objectives

海量数据查询的低效率，缺乏统一的数据规范，以及支持数据的时空特性的缺点是需要解决的严重问题。在本项目中，我们需要对具有时间属性或者空间属性的数据进行分析，在获得的结果中提取出所需要的数据片段，并进行数据的预测、比对之后得出相关结论。

Low efficiency in massive data query, lack of unified data specification, and short in supporting space-time property of data are serious problems to be solved today. In this project, we need to analyze data with space or time characteristics. And use the result to extracts the needed data pieces from the huge amount of data. At last we use the obtained results for projections, comparison and draw the relevant conclusions.

问题1：如何分析传感器采集到的海量的数据？

1、将传感器采集到的数据按照属性分类筛选

2、将处理后的数据拟合出关联函数

Question1：How to analyze the huge amounts of data the sensors collect?

1. Classifying the data according to property characteristic

2. Fitting the treated data to correlation function**[1]**

**[1] "The time-dependent correlation function of the Jordan-Wigner operator as a Fredholm determinant"**

**Zvonarev, MB (Zvonarev, M. B.) ; Cheianov, VV (Cheianov, V. V. ; Giamarchi, T (Giamarchi, T.)**

**JOURNAL OF STATISTICAL MECHANICS-THEORY AND EXPERIMENT**

**文献号: P07035**

**DOI: 10.1088/1742-5468/2009/07/P07035**

**出版年: JUL 2009**

问题2：如何利用数据拟合出较好的关联函数？

1、利用贝塞尔曲线处理数据

2、数据存储的底层采用NoSQL数据库

3、在存储数据时，对数据进行相关分类，给相关类加上语义词义tag

4、在算法中规避数据的遍历算法

5、利用并行计算对算法进行优化

Question2: How to use the data to fit better correlation functions?

1. Handling the data with Bernstein- Bézier curve**[2]**

2. Storing the underlying data with NoSQL database

3. Classifying the data while storage, and add semantic meaning with tags**[3]**

4. Avoiding traversal algorithm

5. Using parallel computing**[4][5]** to optimize the algorithm

**[2]" Two essential properties of (q, h)-Bernstein-Bezier curves"**

**作者:Goldman, R (Goldman, Ron); Simeonov, P (Simeonov, Plamen)**

**APPLIED NUMERICAL MATHEMATICS**

**卷: 96 页: 82-93**

**DOI: 10.1016/j.apnum.2015.04.005**

**出版年: OCT 2015**

**[3]** **Fuzzy semantic tagging and flexible querying of XML documents extracted from the Web**

**作者:Buche, P (Buche, P); Dibie-Barthelemy, J (Dibie-Barthelemy, J); Haemmerle, O (Haemmerle, O); Hignette, G (Hignette, G)**

**JOURNAL OF INTELLIGENT INFORMATION SYSTEMS**

**卷: 26 期: 1 页: 25-40**

**DOI: 10.1007/s10844-006-5449-8**

**出版年: JAN 2006**

**[4] Dynamic evaluation strategy for fine-grain data-parallel computing**

**作者:Muchnick, VB (Muchnick, VB); Shafarenko, AV (Shafarenko, AV)**

**IEE PROCEEDINGS-COMPUTERS AND DIGITAL TECHNIQUES**

**卷: 143 期: 3 页: 181-188**

**DOI: 10.1049/ip-cdt:19960333**

**出版年: MAY 1996**

**(年份有点老but符合度不错）**

**[5] Effective Data Exchange in Parallel Computing**

**作者:Ma, H (Ma, Hui); Li, YQ (Li, Yongqi)**

**书籍团体作者:IEEE(什么鸟玩意？感觉这是个会议文件，所以不用管他，只是为了IEEE的名誉？）**

**2013 INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE AND CLOUD COMPUTING (ISCC)**

**页: 106-112**

**DOI: 10.1109/ISCC.2013.27**

**出版年: 2014**

问题3：如何在海量的数据中找到所需要的数据片段？

1、利用特征提取的思想找到所需要的数据片段

2、利用模式识别的思想找到所需要的数据片段

3、利用拟合出的关联函数判断数据的趋势

Question3:How to search the required data fragments from the big amount of data?

1. Using the idea of feature extraction**[6]** to find pieces of data

2. Using the idea of pattern recognition**[7]** to find pieces of data

3. Determining the trend of the data through the correlation function

**[6] FAST FEATURE-EXTRACTION FOR MACHINING APPLICATIONS**

**作者:FIELDS, MC (FIELDS, MC); ANDERSON, DC (ANDERSON, DC)**

**COMPUTER-AIDED DESIGN**

**卷: 26 期: 11 页: 803-813**

**DOI: 10.1016/0010-4485(94)90094-9**

**出版年: NOV 1994**

**[7] A postprocessing algorithm for the optical recognition of degraded characters**

**作者:Liu, HS (Liu, HS); Wu, MX (Wu, MX); Jin, GF (Jin, GF); Yan, YB (Yan, YB)**

**编者:Lopresti, DP; Zhou, JY**

**DOCUMENT RECOGNITION AND RETRIEVAL VI**

**丛书: PROCEEDINGS OF THE SOCIETY OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS (SPIE)**

**卷: 3651 页: 41-48**

**DOI: 10.1117/12.335820**

**出版年: 1999**

**(这两篇都有点老，但是！！！引用次数很多，也不是只针对图像的，所以就他俩了）**

问题4：得出的结论有什么应用范围？

1、优化数据查询

2、对数据的预测有极大帮助

3、对数据的比对有极大帮助

Question4: Where can we use the result we obtained?

1. Optimizing data query

2. Helpful for the data projection

3. Helpful for the data comparison

创新点与项目特色

本项目的创新点包括以下5个方面：

1、通过贝塞尔曲线等多种方法，对数据进行处理

2、针对数据拟合出关联函数

3、利用并行计算提高优化数据查询

4、在数据存储时对数据进行分类标记

5、利用模式识别以及特征提取的思想，找到所需的数据片段

Innovation and project characteristics:

There are four three aspects of innovation in our project:

1. Analyzing the data with Bernstein- Bézier curve and many other methods

2. Building up the associate function according to the specific data

3. Using parallel computing to improve optimization data query

4. Storing data with classification tag

5. Using the idea of feature extraction and pattern recognition to find pieces of data