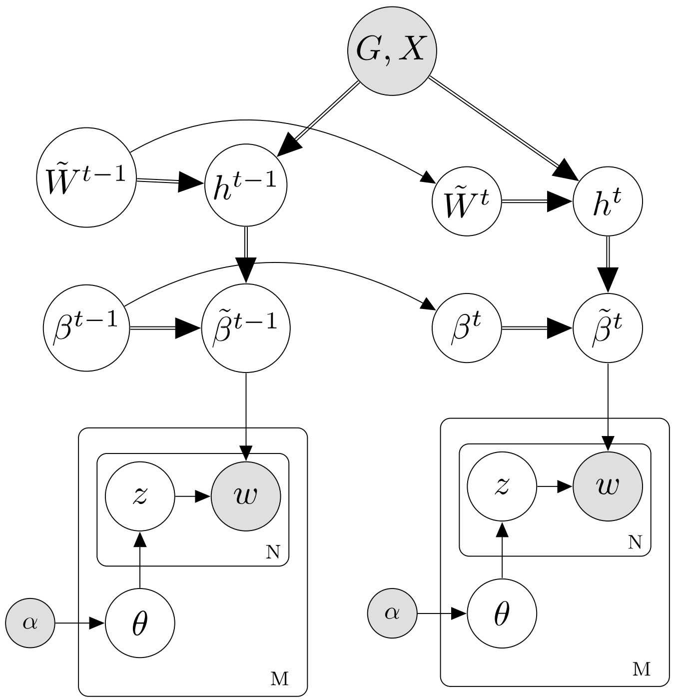






Bach, Tran Xuan, Nguyen Duc Anh, Linh Ngo Van, Khoat Than. “Dynamic transformation of prior knowledge into Bayesian models for data streams.” Conditionally accepted, *IEEE Transactions on Knowledge and Data Engineering*, 2021.

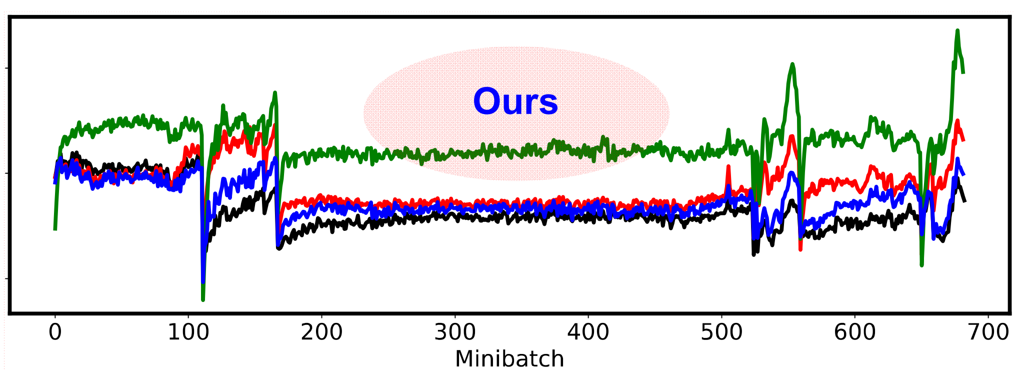
Knowledge graph: GCTM

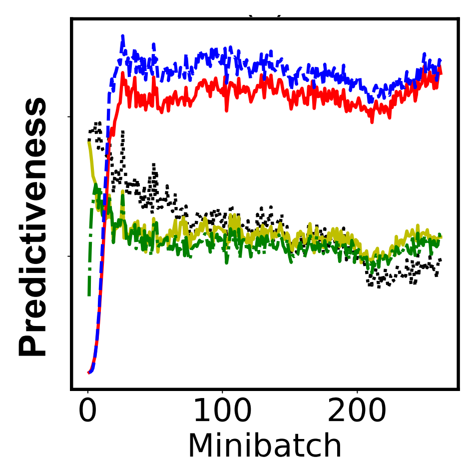


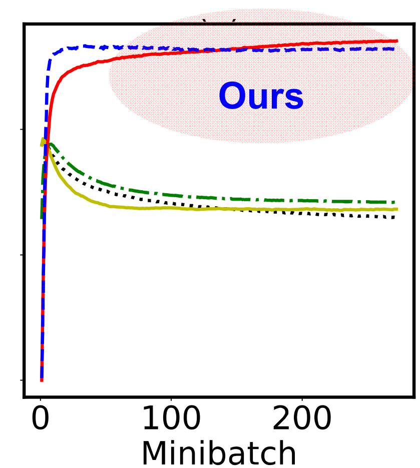
- Many sources of valuable knowledge are graphs
- **GCTM** = Graph convolutional nets + Topic models + Knowledge graph

$$h^t = GCN(G, X; W^t)$$

  - Analyzing hidden topics from text streams







Linh Ngo, Bach Tran, Khoat Than. “Graph Convolutional Topic Model for Data Streams.” Cond. accepted, *Neurocomputing*, 2021.