A Separability-Entanglement Classifier via Machine Learning

Zhengfeng Ji (UTS:QSI)

ADC 2017, Brisbane
A Joint Work With

Sirui Lu, Shilin Huang, Keren Li, Jun Li, Jianxin Chen,
Dawei Lu, Yi Shen, Duanlu Zhou and Bei Zeng
Outline

1. Quantum Entanglement: Basics
2. Entanglement Detection as a Classification Task
3. A Separability-Entanglement Classifier
…, then they can no longer be described in the same way as before, viz. by endowing each of them with a representative of its own. I would not call that one but rather the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought. By the interaction, the two representatives [the quantum states] have become entangled.

— Erwin Schrödinger
Entanglement as Correlation

• EPR paradox (Einstein-Podolsky-Rosen 1935)

Two ways to look at a qubit:

EPR State: \( \frac{|00\rangle + |11\rangle}{\sqrt{2}} = \frac{|++\rangle + |--\rangle}{\sqrt{2}} \).
At the Heart of Quantum Information

• Applications:
  - Quantum cryptography
  - Quantum teleportation
  - Quantum computing
  - Quantum metrology
  - ...

• Understanding the role of entanglement is one of the fundamental open problems in quantum mechanics

China's satellite-based distribution of entangled photon pairs over 1,200km
Quantum States

A quantum state is described by a density matrix $\rho$:

1. Hermitian;
2. Semidefinite positive, $\rho \geq 0$;
3. Normalization: $\text{Tr} \rho = 1$. 
Entangled States (Pure Case)

- A bipartite system \((A, B)\) with the Hilbert space \(\mathcal{H}_A \otimes \mathcal{B}_B\)
- \(\mathcal{H}_A\) has dimension \(d_A\) and \(\mathcal{H}_B\) has dimension \(d_B\)

A pure state \(\left|\psi\right\rangle\) is a **product state** if it can be written as \(\left|\psi\right\rangle = \left|\psi_A\right\rangle \otimes \left|\psi_B\right\rangle\). Otherwise, it is **entangled**.

- Example: The EPR state is entangled.
Entangled States (Mixed Case)

A mixed state $\rho_{AB}$ is a **separable** if

$$\rho_{AB} = \sum_{i} p_{i} \rho_{A,i} \otimes \rho_{B,i}$$

with a probability distribution $p_{i} \geq 0$ and $\sum_{i} p_{i} = 1$

Otherwise, $\rho_{AB}$ is **entangled**.
Example: The Werner states are defined as

\[ \rho = p |\psi\rangle \langle \psi| + (1 - p) \mathbb{I}/4, \]

where \( |\psi\rangle = (|01\rangle - |10\rangle)/\sqrt{2} \).


State \( \rho \) is entangled iff \( p > 1/3 \).
The Set of Separable States

The set of separable states are **convex**.
Separability Problem

To determine whether a given state is separable or entangled.

- This problem is NP-hard in general.
  
  \[ L. \text{ Gurvits, STOC (2003)} \]

- Many entanglement criterions proposed
  - Positive partial transpose criterion (PPT)
  - Linear (or nonlinear) entanglement witness
  - Symmetric extension
  - ...


Positive Partial Transpose Criterion

- Partial transpose as a block matrix operation

\[
\rho = \begin{bmatrix}
\rho_{11} & \cdots & \rho_{1d_A} \\
\vdots & \ddots & \vdots \\
\rho_{d_A1} & \cdots & \rho_{d_Ad_A}
\end{bmatrix}, \quad \rho^{TB} = \begin{bmatrix}
\rho_{11}^T & \cdots & \rho_{1d_A}^T \\
\vdots & \ddots & \vdots \\
\rho_{d_A1}^T & \cdots & \rho_{d_Ad_A}^T
\end{bmatrix}.
\]

- If the partial transpose \(\rho^{TB}\) has non-negative eigenvalues, we say the state \(\rho\) fulfills the positive partial transpose (PPT) condition.
• PPT condition as a separability criterion

1. If $\rho$ is separable, then it is PPT.
2. The converse is true only for $2 \times 2$ or $2 \times 3$ systems.
Entanglement Witness

- For all entangled state $\rho$, there exists an observable $W$ such that
  1. $\text{Tr}(W \rho) < 0$;
  2. $\text{Tr}(W \sigma) \geq 0$, for $\sigma$ separable.

- The observable $W$ is called an entanglement witness.
Machine Learning Approach

- Many criteria only detect a limited set of entangled states. In contrast, a classifier trained with machine learning techniques can handle a variety of input states.

- Use machine learning techniques to feel the geometry of the separable states.
Classification

- Learning (training): learn a model using the training data;
- Testing: test the model using unseen test data to assess the model accuracy.
Feature Vector Representation

- Any quantum state $\rho$ acting on $\mathcal{H}_A \otimes \mathcal{H}_B$ can be represented as a real vector $\mathbf{x} \in X \subseteq \mathbb{R}^{d_A^2 d_B^2 - 1}$.
- We refer to $\mathbf{x}$ as the feature vector of $\rho$ and $X$ the feature space.

Training Dataset

$\mathcal{D}_{\text{train}} = \{ (\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots \}$, $\mathbf{x}_i \in X$ is the $i$-th sample, and $y_i$ is its label:

$$
\begin{cases} 
  y_i = 1, & \text{if } x_i \text{ is entangled,} \\
  y_i = -1, & \text{if } x_i \text{ is separable.}
\end{cases}
$$
Classifier Training

- The aim of supervised learning is to train a classifier $h : X \rightarrow \{-1, 1\}$, where $h$ is expected to be close to the true decision function.

- To evaluate how well $h$ fits the training data $\mathcal{D}_{\text{train}}$, a loss function is defined as

$$
\mathcal{L}(h, \mathcal{D}_{\text{train}}) = \frac{1}{|\mathcal{D}_{\text{train}}|} \sum_{(x_i, y_i) \in \mathcal{D}_{\text{train}}} 1(y_i \neq h(x_i)),
$$

- Generalization

For a generic new input test dataset $\mathcal{D}_{\text{test}}$ that contains previously unseen data, function $\mathcal{L}(h, \mathcal{D}_{\text{test}})$ gives a quantification of the generalization error from $\mathcal{D}_{\text{train}}$ to $\mathcal{D}_{\text{test}}$. 
The error rates of the classifiers trained by different algorithms.

<table>
<thead>
<tr>
<th>Method</th>
<th>Bagging</th>
<th>Boosting</th>
<th>SVM</th>
<th>Decision Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error (%)</td>
<td>12.03</td>
<td>14.8</td>
<td>8.4</td>
<td>23.3</td>
</tr>
</tbody>
</table>

The error rate is difficult to be reduced. This suggests that the boundary of separable states has a very complicated shape.

Maybe we can try to add some prior information to the learning process.
State \( \rho_{AB} \) is \( k \)-symmetric extendible if there is a state \( \rho_{AB_1 \cdots B_k} \) whose marginals on \( A, B_i \) all equal to \( \rho_{AB} \).

- Let \( E_k \) denote the set of \( k \)-symmetric extendible set. Then

\[
E_1 \supset E_2 \supset \cdots E_\infty = \mathcal{S}.
\]
The approximation of the separable set by $\kappa$-extendible states.
Approximate From Inside: CHA

Randomly sample \( m \) product states \( \mathbf{c}_1, \ldots, \mathbf{c}_m \) and form a convex hull \( \mathcal{C} = \text{conv}(\{\mathbf{c}_1, \ldots, \mathbf{c}_m\}) \). \( \mathcal{C} \) approximates the set of separable states.

\[
\max\{\alpha \mid \alpha \mathbf{x} \in \mathcal{C}\}.
\]
A Qutrit-Qutrit Example: Tile States

$$\rho_{\text{tiles}} = \left( \mathbb{I} - \sum_{i=1}^{5} |v_i\rangle\langle v_i| \right) / 4, \text{ where}$$

$$|v_1\rangle = (|00\rangle - |01\rangle) / \sqrt{2},$$

$$|v_2\rangle = (|21\rangle - |22\rangle) / \sqrt{2},$$

$$|v_3\rangle = (|02\rangle - |12\rangle) / \sqrt{2},$$

$$|v_4\rangle = (|10\rangle - |20\rangle) / \sqrt{2},$$

$$|v_5\rangle = (|0\rangle + |1\rangle + |2\rangle) \otimes 2 / 3.$$
Consider a 2-qutrit state:

\[ \rho_p = p\mathbb{I}/(d_A d_B) + (1 - p)\rho_{\text{tiles}}. \quad 0 \leq p \leq 1, \]

- \( \rho_{\text{tiles}} \) is entangled and there must exist a critical point \( p^* \in [0, 1) \)

- In a lecture note by Nathaniel Johnston (2014):

  ... Thus there is a rather large gap of values \( p \in (0.1351, 0.4357) \) where we do not know whether or not \( \rho_p \) is entangled.

- Our result using CHA:

  \[ p^* \in (0.1351, 0.1352) \]
Combining Convex Hull Approximation and Supervised Learning

- To incorporate the convex hull information into the dataset, we extend the feature vector

\[ x \rightarrow (x, \alpha(C, x)). \]

So the new feature vector contains the geometric information of the convex hull.

- Correspondingly, the training dataset is

\[ D_{\text{train}} = \{(x_1, \alpha_1, y_1), \ldots, (x_n, \alpha_n, y_n)\}. \]
Numerical Tests on Two-qubit and Two-qutrit Systems

- Our learning procedure:
  1. build a convex hull approximation
  2. embed information of the convex hull in the training dataset
  3. randomly draw a training subset from the whole dataset
  4. decision trees learning: build a sub-classifier for the training subset
  5. repeat step (3); (4) many times
  6. ensemble learning: combine the sub-classifiers to form the final classifier
Results

(a) $\alpha^{-1}$ vs $\langle H_1 \rangle$

(b) $\langle P_1 \rangle$

(c) Error vs $m \times 10^3$

(d) Error vs $m \times 10^4$
Conclusions

- Our method outperforms the existing popular entanglement detection methods in accuracy and speed in the case of $2 \times 2$ and $3 \times 3$ systems.

- Future work
  - Deal with higher dimensions
  - Use original feature vectors
  - Use raw experimental data
  - Employ more advanced machine learning tools
THANKS!