GENERATING SCORES WITH DEEP LEARNING AND ITEM RESPONSE THEORY

Ananyapam De

INTRODUCTION

Our goal is to generate scores that can accurately predict a child's developmental group

PAPERS SUMMARIES

- Item Response Theory
- Bayesian Knowledge Tracing
- Deep Knowledge Tracing
- Dynamic Key Value Memory Networks for Knowledge Tracing
- Deep IRT

ITEM RESPONSE THEORY (IRT)

- IRT: A framework for modeling the probability of a specific response to an item based on the latent ability of a respondent.

- Latent Trait: Represents the unobserved (hidden) knowledge or ability of a student.

- IRT Parameters:
  - Difficulty: How challenging an item is.
  - Discrimination: How well an item distinguishes between those with and without the skill.
  - Guessing: The chance of answering correctly by guessing.
  - Slipping: The chance of answering incorrectly despite knowing the answer.

ITEM RESPONSE THEORY (IRT)

1 Parameter Logistic Model (1PLM) or Rasch Model

$$P(U_i = 1|\theta) = \frac{1}{1 + e^{-(\theta - b_i)}}$$

- Only the difficulty parameter b_i is considered in this model.

2 Parameter logistic model (2PLM)

$$P(U_i = 1|\theta) = \frac{1}{1 + e^{-a_i(\theta - b_i)}}$$

- This model excludes the guessing parameter c_i.

3 Parameter Logistic Model (3PLM)

$$P(U_i = 1|\theta) = c_i + \frac{(1 - c_i)}{1 + e^{-a_i(\theta - b_i)}}$$

Where,

- U_i = Response to item i (1 for correct, 0 for incorrect).
- θ = Latent ability or trait of the respondent.
- a_i = Discrimination parameter for item i.
- b_i = Difficulty parameter for item i.
- c_i = Guessing parameter for item i.

KNOWLEDGE TRACING (KT)

- *Knowledge Tracing (KT):* A process of tracking and predicting a learner's knowledge over time, often in intelligent tutoring systems.

- Connection to IRT:
  - IRT gives a statistical framework that can enhance the precision of KT.
  - Provides a probabilistic understanding of student's response patterns.

- *Bayesian Knowledge Tracing (BKT)* and *Deep Knowledge Tracing (DKT)* model student's knowledge in concept specific or summarized manner.

- Challenges:
  - Requires large datasets for calibration.
  - Assumes unidimensionality (single latent trait being measured).
  - Complex computational models.

BAYESIAN KNOWLEDGE TRACING (BKT)

- *Bayesian Knowledge Tracing (BKT)* is a specific method within KT.
- Analyses student's knowledge state into different concept states.
- Assumes concept state as binary: known or unknown.
- Uses Hidden Markov model for updating posterior distribution.
- Challenges:
  - Cannot capture relationships between different concepts.
  - Uses discrete variables and simple transition models.
  - Can output mastery level of predefined concepts.
  - Cannot model complex concept state transitions or extract undefined concepts.

BAYESIAN KNOWLEDGE TRACING (BKT)

- $P(K_t)$ = Probability the skill is known at time t.
- $P(L_0)$ = Initial probability the skill is known.
- $P(T)$ = Probability of transitioning from unknown to known state.
- $P(G)$= Guess probability.
- $P(S)$ = Slip probability.

- If the student answers correctly at time t:

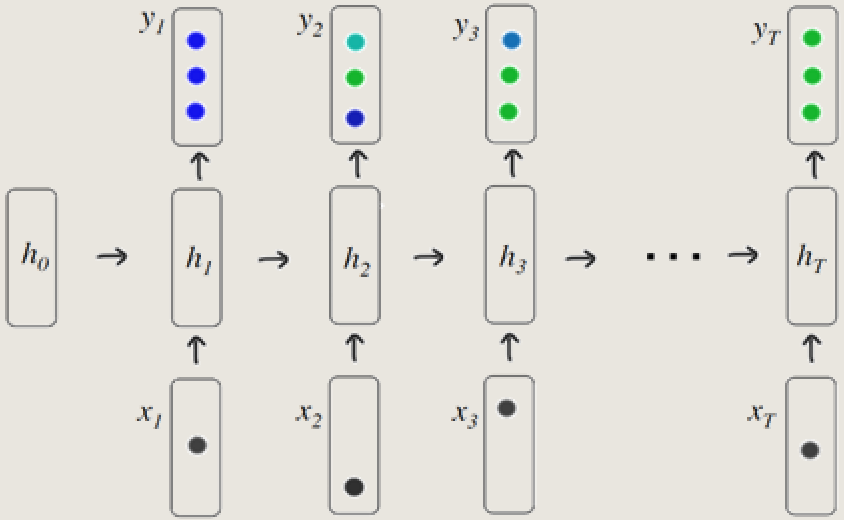
$$P(K_t|\text{Correct}) = \frac{\left(\mathrm{P}(\mathrm{K_t}) \times \left(1 - P(S)\right)\right)}{\mathrm{P}(\mathrm{K_t}) \times \left(1 - P(S)\right) + \left(1 - \mathrm{P}(\mathrm{K_t})\right) \times P(G)}$$

- If the student answers incorrectly at time t:

$$P(K_t|\text{Incorrect}) = \frac{P(K_t) \times P(S)}{\mathrm{P}(\mathrm{K_t}) \times P(S) + \left(1 - \mathrm{P}(\mathrm{K_t})\right) \times \left(1 - P(G)\right)}$$

DEEP KNOWLEDGE TRACING (DKT)

Models Applied



$(x_1, x_2, \dots, x_T) \longrightarrow (y_1, y_2, \dots, y_T)$

- **Vanilla RNN**
  - Uses hidden states $(h_1, \dots, h_T)$ to encode past observations.
- **Long Short Term Memory (LSTM):**
  - A complex variant of RNNs.
  - Retains unit values until 'forget gate' acts.

Data Representation

Small datasets

One hot encoding of student interaction tuple ($h_t = \{q_t, a_t\}$).

Large datasets

- Random vector representation inspired by compressed sensing
- A one hot encoding, being a 1 sparse signal, can be encoded using a fixed random Gaussian input vector of length approx. $\log 2M$.
- Enables efficient representation in high dimensional spaces.

Challenges

- Assumes high dimensional and continuous representation of knowledge state.
- Summarizes all concepts in one hidden state.
- Difficult to trace mastery level of specific concepts or pinpoint student's proficiency.

DYNAMIC KEY VALUE MEMORY NETWORKS FOR KNOWLEDGE TRACING (DKVMN)

- Exploiting relationship between concepts.
- Can learn correlation between exercises and underlying concepts.
- Maintains a concept state for each concept.
- Updates only related concept states at each timestamp.
- Uses:
  - Static matrix (key) for storing concept representations.
  - Dynamic matrix (value) for storing and updating student's understanding of each concept.

DYNAMIC KEY VALUE MEMORY NETWORKS FOR KNOWLEDGE TRACING (DKVMN)

- $M^{v} \in \mathbb{R}^{N \times d_v}$; Value memory matrix (knowledge states)
- $M^{\mathrm{k}} \in \mathbb{R}^{N \times d_k}$; Key memory matrix (latent concepts)
- $A \in \mathbb{R}^{d_k \times Q}$; KC Embedding matrix
- $k_t \in \mathbb{R}^{d_k}$; Embedding vector (key)
- $v_t \in \mathbb{R}^{d_v}$; Response Embedding vector
- $e_t \in \mathbb{R}^{d_v}$; Response erase vector
- $\mathbf{B} \in \mathbb{R}^{Q \times d_v}$; KC response embedding matrix

Updating Weights

$$\mathbf{e}_t = \sigma(\mathbf{W}_e \boldsymbol{v}_t + \mathbf{b}_e),$$

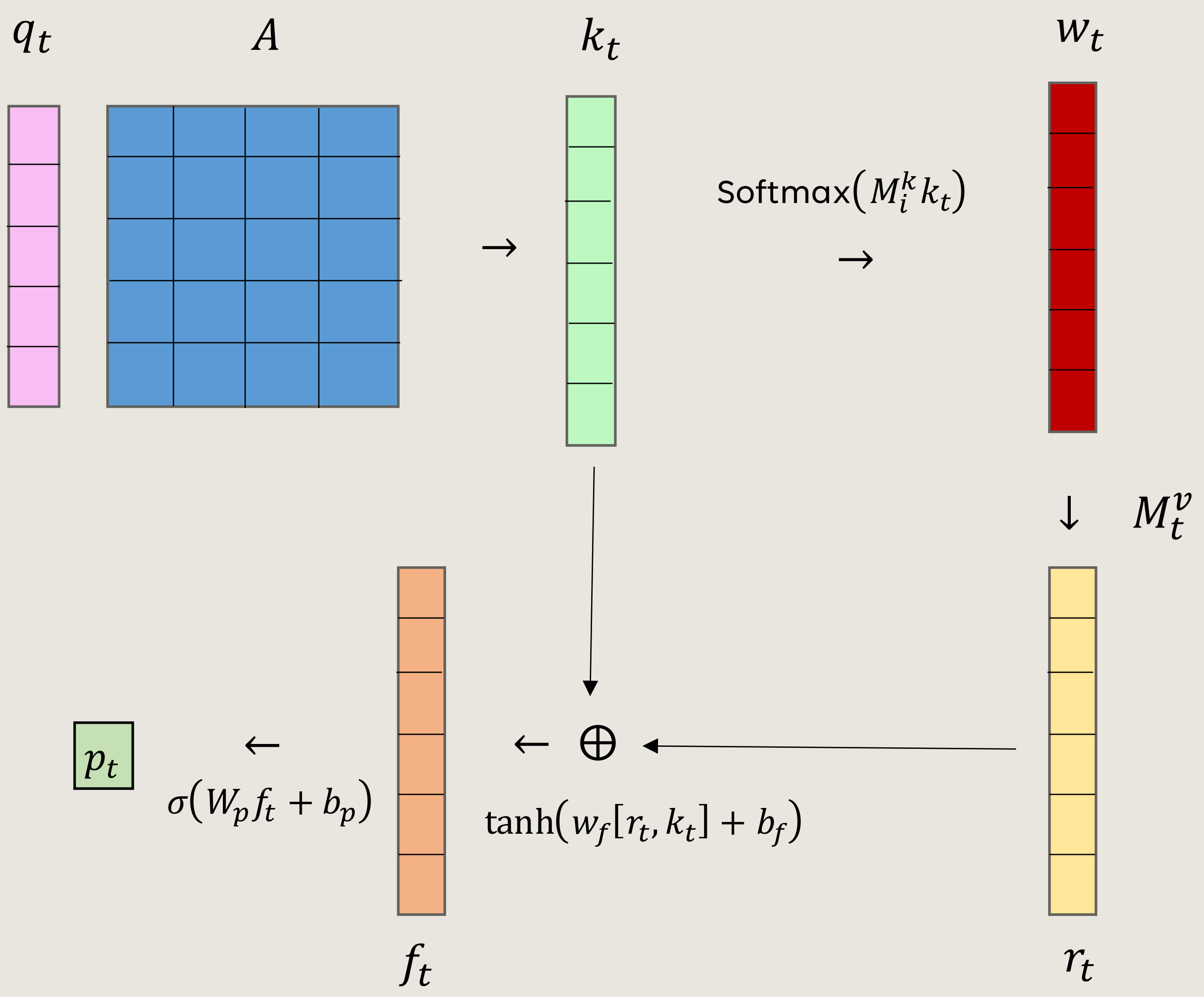
$$\mathbf{a}_t = \tanh(\mathbf{W}_a \boldsymbol{v}_t + \mathbf{b}_a),$$

$$\widetilde{\mathbf{M}}^{v}_{t+1,i} = \mathbf{M}^{v}_{ti} \otimes (1 - w_{ti} \mathbf{e}_t)^T,$$

$$\mathbf{M}^{v}_{t+1,i} = \widetilde{\mathbf{M}}^{v}_{t+1,i} + w_{ti} \mathbf{a}_t^T$$

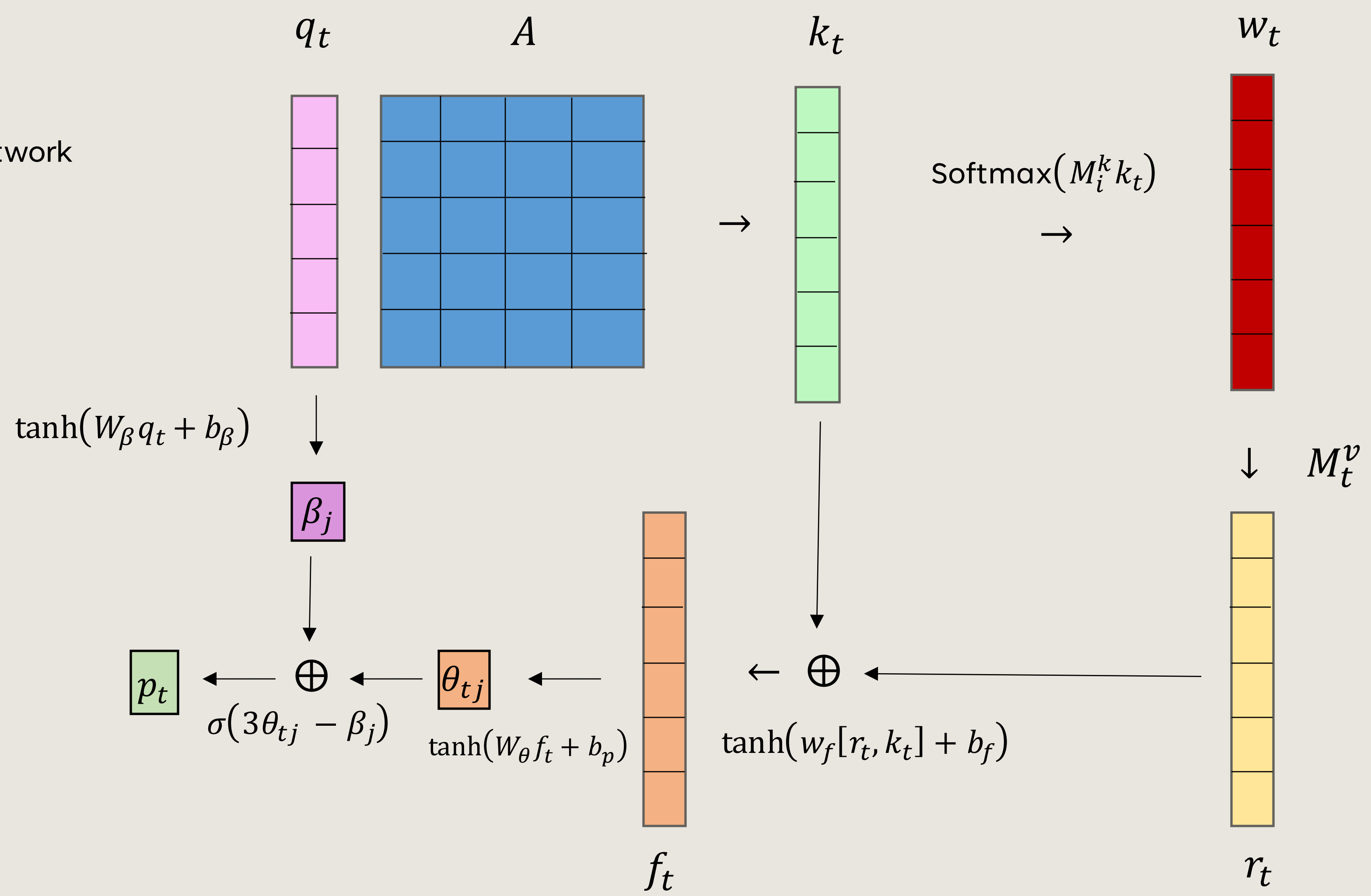
Getting attention weight

Making prediction



DEEP IRT

- Infers student ability & KCs' difficulty.
- Explainable compared to DKVMN, DKT
- Retains predictive power of deep learning
- Alternative to traditional testing uses entire learning trajectory
- Augments students ability and difficulty network to DKVMN
- θ_{tj} ; Student ability on KC j at time t
- β_j ; Difficulty of knowledge component j



THANK YOU