

基于模拟信息转换器的物联网海量数据处理研究*

刘渊, 杨泽林[†]

(宁夏大学物理电气信息学院, 银川 750021)

摘要: 针对大规模物联网应用的海量数据处理信息获取效率低、硬件成本昂贵的问题, 依据压缩感知 (compressed sensing, CS) 理论, 建立了一种模拟信息转换器 (analog to information converter, AIC) 数据处理系统模型。模型以 MATLAB/Simulink 为平台, 分别设计了信号的解调、过滤、采样、重构等功能模块, 并对不同频率分量的信号进行处理。实验结果表明, 该模型可以较低采样率、高压缩比精确重构稀疏信号, 重构效率与观测数 M 、抽取行数 K 以及信号频率分量相关。

关键词: 海量数据; 压缩感知; 模拟信息转换器; 模型仿真; 频率分量; PSNR

中图分类号: TP391.9 **文献标志码:** A **文章编号:** 1001-3695(2013)12-3694-04

doi:10.3969/j.issn.1001-3695.2013.12.045

Research on IOT mass data processing based on analog to information converter

LIU Yuan, YANG Ze-lin[†]

(School of Physics Electrical Information Engineering, Ningxia University, Yinchuan 750021, China)

Abstract: For inefficiency information acquisition and costly hardware of mass data processing in large-scale applications of Internet of Things, this paper built an analog to information converter (AIC) data processing system model based to compressed sensing (CS). It designed signal demodulation module, filtering module, sampling module, and reconstruction module to processing signals which contained different frequency components through MATLAB/Simulink. Experiments show that, the model achieves low speed sampling and accurate reconstruction of sparse signal, and the sparse signal reconstruction efficiency and quality are related to observations value M , rows extraction value K and signal frequency components.

Key words: mass data; compressed sensing; analog to information converter; model simulation; frequency component; PSNR

近些年物联网产业的飞速发展,其应用已深入到生产生活的方方面面。由于物联网数据的海量性,数据处理技术已成为实现物联网大规模普及应用中最为重要的技术之一。传统的信号采样的理论基础是奈奎斯特采样定理,该理论要求信号的采样频率不得低于信号最大频率的两倍。对于海量数据获取和处理,奈奎斯特采样的硬件结构复杂、成本昂贵、信息获取效率低下,数据处理过程中还会造成大量的浪费。因此,传统的奈奎斯特采样定理已经不能满足目前物联网产业进一步发展的需求^[1]。压缩感知理论 (CS) 是现代信息科学领域中一个热门研究方向。该理论表明:只要信号是稀疏的,就可以利用一个与变换基不相关的观测矩阵对信号投影,然后通过求解优化问题,利用重构算法将信号以高概率重构出来。根据 CS 理论产生的模拟信号信息转换器 (AIC) 是该理论的具体硬件实现方式^[2]。

关于 CS 理论的研究较为广泛。石光明等人^[3,4]详细介绍了 CS 理论的原理与框架,对研究中现存的一些难点问题进行了探讨,并指出了进一步研究的方向;李卓凡等人^[5]归纳了 CS 理论在数据压缩、信道编码、逆问题、数据获取方面的应用方式;另外,CS 理论在雷达成像^[6,7]、无线传感网络^[8,9]、图像压缩与重建^[10,11]、超光谱成像^[12,13]等领域得到了广泛应用。CS

理论在国外的医疗成像、光学成像、对地观测等领域得到了具体实现,如 Rice 大学 Baraniuk 等人研制出的单像素相机、耶鲁大学研制的超谱成像仪^[14]。国内很多高校与科研机构已经开始深入研究 CS 理论及其实际应用。但在如何设计有效的硬件电路来实现、解决实际应用中硬件成本过高、信息采样率过低的问题,这方面的研究还远远不够。因此,本文建立了一种 AIC 系统模型,以实现海量稀疏信号的低速率采样和高分辨率重构。

1 压缩感知理论简介

CS 理论主要涉及到信号的稀疏表示、满足等距约束性准则的观测矩阵设计、快速稳健的信号重建算法设计三个核心问题^[1,15]。

1.1 信号的稀疏表示

信号的可稀疏表示是信号压缩感知的初始条件^[3]。假设有信号 $X (X \in \mathbb{R}^N)$, 长度为 N , 基向量为 $\psi_i (i=1, 2, \dots, N)$, 对信号进行变换:

$$X = \sum_{i=1}^N \theta_i \psi_i \quad (1)$$

其中: Θ 是投影系数, $\Theta = [\theta_i] = [X, \psi_i]$ 构成了 $N \times 1$ 的列向量; X 和 Θ 是同一个信号的等价表示, X 是信号在时域的

收稿日期: 2013-04-17; **修回日期:** 2013-05-23 **基金项目:** 科技部国际合作项目 (2011DFA11780); 国家自然科学基金资助项目 (612013830)

作者简介: 刘渊 (1989-), 男, 宁夏银川人, 硕士研究生, 主要研究方向为智能仪器与检测技术、智能信息处理; 杨泽林 (1957-), 男 (通信作者), 教授, 主要研究方向为智能仪器与检测技术 (fktemplar@163.com)。

表示,而 Θ 则是信号在 Ψ 域 的表示。若 X 仅仅是 K 个基矢量的线性组合(即只有 K 个 θ_i 非零系数, $N-K$ 个零系数),则称信号 X 是 K 稀疏的。而信号稀疏表示的关键是找到该信号的最佳稀疏域。只有选择合适的基表示信号,才能令信号稀疏化。

1.2 满足等距约束性准则的测量矩阵设计

利用与变换矩阵不相关的观测矩阵 $\Phi (M \ll N)$ 中的 M 个行向量 $\{\phi_j\}_{j=1}^M$,对稀疏向量进行线性投影,得到 M 个观测值 $y_j \langle \Theta, \phi_j \rangle (j=1,2,\dots,M)$,则观测向量为 $Y = (y_1, y_2, \dots, y_M)^T$,即

$$y = \Phi \Theta = \Phi \Psi^T X = A^{CS} X \tag{2}$$

观测向量 Y 为 M 维,使得测量对象从 N 维降为 M 维。由于观测过程是非自适应的,所以测量矩阵的设计要求在信号从 X 转换为 Y 的过程中,所测量到的 M 个测量值不能破坏原始信号信息,保证信号的精确重构。如果 X 是 K 稀疏的,且 Θ 中 K 个非零系数的位置是已知的,那么只要 $M \ll N$,就可以从 Y 中求出 Θ 。对于具有 K 个非零元素的任意矢量 v ,若 $\varepsilon > 0$,则上述简化后问题的充分必要条件^[16]是

$$(1 - \varepsilon) \|v\|_2 \leq \|\Theta v\|_2 \leq (1 + \varepsilon) \|v\|_2 \tag{3}$$

即矩阵 Θ 必须保持特定 K 稀疏矢量的长度。 K 稀疏压缩信号有稳定解的一个充分条件是:对于任意 $3K$ 稀疏矢量 v , Θ 满足式(3)。该条件即为等距约束性(restricted isometry property, RIP)。

1.3 信号的重构算法

信号重构算法的设计应利用尽可能少的压缩测量,快速、稳健、精确地重建原始信号。如果信号在变换域是 K 稀疏的,且 Θ 的 $2K$ 阶有限等距常数小于 1,那么信号的重建可以转换为约束 l_0 范数最小化的非凸优化问题^[18,19]。研究人员陆续提出了多种近似等价的信号重建算法,主要包括松弛方法、贪婪方法、非凸方法,这三类算法都适用于信号的稀疏表示问题^[20,21]。

2 压缩感知理论的 AIC 结构

基于 CS 理论,一些学者提出了模拟信号的直接信息采样方法,利用宽带伪随机解调器和低速采样器,开发出一套实用的 AIC 的设计框架^[22,23],如图 1 所示。

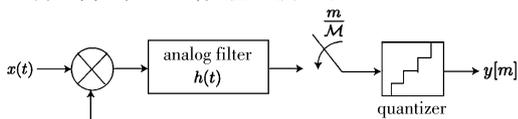


图 1 AIC 结构框架

图 1 表明了稀疏信号通过 AIC 获得观测值的过程,主要包括解调、过滤和采样三个部分。

a)解调是指信号 $x(t)$ 与伪随机序列 $p_c(t)$ 相乘。 $p_c(t)$ 是由线性反馈移位寄存器(LFSR)产生的二进制位序列。一个 n 位的 LFSR 能够在重复之前产生 $2^n - 1$ 位长的伪随机序列。在 AIC 中,伪随机序列 $p_c(t)$ 的幅度为 ± 1 ,且它的频谱非常宽,类似于白噪声。

调制信号 $p_c(t)$ 是由 $p_c[n]$ 组成, $p_c[n] = \pm 1$,则 $p_c(t) = p_c[n], t \in [(n-1)T_e, nT_e], n \in \{0, 1, \dots, N-1\}$ 。

乘法器将稀疏信号 $x(t)$ 与高速调制信号 $p_c(t)$ 相乘,得到

解调输出信号为

$$y(t) = x(t) \times p_c(t), t \in [0, N \times T_e] \tag{4}$$

b)过滤指解调后的信号 $y(t)$ 通过低通滤波器,即对 $y(t)$ 积分。为了保留信号的重要信息,伪随机序列 $p_c(t)$ 的符号转换率应当至少是信号的奈奎斯特采样率。滤波器对信号进行频谱的过滤,把解调后的信号频率信息遍布到所有的谱上,利用低通或带通滤波器来获取到足够的信号信息。

c)解调是对通过滤波器的信号进行采样和量化。设稀疏信号的长度为 N ,ADC 的采样率为 T ,信号的等效采样周期为 T_e 。在时间 NT_e 上对信号进行 M 次采样,则 $MT = NT_e$,因为 $M \ll N$,所以采样周期 T 远大于重构信号的等效采样周期 T_e 。因此,AIC 的采样频率是相对低速的。

对 $y(t)$ 以采样率 T_e 进行采样,可以得到 M 个观测值。

$$y[m] = \int_{-\infty}^{\infty} x(\tau) \times p_c(\tau) \times h(t - \tau) dt \Big|_{t=m\Delta t} \tag{5}$$

对式(5)进行矩阵到向量的变换:

$$y[m] = \sum_{n=1}^N \theta_i \times \int_{-\infty}^{\infty} \psi_i(\tau) \times p_c(\tau) \times h(m\Delta t - \tau) d\tau \tag{6}$$

矩阵向量可以表示为

$$y = x \times P_c \times H \tag{7}$$

式中: y 为观测序列, $y = [y_0, y_1, \dots, y_{M-1}]^T$; $x = [x_0, x_1, \dots, x_{M-1}]^T$ 为原始信号的采样序列; P_c 即为高速调制信号的矩阵,矩阵表达形式为

$$P_c = \begin{bmatrix} p_c(0) & & & \\ & p_c(1) & & \\ & & \ddots & \\ & & & p_c(N-1) \end{bmatrix} \tag{8}$$

由式(6)可以得到等效的观测矩阵 Φ 为

$$\Phi = \int_{-\infty}^{\infty} \psi_i(\tau) \times p_c(\tau) \times h(m\Delta t - \tau) d\tau \tag{9}$$

将式(9)代入优化问题进行求解,可以从观测集合中精确或高概率地重构原始信号 X 。

3 AIC 的模型建立与仿真分析

3.1 AIC 的模型建立

根据图 1 的 AIC 结构,在 MATLAB/Simulink 环境下设计 AIC 模型。分别设计 Signal(信号源)、M Sequences(M 序列信号发生器)、Dot Product(点乘)、Integrator(积分器)和 ADC 模块。AIC 设计的总体结构如图 2 所示。

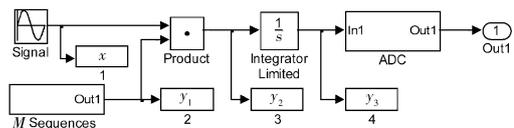


图 2 Simulink 中构建的 AIC 系统模型

通过对各个模块进行相应的参数设置,并结合 MATLAB 编程即可实现 AIC 的功能。

Signal 模块用于输入系统的稀疏信号。 M Sequences 模块用于产生伪随机序列,它是由五个 D 触发器构成的五级移位寄存器。在设计中,用一个 Unit Delay(采样周期延时)来代替一个 D 触发器,设置每一个 Unit Delay 的初始状态与采样周期,在移位脉冲作用下以及 XOR(异或模块)作用下,将产生具有自启动功能的伪随机序列。 M Sequences 模块设计如图 3 所示。

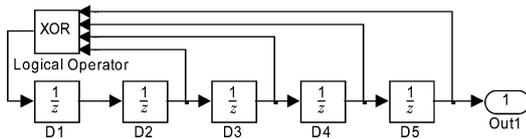


图3 M序列产生器模型设计

ADC 模块的设计如图 4 所示。该模块涉及到的子模块较多,主要有 Pulse Operator(脉冲信号发生器)、Sampled Quantizer Encode(采样量化编码)、Quantizing Decoder(解码器)和 Integer to Bit Converter(二进制转换器)。ADC 模块将在幅度为 $[-1, 1]$ v 的区间里,对信号进行 8 bit 输出,量化间隔为 $1/128$ 、量化码本为 $[-1:2/255:1]$ 。

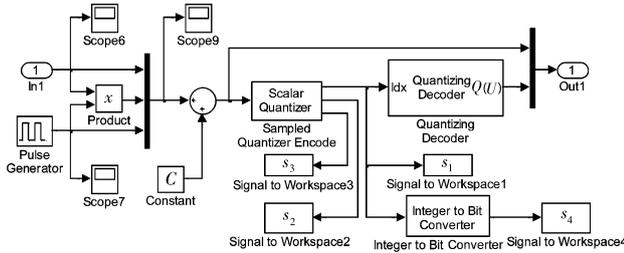


图4 ADC模型设计

3.2 仿真结果及分析

根据在 MATLAB/Simulink 平台上设计的 AIC 模型,结合 MATLAB 编程进行仿真。信号恢复流程如图 5 所示。

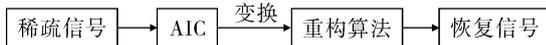


图5 信号通过 AIC 实现重构过程

在程序中,利用 512 阶的 DCT 矩阵产生稀疏信号 $x(n)$ 。设观测数 M 为 50,抽取行数 K 为 8,即稀疏信号的频率分量为 8。根据压缩感知理论,观测数 M 、抽取行数 K 以及信号维数 N 应该符合

$$M \geq K \log(N/K) \quad (10)$$

设信号的奈奎斯特率为 f_s ,信息采样率为 f_m , f_m 是滤波器截止带宽 $h(t)$ 的两倍,则稀疏信号的压缩比 $f_s:f_m = 64:1$ 。

对 AIC 系统模型输出的观测值进行相应变化,并代入优化问题求解得出观测系数,然后使用 BP 算法对信号重构,就可以将原信号以较高的重构概率恢复出来。

图 6~9 分别表示:DCT 矩阵产生的稀疏信号、伪随机序列信号、通过 AIC 变换得到的观测矩阵、利用 BP 算法重构的信号。对重构信号和原稀疏信号计算恢复信噪比为 42 dB,可以说明通过该 AIC 系统模型较好地恢复了原稀疏信号。

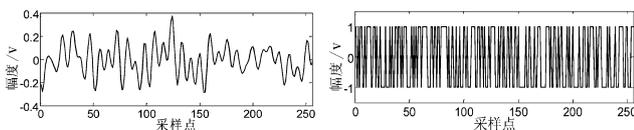


图6 DCT矩阵产生512阶稀疏信号

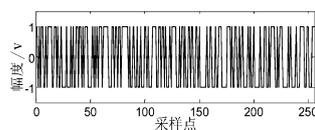


图7 移位寄存器产生伪随机序列

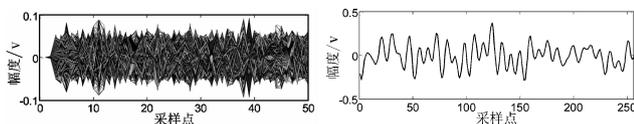


图8 观测值变换得到的观测矩阵

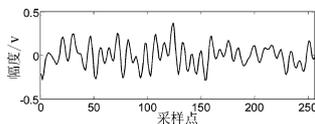


图9 BP算法重构信号

为了进一步验证所设计的 AIC 系统模型是否符合 CS 理

论,令稀疏信号的抽取行数 K 为 8,采用不同的观测数 M ,以不同的压缩比,将同一信号通过 AIC 系统模型进行重构仿真。仿真结果如表 1 和图 10 所示。

表1 当 $K=8$ 时,不同压缩比情况下的重构结果(部分结果)

观测数	压缩比	PSNR/dB	观测数	压缩比	PSNR/dB
30	17.07	11.07	55	9.3	48.47
35	14.63	16.53	60	8.53	69.29
40	12.8	13.26	65	7.88	61.86
45	11.38	17.72	70	7.14	67.08
50	10.24	42.39	75	6.83	56.74

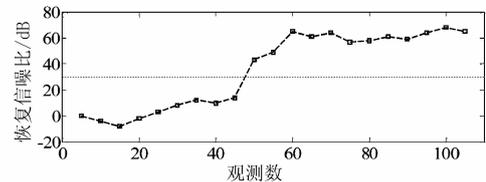


图10 $K=8$ 时,不同压缩比情况下 AIC 系统的重构性能

从表 1 和图 10 中可以看出,当 $M < 45$ 时,即压缩比大于 11 时,信号的恢复信噪比均小于 30 dB,不能够精确重构原始信号;而当 $M > 50$ 时,压缩比小于 10 之后,恢复信噪比均大于 30 dB,可以较好地重构原始信号。

令 $K=1, K=4$,以不同的压缩比,再次将信息通过 AIC 系统模型进行重构仿真,仿真结果如表 2、3 所示。

表2 当 $K=1$ 时,不同压缩比情况下的重构结果

观测数	压缩比	PSNR/dB	观测数	压缩比	PSNR/dB
5	102	0.23	30	8.53	69.42
10	51	43.41	35	7.88	61.26
15	34.13	52.51	40	7.14	67.71
20	25.6	56.29	45	6.83	56.80
25	20.48	64.15	50	6.40	57.41

表3 当 $K=4$ 时,不同压缩比情况下的重构结果

观测数	压缩比	PSNR/dB	观测数	压缩比	PSNR/dB
5	102	2.57	30	8.53	4.45
10	51	-1.02	35	7.88	45.31
15	34.13	-3.72	40	7.14	57.06
20	25.6	1.98	45	6.83	62.94
25	20.48	3.01	50	6.40	61.48

从表 2 中可以看出,当观测数 $M > 10$ 时,即压缩比小于 51,信号的恢复信噪比大于 30 dB,能够重构原始信号;当观测数 $M=5$ 时,信号的恢复信噪比小于 30 dB,不能够重构原始信号。从表 3 中可以看出,当观测数 $M \geq 35$ 时,信号的恢复信噪比大于 30 dB,能够重构原始信号;反之亦然。

根据以上实验可以得出如下结论:

a) 根据 CS 理论的要求, N, K 和 M 的关系必须符合式(10)的要求,即当 $N=512, K$ 分别为 1、4、8 时, M 必须大于等于 10、32、48。从图 10 与表 1~3 的结果中可以看出,实验结果符合该要求。由此说明设计的 AIC 系统模型符合 CS 理论。

b) 对于单频稀疏信号($K=1$),AIC 对信号的压缩能力突出,在保证信号精确重构的前提下,压缩比可以达到 50:1 以上;对于混频稀疏信号,其压缩效果与频率分量和观测数的数量有关。但是为了达到较高的数据压缩率,减少数据压缩时间,缓解数据的处理与储存的压力,观测数 M 不宜太大。

c) 对于宽带信号的处理,只要能将其稀疏化,并选择合适的观测数 M 和抽取行数 K 令稀疏信号通过 AIC 器件,就可以以远低于奈奎斯特采样定理所要求的速率,高概率重构原始信号。

4 结束语

本文在 CS 理论的基础上对 AIC 进行了研究。阐述了 CS 理论的三个核心问题,分析了 AIC 的结构与功能,然后在 MATLAB/Simulink 平台上建立了 AIC 系统模型,并进行了相关实验与仿真。实验结果符合 CS 理论要求,验证了建模的可行性。因此,在处理物联网海量数据时,用 AIC 器件代替传统的 ADC 能够有效解决 ADC 采样速率不足等问题,实现对宽带信号的低速率采样、高概率重构。

参考文献:

- [1] 周开乐,丁帅,胡小建. 面向海量数据应用的物联网信息服务系统研究综述[J]. 计算机应用研究, 2012, 29(1): 8-11.
 - [2] CANDES E, ROMBERG J, TAO T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information[J]. *IEEE Trans on Information Theory*, 2006, 52(2): 489-509.
 - [3] 石光明,刘丹华,高大化,等. 压缩感知理论及其研究进展[J]. 电子学报, 2009, 37(5): 1070-1081.
 - [4] 戴琼海,付长军,季向阳. 压缩感知研究[J]. 计算机学报, 2011, 34(3): 425-434.
 - [5] 李卓凡,闫敬文. 压缩感知及应用[J]. 微计算机应用, 2010, 31(3): 12-16.
 - [6] ANITORI L, MALEKI A, OTTEN M, *et al.* Design and analysis of compressed sensing radar detectors[J]. *Signal Processing*, 2013, 61(4): 813-827.
 - [7] 刘记红,徐少坤,高勋章,等. 压缩感知雷达成像技术综述[J]. 信号处理, 2011, 27(2): 251-260.
 - [8] MENG J, LI H, HAN Z. Sparse event detection in wireless sensor networks using compressive sensing[C]//Proc of the 43rd Annual Conference on Information Sciences and Systems. 2009.
 - [9] LING Q, TIAN Z. Decentralized sparse signal recovery for compressive sleeping wireless sensor networks[J]. *IEEE Trans on Signal Processing*, 2012, 58(7): 3816-3827.
 - [10] BOURQUARD A. Binary compressed imaging[J]. *IEEE Trans on Signal Processing*, 2013, 22(3): 1042-1055.
 - [11] LI Y, SHA X, WANG K, *et al.* The seight-block compressed sensing and its application to image reconstruction[C]//Proc of the 2nd International Conference on Instrumentation, Measurement, Computer, Communication and Control. 2012.
 - [12] HUO C, ZHANG R, YIN D. Compression technique for compressed sensing hyperspectral images[J]. *International Journal of Remote Sensing*, 2012, 33(5): 1586-1604.
 - [13] RIVENSON Y, STERN A, JAVIDI B. Overview of compressive sensing techniques applied in holography[J]. *Applied Optics*, 2013, 52(1): 423-432.
 - [14] 焦李成,杨淑媛,刘芳,等. 压缩感知回顾与展望[J]. 电子学报, 2011, 39(7): 1651-1662.
 - [15] 邵文泽,韦志辉. 压缩感知基本理论: 回顾与展望[J]. 中国图象图形学报, 2012, 17(1): 1-12.
 - [16] CANDES E. The restricted isometry property and its implications for compressed sensing[J]. *Académie des Sciences*, 2006, 346(1): 592-598.
 - [17] PROVOST J, LESAGE F. The application of compressed sensing for photoacoustic tomography[J]. *IEEE Trans on Medical Imaging*, 2009, 28(4): 585-594.
 - [18] CANDES E, ROMBERG J. Practical signal recovery from random projections[C]//Proc of SPIE. 2005: 76-86.
 - [19] CHEN S, DONOHO D, SAUNDERS M. Atomic decomposition by basis pursuit[J]. *SIAM Rev*, 2001, 43(1): 129-159.
 - [20] ELAD M. Optimized projections for compressed sensing[J]. *IEEE Trans on Signal Processing*, 2007, 55(12): 5695-5702.
 - [21] GILBERT A, MUTHUKRISHNAN S, STRAUSS M. Improved time bounds for near optimal sparse fourier representation via sampling[C]//Proc of SPIE. 2005.
 - [22] LASKA J. Random sampling for analog-to-information conversion of wideband signals[J]. *IEEE Trans on Information*, 2006, 58(6): 457-579.
 - [23] BARANIUK R, CEVHER V, DUARTE M, *et al.* Model-based compressive sensing[J]. *IEEE Trans on Information of Theory*, 2010, 56(4): 1982-2001.
-
- (上接第 3693 页)
- of H5N1: a clustering and association approach[J]. *Knowledge and Information Systems*, 2011, 27(2): 227-251.
 - [4] LI Sha-sha, ZHOU Yuan-chun, KOU Zheng. A wavelet packet based approach for the research of the Avian influenza virus cross-species infection[C]//Proc of the 1st International Conference on Networking and Distributed Computing. 2010: 301-304.
 - [5] CARNEIRO C, ALP A, MACEDO J. Advanced data mining method for discovering regions and trajectories of moving objects: "Ciconia Ciconia" scenario[C]//Lecture Notes in Geoinformation and Cartography. Berlin: Springer-Verlag, 2008: 201-224.
 - [6] PRIYONO A, RIDWAN M, ALIAS J A, *et al.* Generation of fuzzy rules with subtractive clustering[J]. *Journal Teknologi*, 2012, 43(D): 143-153.
 - [7] ZHENG Yu, XIE Xing. Learning location correlation from GPS trajectories[C]//Proc of the 11th International Conference on Mobile Data Management. 2010: 27-32.
 - [8] LI Quan-nan, ZHENG Yu, XIE Xing, *et al.* Mining user similarity based on location history [C]// Proc of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2008: 1-10.
 - [9] ZHENG Yu, ZHANG Li-zhu, MA Zheng-xin. Recommending friends and locations based on individual location history[J]. *ACM Transaction on the Web*, 2011, 5(1): 1-44.
 - [10] HAN Jia-wei, KAMBER M, PEI Jian. Data mining concepts and techniques[M]. 3rd ed. [S. l.]: Morgan Kaufmann, 2011.
 - [11] WARD S, BISHOP C M, WOAKES A J, *et al.* Heart rate and the rate of oxygen consumption of flying and walking barnacle geese and bar-headed geese [J]. *Journal of Experimental Biology*, 2002, 205(21): 3347-3356.
 - [12] GIANNOTTI F, NANNI M, PINELLI F, *et al.* Trajectory pattern mining[C]//Proc of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2007: 330-339.
 - [13] VERHEIN F, CHAWLA S. Mining spatio-temporal patterns in object mobility databases [J]. *Data Mining and Knowledge Discovery*, 2008, 16(1): 5-38.
 - [14] LI Xiao-lei, HAN Jia-wei, KIM S, *et al.* ROAM: rule-and motif-based anomaly detection in massive moving object data sets[C]//Proc of the 7th SIAM International Conference on Data Mining. 2007.