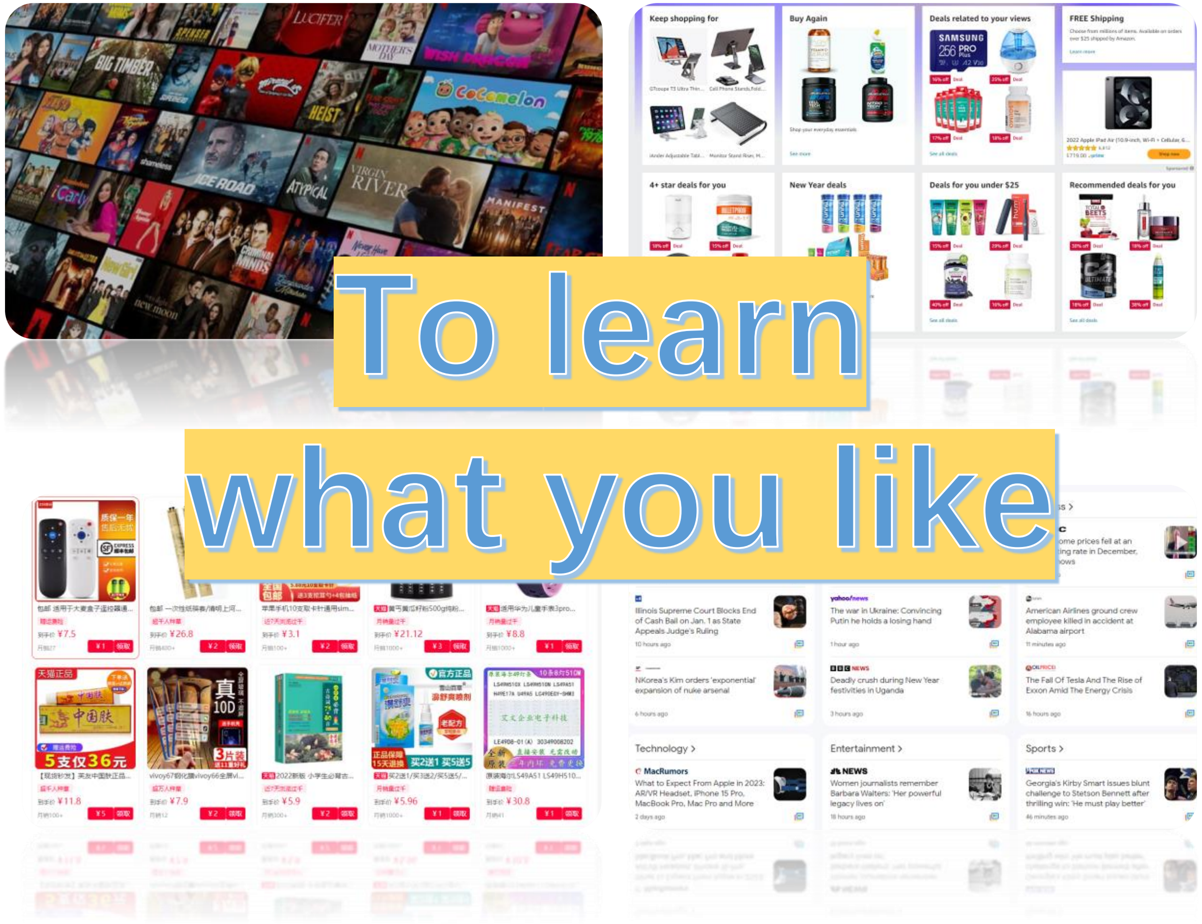
Track user preference in Recommendation via Quantum Jump

Jinkun Han

User preference & Personalization

- User's preference is important in recommendation systems
- Recommend based on user preference
  - Online shopping
  - Advertisement
  - News
  - Games
  - Micro-video
  - …



Introduction

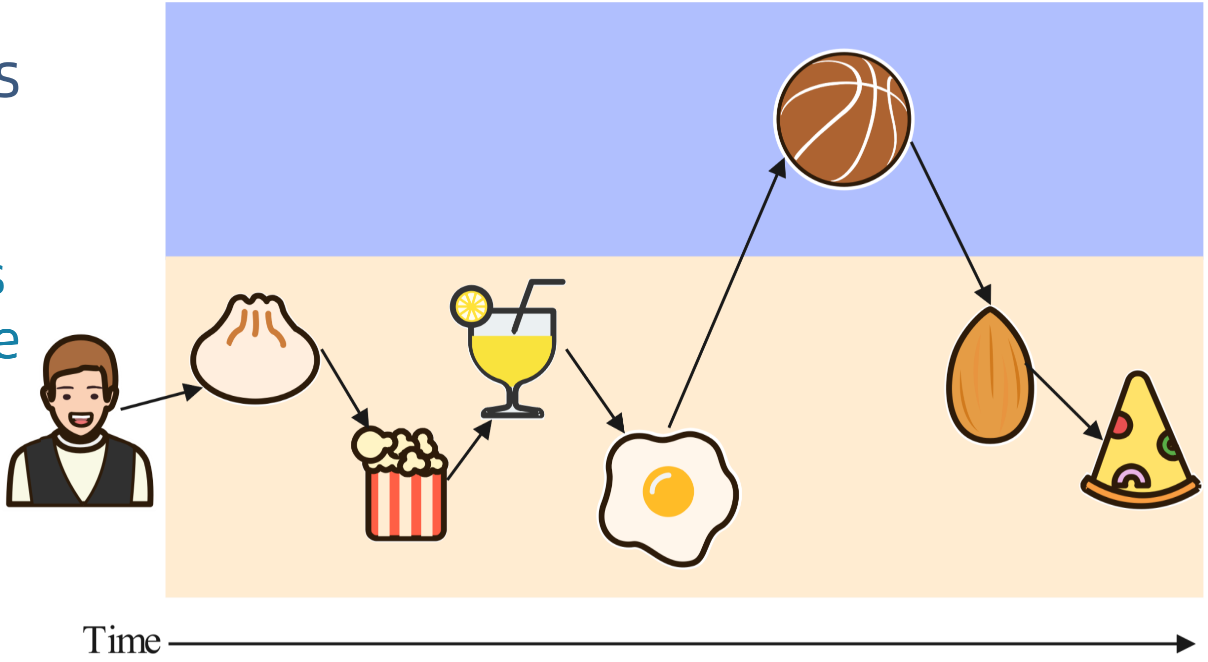
User preference & Personalization

- AI assistant
  - Google Assistant
  - Amazon Alexa
  - Apple Siri
  - …

- Robots with personalized service

People are acting like this ...

- Want to buy something new suddenly
- Viewed items influence user's choices
- Viewed items:
  - The user might be interested in the items but stops rating or buying items for some reason
    - Shopping: price, attributes, waiting for coupon
    - Video: time, can not satisfy full interests



Two assumptions

1. Viewed items influence users' choices
2. Viewed items contribute to suddenly changed preference

Introduction

Viewed items influence users' choices

"Energy+"

Buy

Green

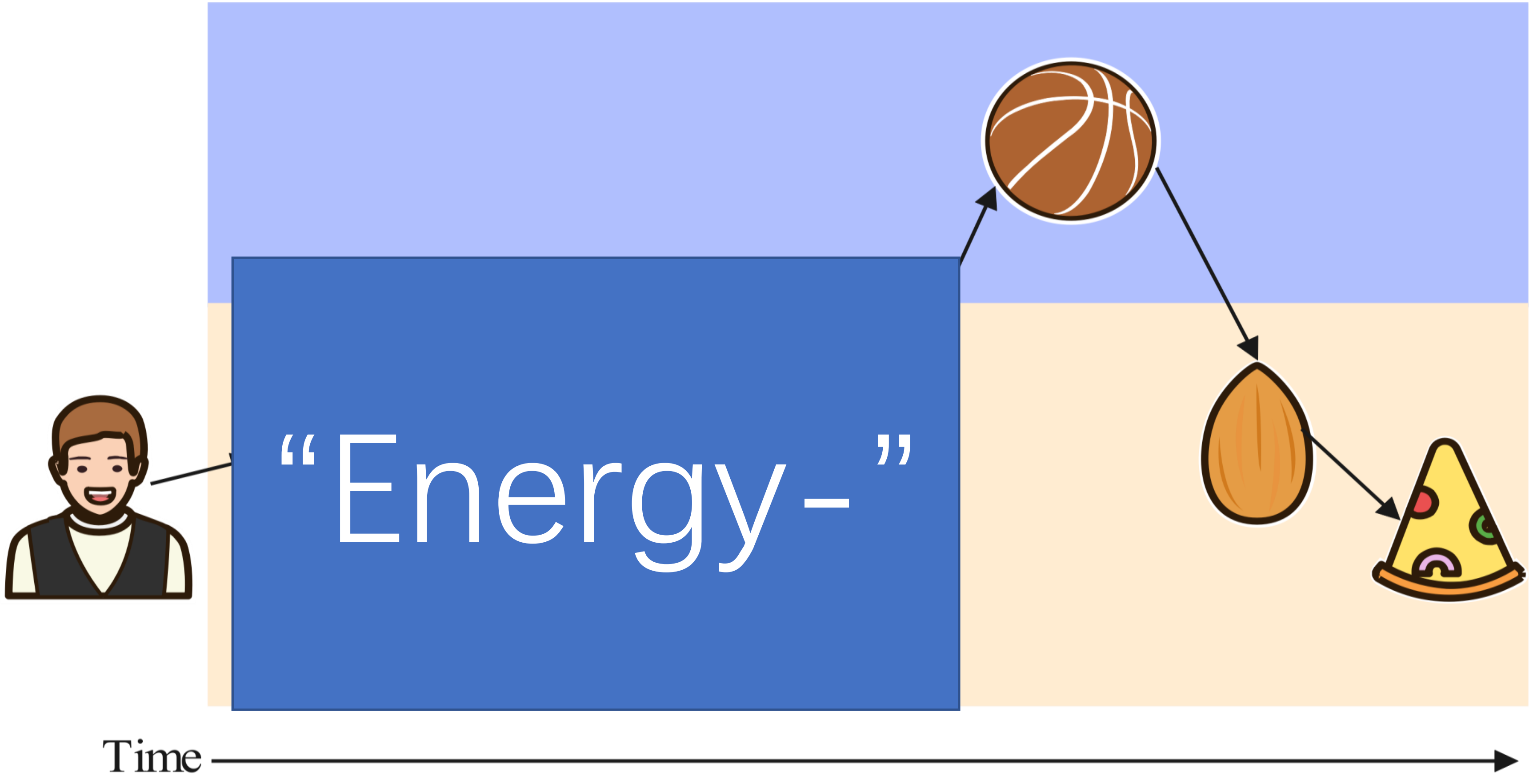
Dense Leaves

Round flowerpot

Computer Science
College of Arts & Sciences
Georgia State University

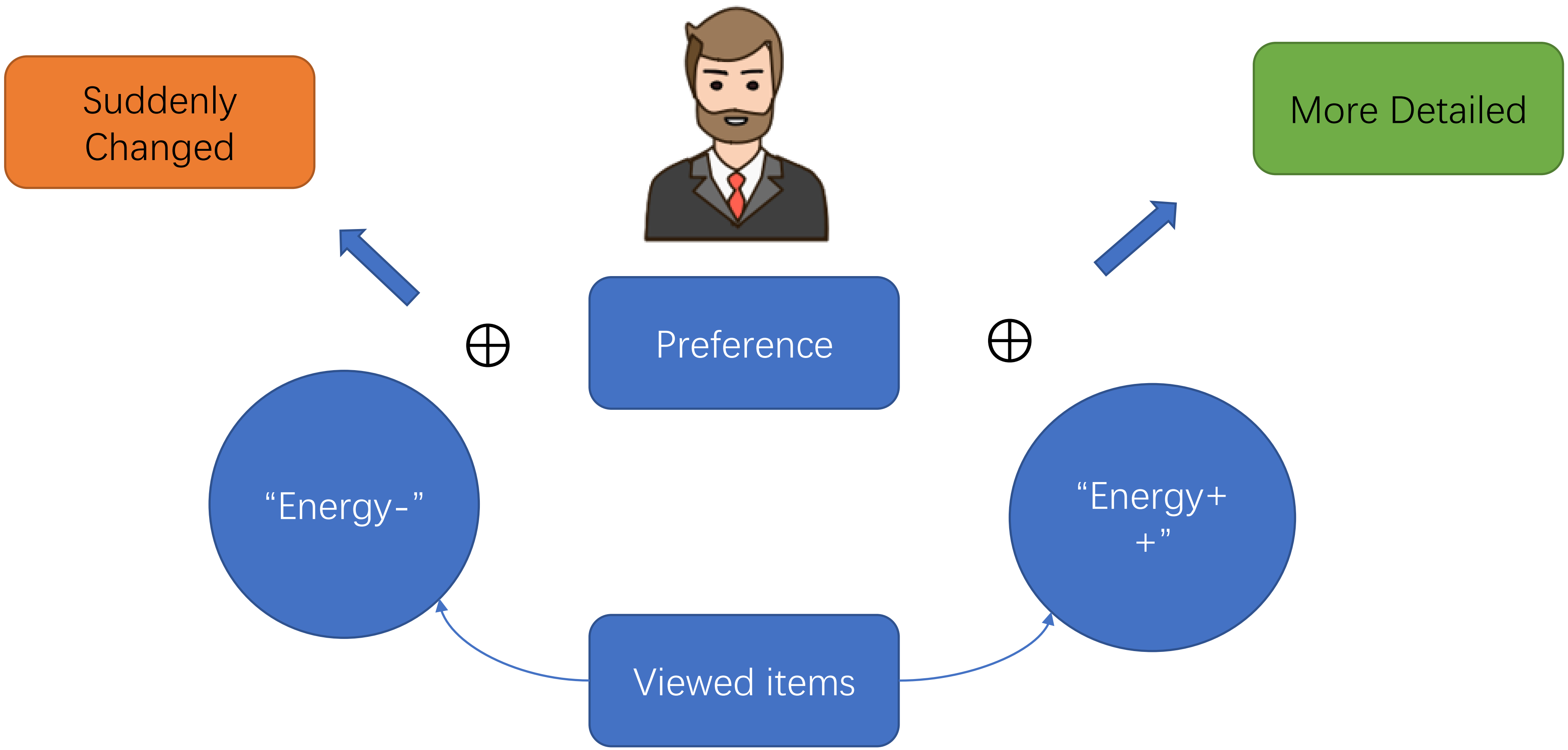
Suddenly changed preference

- Viewed items contribute to suddenly changed preference



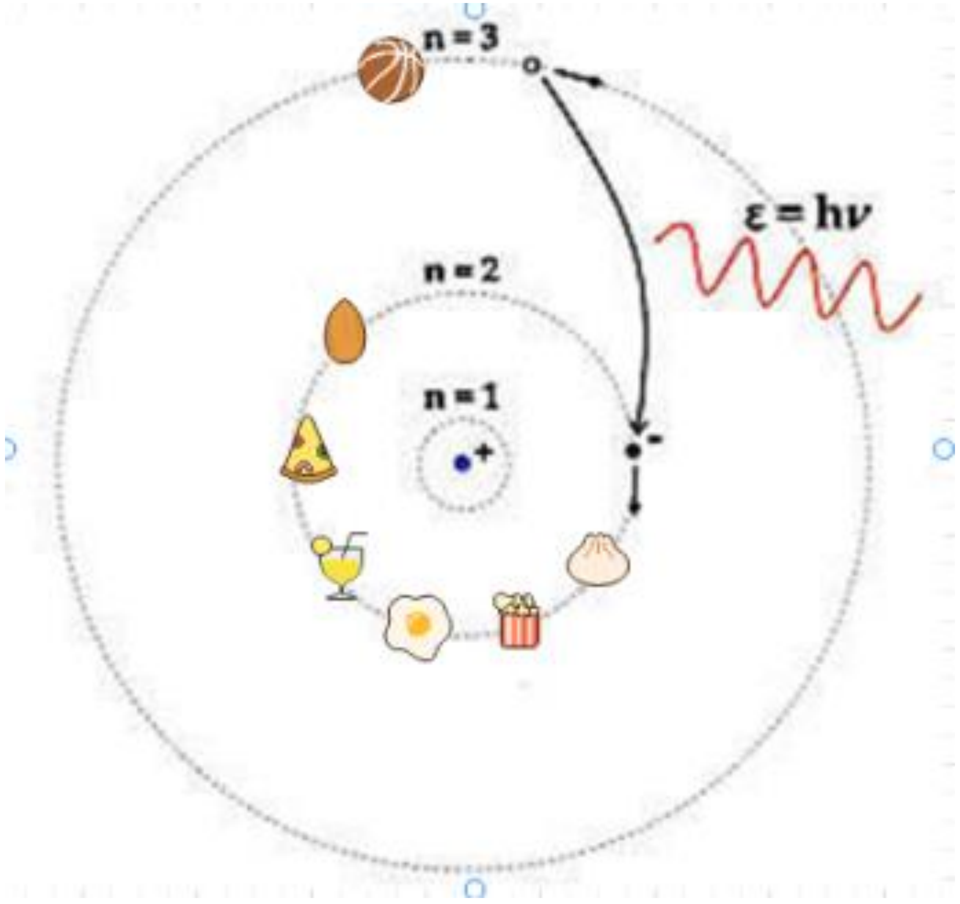
Computer Science
College of Arts & Sciences
Georgia State University

“Energy” influences the preference



It looks like Quantum Jump

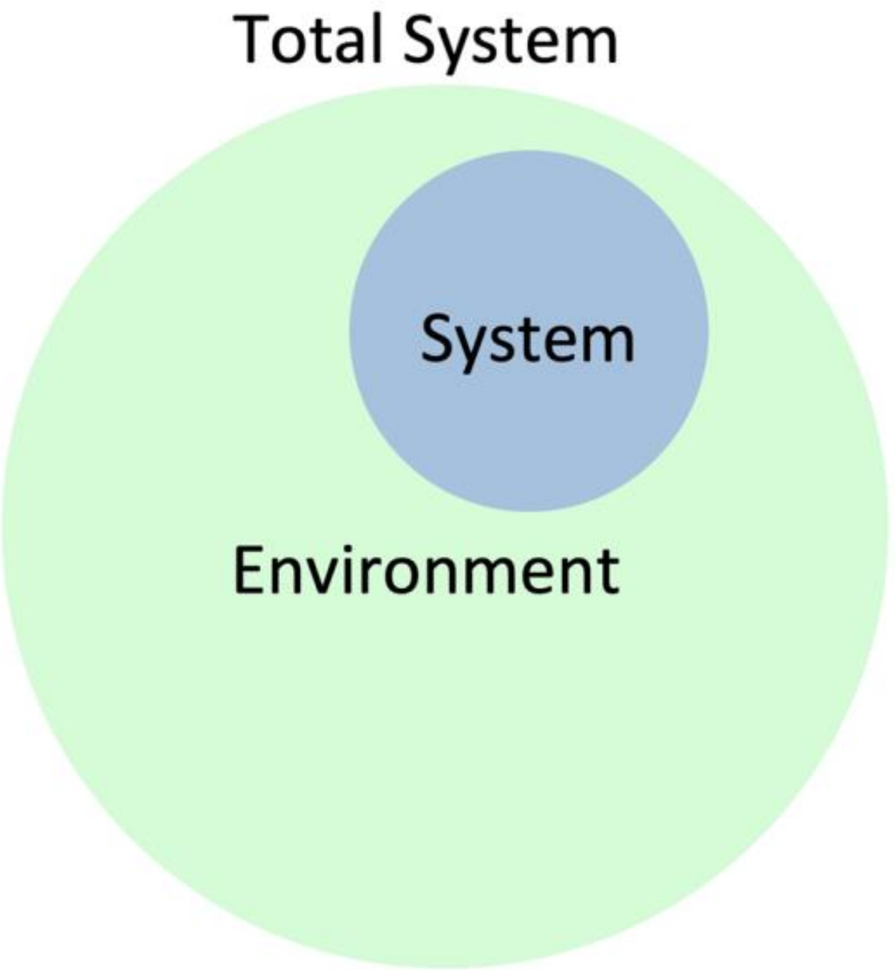
- From a quantum state to another quantum state
- "Sudden change" from one energy level to another
  - Excitation
  - Deexcitation



Open quantum system

- The whole system is divided →
  - Our system of interest (interacted items)
  - An environment (viewed but not interacted items)

- In this case, the most general quantum dynamics is generated by the Lindblad equation
  - Also called Gorini-Kossakowski-Sudarshan-Lindblad equation



Density matrix

- In quantum mechanics, a density matrix (or density operator) is a matrix that describes the quantum state of a physical system
  - A density matrix ρ has unit trace ($\mathrm{Tr}[\rho] = 1$)

$$\rho \equiv \sum_i p_i |s_i\rangle\langle s_i| \qquad (1)$$

- Density matrix represent a user's preference

Preliminaries

Hamiltonian

The Schrödinger equation can be formally solved in the following way. If at $t = 0$, the state of a system is given by $|\psi(0)\rangle$; at time t, it will be

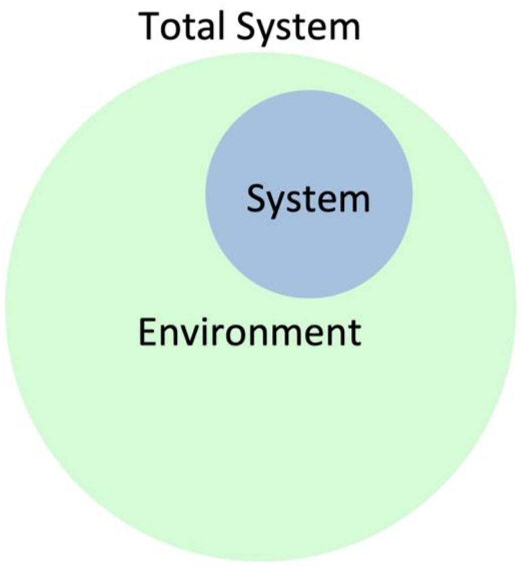
$$|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle. \quad (11)$$

- In quantum mechanics, the Hamiltonian (H) of a system is an operator corresponding to the total energy of that system
- Schrödinger equation

$$ih\frac{d}{dt}|s(t)\rangle = H|s(t)\rangle \quad (2)$$

- So far, we know what is happening in a closed system
- The Schrödinger equation can be formally solved in following way. If at $t = 0$, the state of a system is given by $|s(0)\rangle$; at time t is will be:

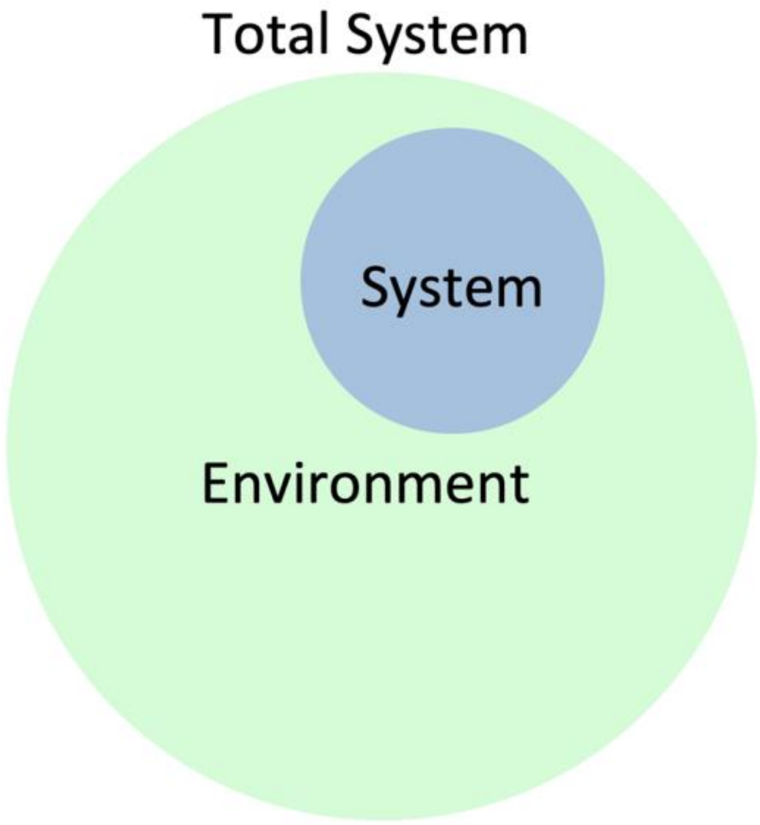
$$|s(t)\rangle = f(H,t)|s(0)\rangle$$



Derivation

- $\rho \equiv \sum_i p_i |s_i\rangle\langle s_i|$ (1)
- $|s(t)\rangle = f(H,t)|s(0)\rangle$ (2)
- Liouville's theorem: $\frac{d}{dt}\rho = \frac{1}{ih}[H,\rho]$

where the commutator $[H,\rho] = H\rho - \rho H$



Quantum state and Hamiltonian

- All unitary evolutions of a state belonging to a finite Hilbert space can be constructed

$$\dot{\rho} = \frac{1}{ih}[H, \rho] \quad (3)$$

where the commutator $[H, \rho] = H\rho - \rho H$

Lindblad master equation

Conjugate operation

$$\dot{\rho} = \boxed{[H, \rho]} + \boxed{\sum_{i=1}^{n} g_{i,j}(o_i \rho o_i^{\dagger} - \frac{1}{2}\{o_i^{\dagger} o_i, \rho\})} \quad (4)$$

- Term 1: the closed system (no noise)
- Term 2: the open environment
- F_i is usually called jump operator

Computer Science

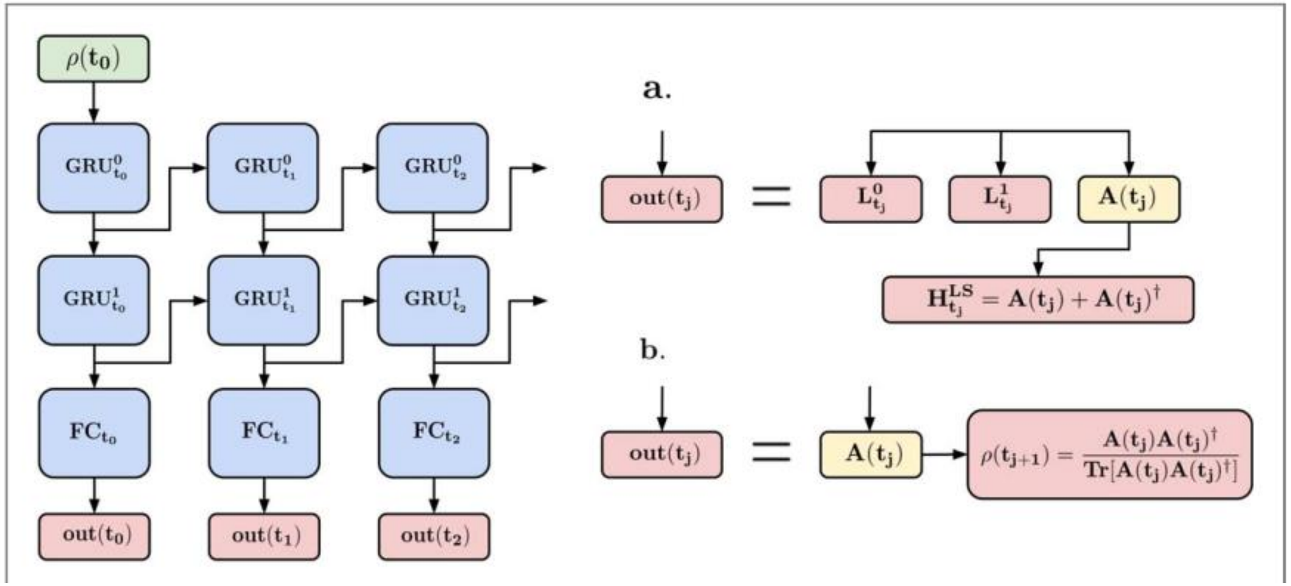
College of Arts & Sciences

Georgia State University

Current model

$$\mathcal{L}_{\leqslant t}[\rho] = -\mathrm{i}[H + H^{LS}_{\leqslant t}, \rho] + \sum_{\mu}\left[L^{\mu}_{\leqslant t}\rho L^{\mu\dagger}_{\leqslant t} - \frac{1}{2}\{L^{\mu\dagger}_{\leqslant t}L^{\mu}_{\leqslant t}, \rho\}\right],$$

- H^{LS} is a 'Lamb-shift' term
  - Namely a correction to the Hamiltonian induced by the environment



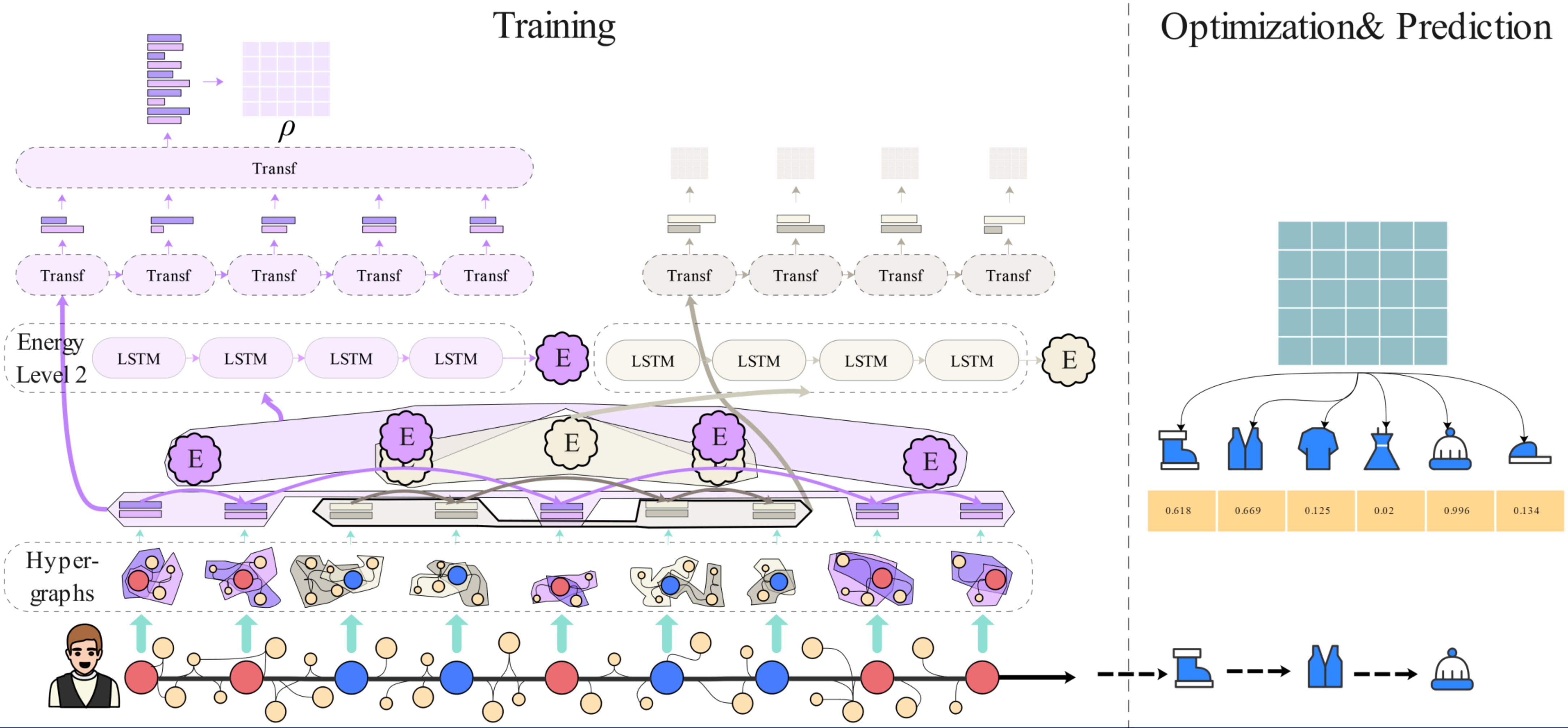
Components

- Interacted items → energy & density
  - Transformer or LSTM to learn latent representation to construct density matrix
  - Another component to evaluate energy
- Non-interacted items → energy & outside environment
  - Transformer or LSTM to learn jump operator
  - Another component to evaluate energy
- Hypergraphs
  - High-order relations between nodes (users and items)
  - Help to find energy (+/-)

Computer Science
College of Arts & Sciences
Georgia State University

Proposed Model

$$\dot{\rho} = [H, \rho] + \sum_{i=1}^{n} g_{i,j}\left(o_i \rho o_i^{\dagger} - \frac{1}{2}\{o_i^{\dagger} o_i, \rho\}\right)$$



Computer Science
College of Arts & Sciences
Georgia State University

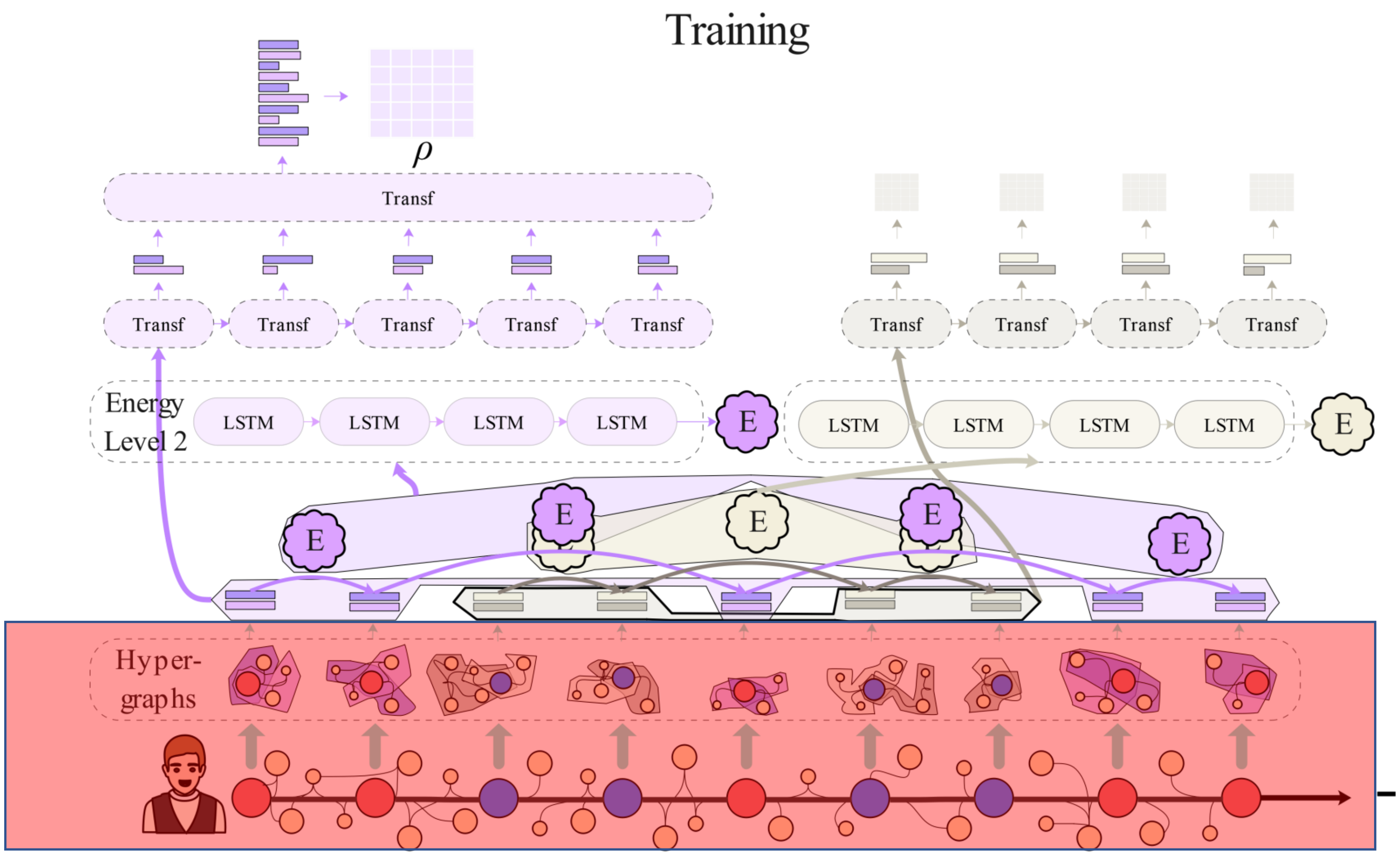
Definition

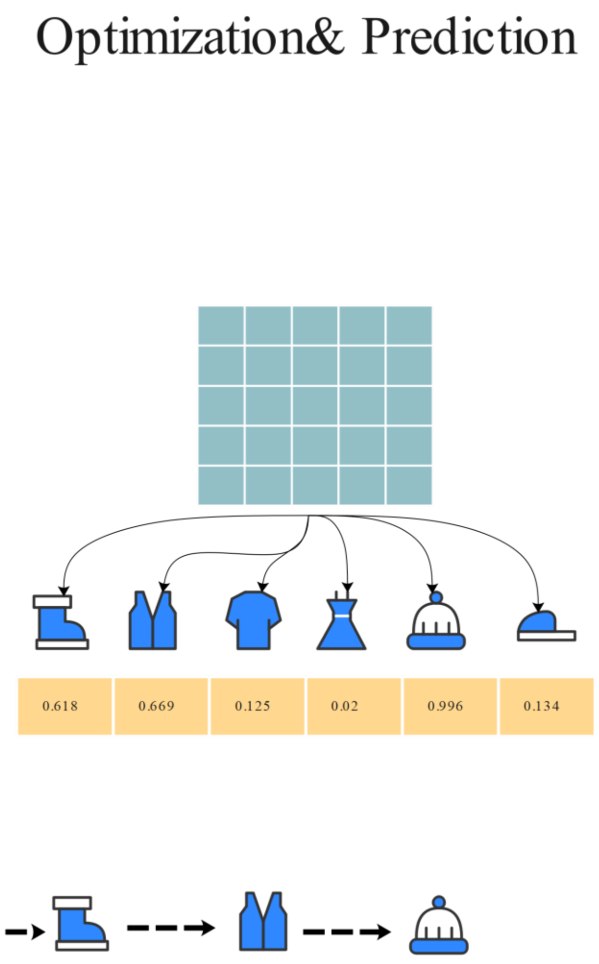
- User $u \in U$ and u's viewed item $v \in V^u = \{v_1, v_2 \dots\}$.
- $V^{u,+} = \{v_1^+, v_2^+ \dots v_\pi^+\} \subseteq V^u$ includes the interacted items
- $V^{u,-} = \{v_1^-, v_2^- \dots v_\tau^-\} \subseteq V^u$ includes the non-interacted items
  - $V^u \subseteq V$, τ and π the indexes
- A is adjacency matrix of a normal graph
- Hypergraph $\dot{G} = (V, \dot{E}, \dot{W})$
  - $\dot{E}$ - hyperedges
  - $\dot{W}$ - hyperweights
- d is the dimension, e.g. 64
- $\varphi(x) \coloneqq \frac{x}{\sqrt{\sum_{j \in x} j^2}}$ from a vector to a state

Framework

Proposed Model

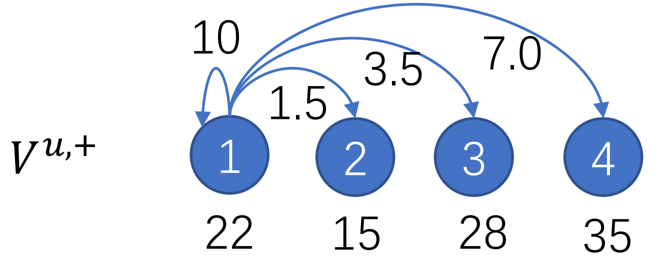
$$\dot{\rho} = [H, \rho] + \sum_{i=1}^{n} g_{i,j}\left(o_i \rho o_i^{\dagger} - \frac{1}{2}\{o_i^{\dagger} o_i, \rho\}\right)$$





1.1 Graph to hypergraph – item weights

- Motivation: capture high-order vertices
  - Contain more density
  - Contain more information
- Hypergraph: nodes, hyperedges, hyperweights
- Given any user u and u's items $V^{u,*}$
  - $V^{u,*}$ representing either $V^{u,+}$ or $V^{u,-}$
  - t representing either τ and π
  - Item similarities $\omega_{i,j} = v_i \odot (v_j)^T$, $i, j \leq t$ and $v_i \in \mathbb{R}^{1\times d}$
  - For each item: $\Omega_i = \frac{\sum_{j\leq t} \omega_{i,j}}{\sum_{i\leq t}\sum_{j\leq t} \omega_{i,j}}$ s.t. $\sum_{i\leq t} \Omega_i = 1$

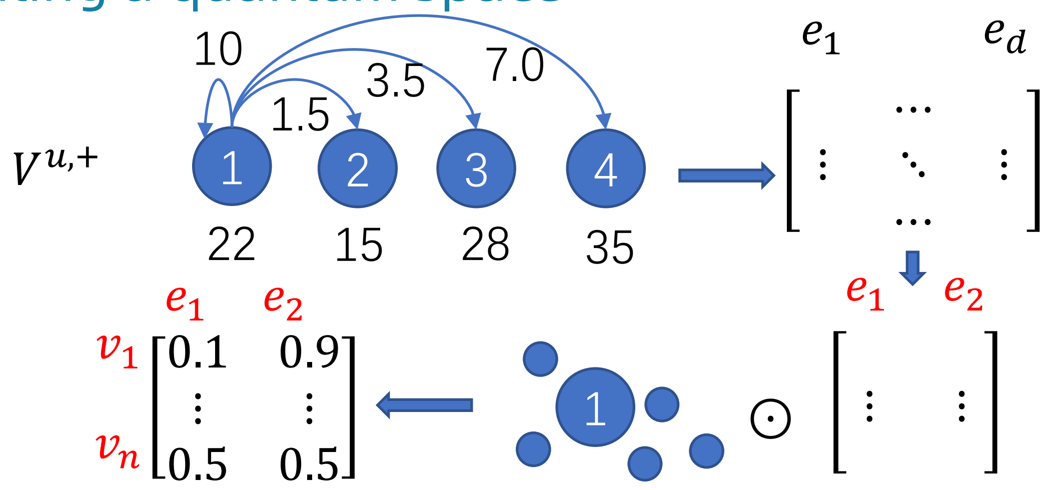


Importance of 1:
22/(22+15+28+35)=22%

Computer Science
College of Arts & Sciences
Georgia State University

1.2 Graph to hypergraph - hyperedges

- For each item $v \in V^{u,*}$, a hypergraph $\dot{g}_i^{u,*} = (V_i^{u,*}, \dot{E}, \dot{W})\ for\ i \leq t$ is constructed based on hyperedges and hyperweights
- Quantum-based hyperedge learning
  - Density matrix $\rho_{1\ldots t} = \sum_{i\leq t} \Omega_i \, |\varphi(v_i)\rangle\langle\varphi(v_i)| \in \mathbb{R}^{d\times d}$
  - There is a set of basis vector (span the event space) that makes $\rho_{1\ldots t}$ exists.
  - $B_{V^{u,*}} = \Re(\sigma(\rho_{1\ldots t})) \in \mathbb{R}^{d\times|\dot{E}|}$ is a matrix representing a quantum space
    - $\Re$ - ReLU
    - σ – Linear
    - $\dot{E}$ decides how many hyperedges in $\dot{g}_i^u$
  - $\dot{E} = V_t^{u,*} \odot B_{V^u} \in \mathbb{R}^{|V^u|\times|\dot{E}|}$
  - softmax($\dot{E}$) to range(0,1)



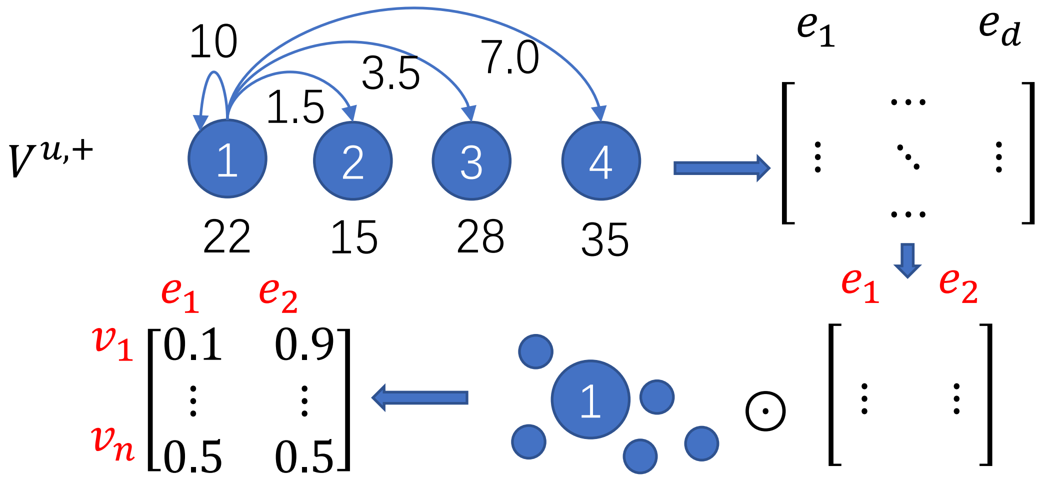


Figure 1

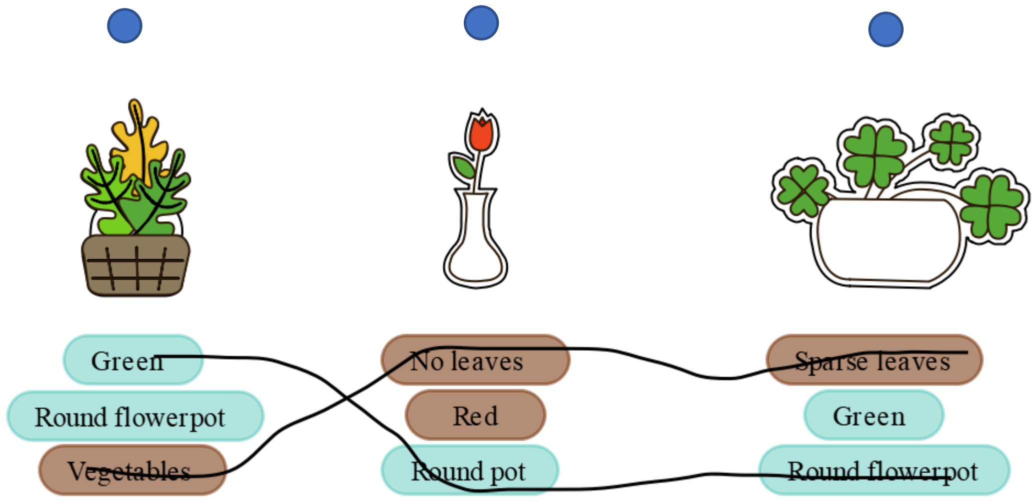


Figure 2

1.3 Graph to hypergraph - hyperweights

- One hyperedge to adjacency (learned) matrix $\ddot{E}_i = \dot{E}_{:,j} \otimes \left(\dot{E}_{:,j}\right)^T$
  - $\ddot{E} \in \mathbb{R}^{|\dot{E}| \times |V^{u,*}| \times |V^{u,*}|}$
- L_1 distance between any learned adjacency and real adjacency $\dot{W}_{i,i} = \sum\left(\left|\ddot{E}_i - A\right|\right)$
- So far, obtain $\dot{g}_t^{u,*} = (V_t^{u,*}, \dot{E}, softmaxDiag\left(\dot{W}\right))$

$$\dot{E} = \begin{array}{c c} & \begin{array}{cc} e_1 & e_2 \end{array} \\ \begin{array}{c} v_1 \\ \\ v_n \end{array} & \begin{bmatrix} 0.1 & 0.9 \\ \vdots & \vdots \\ 0.5 & 0.5 \end{bmatrix} \end{array} \Rightarrow$$

$$\ddot{E}_1 = \begin{array}{c c} & \begin{array}{ccc} v_1 & & v_n \end{array} \\ \begin{array}{c} v_1 \\ \\ v_n \end{array} & \begin{bmatrix} 0.01 & \cdots & 0.05 \\ \vdots & \ddots & \vdots \\ 0.05 & \cdots & 0.25 \end{bmatrix} \end{array} \qquad \ddot{E}_2 = \begin{array}{c c} & \begin{array}{ccc} v_1 & & v_n \end{array} \\ \begin{array}{c} v_1 \\ \\ v_n \end{array} & \begin{bmatrix} 0.81 & \cdots & 0.45 \\ \vdots & \ddots & \vdots \\ 0.45 & \cdots & 0.25 \end{bmatrix} \end{array}$$

$$- \; A \; \begin{array}{c c} & \begin{array}{ccc} v_1 & & v_n \end{array} \\ \begin{array}{c} v_1 \\ \\ v_n \end{array} & \begin{bmatrix} 0.01 & \cdots & 0.05 \\ \vdots & \ddots & \vdots \\ 0.05 & \cdots & 0.25 \end{bmatrix} \end{array} = \dot{W}_{i,i}$$

$$\dot{W} = \begin{bmatrix} 0.15 & 0 \\ 0 & 0.85 \end{bmatrix}$$

1.4 Hypergraph learning

- Update graph information via

$$\ddot{V}^{v} = D^{-1}\dot{E}\,WB^{-1}(\dot{E}\,)^{T}V_{t}^{u,*}\Theta$$

  - D and B are degree matrices of $\dot{g}_{t}^{u,*}$

1.5 Learn edge information

- Updated hypergraph

$$\ddot{g}_t^{u,*} = (\ddot{V}_t^{u,*}, \dot{E}, \dot{W})$$

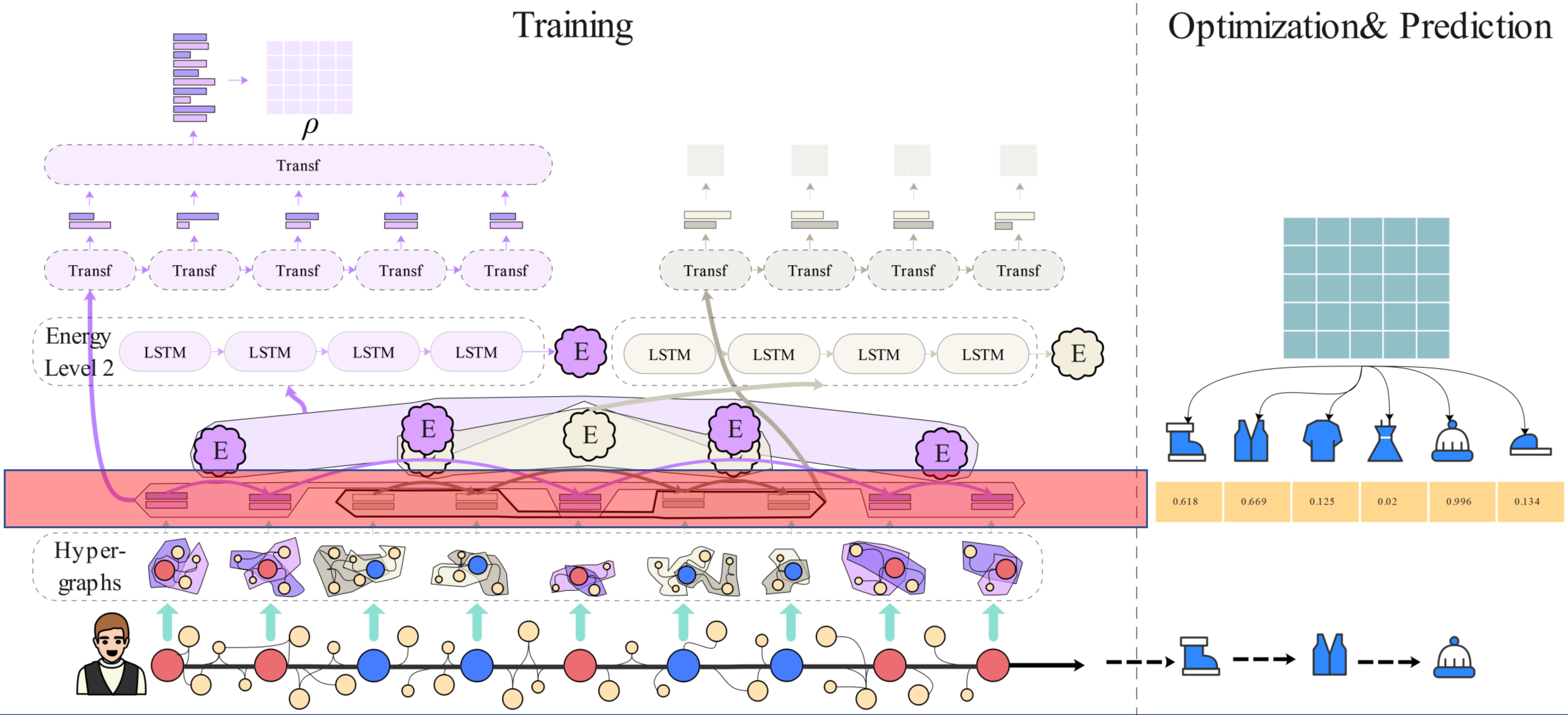
- Edge information

$$\ddot{s}_{t,j}^{u,*} = \ddot{e}_{t,j}^{u,*} = \frac{\sum_{v_i \in \dot{E}_{:,j}} \dot{E}_{i,j} v_i}{sum(\dot{E}_{:,j})}$$

Computer Science
College of Arts & Sciences
Georgia State University

Proposed Model

$$\dot{\rho} = [H, \rho] + \sum_{i=1}^{n} g_{i,j}(o_i \rho o_i^{\dagger} - \frac{1}{2}\{o_i^{\dagger} o_i, \rho\})$$



1.6 Hamiltonian

- New states:

$$\dddot{s}_t^{u,*} = \sum_{e \in \dot{E}_t^v} \ddot{s}_{t,j}^{u,*}$$

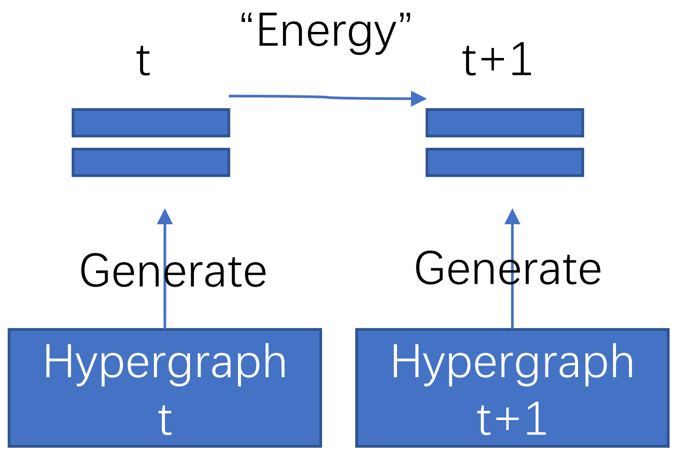
- Hamiltonian H is defined as:

$$h_{t,t+1}^{u,*} = (\dddot{s}_{t+1}^{u,*})^{-1} \odot \dddot{s}_t^{u,*} \in \mathbb{R}^{d*d}$$

- LSTM to learn final Hamiltonian:

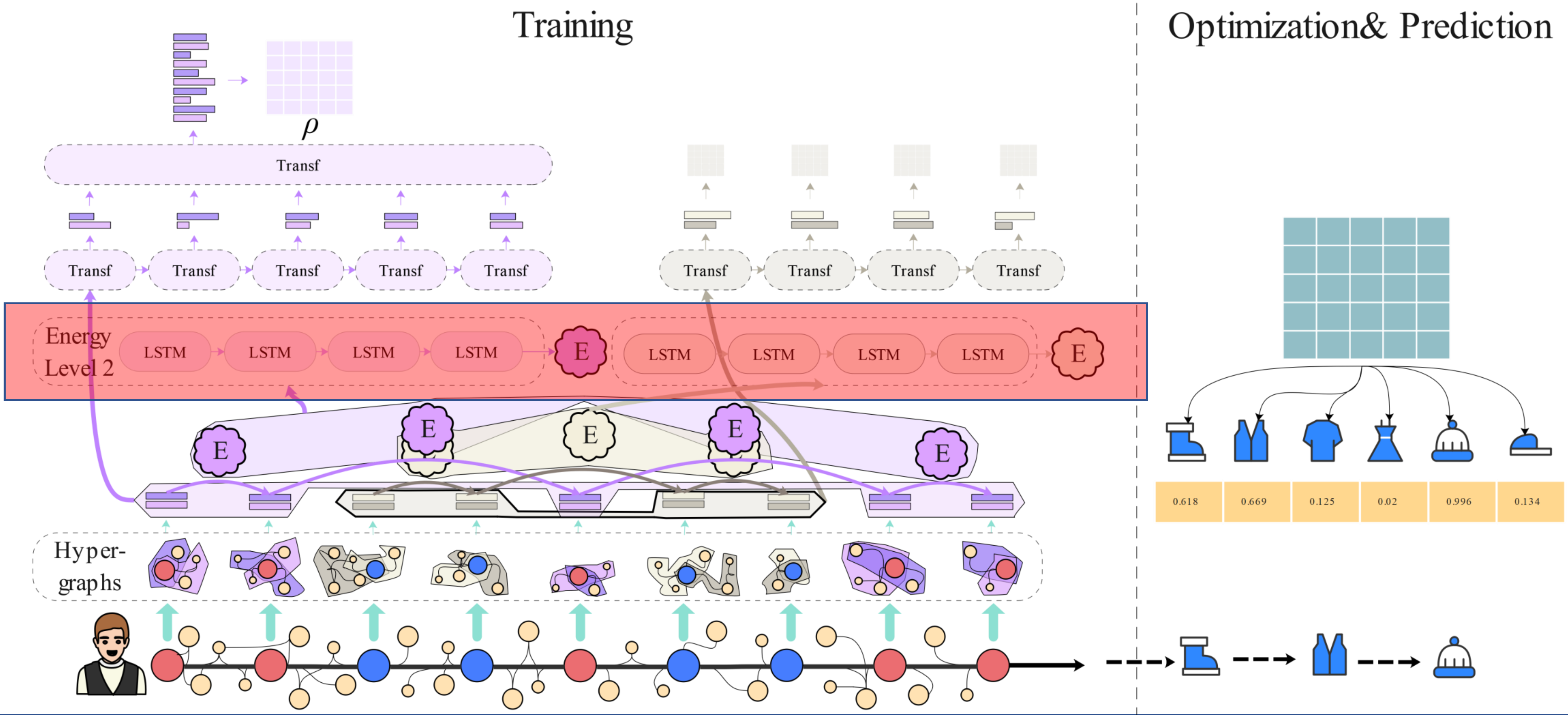
$$h_{t,t+1}^{u,*}{}' = LSTM\left(h_{t,t+1}^{u,*}, h_{t-1,t}^{u,*}{}'\right)$$

  - $h_{0,1} = zero\ matrix$



Proposed Model

$$\dot{\rho} = [H, \rho] + \sum_{i=1}^{n} g_{i,j}\left(o_i \rho o_i^{\dagger} - \frac{1}{2}\{o_i^{\dagger} o_i, \rho\}\right)$$



1.7 Transformer for all

- Learn density matrix

$$\ddot{s}_t^{u,+\,\prime} = Transformer(\ddot{s}_t^{u,+}, \ddot{s}_{t-1}^{u,+\,\prime})$$

  - $\ddot{s}_0^{u,+}$ =zero matrix

- Final states

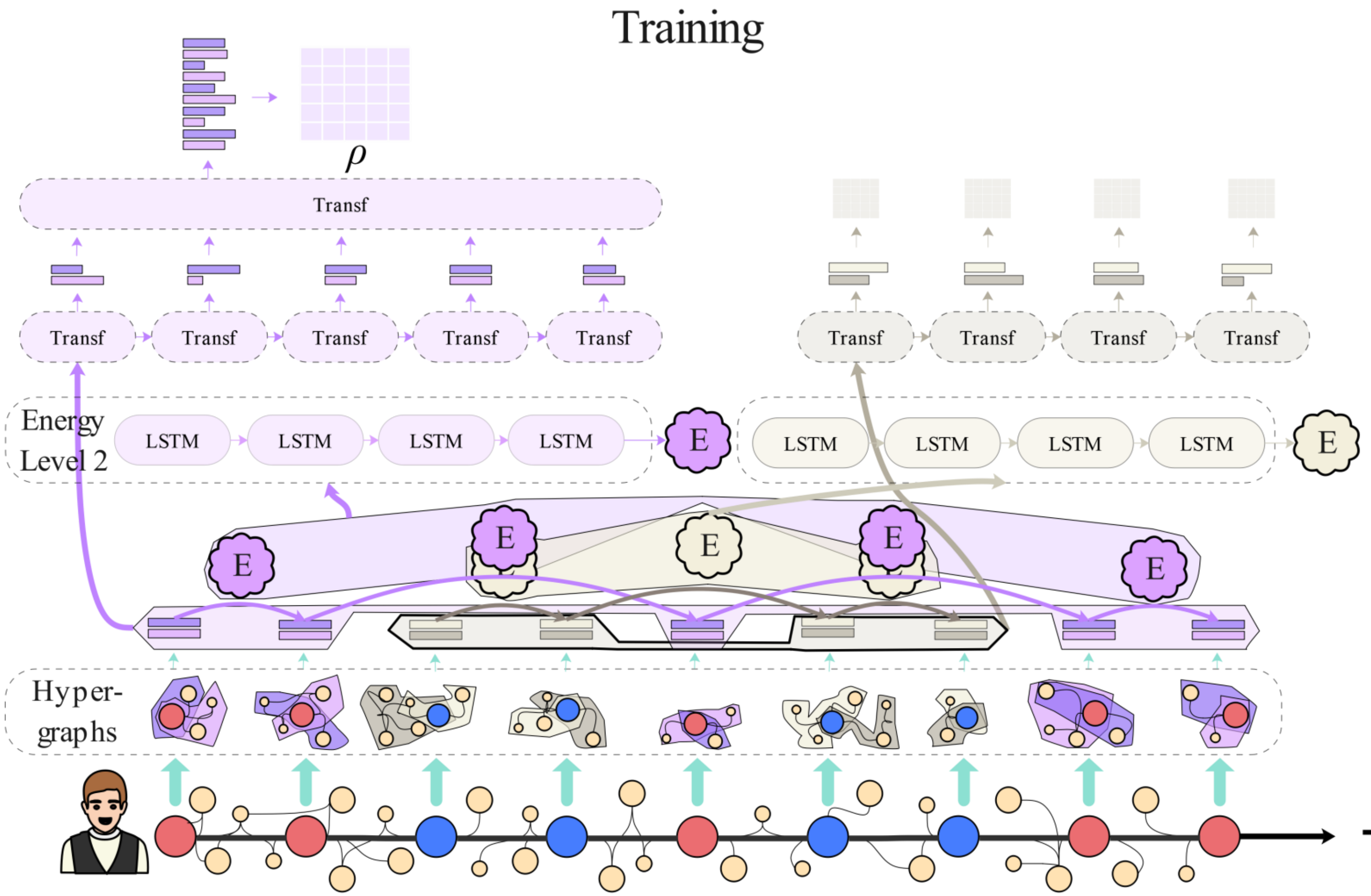
$$S^{u,+} = stack(\ddot{s}_1^{u,+\,\prime}, \ddot{s}_2^{u,+\,\prime}, \dots, \ddot{s}_\pi^{u,+\,\prime})$$

- Final attention

$$\beta^{u,+} = \frac{\sum_{\#\ of\ heads} Attention(S^{u,+})}{sum(\sum_{\#\ of\ heads} Attention(S^{u,+}))}$$

- Density matrix

$$\rho^{u,+} = \sum_{i \le |\beta^{u,+}|} \beta_i^{u,+} |\varphi(S_i^{u,+})\rangle\langle\varphi(S_i^{u,+})|$$



1.7 Transformer

Transformer

$transformer(x)$ is defined as:

1. Add position encoding $x + p$
2. Multi-Head Attention
   - $z_1 = selfAttention(X + P) = softmax(\frac{(x+p)M_1 \odot [(x+p)M_2]^T}{\sqrt{d}})(x + p)M_3$
   - Calculate $z_2 \dots z_k$
   - $\hat{z} = \sigma(z_1 \dots z_k)$
   - Note: Black part is function $Attention(\cdot)$
3. ADD & norm $x + z$
   - Norm is Layer Normalization

1.8 Jump operators

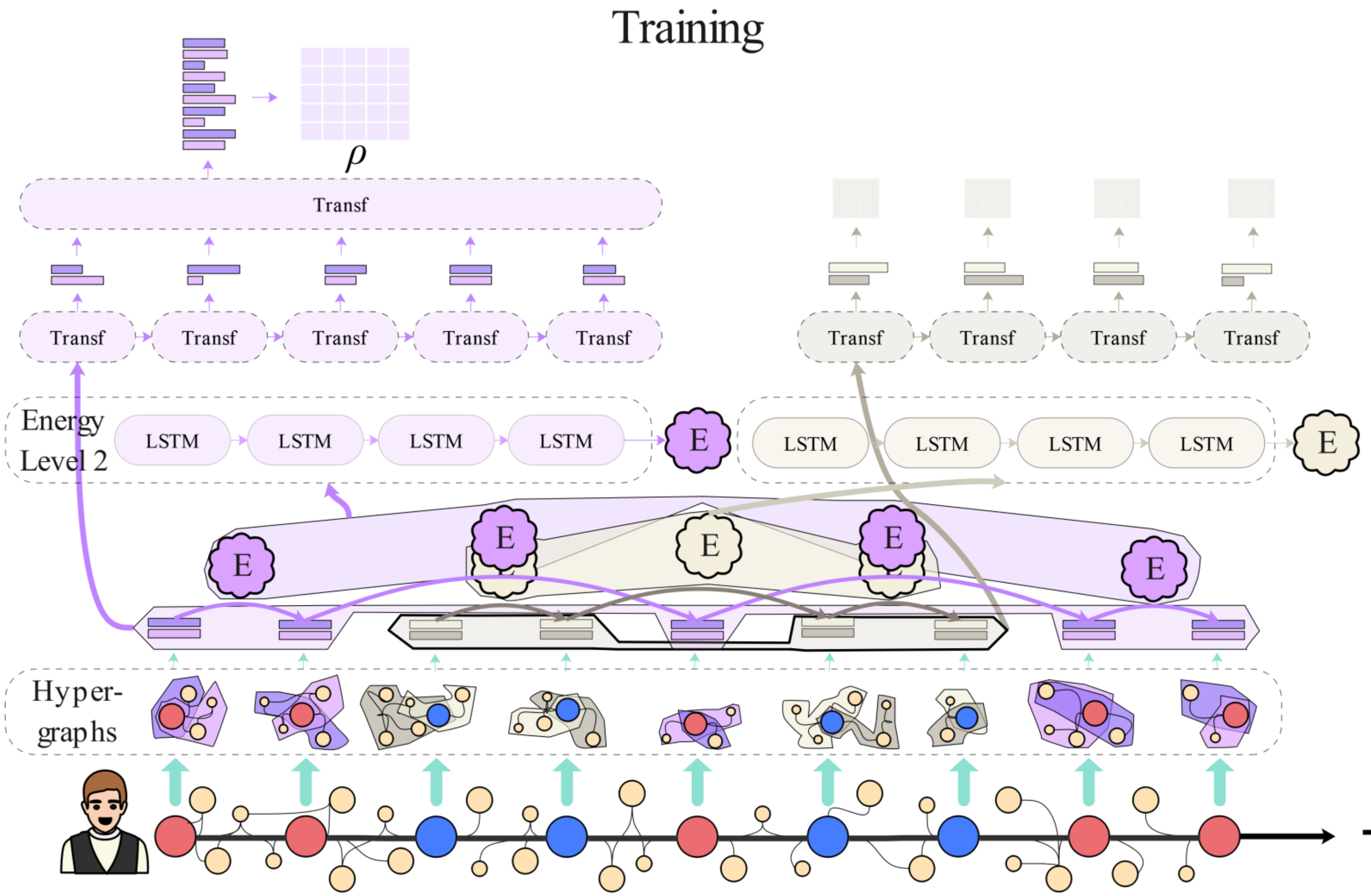
- Learn density matrix

$$\ddot{s}_t^{u,-\,\prime} = Transformer(\ddot{s}_t^{u,-}, \ddot{s}_{t-1}^{u,-\,\prime})$$

  - $\ddot{s}_0^{u,+}$ =zero matrix

- Jump operator

$$O_t^{u,-} = \sum_{x \in |\ddot{s}_t^{u,-\,\prime}|} |\varphi(x)\rangle\langle\varphi(x)|$$



1.9 preference matrix and score

- Preference direction matrix

$$P^u = -[H^+ + \delta H^-, \rho^{u,+}] + \lambda \sum_{t \le \tau} [O_t^{u,-} \rho^{u,+} O_t^{u,-} - \frac{1}{2} \{O_t^{u,-} O_t^{u,-}, \rho^{u,+}\}]$$

- Score: for an item v

$$y_v^u = [P^u \boxed{\sum_{i \le t} \Omega_i \, |\varphi(v_i)\rangle}] v_j$$

Direction that preference goes to   Mixed states

Where i is the index of interacted items and j is the index of candidate items

Computer Science
College of Arts & Sciences
Georgia State University

2.0 Optimization

- Bayesian personalized ranking
  - $L_{main} = \frac{1}{|U|} \sum_{u \in U} \sum_{v^+ \in V^+} \sum_{v^- \in V^-} -logSigmoid(y^u_{v^+} - y^u_{v^-})$
- Hermitian loss
  - $L_H = \frac{1}{|U|} \sum_{u \in U} [sum(|H^{u,+} - (H^{u,+})^\dagger| + |H^{u,-} - (H^{u,-})^\dagger|)]$

Experiments

Plan

Dataset

- MovieLen
- Amazon shopping
- Yelp

Experiments

- Accuracy & NDCG comparison
- Ablation study
- Case study
  - "Energy" visualization and analysis