



同济大学

TONGJI UNIVERSITY

# 基于KNN优化的经济预测混合模型 及其在智慧城市中的应用

报告人：方志晗 学号：1152703

指导老师：穆斌（教授）



同济大学

TONGJI UNIVERSITY

研究背景及问题定义

预测模型及其工程实践

预测结果分析



同济大学

TONGJI UNIVERSITY

# 研究背景及问题定义



# 研究背景

- 国民生产总值GDP是衡量一个国家或者地区经济发展的重要指标；
- 税收收入是政府财政主要来源，是制定城市规划方案的基础；
- GDP和税收收入作为城市发展的重要指标，能够代表一个地区经济发展水平；
- 因此，科学的GDP和税务预测对于城市管理者来说至关重要。



# 研究现状

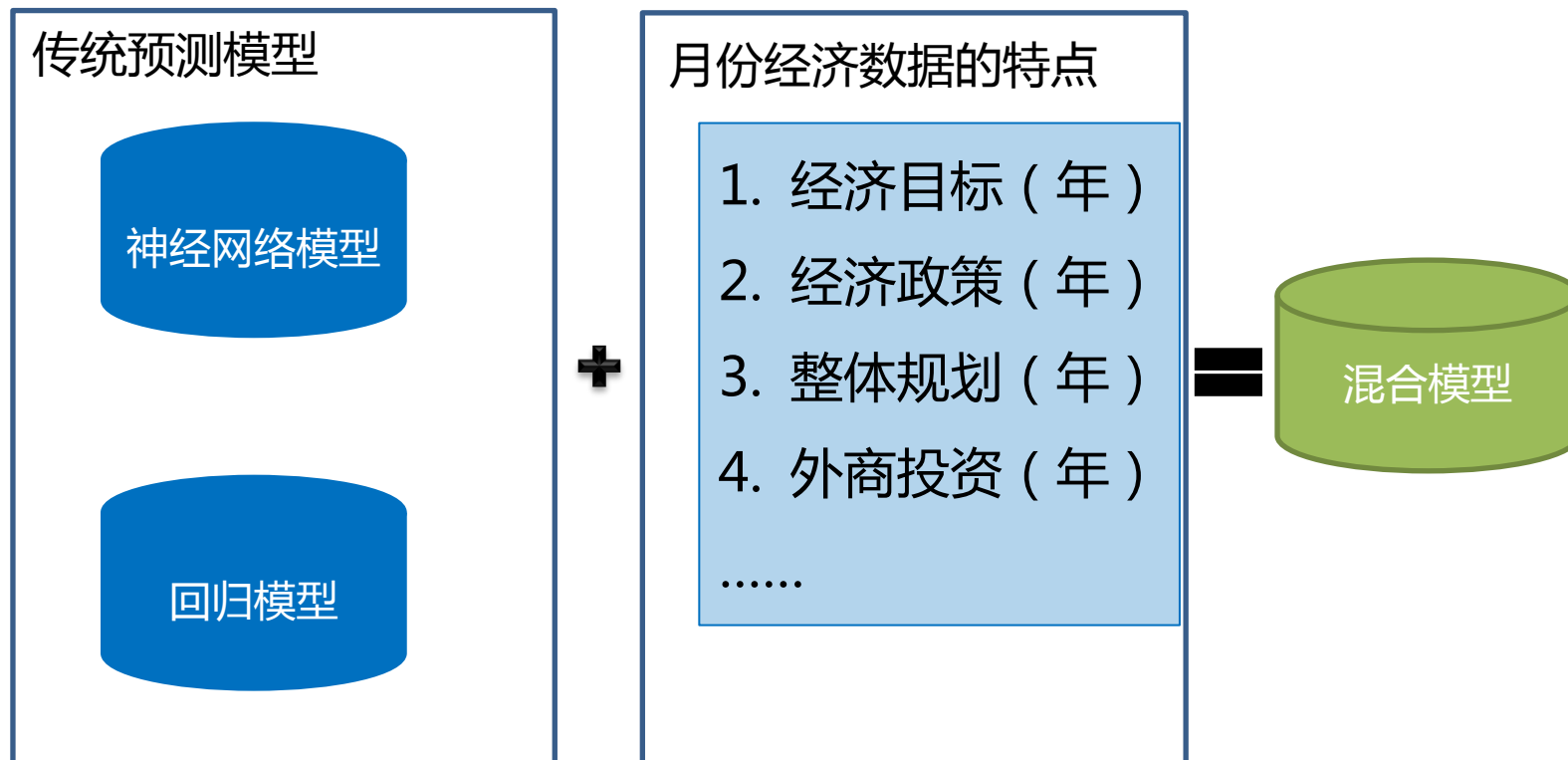
## 当前研究现状：

1. 时间序列预测方法
2. 马尔科夫预测模型
3. 回归模型
4. 灰色理论模型
5. 神经网络模型

## 研究特点：

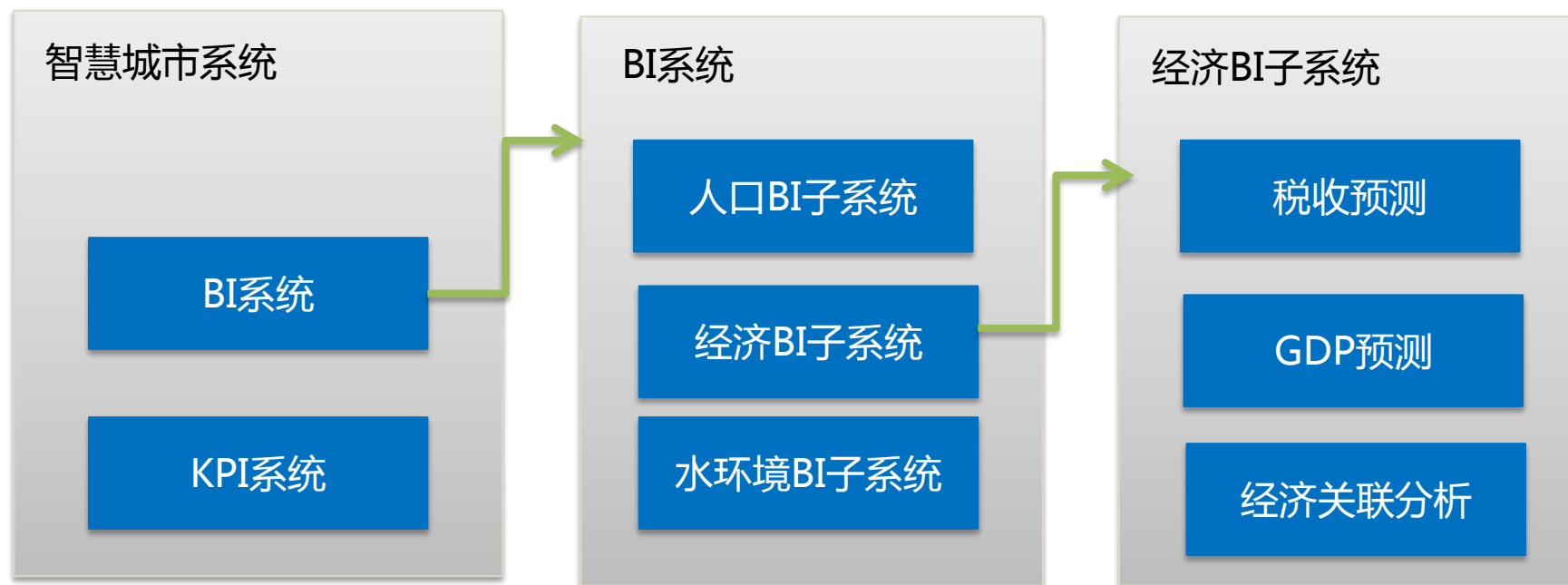
1. 按年分析预测
2. 按月分析预测没有考虑月份数据与年份数据的特点

# 研究概述





# 工程概述





同济大学

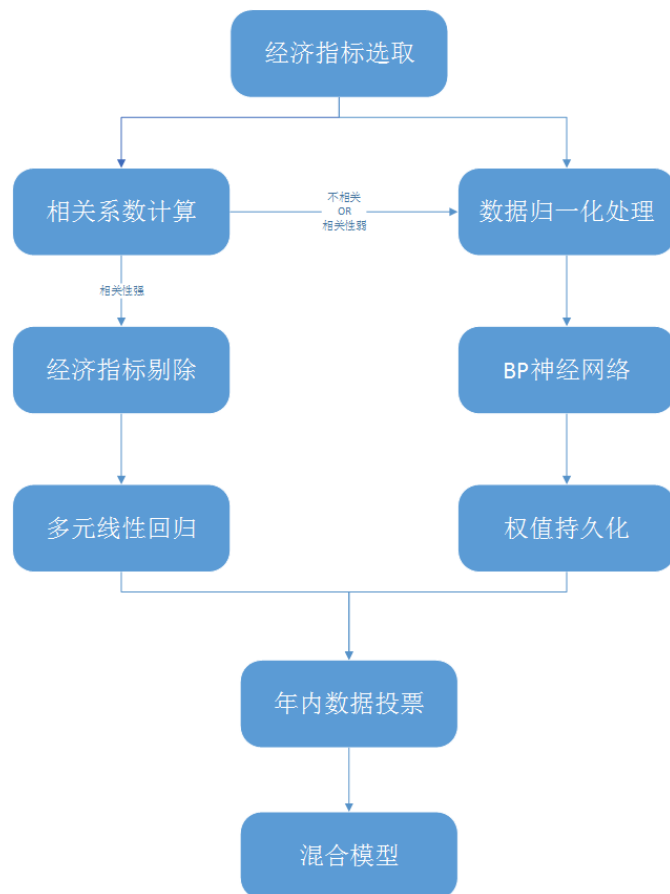
TONGJI UNIVERSITY

# 预测模型及其工程实践



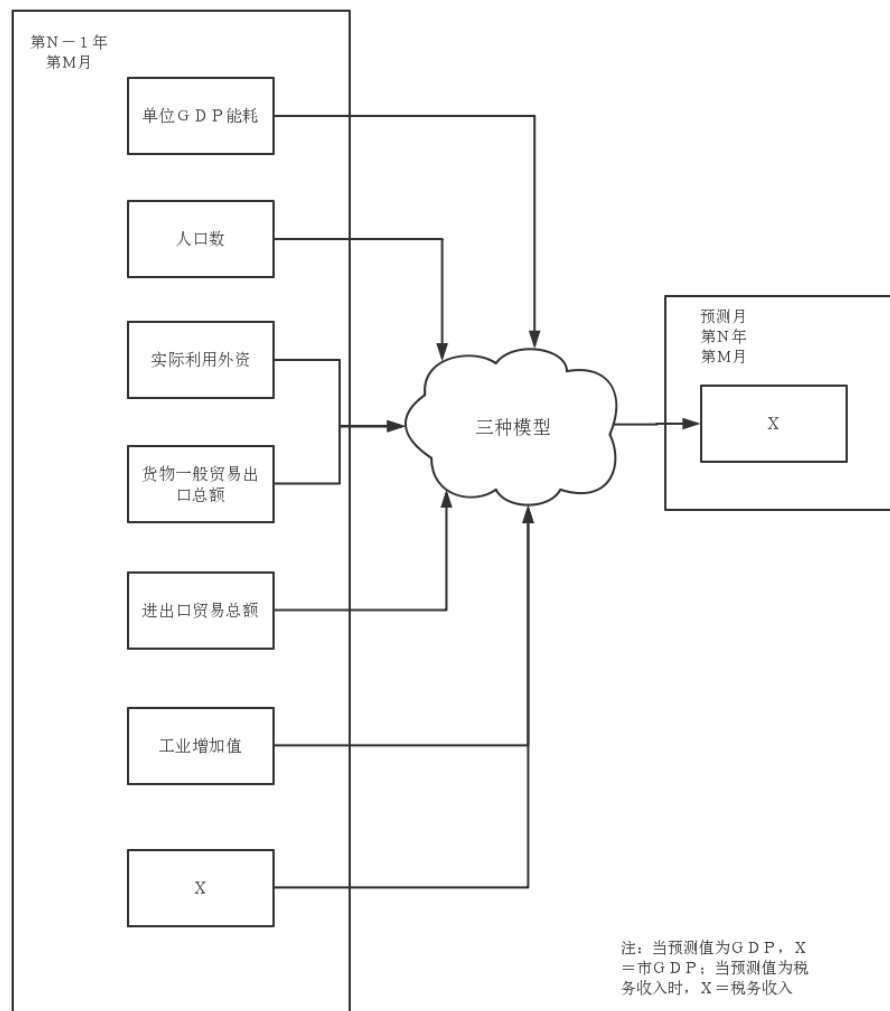


# 模型概论



1. 两种成熟模型
2. 年内月份投票
3. 优化混合模型

# 经济指标选取



1. 神经网络模型
2. 多元线性回归
3. 优化混合模型



# 多元线性回归模型

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k + u$$

$$\begin{cases} y = b_0 + b_1x_{11} + b_2x_{21} + \cdots + b_kx_{k1} + u_1 \\ y = b_0 + b_1x_{12} + b_2x_{22} + \cdots + b_kx_{k2} + u_2 \\ \cdots \\ y = b_0 + b_1x_{1n} + b_2x_{2n} + \cdots + b_kx_{kn} + u_n \end{cases}$$

1. 计算样本估计与实际值的方差和
2. 在方差和上对每个参数估计求偏导
3. 当偏导值为0时，表示误差最小

$$Q = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (\hat{b}_0 + \hat{b}_1x_{1i} + \cdots + \hat{b}_kx_{ki}))^2$$



# 多元线性回归模型

$$\hat{B} = (X'X)^{-1} X'Y$$

$$X'X = \begin{pmatrix} n & \sum x_{1i} & \sum x_{2i} & \cdots & \sum x_{ki} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{2i}x_{1i} & \cdots & \sum x_{ki}x_{1i} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum x_{ki} & \sum x_{1i}x_{ki} & \sum x_{2i}x_{ki} & \cdots & \sum x_{ki}^2 \end{pmatrix}$$

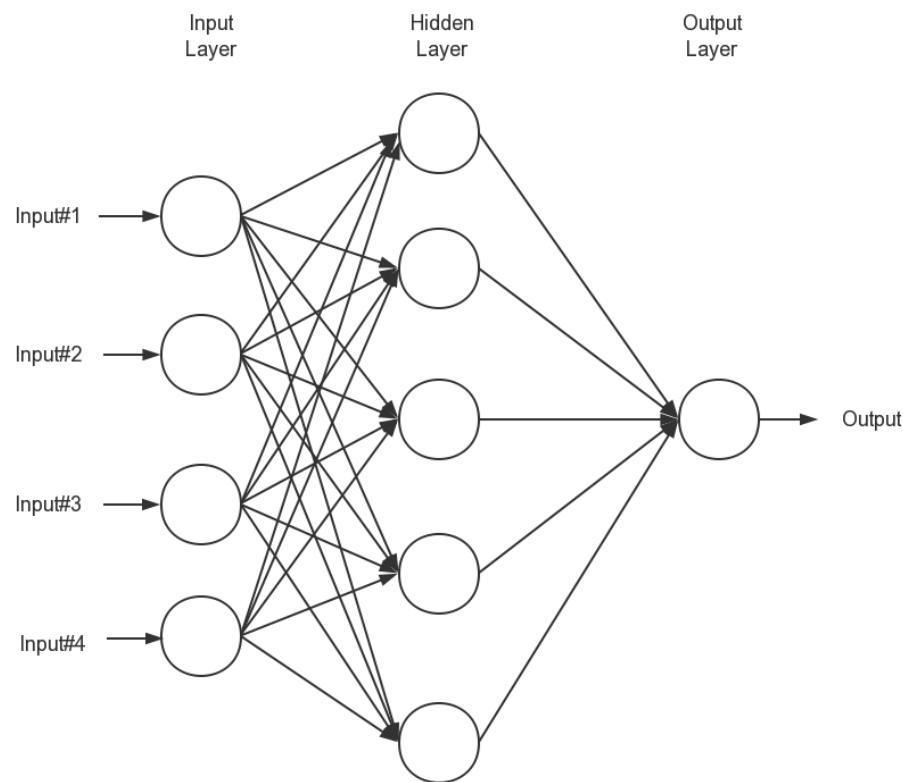
$$\hat{B} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \cdots \\ \hat{b}_k \end{pmatrix}$$

$$X'Y = \begin{pmatrix} \sum y_i \\ \sum x_{1i}y_i \\ \cdots \\ \sum x_{ki}y_i \end{pmatrix}$$

模型特点：

1. 当多个变量与因变量之间的关系为线性关系时，回归效果较好
2. 对于非线性数据拟合度较差

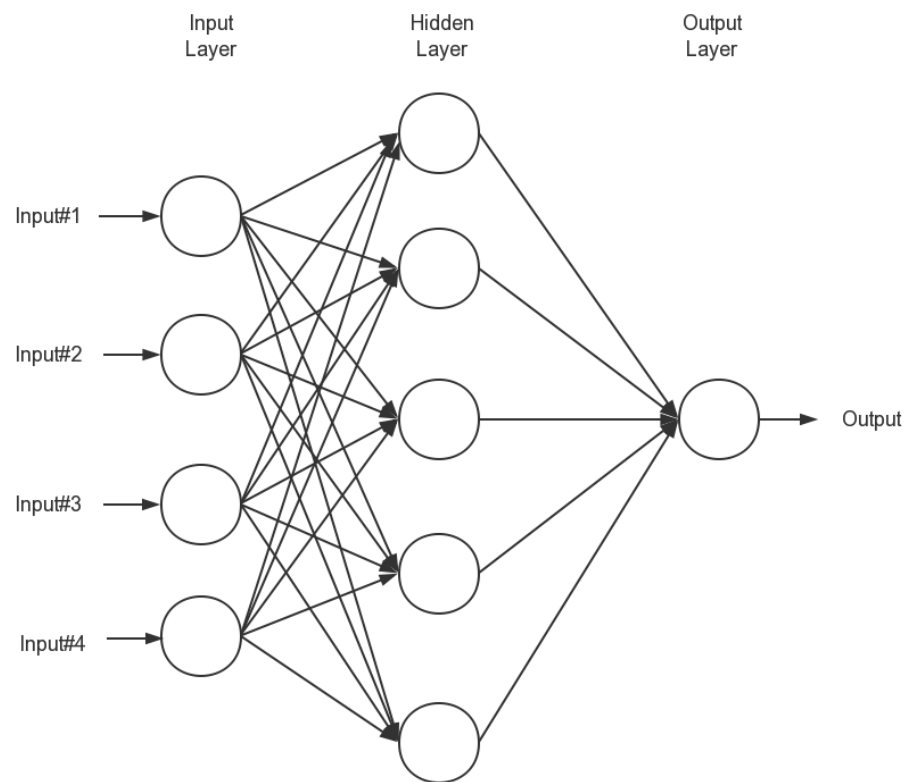
# 神经网络模型



## 参数设置

1. 三层神经网络；
2. 输入层神经元个数为指标个数；
3. 隐藏层神经元个数为指标个数/2;
4. 激活函数采用sigmoid函数；
5. 对数据进行归一化处理；
6. 缩小数据归一后倍数，以便处理预测值大于熟练数据中最大值的情况；

# 神经网络模型



模型特点：

1. 对非线性数据拟合度也很好；
2. 收敛慢，容易出现过度拟合或者拟合不充分的情况；
3. 我们采用权值持久化的方法处理数据。

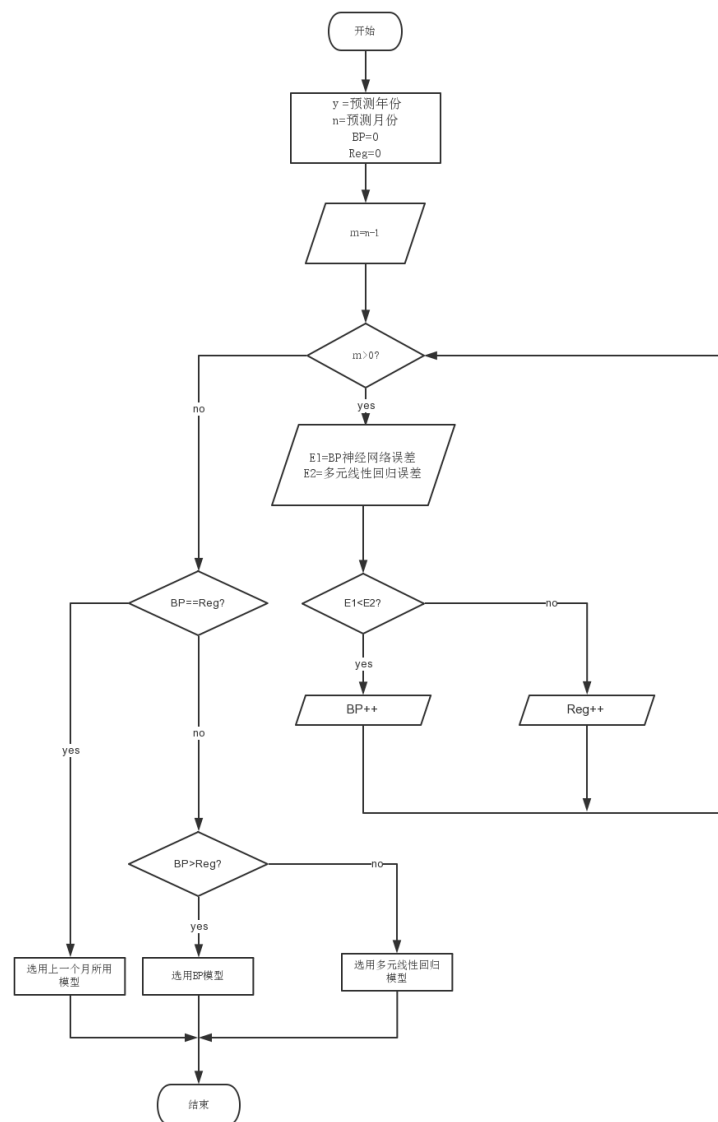
# 年内临近投票算法

算法描述如下：

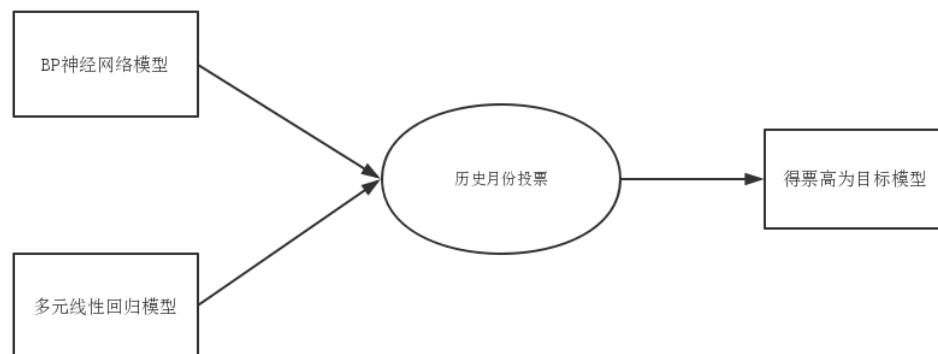
(1) 如果预测月份为当年第一月，计算前一年12个月中BP神经网络模型和多元线性回归模型作为更好模型的月数，月数高者为该年的最优模型，将该模型作为一月的预测模型；

(2) 如果预测月份为当年第 $n$ 月 ( $1 < n \leq 12$ )，则我们利用1到 $n-1$ 的预测误差来对第 $n$ 月的预测模型进行投票，误差较小的模型在该月拥有一票；

(3) 如果在前 $n-1$ 个月中，多元线性回归模型和BP神经网络模型拥有的票数相等，则在此时 $n-1$ 月有2票的权利，即 $n$ 的前一个月在此时有一票决定权。



# 经济预测混合模型



理论前提：

1. 同一年的整体经济环境改变不大；
2. 经济政策按年发布，经济发展目标按年制定；
3. 人口流动按年循环；
4. 因此年内经济发展可以互相借鉴；





# GDP与税务关联度变化

相关系数的计算公式计算如下：

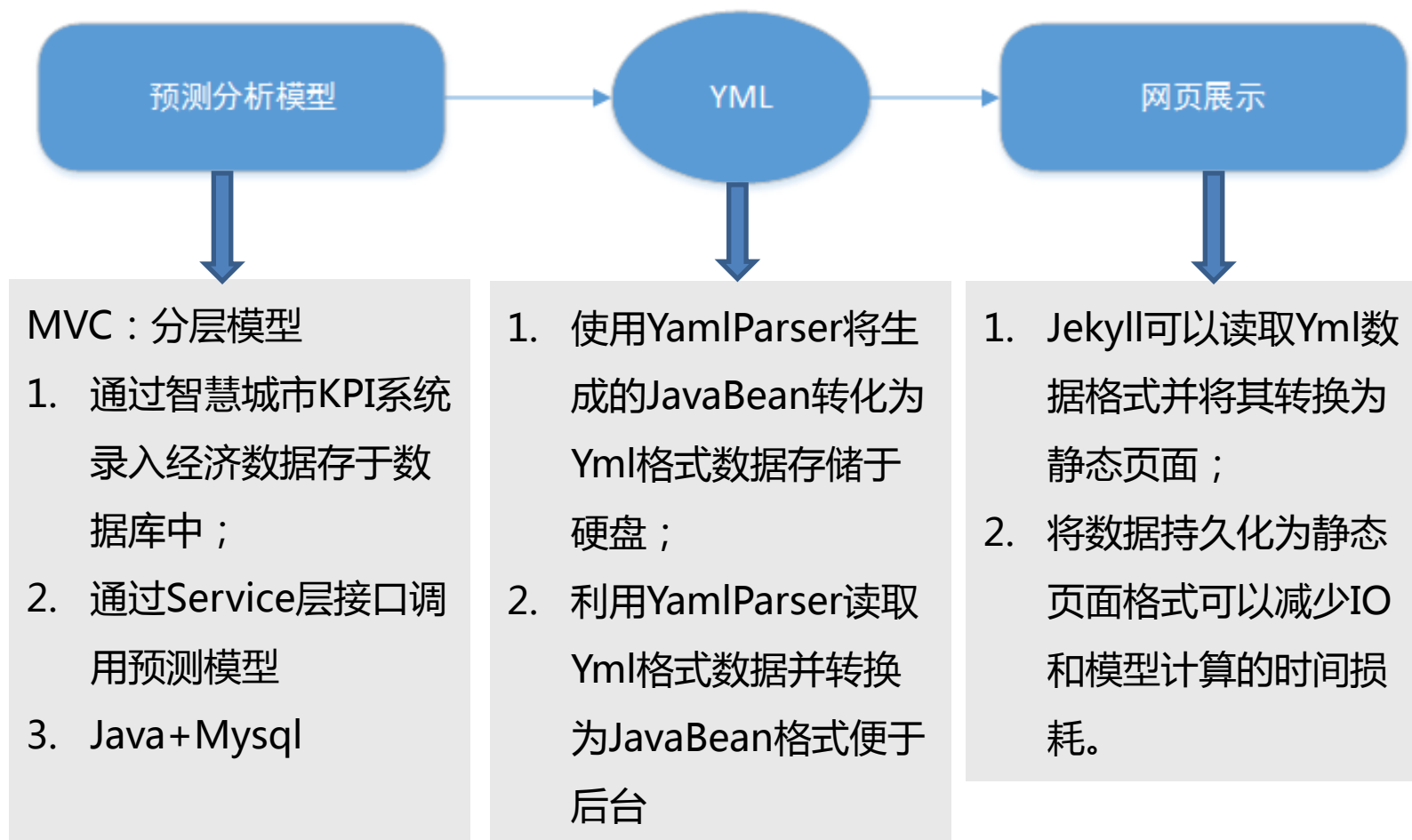
$$\rho_{X,Y} = \text{corr}(X,Y) = \frac{E[(X - \bar{X})(Y - \bar{Y})]}{\sigma_X \sigma_Y}$$

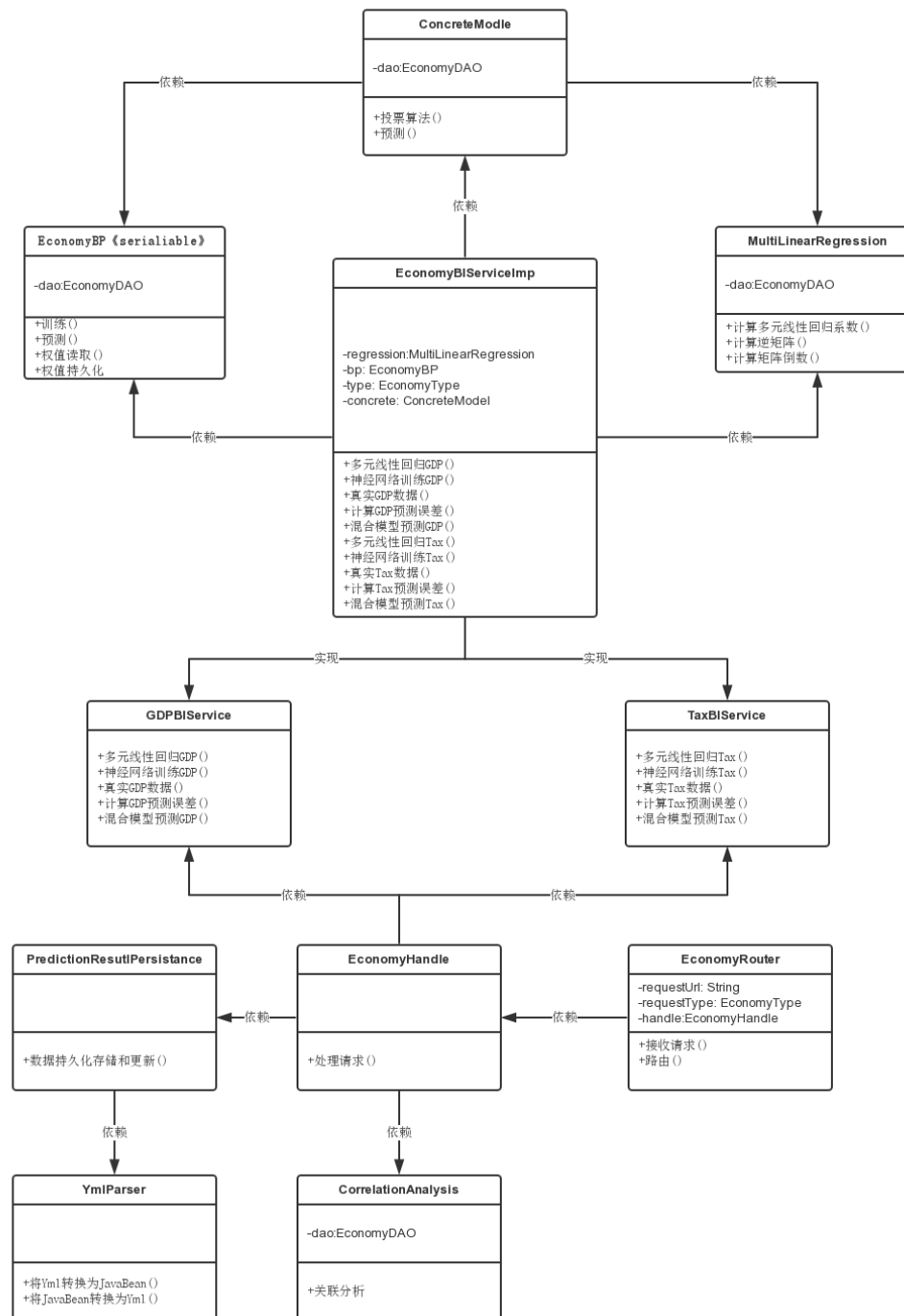
即：

$$\rho_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$



# 智慧城市工程实现







同济大学

TONGJI UNIVERSITY

# 预测结果分析



# 模型测试数据集及验证方法

## 数据集：

1. 将2001年到2012年12年144个月的GDP和税务收入作为预测目标；
2. 计算预测目标的预测值；

## 验证方法：

1. 计算三种模型预测值与当月实际数据的差值；
2. 差值最小的模型记为当月的最优模型；
3. 计算一年中三种模型每个模型最为最优模型的月份数目，月份最多者为该年的最优模型。
4. 计算这144个月来三种模型最为最优模型的月份，以此评价三种模型的好坏。



# 三种模型GDP预测

| 年月    | BP最优月数 | 回归模型最优月数 | 混合模型最优月数 |
|-------|--------|----------|----------|
| 2001年 | 9      | 3        | 9        |
| 2002年 | 6      | 6        | 6        |
| 2003年 | 0      | 12       | 12       |
| 2004年 | 0      | 12       | 12       |
| 2005年 | 10     | 2        | 10       |
| 2006年 | 0      | 12       | 11       |
| 2007年 | 0      | 12       | 12       |
| 2008年 | 11     | 1        | 11       |
| 2009年 | 12     | 0        | 12       |
| 2010年 | 4      | 8        | 8        |
| 2011年 | 0      | 12       | 12       |
| 2012年 | 12     | 0        | 11       |
| 合计    | 64     | 80       | 126      |
| 命中率   | 44.44% | 55.56%   | 87.5%    |



# 三种模型税务预测

| 年月    | BP最优月数 | 回归模型最优月数 | 混合模型最优月数 |
|-------|--------|----------|----------|
| 2001年 | 0      | 12       | 12       |
| 2002年 | 0      | 12       | 12       |
| 2003年 | 9      | 3        | 9        |
| 2004年 | 0      | 12       | 11       |
| 2005年 | 12     | 0        | 11       |
| 2006年 | 12     | 0        | 12       |
| 2007年 | 0      | 12       | 11       |
| 2008年 | 0      | 12       | 12       |
| 2009年 | 12     | 0        | 11       |
| 2010年 | 1      | 11       | 10       |
| 2011年 | 8      | 4        | 7        |
| 2012年 | 0      | 12       | 11       |
| 合计    | 54     | 90       | 118      |
| 命中率   | 37.5%  | 62.5%    | 81.94%   |



# 混合模型预测结果

Smart City System智慧城市系统是上海市信息化发展的重要一步，其中的经济分析子系统主要关注与GDP预测分析，税收收入预测分析以及GDP与税收的关联分析

GDP预测

税务预测

关联分析

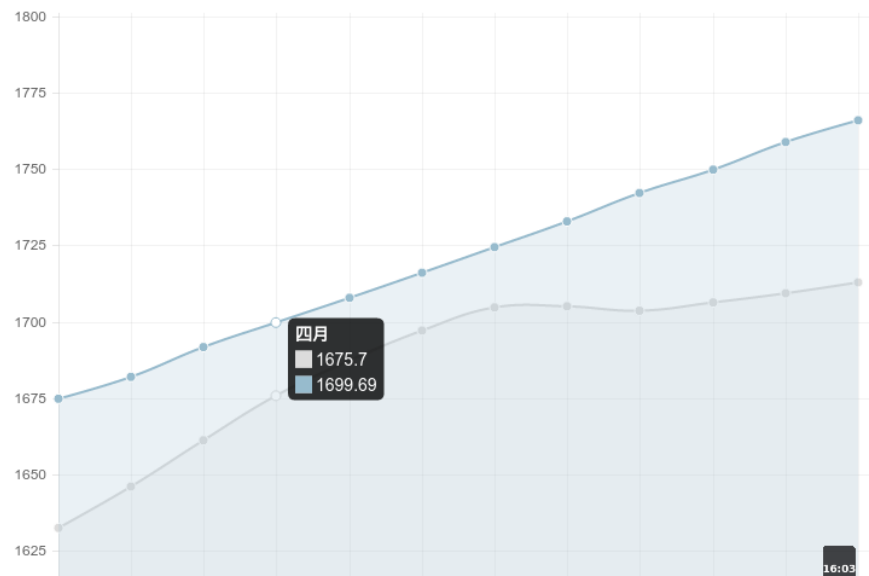
Currently v1.0.0

© 2015. All rights reserved.



SmartCity 经济预测分析系统

2012年上海市GDP预测分析图 单位:亿元







# 关联分析结果

Smart City System智慧城市系统是上海市信息化发展的重要一步，其中的经济分析子系统主要关注与GDP预测分析，税收收入预测分析以及GDP与税收的关联分析

GDP预测

税务预测

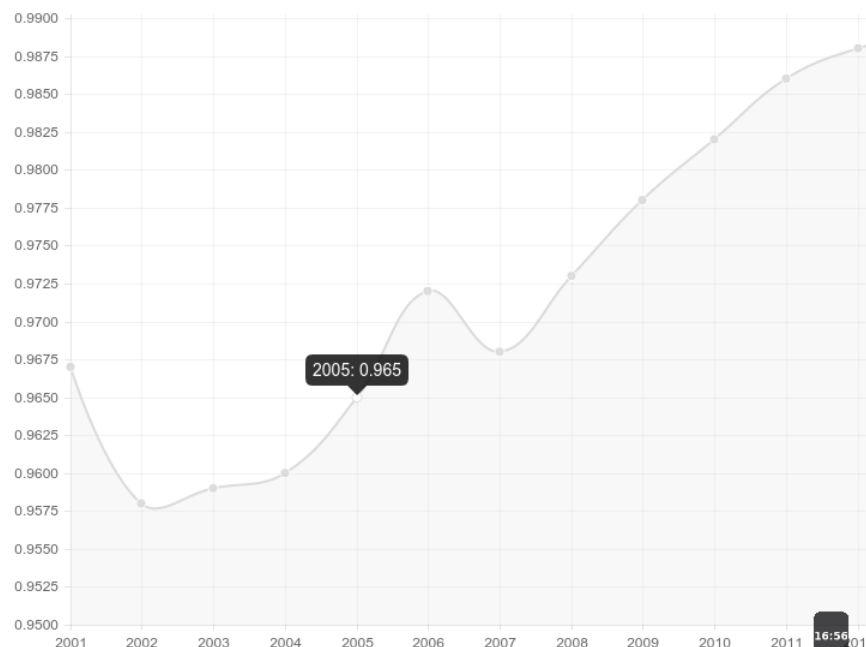
关联分析

Currently v1.0.0

© 2015. All rights reserved.



2012年上海市经济关联分析图 GDP与税收



THANK YOU