

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200810156588.4

[51] Int. Cl.

G06F 17/22 (2006.01)

G06F 17/27 (2006.01)

G06F 17/30 (2006.01)

[43] 公开日 2009 年 3 月 25 日

[11] 公开号 CN 101393544A

[22] 申请日 2008.10.7

[21] 申请号 200810156588.4

[71] 申请人 南京师范大学

地址 210046 江苏省南京市仙林新城文苑路 1 号

[72] 发明人 张雪英 申琪君 李伯秋 陈文君

[74] 专利代理机构 南京知识律师事务所

代理人 程化铭

权利要求书 3 页 说明书 14 页 附图 10 页

[54] 发明名称

面向地址编码的中文地址语义解析方法

[57] 摘要

本发明公开了一种面向地址编码的中文地址语义解析方法，其步骤如下：第一步：根据样本数据，构建地址特征字库；a. 建立样本数据；b. 筛选特征字；c. 筛选附属特征字；所筛选的特征字和附属特征字构成了特征字库；第二步：根据特征字库，按照地址表示规则，将中文地址转换为数字形式的字符串；第三步：构建地址解析规则库；第四步：语义解析，包括：地址表示，将原始地址转换为数字表示、地址解析，将数字表示的地址拆分为地址要素；地址还原，将数字表示的解析结果还原为与原始地址对应的字符串。

1、一种面向地址编码的中文地址语义解析方法，其步骤如下：

第一步：根据样本数据，构建地址特征字库

a、建立样本数据：将原始地址数据中的各个地址要素分离出来，形成样本数据；

b、筛选特征字：将样本数据中所有地址要素的最后一个字符和两个字符的频率分别进行统计，并按照由大到小排序；将累积频率占百分比80%以上的单个字符筛选为特征字；将累积频率占80%以上两个字符筛选为特征字，必须最后一个字符不是单特征字；

c、筛选附属特征字；

所筛选的特征字和附属特征字构成了特征字库；

第二步：根据特征字库，按照地址表示规则，将中文地址转换为数字形式的字符串，其中1表示特征字，2表示附属特征字，3表示两个连续重复特征字的后一个字符，0表示普通字符，9表示结束符；将连续的0字符压缩为一个0字符；

第三步：构建地址解析规则库

将中文地址转换为数字串之后，其构成均遵循以下规则：

“0”后只能是“1”、“2”、“9”中的一个数；

“1”后只能是“0”、“1”、“2”、“3”、“9”中的一个数；

“2”后只能是“0”、“1”、“2”、“9”中的一个数；

“3”后只能是“0”、“1”、“2”、“9”中的一个数；

只能以“0”、“1”、“2”、“3”中的一个数开始；

只能以“9”结束;

按照上述规则,将地址表示为树结构,每一条路径代表一条解析规则,树的第一级节点分别为“0”、“1”、“2”、“3”,其后裔结点按照上述规则来组织;当各条路径到达一定长度时,可以确定地址要素的拆分点,从而终止该路径的继续扩展;同时,每条解析规则规定具体的拆分位置,并用“f+拆分位置”表示;

第四步:语义解析

输入:一条原始地址数据,用 Address-Before 表示;

A、地址表示:将原始地址转换为数字表示,结果为 Numbers-Before,

a、Numbers-Before 置为空串;用 n 表示当前待解析原始地址的长度

b、i 从 1 直到 n,循环执行:

如果原始地址的第 i 个字符为主特征字,则 Numbers-Before[i]表示为 1;

如果原始地址的第 i+1 个字符为主特征字,则 Numbers-Before[i]表示为 3;

如果原始地址的第 i 个字符为附属特征字,则 Numbers-Before[i]表示为 2;

如果原始地址的第 i 个字符为普通字符,则 Numbers-Before[i]表示为 0;

i 赋值为 i+1;

结束循环;

c、在 Numbers-Before 的末尾添加 9;

d、将 Numbers-Before 中的连续多个 0 压缩为一个 0;

B、地址解析:将 Numbers-Before 根据解析规则拆分为地址要素,结果为 Numbers-After;

a、Numbers-After 置为空串;用 k 表示 Numbers-before 的长度;

b、m 从 1 直到 k，执行循环：

如果 Numbers_Before 的左边 m 个字符与某条解析规则匹配，则按照规则将 Numbers_before 拆分左右两个子串；左子串 Numbers_Left 保存为解析结果中的一个地址要素，不再进行拆分；右子串 Numbers-right 继续进行拆分；将 Numbers_Before 定义为 Numbers-right；

m 赋值为 Numbers_Left 的长度+1；

否则

m 赋值为 m+1；

结束循环；

C、地址还原：将数字表示的解析结果还原为与原始地址对应的字符串，结果为 Address_After。

面向地址编码的中文地址语义解析方法

技术领域

本发明提出了一种不依赖地名词典的中文地址语义解析方法,适用于房地产管理、土地管理、城市规划、公安、邮政、税收、电讯、公共卫生和号码百事通等领域地理信息系统(GIS)中的地址编码。

背景技术

在日常生产与生活中,地址是最常用的使用自然语言描述地理位置的参考系统之一。地址描述是当前各类业务系统中描述空间位置的最常用手段,利用地址编码技术能够使大量的原来已经存在于管理信息系统(MIS)中的地址定位信息转换成可以被用于地理信息系统(GIS)的地理坐标方式,使得GIS可以通过对地理数据的集成、存储、检索、操作和分析,将分散在各个部门的数据通过空间参照系联系起来,为土地利用、资源管理、环境监测、交通运输、城市规划等提供决策服务,从而大大促进GIS技术的应用。

地址编码是指将自然语言描述的地址信息,根据地址模型和编码规则进行智能语义解析,通过与数据库中匹配,建立与对应的空间坐标信息和地理编码关联的过程,其基本原理如图1所示。

地址编码需要解决以下三个关键技术问题:①地址语义解析:指将自然语言描述的地址拆分为在某一限定区域内,可以指定某一具体地理范围的地址要素。比如“南京市鼓楼区宁海路122号”解析为“南京市”、“鼓楼区”、“宁海路”、“122号”四个地址要素。地址中各个地址要素按照从大到小的关系排列,后面的地址要素必须相对于前面的地址要素才有意义。②地址模型:地址

模型用于描述各种类型地址中地址要素的构成规则；③地址匹配：是指根据既定的地址模型和编码规则，将通过计算机语义解析的地址与 GIS 中标准地址进行匹配，并给出地理坐标值的智能化过程。

从 20 世纪 70 年代起，美国就开始建立全国地理编码系统——“双重独立地图编码系统”（Dual Independent Map Encoding, DIME），DIME 的开发在 GIS 技术的发展史上具有里程碑的意义。80 年代后期，美国国情普查局将 DIME 系统发展为拓扑集成的地址编码参照系统（Topologically Integrated Geographic Encoding and Referencing, TIGER）。由于 TIGER 数据库覆盖范围广，精度好，更新有保证，而且费用低廉，已经成为美国的公认地址编码参考标准。目前，国外地理编码库以及地理编码软件工具已经商品化，有很多关于地址数据的内容标准和规范说明，例如 FGDC 地址数据内容标准公共草案。绝大多数国外 GIS 软件中都有地址编码功能，比如 Mapinfo 的 MapMaker，ArcGIS 的 Geocoding 和 GeoMedia 的 Geocodes Addresses，在具体应用中具有很高的响应速度和准确率。国外地址编码技术成功推广应用的关键因素在于：一是从语言角度看，英文等西方语种地址描述中单词之间存在空格分隔符；从标准化角度看，地址的数据命名和表述，以及软件开发和应用服务都遵循标准化和规范化原则。因此，国外地址编码中地址语义解析都采用与标准地址库中地名进行简单字符匹配的方法（统称“词典匹配法”）。

具体地讲，词典匹配法是以 GIS 中地名数据为地名词典，运用字符串匹配算法进行词法和语法分析，将地址串与词典中的地名进行匹配，达到地址解析的目的。词典匹配法只能将地址串中与词典中完全相同的地名进行解析。例如，假设词典中收录有“南京市”、“鼓楼区”、“宁海路”三个地名而没有“文

苑路”，就可以将地址“南京市鼓楼区宁海路122号”解析为“南京市”、“鼓楼区”、“宁海路”、“122号”四个地址要素，而将“南京市文苑路12号”解析为“南京市”和“文苑路12号”。因此，词典匹配法的效果与词典规模和更新速度呈正比，效率却与词典规模呈反比。然而，由于汉语言文字自身的历史和文化的特點，以及地址编码规范的严重匮乏，导致中文地址存在较为严重的不规范性，不仅地址命名的规律性差、格式复杂、存在一地多名的现象，而且增加、删除、修改比较随意。很显然，词典匹配法在适用范围、更新维护，准确率和响应速度等方面都不能满足大规模数据处理的需求。中文地址语义解析已成为目前中文地址编码技术需要解决的重点问题（黄颂. 中文地址编码技术的研究[D]. 北京大学: 硕士学位论文，2005）。

从上世纪80年代开始，国内地址编码技术研究侧重于地址标准库和地址匹配算法，例如，北京市和上海市先后颁布了一系列城市道路、道路交叉口等的编码标准（朱建伟，王泽民. 地理编码原理及其本地化解决方案[J]. 北京测绘, 2004(2), 24-27.）。地址语义解析都采用字典匹配法，如北京长地计算机公司的“寻址神”，北大方正的“Map Searcher”，朝夕科技的“北京地址编码数据库系统及标准地址匹配引擎”，北京超图公司的“客户关系管理系统”、山海易绘的EzGeoCoding等系统。

发明内容

本发明所要解决的技术问题在于，克服现有技术存在的缺陷，提供一种面向地址编码的中文地址语义解析方法，而不依赖于地名词典进行地址解析。

本发明面向地址编码的中文地址语义解析方法，其具体技术流程如下：

第一步：根据样本数据，构建地址特征字库

中文地址包括行政区划、街道、门楼牌号和补充信息四个类型的地址要素，按照地址要素的地理区域范围由大到小排列（最后为补充信息）构成中文地址串；中文地址串的构成中会有部分类型的地址要素缺失；

所述行政区划，按照<中华人民共和国行政区代码>（GB 2260-1995），分为四级，由村以上的行政区域由大到小（会有缺失）排序：第一级为省、自治区、直辖市和特别行政区；第二级为市、地区、自治州、盟及国家直辖市所属市辖区和县；第三级为县、市辖区、县级市、旗；第四级为乡、镇、村；

一般说来，一个地址中往往包含多个不同级别的行政区划名称。例如“南京市鼓楼区宁海路122号”中包括“南京市”（第二级）和“鼓楼区”（第三级）两个不同级别的行政区划名称。

所述街道是指路名和/或街道名；

所述门楼牌号是指门牌号、楼牌号、楼名和/或房间号；

所述补充信息，是指门楼牌号之后加上的机构名称或者表示空间关系的词汇（东、西、南、北等），比如“南京市鼓楼区江东北路301号滨江市场”中的“滨江市场”就是一个机构名称，“南京市江浦县永宁镇西葛街西”中的“西”就是一个表示空间方向关系的词汇。

一个中文地址串可以拆分为多个不同类型的地址要素；地址要素为普通字符+特征字的组合（补充信息除外）；其中

行政区划的特征字为：省、自治区、直辖市、特别行政区、市、地区、自治州、盟、区、县、旗、乡、镇、村、屯、庄等；

街道的特征字为：路、街道、街、大街、大道、马路、里、弄、胡同、巷、条等；

门楼牌号的特征字为：号、楼、宿舍、斋、馆、堂等；

构建地址特征字库包括以下几个步骤：

1、建立样本数据：将原始地址数据中的各个地址要素分离出来，形成样本数据。

表 1 原始数据样例

原始数据
南京市玄武区后宰门西村 87 号
六合区竹镇镇仕林路 2 号
南京市浦口区桥林镇南二村山根组
南京市江浦县龙山乡龙南村
南京市雨花台区西云村 3 号
南京杭州环北市场负一楼 037-054 号
秦淮区红花街道翁家营村翁家营 153 号

表 2 样本数据样例

样本数据	地址要素类型	最后一个字符	最后两个字符
南京市	行政区划	市	京市
玄武区	行政区划	区	武区
后宰门西村	街道	村	西村
87 号	门楼牌号	号	7 号
六合区	行政区划	区	合区
竹镇镇	行政区划	镇	镇镇
仕林路	街道	路	林路
2 号	门楼牌号	号	2 号
南京市	行政区划	市	京市
浦口区	行政区划	区	口区
桥林镇	行政区划	镇	林镇
南二村	行政区划	村	二村
山根组	行政区划	组	根组
南京市	行政区划	市	京市
江浦县	行政区划	县	浦县
龙山乡	行政区划	乡	山乡
龙南村	行政区划	村	南村
南京市	行政区划	市	京市
雨花台区	行政区划	区	台区

西云村	行政区划	村	云村
3 号	街道	号	3 号
南京	行政区划	京	南京
杭州环北市场	补充信息	场	市场
负一楼	门楼牌号	楼	一楼
037-054 号	门楼牌号	号	4 号
秦淮区	行政区划	区	淮区
红花街道	街道	道	街道
翁家营村	行政区划	村	营村
翁家营	街道	营	家营
153 号	门楼牌号	号	3 号

2、筛选特征字：特征字表示一个地址要素的结尾，可以看作是地址要素的单位；大多数情况下，根据特征字就可以比较准确地将地址划分成独立的语义单元。特征字筛选过程：将样本数据中所有地址要素的最后一个字符和两个字符的频率分别进行统计，并按照由大到小排序；将累积频率占百分比 80%以上的单个字符筛选为特征字（称为“单特征字”）；将累积频率占 80%以上两个字符（必须最后一个字符不是单特征字）筛选为特征字（称为“复特征字”）；

3、筛选附属特征字：中文地址中通常包含一些表达空间关系的词汇，如东、南、西、北等，可用于辅助判断地址要素的拆分位置，将这些词汇筛选为附属特征字；

所筛选的特征字和附属特征字构成了特征字库；

第二步：根据特征字库，按照地址表示规则，将中文地址转换为数字形式的字符串；

为了便于计算机处理，需要将中文地址字符串转换为数字表示，其中 1 表示特征字，2 表示附属特征字，3 表示两个连续重复特征字的后一个字符，0 表示普通字符，9 表示结束符。普通字符对于拆分规则的制定没有意义，可将连续

的 0 字符压缩为一个 0 字符。例如，“江苏省六合县八百桥镇冶东村小林 32 号”表示为“01010110212019”，“建邺区应天路叶圩村村部”表示为“0101011309”。

第三步：构建地址解析规则库

将中文地址转换为数字串之后，其构成均遵循以下规则：

- “0”后只能是“1”、“2”、“9”中的一个数；
- “1”后只能是“0”、“1”、“2”、“3”、“9”中的一个数；
- “2”后只能是“0”、“1”、“2”、“9”中的一个数；
- “3”后只能是“0”、“1”、“2”、“9”中的一个数；
- 只能以“0”、“1”、“2”、“3”中的一个数开始；
- 只能以“9”结束。

按照上述规则，地址可以表示为树结构，每一条路径代表一条解析规则。树的第一级节点分别为“0”、“1”、“2”、“3”，其后裔结点按照上述规则来组织。但是，当各条路径到达一定长度时，可以确定地址要素的拆分点，从而终止该路径的继续扩展。同时，每条解析规则必须规定具体的拆分位置，并用“f+拆分位置”表示。以样本数据中的地址为例，对规则的应用频率进行统计，将累积频率占 95%以上的解析路径筛选为解析规则（如表 3 所示）。解析规则的树状结构如图 3 所示。

表 3 地址解析规则

序号	解析规则	序号	解析规则
1.	010f2	93.	0202019f6
2.	019f2	94.	02110101f4
3.	2010f3	95.	011122101f3
4.	0110f3	96.	112f2
5.	09f1	97.	020210101f5
6.	0210f3	98.	309f2
7.	012f2	99.	02021019f7

8.	0120f2	100.	012101021f2
9.	0112f3	101.	2020109f2
10.	0119f3	102.	21119f4
11.	2110f3	103.	012101011f2
12.	21010f2	104.	202010101f2
13.	21019f2	105.	211111f3
14.	21210f4	106.	21112f3
15.	2012f3	107.	119f2
16.	2019f3	108.	3101f2
17.	20210f4	109.	202021f2
18.	01110f4	110.	012101010109f2
19.	10f2	111.	02221f1
20.	0219f3	112.	21211019f5
21.	02010f4	113.	11109f4
22.	201101f4	114.	2209f3
23.	219f2	115.	0111219f3
24.	0212f3	116.	0202109f6
25.	02210f4	117.	202020101f2
26.	012101019f2	118.	2120f2
27.	11101f3	119.	2013f3
28.	011110f2	120.	1211f1
29.	2210f3	121.	0121010101010f4
30.	2109f3	122.	02211f5
31.	02019f4	123.	210111f2
32.	2119f3	124.	2111101f2
33.	209f2	125.	020119f5
34.	2112f3	126.	020209f5
35.	013f2	127.	021119f3
36.	201102f4	128.	202120f4
37.	301f3	129.	020219f5
38.	2219f3	130.	20111019f3
39.	1210f3	131.	1212f3
40.	0209f3	132.	2021109f4
41.	2201f4	133.	129f2
42.	201109f3	134.	201121019f3
43.	2221f4	135.	2021101f2
44.	21021f2	136.	020210109f8
45.	01210101019f4	137.	02110109f3
46.	21219f4	138.	22119f4
47.	19f1	139.	0113f3
48.	2101101f2	140.	02111019f3
49.	21012f2	141.	2102019f2
50.	29f1	142.	213f2
51.	110f2	143.	20221f5

52.	0202010f6	144.	0221201f4
53.	20119f4	145.	212112f4
54.	01119f4	146.	229f2
55.	01112101f3	147.	021120f4
56.	02012f4	148.	212119f2
57.	20201019f2	149.	01112019f3
58.	1119f3	150.	0202210f6
59.	021109f3	151.	2011109f4
60.	20212101f2	152.	21020109f2
61.	211101f4	153.	011129f3
62.	022010f5	154.	021102f3
63.	2122f2	155.	3109f3
64.	2212f3	156.	021101101f4
65.	0211019f3	157.	022019f5
66.	21212f4	158.	2121109f4
67.	1112f3	159.	021110101f5
68.	0121010109f2	160.	021111f3
69.	22110f4	161.	02209f4
70.	20219f4	162.	210201019f6
71.	02119f4	163.	2102210f2
72.	11102f3	164.	01112010f4
73.	029f2	165.	11110f4
74.	2202f5	166.	20201109f5
75.	202010109f5	167.	211119f2
76.	011119f2	168.	3219f3
77.	120f1	169.	39f1
78.	3210f3	170.	0111211f2
79.	210119f2	171.	201119f3
80.	011111f3	172.	3021f4
81.	211109f5	173.	0202021f7
82.	2101109f4	174.	021110109f3
83.	02219f4	175.	0211109f3
84.	0201101f5	176.	20201011f2
85.	0201109f6	177.	20212109f2
86.	011112f2	178.	210201011f6
87.	0121010101019f2	179.	210209f2
88.	2011201f4	180.	0121010201010f4
89.	1219f3	181.	201111019f3
90.	12210f4	182.	202011019f2
91.	202019f5	183.	2101102f2
92.	20209f4	184.	3221f4

图3中,规则“0120f2”表示在各层上的节点分别为“0”、“1”、“2”、“0”。当扫描地址串时,如果有字符串序列与“0120”匹配,即可确定在第二个数字后面拆分,用“f2”表示拆分位置。例如,“白下区南台巷”表示为“01201”,与规则“0120”匹配,即在左边“01”之后拆分,将其解析为“白下区”和南台巷。规则“029f2”表示当地址串中的部分字符序列与“029”匹配时,即可确定在第二个数字后面拆分。由于“9”即表示数字串结束,所以规则也表示在末尾拆分。

第四步:语义解析

在特征字和解析规则制定的基础上,本发明设计了一种中文地址解析算法(简称“RBAI算法”)。该算法包括三个部分:地址表示、地址解析和地址还原。具体解析过程如下:

输入:一条原始地址数据,用 Address-Before 表示;

输出:该原始地址的解析结果,用 Address-After 表示。

(1)地址表示:将原始地址转换为数字表示,结果为 Numbers-Before。

第1步:Numbers-Before 置为空串;用 n 表示当前待解析原始地址的长度

第2步:i 从 1 直到 n,循环执行:

如果原始地址的第 i 个字符为主特征字,则 Numbers-Before[i]表示为 1;

如果原始地址的第 i+1 个字符为主特征字,则 Numbers-Before[i]表示为 3;

如果原始地址的第 i 个字符为附属特征字,则 Numbers-Before[i]表示为 2;

如果原始地址的第 i 个字符为普通字符,则 Numbers-Before[i]表示为 0;

i 赋值为 i+1;

结束循环;

第3步: 在 Numbers_Before 的末尾添加 9; 将 Numbers_Before 中的连续多个 0 压缩为一个 0。

(2) 地址解析: 将 Numbers_Before 根据解析规则拆分为地址要素, 结果为 Numbers_After。

第4步: Numbers_After 置为空串; 用 k 表示 Numbers_before 的长度

第5步: m 从 1 直到 k, 执行循环:

如果 Numbers_Before 的左边 m 个字符与某条解析规则匹配, 则按照规则将 Numbers_before 拆分左右两个子串; 左子串 Numbers_Left 保存为解析结果中的一个地址要素, 不再进行拆分; 右子串 Numbers-right 继续进行拆分; 将 Numbers_Before 定义为 Numbers-right;

m 赋值为 Numbers_Left 的长度+1;

否则

m 赋值为 m+1;

结束循环;

(3) 地址还原: 第6步: 将数字表示的解析结果还原为与原始地址对应的字符串, 结果为 Address_After。

相对于现有中文地址编码中采用的地址解析技术, 本发明主要以下几个优点:

- a. 不依赖于词典: 避免词典构建和更新, 而且可以解析出 GIS(或词典)中没有收录的地址要素名称;
- b. 不依赖中文分词等自然语言处理技术;
- c. 效率高: 由于没有采用基于词黄的字符串匹配算法, 而是对单个地址进

行操作，效率显著提高；

d. 适用性强：可以根据实际应用情况，通过更新样本数据，快速更新特征字库和解析规则，而解析算法则不需要更新。

e. 实现简单，易于推广：特征字库和解析规则重用性强，算法简单，方便嵌入各类应用系统中。

附图说明

图 1、是地址编码基本原理与过程示意图；

图 2、是本发明方法流程示意图；

图 3、表示基于树结构的解析规则；

图 4、地址解析示意图（地址：六合县雄州镇朝天街 108 号）；

图 5、地址解析示意图（地址：江苏省六合县八百镇金山村）；

图 6、地址解析示意图（地址：六合县六城镇泰山村 82 号）；

图 7、地址解析示意图（地址：六合区八百桥镇街道）；

图 8、地址解析示意图（地址：六合区雄州镇健康巷 1 号-2）；

图 9、地址解析示意图（地址：南京市玄武区明故宫 4 号）；

图 10、地址解析示意图（地址：六合区雄州镇中心农贸市场）；

图 11、地址解析示意图（地址：北门桥路 5 号 302 室）；

图 12、地址解析示意图（地址：六合区程桥镇东大桥边）；

图 13、地址解析示意图（地址：玄武区相府营 14 号 104 室）；

图 14、中文地址语义解析算法流程图；

图 15、地址表示算法流程图；

图 16、地址解析算法流程图。

具体实施方式

下面结合附图和实施例对本发明方法作进一步详细说明。

在主频 1.73GHz 和内存 1GB 的 PC 机上，用 Visual C#.NET2003 平台开发的实现本发明技术的软件系统，以江苏省南京市 194719 户企业地址数据为样本数据，根据样本数据，构建地址特征字库，根据特征字库，按照地址表示规则，进行地址语义解析操作。

实例 1：中文地址字符串：六合县雄州镇朝天街 108 号，如图 4 所示，解析为：六合县/雄州镇/朝天街/108 号（“/”表示切分位置）。

实例 2：中文地址字符串：江苏省六合县八百镇金山村，如图 5 所示，解析为：江苏省/六合县/八百镇/金山村。

实例 3：中文地址字符串：六合县六城镇泰山村 82 号，如图 6 所示，解析为：六合县/六城镇/泰山村/82 号。

实例 4：中文地址字符串：六合区八百桥镇街道，如图 7 所示，解析为：六合区/八百桥镇街道。

实例 5：中文地址字符串：六合区雄州镇健康巷 1 号-2，如图 8 所示，解析为：六合区/雄州镇/健康巷/1 号/-2。

实例 6：中文地址字符串：南京市玄武区明故宫 4 号，如图 9 所示，解析为：南京市/玄武区/明故宫/4 号。

实例 7：中文地址字符串：六合区雄州镇中心农贸市场，如图 10 所示，解析为：六合区/雄州镇/中心农贸市场。

实例 8: 中文地址字符串: 北门桥路 5 号 302 室, 如图 11 所示, 解析为: 北门桥路/5 号/302 室。

实例 9: 中文地址字符串: 六合区程桥镇东大桥边, 如图 12 所示, 解析为: 六合区/程桥镇/东大桥/边。

实例 10: 中文地址字符串: 玄武区相府营 14 号 104 室, 如图 13 所示, 解析为: 玄武区/相府营/14 号/104 室。

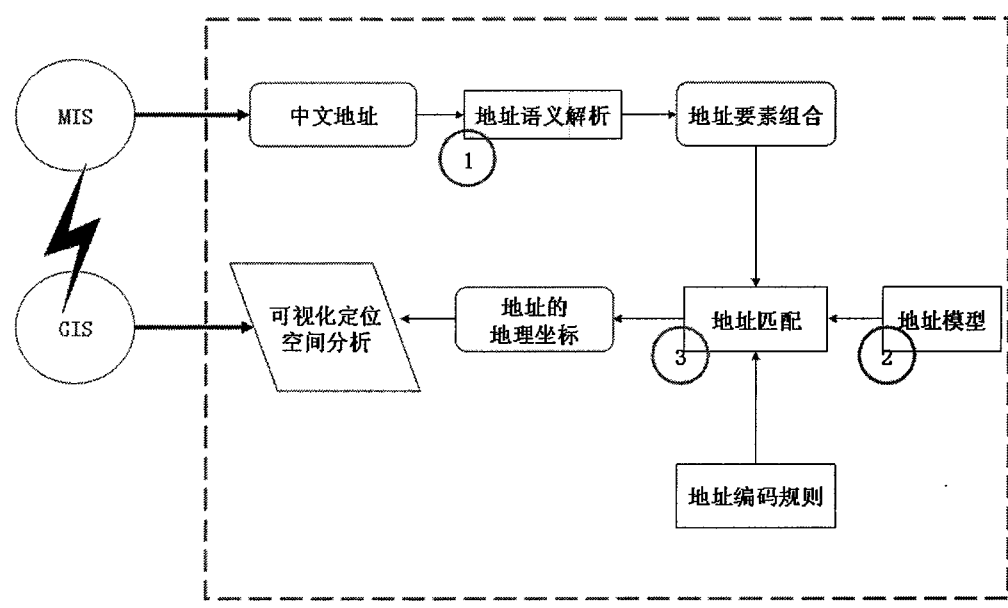


图 1

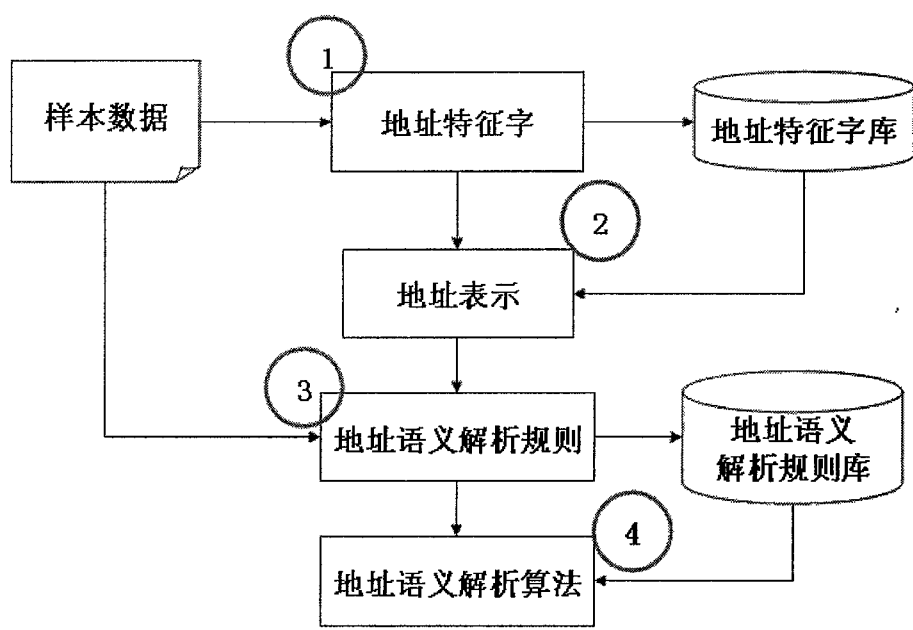


图 2

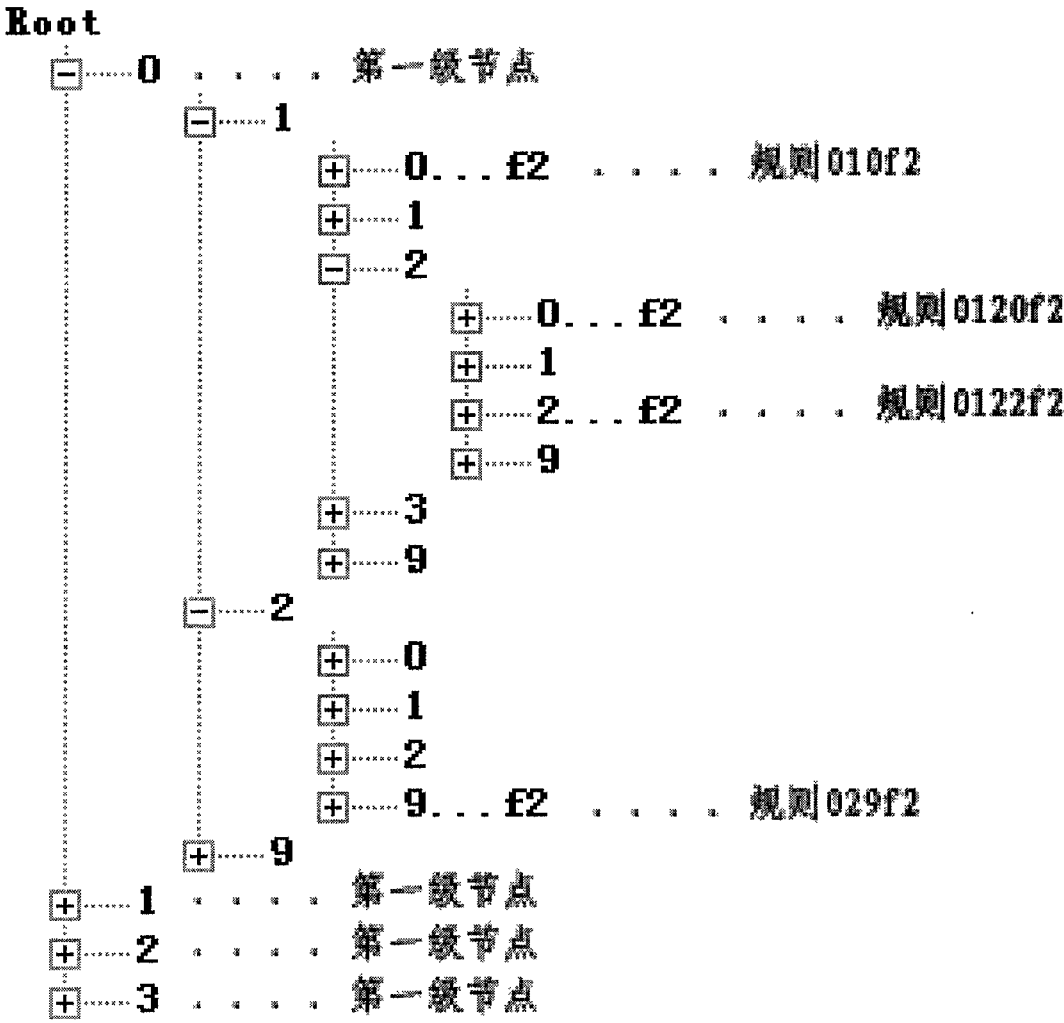


图 3

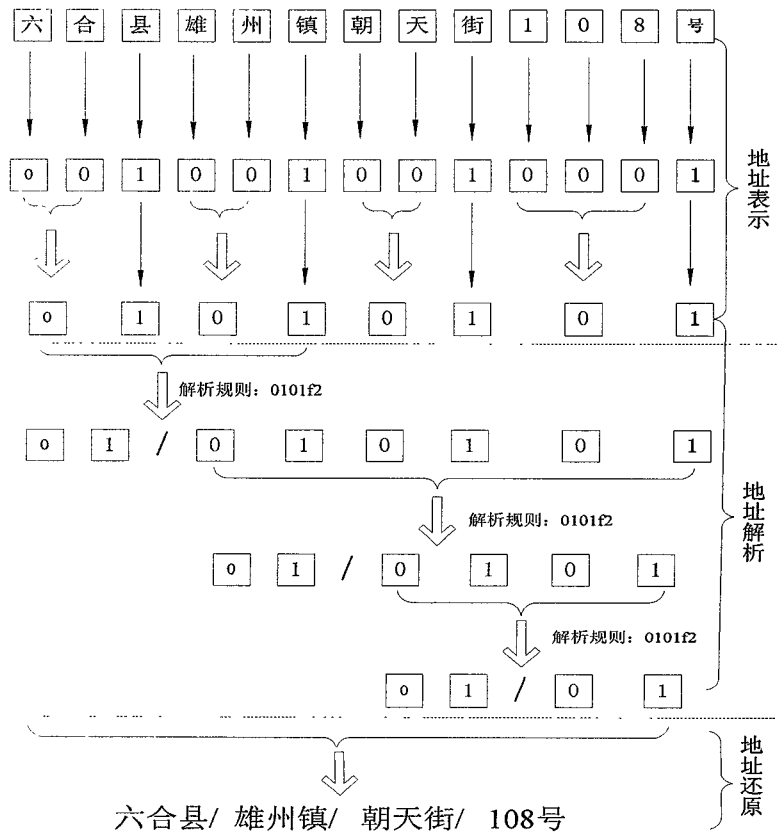


图 4

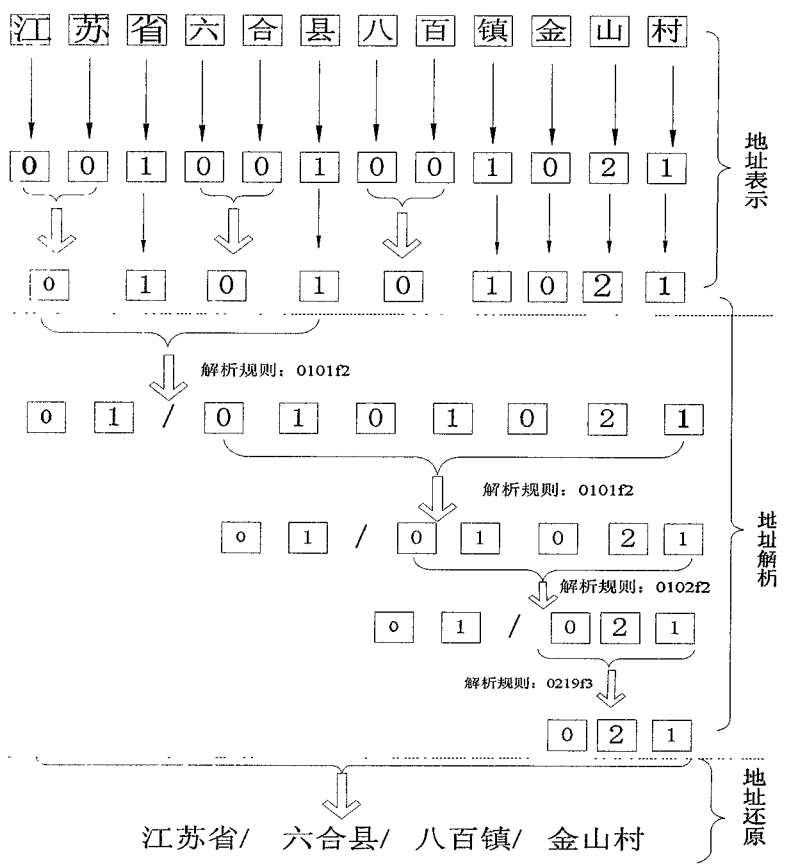


图 5

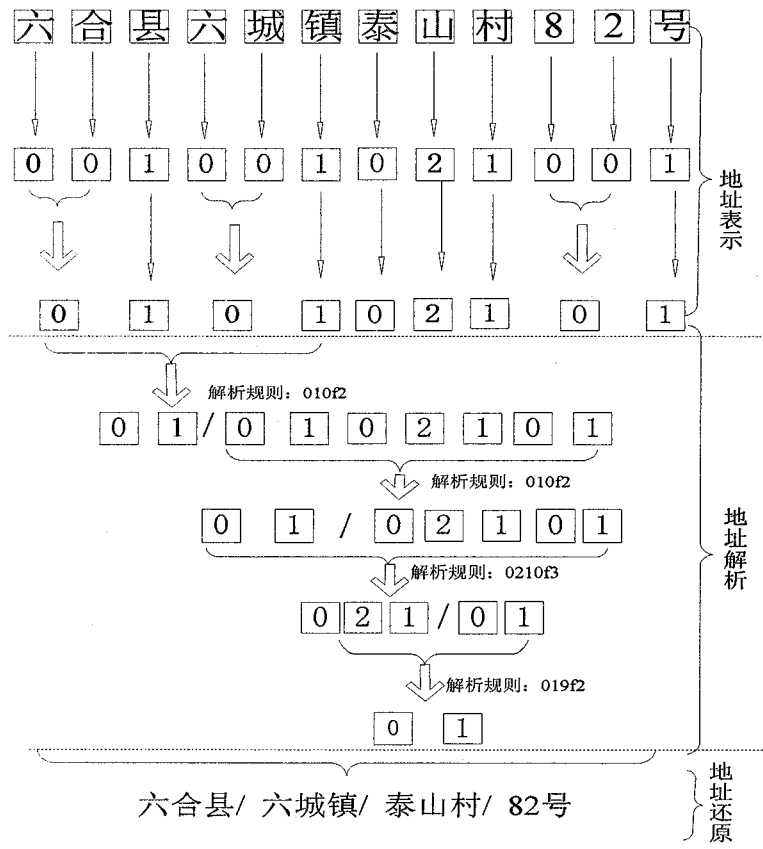


图 6

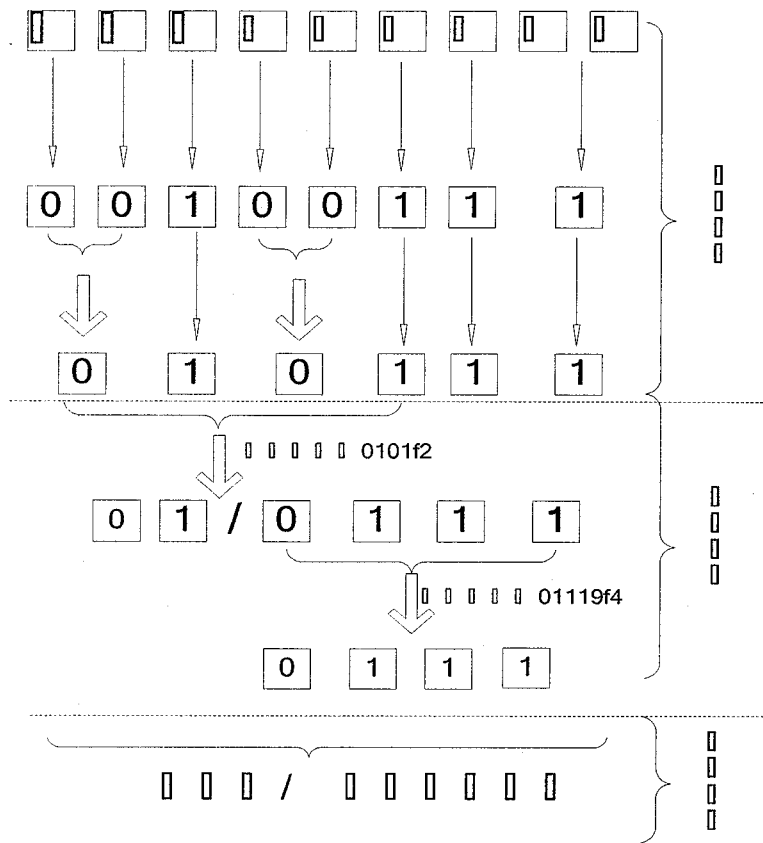


图 7

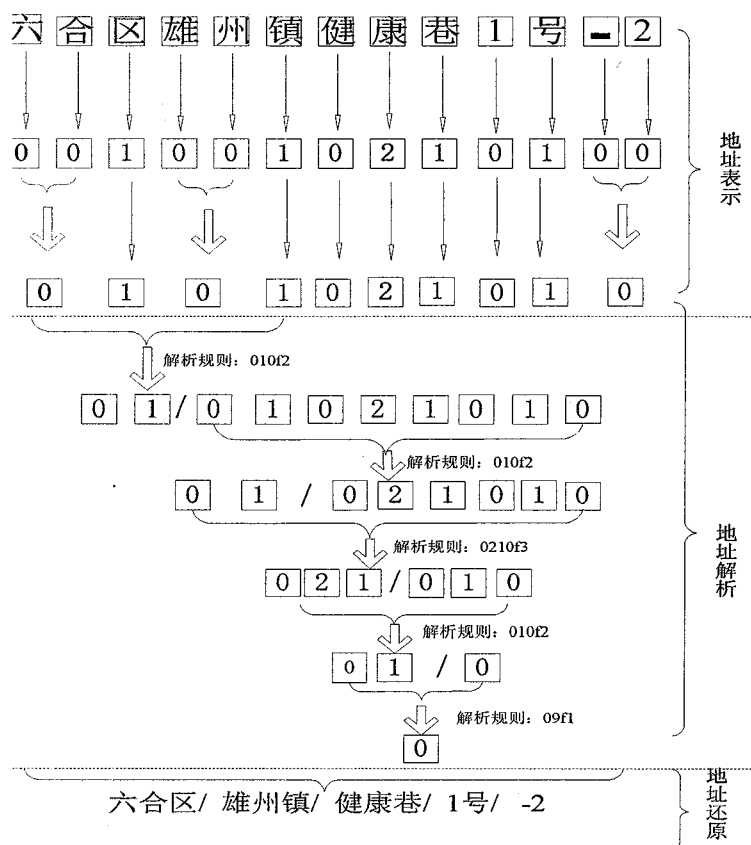


图 8

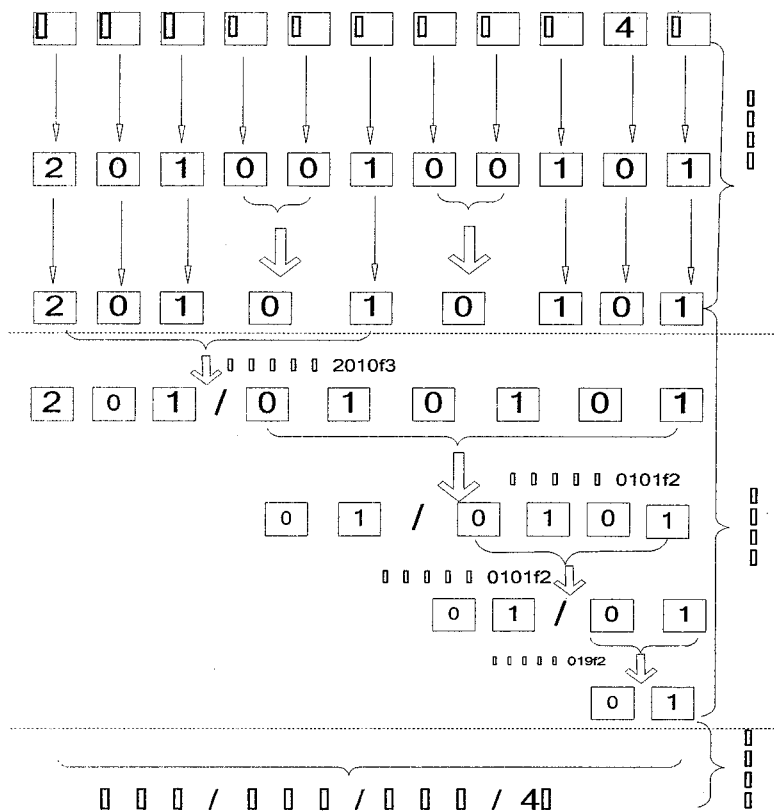


图 9

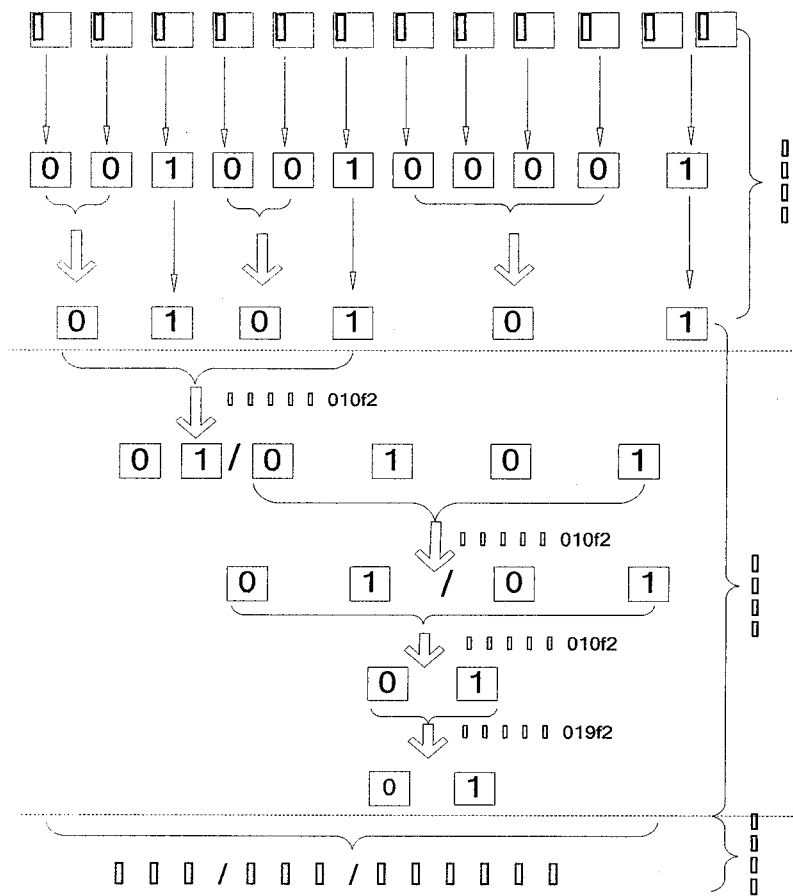


图 10

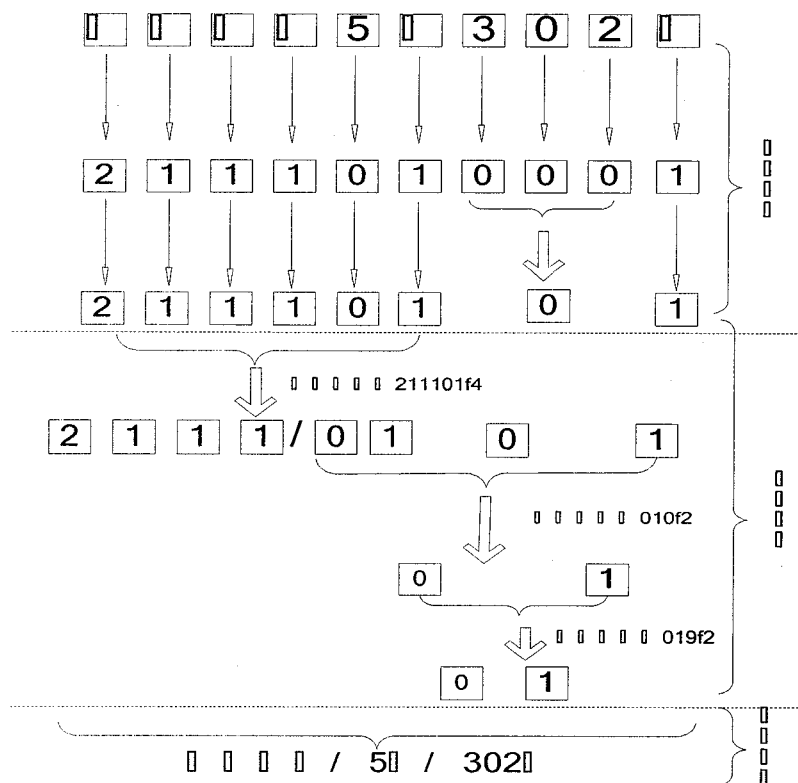


图 11

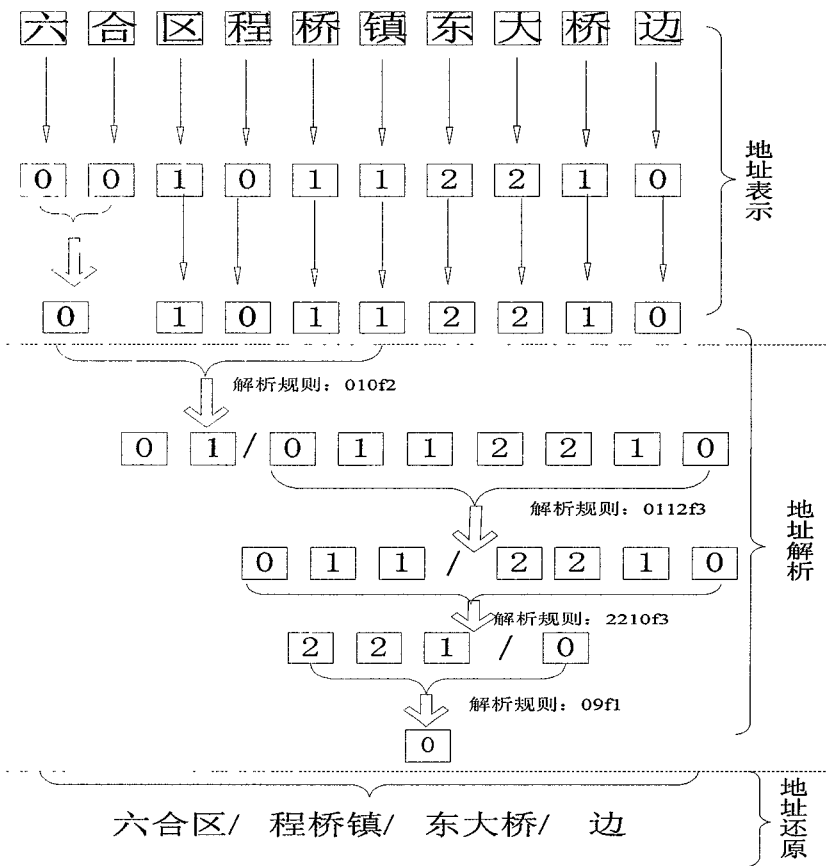


图 12

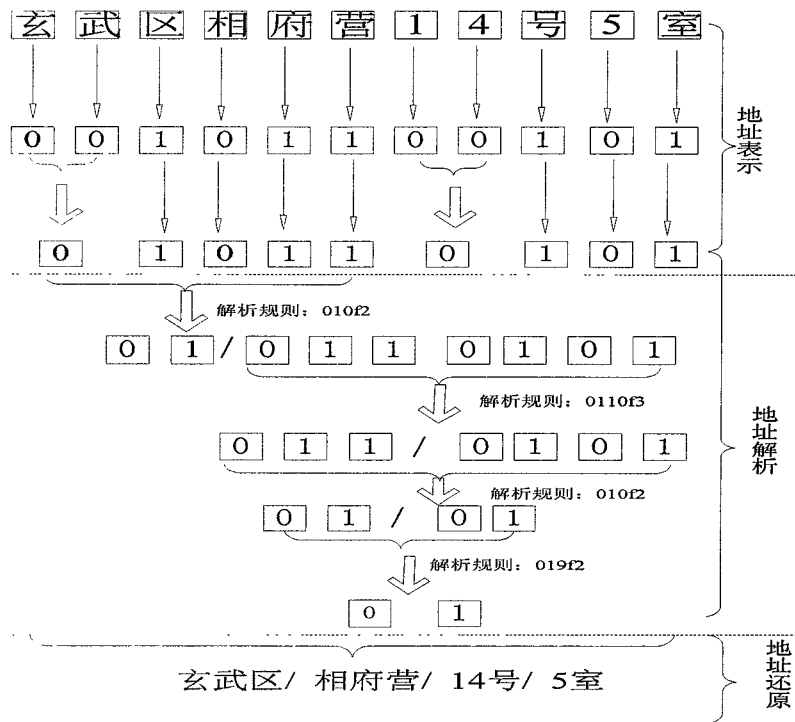


图 13

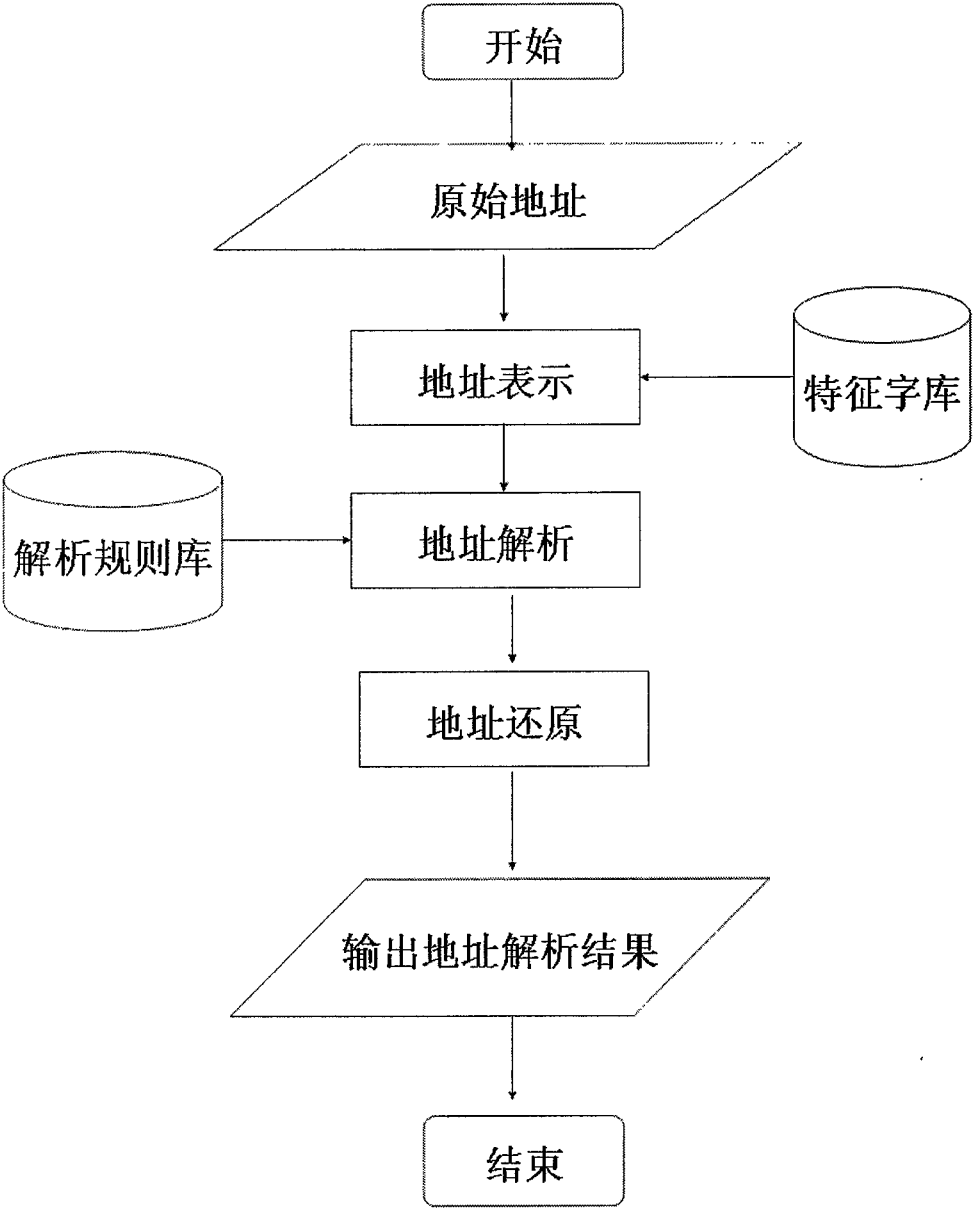


图 14

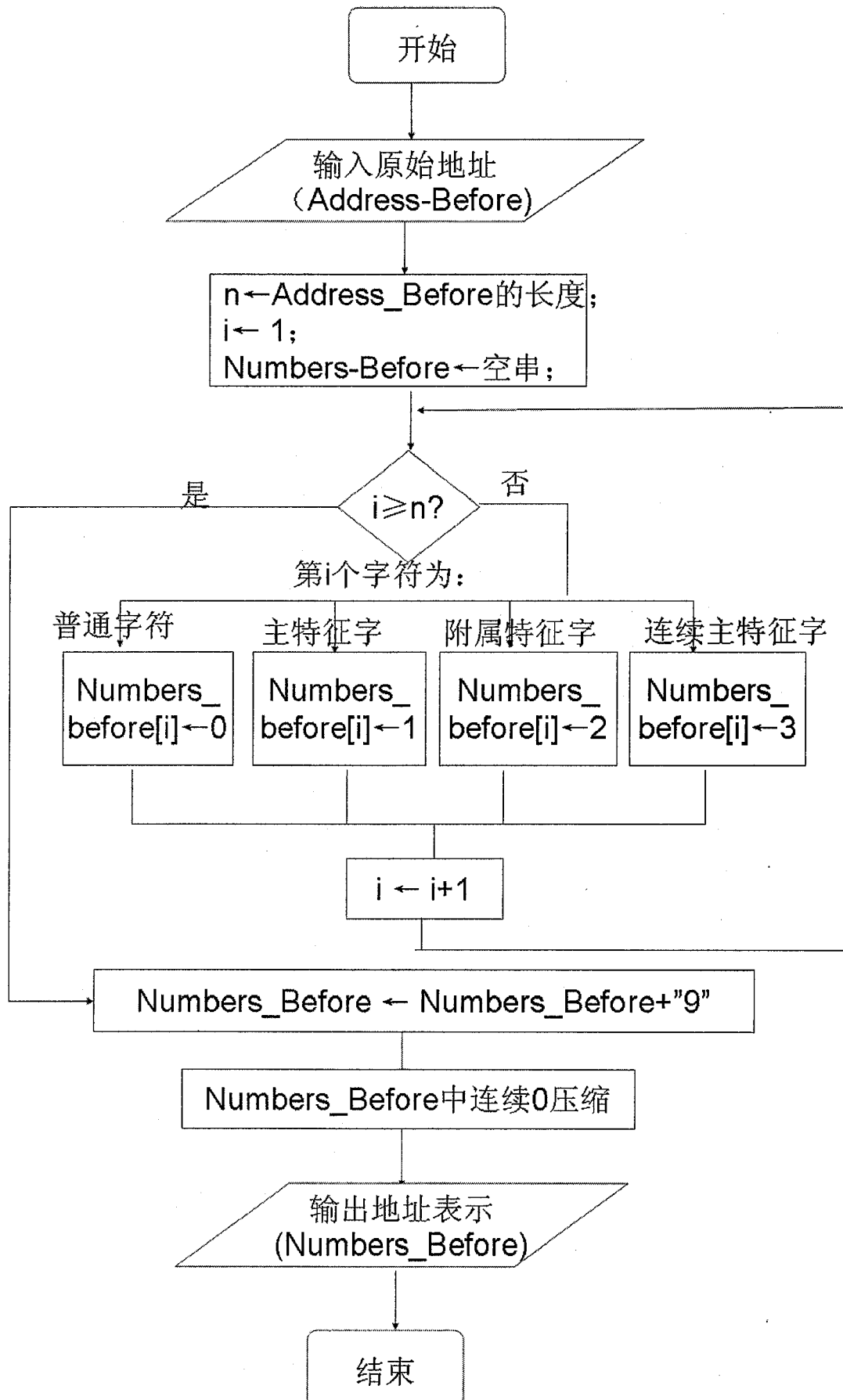


图 15

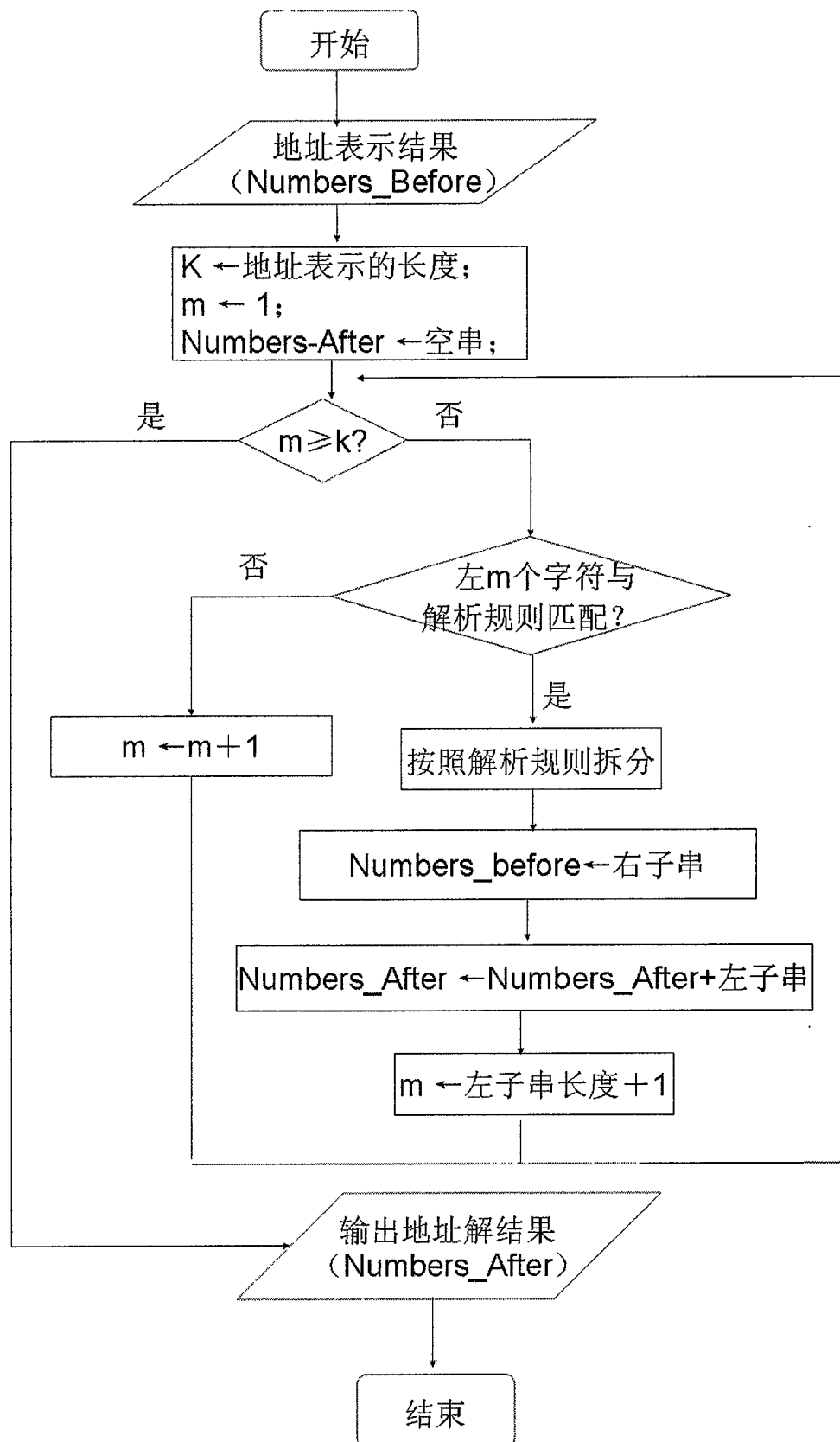


图 16