Presentation on March 20, 2025

Theorem 5.8: Embedding Large Subsets of Finite Metric Spaces into Euclidean Space

Quan Jin

School of Mathematical Sciences
Shanghai Jiao Tong University
remotable@sjtu.edu.cn

Various ways to measure distance

Standard way to measure distance: Euclidean distance, which is defined as

$$\rho(A, B) := \|A - B\| = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

It describes the shortest path between two points on a plane, as shown in Figure1(a).

But in reality, sometimes we can't go directly from point A to point B. For example, in a city, we can only go along the streets. In this case, the distance between two points will be like Figure1(b).

Also, sometimes we might feel the longer route "shorter" if we take a faster kind of transportation, as shown in Figure1(c).

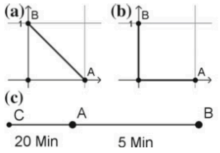

Figure 1: Various ways to measure distance

Metric Space

From the above examples, we can see that the distance in (a) and (b) have some properties in common, but the distance in (c) is different.

In mathematics, we define a metric space as a set with a distance function that satisfies the following properties:

i. $\rho(A, B) \geq 0$

ii. $\rho(A, B) = 0$ if and only if $A = B$

iii. $\rho(A, B) = \rho(B, A)$

iv. $\rho(A, B) + \rho(B, C) \geq \rho(A, C)$ for all A, B, C

Measure the "distance" between people

With the definition of metric space, we can define a "distance" between people.

We can define all your friends to be at "distance" 1 to you, then friends of your friends at "distance" 2, their friends at "distance" 3, and so on. Assuming the world is "connected", that is, there is a chain of friends between any two people A and B, we can define $\rho(A, B)$ to be the shortest "length" of such a chain.

Once again, this "distance" ρ satisfies all the properties (i)–(iv) listed above, which means it is a metric.

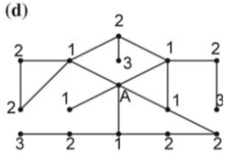

Figure 2: An Example of "distance" between people

The impossibility of an exact representation

Ideally, if we draw a picture on a plane, with each person represented by a point, the (standard Euclidean) distance between points representing A and B should be exactly (A, B).

However, this kind of exact representation is rarely possible.

For example, imagine four people A, B, C, D, such as A is a friend of everyone else, but B, C, D are not friends with each other.

In this case, $\rho(A, B) = \rho(A, C) = \rho(A, D) = 1$ but $\rho(B, C) = \rho(C, D) = \rho(D, B) = 2$, which is impossible to represent in a Euclidean Space.

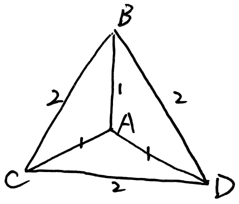

Figure 3: The impossible picture

Distortion

Now that we can't draw an exact picture, we can try an approximate one. We can find that if we can tolerate some errors, such as changing the distance of friends to 1.2, or changing the distance of friends of friends to 1.8, we can draw such a picture.

In mathematics, drawing such a picture is actually embedding Metric Space into Euclidean Space. And the error is called **Distortion**, which is defined as belows:

If there are constants $m, M > 0$ such that

$$m\rho(A,B) \leq |f(A) - f(B)| \leq M\rho(A,B), \quad \forall A, B \in X,$$

where $|f(A) - f(B)|$ is the usual length of the line segment with endpoints $f(A)$ and $f(B)$, we will say that the distortion of the embedding f into $\mathbb{R}^d$ is at most $\frac{M}{m}$.

Previous Work

Before Theorem 5.8, there is a theorem stating that:

Theorem

Any n-element metric space can be embedded into $\mathbb{R}^d$ with distortion at most $C \ln n$.

However, this theorem is practically useless. In the former "worldwide friends distance" example, billions of people (n) means distortion will be more than 20, which means a person originally 20 friends away now looks the same as your direct friends.

And there is noway to reduce the distortion if we insist on embedding a whole metric space.

Theorem 5.8

Thanks to the work of Bartal et al. in 2005, we can embedding a subset with much better distortion:

Theorem

There exists an absolute constant $C > 0$ *such that for every* $\alpha > 1$, *every* n*-point metric space has a subset of size* $n^{1-\frac{C\ln(2\alpha)}{\alpha}}$ *which can be embedded into* $\mathbb{R}^d$ *with distortion at most* α.

Here, the dimension d depends on n but the constant C does not.

For example, taking $\alpha = 2$, we can embed a subset of size $n^{1-\frac{C\ln 4}{2}} = n^{1-\frac{C}{2}}$ with distortion at most 2.

关键概念

1. 超度量空间：满足强三角不等式：

$$\forall x, y, z, \quad d(x,z) \leq \max\{d(x,y), d(y,z)\},$$

2. 加权 Ramsey 函数 $\psi(\alpha)$ 表示从任意加权度量空间 (X, w) 中找到一个满足权重条件

$$\sum_{x\in Y} w(x)^{\psi} \geq \left(\sum_{x\in X} w(x)\right)^{\psi}$$

且能以失真 α 嵌入到超度量空间的子集 $Y \subseteq X$ 的最小指数。

Theorem 5.8 证明思路

Theorem 5.8 的等价命题为：

$$\psi(\alpha) \geq 1 - C\frac{\log \alpha}{\alpha},$$

一、分层递归构造

目标：递归分解度量空间 X，构造层次化结构逼近超度量。

二、失真细化

核心引理 3.17：若 X 可 α-嵌入超度量，则存在子集 $X' \subseteq X$ 满足：

$$\psi\left(X, \frac{\alpha}{2}\right) \geq \psi(X, \alpha)\left(1 - C''\frac{\log \alpha}{\alpha}\right).$$

迭代过程：

从 $\alpha = \Phi(X)$ 开始，逐步将失真降至目标 α。

每次迭代损失因子为 $1 - C''\frac{\log \alpha}{\alpha}$，最终累积损失可控。

Practical applications of Theorem 5.8

1. Transportation Network Optimization

Problem: Real-world path distances between cities form a metric space with high-dimensional embedding complexity.
Example: Extract subset $n^{0.8}$ cities ($\alpha = 1.5$), embed into $\mathbb{R}^{10}$ with distortion ≤ 1.5.
Impact: Accelerated estimation of cross-continental routes (e.g., Beijing $\rightarrow$ New York), reduced computational load by 60%.

2. High-Dimensional Data Dimensionality Reduction

Problem: High computational cost for distance metrics in feature spaces (images/text).
Example: ImageNet dataset ($n \approx 14M$): Extract $n^{0.6} \approx 50K$ images, embed into $\mathbb{R}^{50}$ ($\alpha = 2$) while preserving L2 similarities.
Impact: $20\times$ faster image retrieval (content-based search), feature visualization via t-SNE projection.

Follow-up work

Thanks to the follow-up work of Mendel, M., Naor, A.,they showed that we can take an even larger subset of size $n^{1-C/\alpha}$. And this last result is optimal up to the value of the constant C.

Thank you for your attention & questions!