Quantum Recommendation Systems

Irene Dovichi

Introduction to Quantum Computing

presentation a.y. 22/23

Università di Pisa

The Recommendation Problem

- Purpose: provide personalized recommendations to users based on the purchases and ratings they have made.
- Setting:
  - m users
  - n products
  - $P \in \mathcal{M}(m, n, \mathbb{R})$ preference matrix with a good rank-k approximation

  in practice: $m \approx 100$ million, $n \approx 1$ million, $k \approx 100$
- Classical algorithms run in time polynomial in the matrix dimension
- Here: quantum algorithm that runs in time $O(poly(k)polylog(mn))$ and which solves approximately the same problem as the classic version

Our model for the problem

	P_1	P_2	P_3	P_4	$\cdots$	$\cdots$	P_{n-1}	P_n
U_1	.8	.4	?	?	$\cdots$	$\cdots$	?	.9
U_2	.2	?	.6	?	$\cdots$	$\cdots$	.85	?
U_3	?	?	.8	.9	$\cdots$	$\cdots$	?	.2
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
U_m	?	.75	?	?	$\cdots$	$\cdots$	?	.2

Figure: Preference matrix P

	P_1	P_2	P_3	P_4	$\cdots$	$\cdots$	P_{n-1}	P_n
U_1	1	0	0	0	$\cdots$	$\cdots$	0	1
U_2	0	0	0	0	$\cdots$	$\cdots$	1	0
U_3	0	0	1	1	$\cdots$	$\cdots$	0	0
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
U_m	0	1	0	0	$\cdots$	$\cdots$	0	0

Figure: Rounded preference matrix T

Observation

The rounding process can be done in different ways.

The low-rank assumption

Observation

It is reasonable to assume that the matrices P and T have a good low-rank approximation.

- There are k types of users, and the users of each type agree on the items of greatest value.
- There are some basic parameters that determine the preference for a product: price, quality, brand, popularity.
- Empirical evidence.

The low-rank approximation

The low-rank approximation of T can be computed as follows:

- define the subsample matrix $\widehat{T}$ as:

$$\widehat{T}_{ij} = \begin{cases} T_{ij}/p & p \\ 0 & 1-p \end{cases}$$

- perform the SVD and project $\widehat{T}$ to its top-k right singular vectors, obtaining $\widehat{T}_k$

Theorem (SVD)

A matrix $A \in \mathcal{M}(m, n, \mathbb{R})$ can be decomposed as:

$$A = U\Sigma V^t = \sum_{i=1}^{rnkA} \sigma_i u_i v_i^t$$

where $U \in \mathcal{M}(m, m, \mathbb{R})$ and $V \in \mathcal{M}(n, n, \mathbb{R})$ are orthogonal, and $\Sigma \in \mathcal{M}(m, n, \mathbb{R})$ is diagonal with nonnegative entries.

Sampling vs Reconstructing

Definition

Given a matrix $A \in \mathcal{M}(m, n, \mathbb{R})$, we *sample* from A when we pick an element (i, j) with probability $|A_{ij}|^2/||A||_F^2$. We write: $(i, j) \sim A$.

Observation

Sampling from T always gives a good recommendation.

Sampling vs Reconstructing

Definition

Given a matrix $A \in \mathcal{M}(m, n, \mathbb{R})$, we *sample* from A when we pick an element (i, j) with probability $|A_{ij}|^2/||A||_F^2$. We write: $(i, j) \sim A$.

Observation

Sampling from T always gives a good recommendation.

Lemma

Let $\widetilde{T}$ be a matrix such that $||T - \widetilde{T}||_F \leq \varepsilon ||T||_F$. The probability that $(i, j) \sim \widetilde{T}$ is bad is:

$$Pr[(i, j)\ bad] \leq \left(\frac{\varepsilon}{1 - \varepsilon}\right)^2$$

Sampling vs Reconstructing

Definition

Given a matrix $A \in \mathcal{M}(m, n, \mathbb{R})$, we *sample* from A when we pick an element (i, j) with probability $|A_{ij}|^2/||A||_F^2$. We write: $(i, j) \sim A$.

Observation

Sampling from T always gives a good recommendation.

Lemma

Let $\widetilde{T}$ be a matrix such that $||T - \widetilde{T}||_F \leq \varepsilon ||T||_F$. The probability that $(i, j) \sim \widetilde{T}$ is bad is:

$$Pr[(i, j)\ bad] \leq \left(\frac{\varepsilon}{1 - \varepsilon}\right)^2$$

But which $\widetilde{T}$ should we choose?

Sampling matrix

The matrix from which we will sample is:

$$\widehat{T}_{\geq\sigma,\nu}$$

which is the projection of $\widehat{T}$ onto the space spanned by the right singular vectors v_i which correspond to the singular values $\geq \sigma$, and some in the range $[(1-\nu)\sigma, \sigma)$.

Observation

Sampling from $\widehat{T}_{\geq\sigma,\nu}$ is enough to have good recommendations for 'typical' users, as implied by the following Theorem.

Theorem

For a certain value of the probability p, and for $\nu = 1/3$:

$$||T - \widehat{T}_{\geq\sigma,\nu}||_F \ \leq \ 9\,\varepsilon\,||T||_F$$

How to sample?

- We use a quantum procedure that given $|\widehat{T}_i\rangle$ outputs $|(\widehat{T}_{\geq\sigma,\nu})_i\rangle$.
- We measure in the computational basis.

The procedure is called Quantum Projection Algorithm and it requires two tools:

1. an appropriate data structure
2. an efficient quantum algorithm for Singular Value Estimation

and it runs in time $O(polylog(mn))$.

Quantum Singular Value Estimation

Theorem (SVE)

Let $A \in \mathcal{M}(m, n, \mathbb{R})$ be a matrix with SVD decomposition $A = \sum_{i=1}^{rnkA} \sigma_i u_i v_i^t$ stored in an appropriate data structure, and $\varepsilon > 0$. There is an algorithm (SVE algorithm) that performs the mapping:

$$\sum_{i=1}^{n} \alpha_i |v_i\rangle \mapsto \sum_{i=1}^{n} \alpha_i |v_i\rangle |\bar{\sigma}_i\rangle$$

where $|\bar{\sigma}_i - \sigma_i| \leq \varepsilon ||A||_F$, in $O(polylog(mn)/\varepsilon)$ time.

SVE - idea

- Factorize A with two isometries: $\frac{A}{||A||_F} = P^t Q$
- Use P, Q to define a unitary W such that: $WQv_i = e^{i\theta_i} Qv_i$, with $\sigma_i = ||A||_F \, cos(\theta_i/2)$
- Use QPE to estimate the θ_i

Theorem (QPE)

Let U be a unitary operator such that $U|v_j\rangle = e^{i\theta_j}|v_j\rangle$, and $\varepsilon > 0$. There is an algorithm (QPE algorithm) that performs the mapping:

$$\sum_{j=1}^{n} \alpha_j |v_j\rangle \mapsto \sum_{j=1}^{n} \alpha_j |v_j\rangle |\bar{\theta}_j\rangle$$

where $|\bar{\theta}_j - \theta_j| \leq \varepsilon$, in $O(T(U) log\, n/\varepsilon)$ time.

SVE - components

- Factorize A with two isometries: $\frac{A}{||A||_F} = P^t Q$
  - Take $P \in \mathcal{M}(mn, m, \mathbb{R})$ with columns

$$P^i = e_i \otimes \frac{A_i}{||A_i||}$$

  - Take $Q \in \mathcal{M}(mn, n, \mathbb{R})$ with columns

$$Q^j = \frac{\widetilde{A}}{||A||_F} \otimes e_j$$

where $\widetilde{A} \in \mathbb{R}^m$ with components $\widetilde{A}_i = ||A_i||$

SVE - components

- Factorize A with two isometries: $\frac{A}{||A||_F} = P^t Q$
  - Take $P \in \mathcal{M}(mn, m, \mathbb{R})$ with columns

$$P^i = e_i \otimes \frac{A_i}{||A_i||}$$

  - Take $Q \in \mathcal{M}(mn, n, \mathbb{R})$ with columns

$$Q^j = \frac{\widetilde{A}}{||A||_F} \otimes e_j$$

    where $\widetilde{A} \in \mathbb{R}^m$ with components $\widetilde{A}_i = ||A_i||$

- Use P, Q to define a unitary W such that: $WQv_i = e^{i\theta_i} Qv_i$
  - Take $W = (2PP^t - I)(2QQ^t - I)$

SVE - components

- Factorize A with two isometries: $\frac{A}{||A||_F} = P^t Q$
  - Take $P \in \mathcal{M}(mn, m, \mathbb{R})$ with columns

$$P^i = e_i \otimes \frac{A_i}{||A_i||}$$

  - Take $Q \in \mathcal{M}(mn, n, \mathbb{R})$ with columns

$$Q^j = \frac{\widetilde{A}}{||A||_F} \otimes e_j$$

    where $\widetilde{A} \in \mathbb{R}^m$ with components $\widetilde{A}_i = ||A_i||$

- Use P, Q to define a unitary W such that: $WQv_i = e^{i\theta_i} Qv_i$
  - Take $W = (2PP^t - I)(2QQ^t - I)$
- It holds that $\sigma_i = ||A||_F \, cos(\theta_i/2)$
  - $cos(\theta_i/2) = Pu_i \cdot Qv_i/||Pu_i|| \, ||Qv_i|| = u_i^t A v_i/||A||_F = \sigma_i/||A||_F$

SVE - procedure

<u>Input</u> : $A \in \mathcal{M}(m, n, \mathbb{R})$, $x \in \mathbb{R}^n$ stored in an appropriate data structure, $\varepsilon > 0$.

<u>Output</u> : The state $\sum \alpha_i |v_i\rangle |\bar{\sigma}_i\rangle$.

1: Create $|x\rangle = \sum_{i=1}^{n} \alpha_i |v_i\rangle$

2: Append the register $|0^{\lceil log\, m \rceil}\rangle$ and create $|Qx\rangle = \sum_{i=1}^{n} \alpha_i |Qv_i\rangle$:

$$|0^{\lceil log\, m \rceil}\, x\rangle = \sum_{j=1}^{n} x_j \, |0^{\lceil log\, m \rceil}\, j\rangle \longrightarrow \sum_{j=1}^{n} x_j \, |\widetilde{A}\, j\rangle = |Qx\rangle$$

3: Apply the QPE for W on $|Qx\rangle$: $\sum_{i=1}^{n} \alpha_i |Qv_i\rangle |\bar{\theta}_i\rangle$ (with 2ε)

4: Compute $\bar{\sigma}_i = ||A||_F \, cos(\bar{\theta}_i/2)$ and apply the IQPE: $\sum_{i=1}^{n} \alpha_i |Qv_i\rangle |\bar{\sigma}_i\rangle$

5: Apply the inverse transformation of line 2: to get: $\sum_{i=1}^{n} \alpha_i |v_i\rangle |\bar{\sigma}_i\rangle$

SVE - analysis

- $|\bar{\sigma}_i - \sigma_i| \leq \varepsilon ||A||_F$:

$$\begin{aligned} |\bar{\sigma}_i - \sigma_i| &= ||A||_F \, |cos(\bar{\theta}_i/2) - cos(\theta_i/2)| \\ &\leq |sin(\phi)| \frac{|\bar{\theta}_i - \theta_i|}{2} ||A||_F \\ &\leq \varepsilon ||A||_F \end{aligned}$$

  where we applied the Mean Value Theorem to $f(t) = cos(t)$ (ϕ is between $\bar{\theta}_i/2$ and $\theta_i/2$), and we used that $|\bar{\theta}_i - \theta_i| \leq 2\varepsilon$ (we performed the QPE with 2ε).

- $O(polylog(mn)/\varepsilon)$ time :

  the unitary W can be implemented in time $O(polylog(mn))$ and the QPE runs in time $O(T(W) log\, n/\varepsilon)$.

Pseudo-inverse matrix

Definition

Let $A = U\Sigma V^t = \sum_{i=1}^{rnkA} \sigma_i u_i v_i^t \in \mathcal{M}(m, n, \mathbb{R})$. The Moore-Penrose inverse of A is the matrix:

$$A^+ = V\Sigma^+ U^t = \sum_{i=1}^{rnkA} \frac{1}{\sigma_i} v_i u_i^t$$

Observation

- $A^+ A$ is the projection onto the row space $Row(A)$.
- $Row(A) = Span\{v_i\}$.

Quantum Projection Algorithm

Input : $A \in \mathcal{M}(m, n, \mathbb{R}),\ x \in \mathbb{R}^n$ stored in an appropriate data structure, and the parameters $\sigma,\ \nu > 0$.

Output : The state $|A^{+}_{\geq\sigma,\nu} A_{\geq\sigma,\nu}\, x\rangle$ with probability $\geq 1 - 1/poly(n)$.

1: Create $|x\rangle = \sum_{i=1}^{n} \alpha_i |v_i\rangle$

2: Apply the SVE on $|x\rangle$: $\sum_{i=1}^{n} \alpha_i |v_i\rangle |\bar{\sigma}_i\rangle$ (with $\varepsilon = \nu\sigma/2||A||_F$)

3: Apply the unitary operator $|t\rangle|0\rangle \mapsto \begin{cases} |t\rangle|0\rangle & \text{if } t \geq (1 - \frac{\nu}{2})\,\sigma \\ |t\rangle|1\rangle & \text{otherwise} \end{cases}$ on a second register : $\sum_{i \in S} \alpha_i |v_i\rangle |\bar{\sigma}_i\rangle |0\rangle + \sum_{i \in S^C} \alpha_i |v_i\rangle |\bar{\sigma}_i\rangle |1\rangle$

4: Apply the ISVE on the state in line 3: to erase the $|\bar{\sigma}_i\rangle$s :

$$\sum_{i \in S} \alpha_i |v_i\rangle |0\rangle + \sum_{i \in S^C} \alpha_i |v_i\rangle |1\rangle = \beta |A^{+}_{\geq\sigma,\nu} A_{\geq\sigma,\nu}\, x\rangle |0\rangle + \sqrt{1-\beta^2}\, |A^{+}_{\geq\sigma,\nu} A_{\geq\sigma,\nu}\, x\rangle^{\perp} |1\rangle$$

5: Post-select on getting outcome $|0\rangle$: $|A^{+}_{\geq\sigma,\nu} A_{\geq\sigma,\nu}\, x\rangle$ in the first register.

Quantum Projection Algorithm - analysis

Theorem

The Quantum Projection Algorithm outputs $|A^{+}_{\geq\sigma,\nu} A_{\geq\sigma,\nu}\, x\rangle$ with probability $\geq 1 - 1/poly(n)$ in time

$$O\Big(\frac{polylog(mn)||A||_F\, ||x||^2}{\sigma||A_{\geq\sigma} A^{+}_{\geq\sigma} x||^2}\Big).$$

Quantum Recommendation Algorithm

Input : A subsample matrix $\widehat{T}$ stored in an appropriate data structure, and a user index i.

Output : A product index j.

1: Applying the Quantum Projection Algorithm with input

$$A = \widehat{T} \qquad \sigma = \sqrt{\varepsilon^2 p/2k}\,||\widehat{T}||_F$$
$$x = \widehat{T}_i \qquad \nu = 1/3$$

we will get the state $|\widehat{T}^{+}_{\geq\sigma,\nu}\widehat{T}_{\geq\sigma,\nu}\widehat{T}_i\rangle$ with probability at least $1 - 1/poly(n)$.

2: Measure the state $|\widehat{T}^{+}_{\geq\sigma,\nu}\widehat{T}_{\geq\sigma,\nu}\widehat{T}_i\rangle$ in the computational basis.

Observation

The state $|\widehat{T}^{+}_{\geq\sigma,\nu}\widehat{T}_{\geq\sigma,\nu}\widehat{T}_i\rangle$ is exactly the state $|(\widehat{T}_{\geq\sigma,\nu})_i\rangle$.

References

[1] I. Kerenidis and A. Prakash. *Quantum Recommendation Systems*. 2016. DOI: 10.48550/arXiv.1603.08675.

[2] C. Shao and H. Xiang. *Quantum Circulant Preconditioner for Linear System of Equations*. 2018. DOI: 10.48550/arXiv.1807.04563.

[3] L. N. Trefethen and D. Bau. *Numerical Linear Algebra*. SIAM, 1997.