Progressive Layered Extraction (PLE)

潘微科
深圳大学计算机与软件学院

Hongyan Tang, Junning Liu, Ming Zhao, Xudong Gong. **Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations**. RecSys 2020.

特别说明：该PPT中的绝大部分文字和图示等内容来自参考文献。

研究动机

- There has been an increasing trend to apply Multi-Task Learning (MTL) in RS to model the multiple aspects of user satisfaction or engagement simultaneously. And in fact, it has been the **mainstream** approach in major industry applications.

- However, MTL models often suffer from performance degeneration with **negative transfer** due to the complex and competing task correlation in real-world recommender systems.
  - Tasks in real-world recommender systems are often loosely correlated or even conflicted, which may lead to **performance deterioration**.

- It is critical to design a more powerful and efficient model to handle complicated correlations and eliminate the challenging **seesaw phenomenon**.
  - Existing MTL models often improve some tasks **at the sacrifice** of the performance of others, when task correlation is complex and sometimes sample dependent, i.e., multiple tasks **could not be improved simultaneously** compared to the corresponding single-task model.

介绍

- 研究问题：多任务学习（multi-task learning）
  Build a single model that learns multiple goals and tasks simultaneously

- 模型名称：Progressive Layered Extraction (PLE)

相关工作

- Robert A Jacobs, Michael I Jordan, Steven J Nowlan, Geoffrey E Hinton. **Adaptive mixtures of local experts**. Neural Computation 1991.

  Mixture-of-Experts (MoE)

  One-gate Mixture-of-Experts (OMoE)

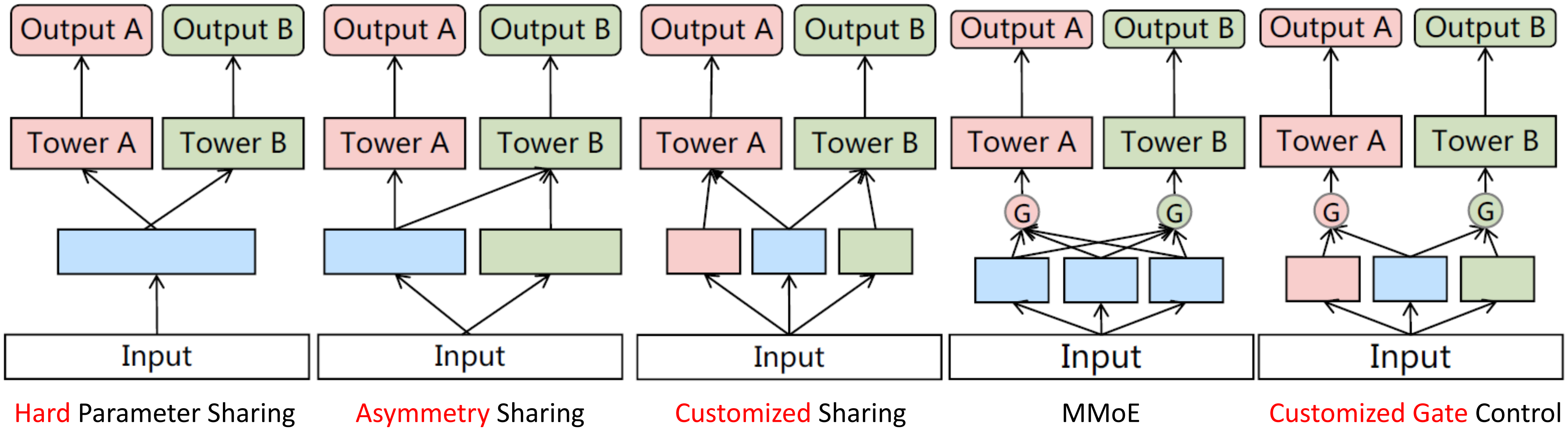
- Rich Caruana. **Multitask learning**. In *Learning to learn*. Springer, 95-133. 1998.

  Shared-Bottom multi-task DNN structure

- Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, Ed H. Chi. **Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts**. KDD 2018.

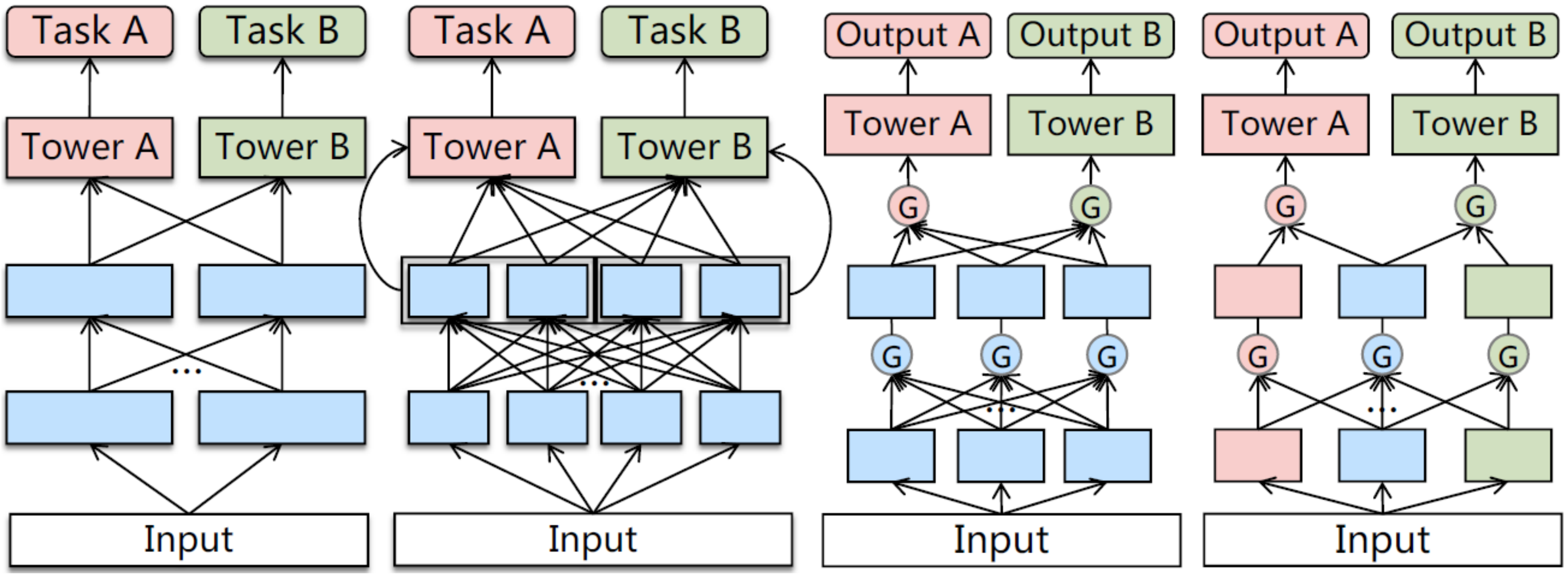
  Multi-gate Mixture-of-Experts (MMoE)

Single-Level MTL Models



- 注： MMoE和CGC使用了gating network

Multi-Level MTL Models

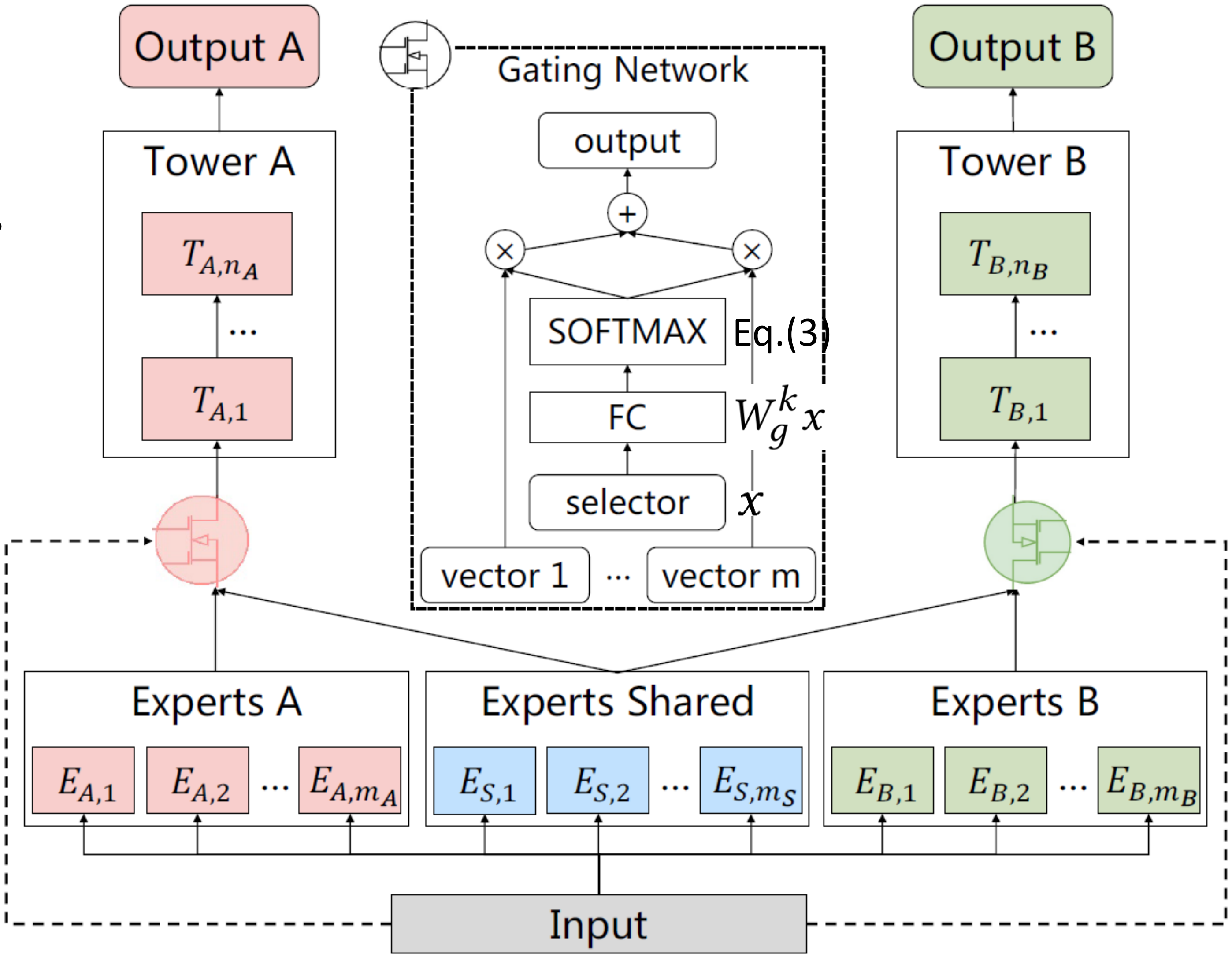


Cross-Stitch Network　Sluice Network　ML-MMOE　PLE

- 注：Progressive Layered Extraction (PLE)中的Progressive体现在gating network的设计（lower layer与upper layer中的有所不同）

Customized Gate Control (CGC) Model (1/2)

- A novel sharing structure
- Separates shared and task-specific experts
  - shared experts
  - task-specific experts
- Each expert module contains multiple sub-networks (experts)
- 没有Experts B与Tower A的connection:
  - Different types of experts concentrate on learning different knowledge efficiently without interference



Customized Gate Control (CGC) Model (2/2)

- Shared experts and task-specific experts are combined through a gating network for selective fusion

i.e., the outputs of experts. More precisely, the output of task k's gating network is formulated as:

$$g^k(x) = w^k(x)S^k(x), \tag{2}$$

where x is the input representation, and $w^k(x)$ is a weighting function to calculate the weight vector of task k through linear transformation and a SoftMax layer:

$$w^k(x) = \text{Softmax}(W_g^k x), \tag{3}$$

where $W_g^k \in R^{(m_k+m_s)\times d}$ is a parameter matrix, m_s and m_k are the number of shared experts and task k's specific experts respectively, d is the dimension of input representation. $S^k(x)$ is a selected matrix composed of all selected vectors including shared experts and task k's specific experts:

$$S^k(x) = [E_{(k,1)}^T, E_{(k,2)}^T, \ldots, E_{(k,m_k)}^T, E_{(s,1)}^T, E_{(s,2)}^T, \ldots, E_{(s,m_s)}^T]^T \tag{4}$$

Finally, the prediction of task k is:

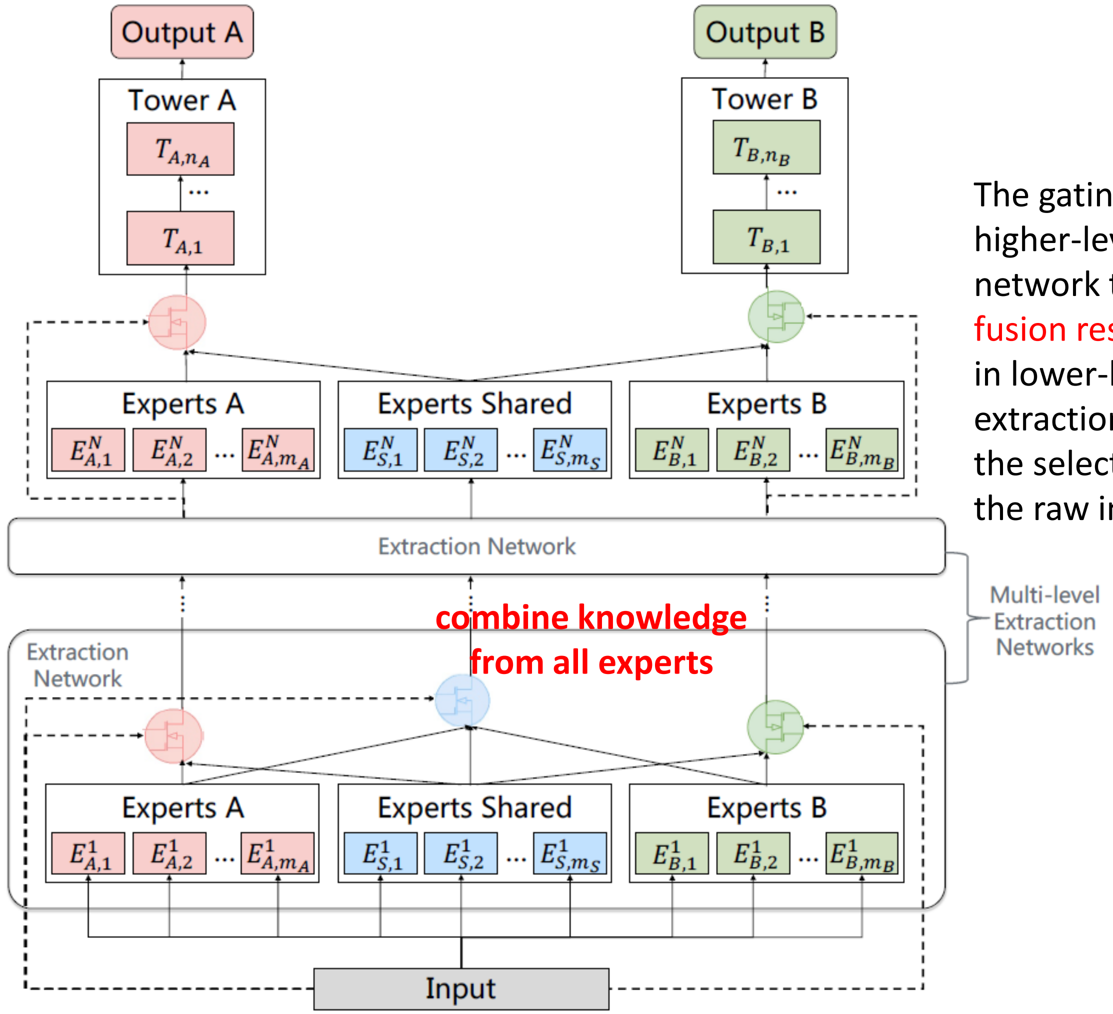
$$y^k(x) = t^k(g^k(x)), \tag{5}$$

where t^k denotes the tower network of task k.

Progressive Layered Extraction (PLE) Model

- Parameters of different tasks are separated progressively in upper layers

- Parameters of different tasks are NOT fully separated in the early layer



The gating networks in higher-level extraction network take the fusion results of gates in lower-level extraction network as the selector instead of the raw input

总结

- 主要视角： joint representation learning and information routing

- 主要思路：
  - Explicitly separates shared components and task-specific components to alleviate harmful parameter interference between common and task-specific knowledge
  - Adopts a progressive routing mechanism to extract and separate deeper semantic knowledge from lower-layer experts and separate task-specific parameters in higher levels gradually
  - Introduces multi-level experts and gating networks

- 主要优点： With multi-level experts and gating networks, PLE extracts and combines deeper semantic representations for each task to improve generalization.

参考资料

- Hongyan Tang, Junning Liu, Ming Zhao, Xudong Gong. **Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations**. RecSys 2020.