

## 1、简答题：什么是知识图谱？它的主要用途是什么？

知识图谱是一种包含多种关系的图数据，是一种规模庞大的语义网络。在图书情报界称为知识域可视化或知识领域映射地图，是显示知识发展进程与结构关系的一系列各种不同的图形，用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。

知识图谱中的知识常以 (subject, relation, object) 的形式进行表示，其中subject 和 object 通常表示一个实体或类别，relation 表示关系。将这种表示方式对应到图上，实体即为图中的节点，关系即为图中连接节点的边，它们共同构成一个带标签的有向图。

主要作用是描述通用或专用场景下实体间的关联关系，主要应用场景为搜索引擎、语音助手、智能问答，电子商务，公共安全，金融以及医疗等领域。

## 2、简答题：什么是嵌入空间？为什么要用嵌入空间？

利用神经网络将高维的局部表示空间，映射到一个非常低维的分布式空间中（分布式空间的维度远远小于局部表示空间），在这个低维空间中，每一个特征不再是坐标轴上的点，而是分散在整个低维空间中，这一映射过程即为嵌入，被映射到的低维空间即为嵌入空间。

通过嵌入空间可以实现分布式的编码方式，相比传统的 one-hot 编码维度大大降低，可以提高编码的效率，降低时间和空间复杂度。

## 3、仿真题：假设有3个工作原理不同、测量精度未知的距离传感器，请设计一个传感器融合算法，实现两个静止物体之间距离的最优估计。要求借助数学仿真，获得融合算法的估计精度，并给出算法说明和性能评价。程序代码需要作为附录提交。

### 算法说明

将三个传感器安装在同一静止物体上，并标定时空同步，忽略三个传感器的系统误差，那么测量误差由随机误差决定。传感器的噪声满足正态分布，而其标准差的大小则为随机数。可以通过加权平均法获得测量距离的最优估计，每个传感器的权值可由其噪声的标准差决定。

已知仅有两个传感器时的最优权重计算算法，由多元函数求极值的理论，可推广至多个传感器。假设总共有 $n$ 个传感器，那么对于第 $m$ 个传感器，其对应的权重为

$$w_m = \frac{\frac{1}{\sigma_m^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \quad (1)$$

通过加权平均法获得测量距离的最优估计:

$$y = w_1 * y_1 + w_2 * y_2 + w_3 * y_3 \quad (2)$$

在仿真实验中，在(0,100)内随机选取1个浮点数作为距离的真值，之后再在(0,1)内随机选择三个小数分别作为三个传感器噪声的标准差，将之前选的浮点数分别与这三个小数组成三个传感器的观测模型，满足的正态分布。将该过程重复100次，得到100个真值与估计值。

## 性能评价

将融合估计得到的距离与真实得到的距离用不同颜色的散点进行表示，得到的散点图如图1所示，横坐标为仿真的轮次，可以看出无论随机真值大小为多少，真实值与理论值都非常接近。

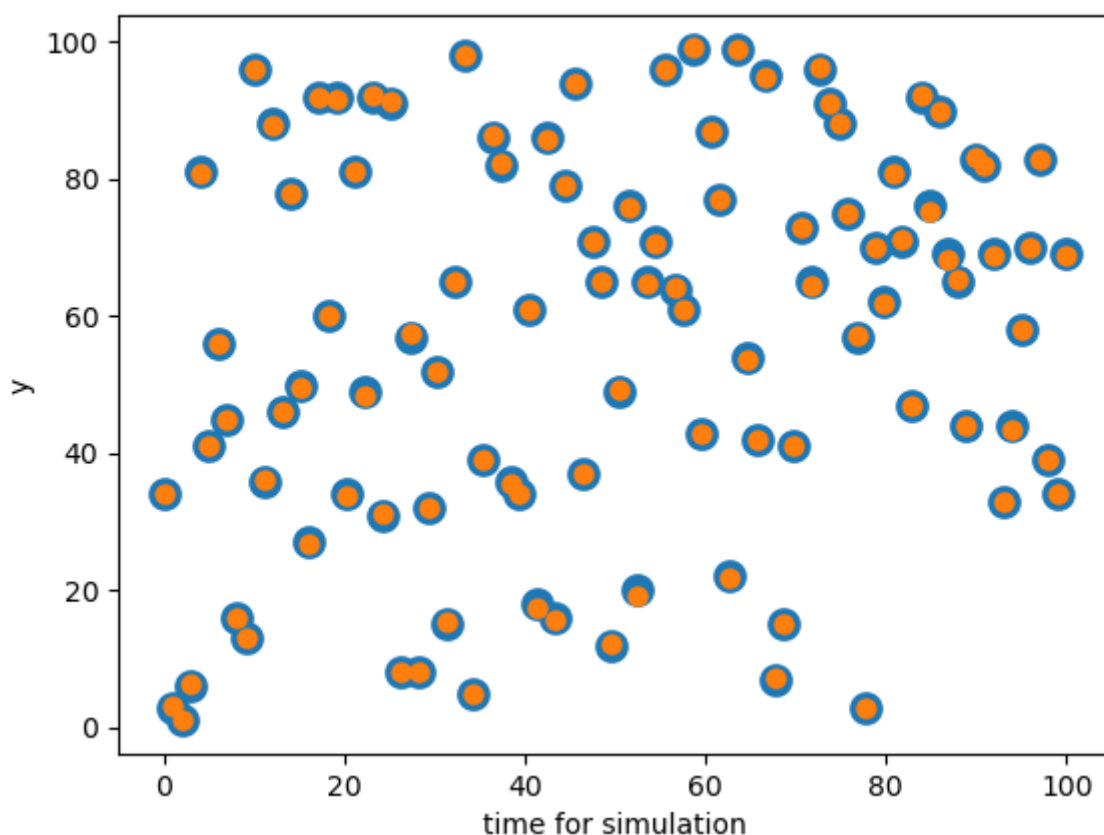


图1 真实值与预测值的散点图

将真实值与理论值画在同一张图上并进行拟合，可以看出其形态非常接近 函数 $y=x$ 。

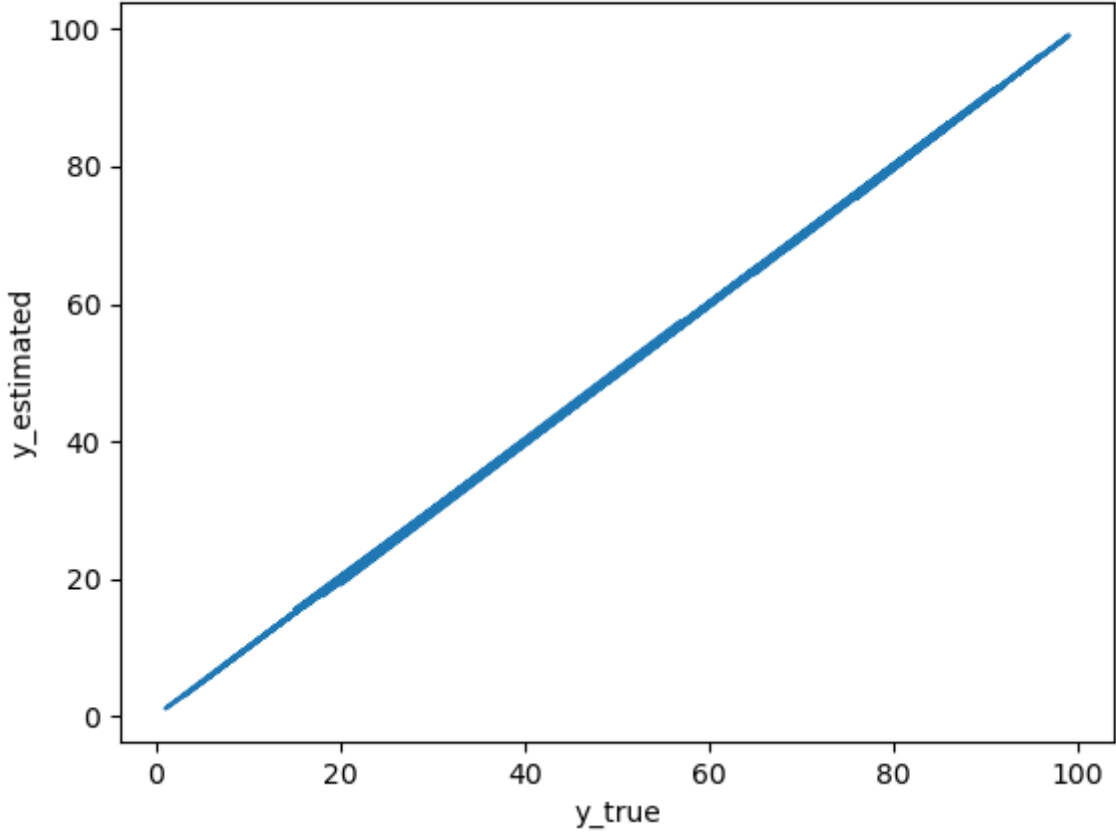


图2 真实值与估计值的回归曲线

通过均方误差(RMSE), 平均绝对误差(MAE) 以及平均绝对百分比误差(MAPE) 来进行算法的性能评价并给出量化结果。

计算每个估计值与真实值的均方误差, 平均绝对误差以及平均绝对百分比误差:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (5)$$

单次实验的误差的计算结果由下图所示，可以看出大部分的估计结果的精度都非常高。

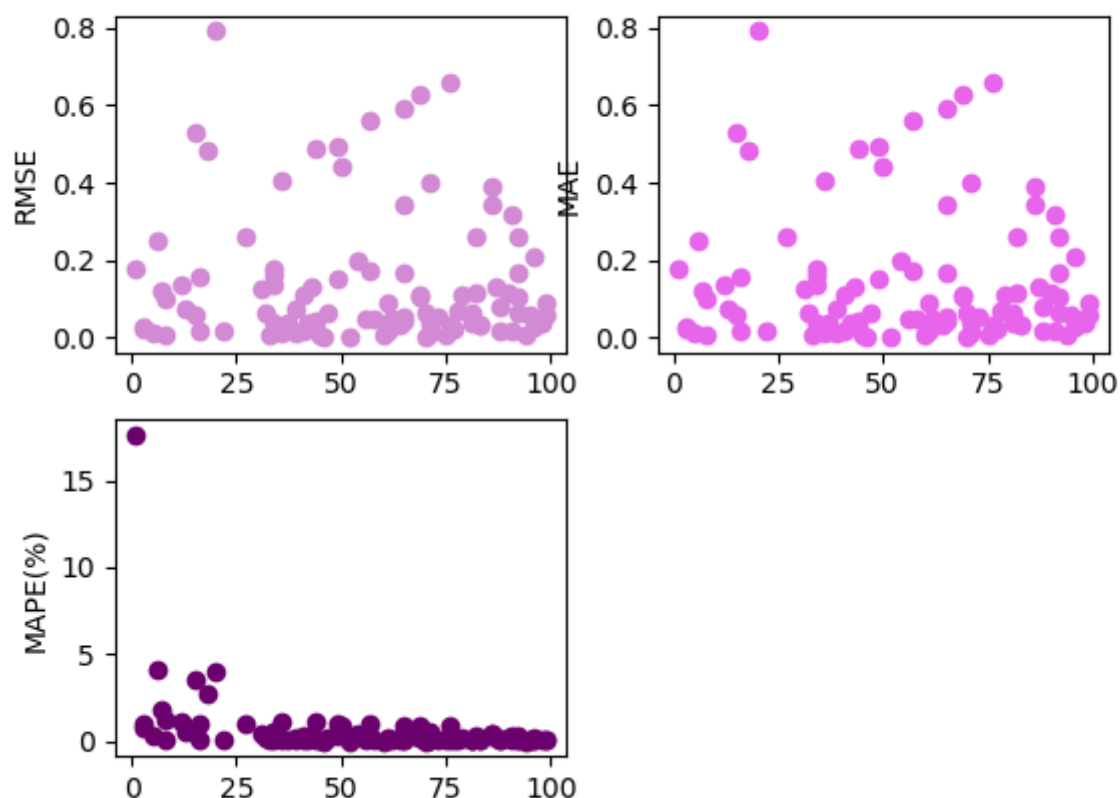


图3 单次实验误差

经过计算可以得到RMSE为0.05077658123200085 MAE为0.1537521358842869  
MAPE为 0.4680376889216562 %，综合多种评判标准，可以看出该融合算法具备较高的精度。