

词义排歧方法研究

◇冯志伟(教育部语言文字应用研究所)

关键词:自然语言处理 ;词义排歧 ;选择限制 ;有指导的学习方法 ;无指导的方法

摘 要:本文全面总结了 40 多年来在词义排歧方面的研究成果 ,分别讨论和比较各种词义排歧的方法 :选择最常见义项的方法、利用词类进行词义排歧的方法、基于选择限制的方法、鲁棒的词义排歧方法、有指导的学习方法、自力更生的词义排歧方法、无指导的词义排歧方法、基于词典的词义排歧方法等。这些方法对于自然语言的计算机处理都是非常重要的。

The Approaches for Word Sense Disambiguation (WSD)

FENG Zhiwei

Keywords: natural language processing, word sense disambiguation (WSD), selectional restriction, supervised learning approach, unsupervised approach

Abstract:The author discussed in detail the various approaches for the Word Sense Disambiguation (WSD): the approach for selection of commonly-used sense, the approach using POS for disambiguation, selectional restriction based approach, robust WSD, supervised learning approach, bootstrapping approach, unsupervised approach, dictionary-based approach. These WSD approaches are all very important for the natural language processing by computer.

词义排歧(Word Sense Disambiguation, WSD)是自然语言计算机处理中的一个困难问题。由于多义词是任何语言中都普遍存在的现象,而多义词中诸多的词义分布又很不容易找到一般的规律,多义词的排歧涉及到上下文因素、语义因素、语境因素,甚至还涉及到日常生活中的常识,而这些因素的处理,恰恰是计算机最感棘手的问题。

早在机器翻译刚刚问世的时候,美国著名数理逻辑学家 Bar-Hillel 在 1959 年就指出,全自动高质量的机器翻译(Fully Automatic, High Quality, MT,简称 FAHQMT)是不可能的,他说明,FAHQMT 不仅在当时的技术水平下是不可能的,而且,在理论原则上也是不可能的。他举出了如下简单的英语片段,说明要在上下文中发现多义词 pen 的正确译文是非常困难的事情。

John was looking for his toy box. Finally he found it. The box was in the pen. John was very

happy.

他的理由如下:

- (i)pen 在这里只能翻译为 play-pen (“游戏的围栏”),而绝对不能翻译为书写工具“钢笔”。
- (ii)要确定 pen 的这个正确的译文是翻译好这段短文的关键所在。
- (iii)要确定这样的正确译文依赖于计算机对于周围世界的一般知识。
- (iv) 我们没有办法把这样的知识加到计算机中去。

可见,词义排歧问题一开始就困扰着刚刚萌芽的机器翻译研究。

从 1959 年到现在,学者们在探索多义词排歧的研究中做了大量的工作。尽管词义排歧的问题距离彻底解决还非常遥远,但是,从这 40 多年的成就已经可以让我们看到希望的曙光。

词义排歧的方法归纳起来可以有如下几种:

1. 选择最常见义项的方法

早期的机器翻译系统没有词义排歧的功能,虽然机器词典中的多义词都列举出各种不同的义项,但实际上只是选择排列在第一位的那个最常见的义项。这样的办法虽然能够处理一些多义词,达到一定的排歧目的,但是,词义排歧的效率不高,这是早期机器翻译系统的译文质量低劣的重要原因之一。例如,在上面的例子中,由于 pen 最常见的词义是“钢笔”,因此就把 pen 翻译成“钢笔”,结果,The box was in the pen 就翻译成“盒子在钢笔中”,弄出了笑话。

2. 利用词类进行词义排歧的方法

有些多义词的词义与它们所属的词类有关。不同的词义往往属于不同的词类。因此,如果我们能够确定这些多义词的词类,词义排歧的问题也就迎刃而解了。例如:

face:当 face 是动词时,它的词义是“面对”;当 face 是名词时,它的词义是“面孔”。在“The house faces the park”中 faces 前面为名词词组,后面也为名词词组,可判定为动词,因而它的词义是“面对”,整句的意思是“房子面对公园”。在“She pulled a long face”中 face 前面是形容词,可判定为名词,其词义是“面孔”,整句的意思是“她拉长了面孔”。

May:当 May 是助动词时,它的词义是“可以”(在句子开头,第一个字母大写,在其他情况下,第一个字母不大写),当 May 是名词并且第一个字母大写时,它的词义是“五月”。在“May I help you?”中,May 是助动词。因而它的词义是“可以”,整个句子的意思是“我可以帮助你吗?”在“May Day is first day of May”中,May 是名词,因而它的词义是“五月”,整个句子的意思是“五月一日是五月的第一天”。

如果我们有一个高效率的词性标注系统,可以正确地决定兼类的多义词的词类,那么,我们就可以利用标注正确的词类,来决定多义词的词义,从而达到词义排歧的目的。可是,当同一个词类的多义词还存在多个不同的词义的时候,这种“以词类决定词义”的方法就显得无能为力了,因为在判定了词类之后,还需要对不同的词义进行选择。

例如,works 这个多义词可兼属动词和名词,当它是动词的时候,它的词义是“工作”,当它是名

词的时候,它的词义可以是“工厂”,也可以是“著作”。在句子“My daughter works in an office”,works 处于名词词组之后,介词之前,可判定为动词,因而它的词义是“工作”,整个句子的意思是“我女儿在一个办公室工作”。可是,当判定 works 为名词的时候,它的词义还没有最后决定,这就会出现两难的尴尬局面。在“It is a gas works”和“I read the works of Shakespears”中,works 都可以判定为名词,可是,我们还决定不了前句中 works 的词义是“工厂”,后句中的 works 的词义是“著作”。这时,我们还需要根据上下文的选择限制来排歧。比如说,如果我们规定,works 与表示燃料的名词连用,可判定其词义是“工厂”,当 works 与作家的名字连用,可判定其词义是“著作”,那么,我们就可根据这样的选择限制来进行词义排歧。

3. 基于选择限制的词义排歧方法

选择限制(selectional restriction)和语义类型的分类(type hierarchies)是词义排歧的主要的知识源。在语义分析中,它们被用来删除不恰当的语义从而减少歧义的数量。最早研究选择限制的是 Katz 和 Fodor(1963)。把选择限制应用于计算机处理的是 Hirst(1987)。

例如,dish 的排歧,试研究下面的一段话:

“In our house, every body has a career and none of them includes washing dishes,” he says. In her tiny kitchen at home, Mr. Chen works efficiently, stir-frying several simple dishes, including braised pig's ears and chicken livers with green peppers.(他说道:“在我们的房子里,每一个人都有自己的事情,可这些事情不包括洗碟子。”在她的小厨房里,陈先生干得很有成效,他炒炸几个简单的菜肴,包括炖猪耳朵和炒胡椒鸡肝。)

前句中的 dishes 是用于吃饭的物理物(physical object),后句中的 dishes 则是菜肴。它们的选择限制各不相同,前者是 wash 的 PATIENT(受事),它应该具有可洗性(washable),后者是 stir-fry 的 PATIENT(受事),它应该具有可食性(edible)。谓词选择其歧义论元的正确含义,删除不能匹配的含义。由此可见,使用选择限制实际上是一种“观其伴而知其意”(You shall know the a word by the company it keeps)的方法。

当谓词有歧义时,可以根据其论元(argument)

的语义来消除歧义。例如：

Well, there was the time served green-lipped mussels from New Zealand.(好,有时间来品尝从新西兰来的绿唇蚌。)

Which airlines serve Denver?(哪一个航班到 Denver 去?)

Which ones serve breakfast?(哪一个航班提供早餐?)

前句中的 serve 要求某种食物作为其 PATIENT, 中句中的 serve 要求地名或者团体作为其 PATIENT, 后句中的 serve 要求某种饭局作为其 PATIENT。如果我们确信 mussel, Denver 和 breakfast 都是无歧义的, 那么, 就可以通过它们的语义来消除 serve 的歧义。

4. 鲁棒的词义排歧方法

前面的方法是所谓“规则对规则”(rule-to-rule approach)的方法, 另外, 还有自立的方法(stand-alone approach)。自立的方法是一种鲁棒(robust)的词义排歧方法。鲁棒的自立的方法依靠词类标注来工作, 力求把对于信息的要求减低到最低限度, 从而做到“自立”, 让机器自己学习而获得信息。

这种机器学习的方法, 要求对系统积极性训练, 使得系统能够自行进行词义排歧。

要进行词义排歧的词叫做目标词(target word), 目标词所嵌入的文本, 叫做上下文(context)。输入按下面方式进行初始化的处理:

● 输入文本一般应该是经过词类标注的;

● 上下文可以看成是围绕目标词的长短不一的语言片段;

● 上下文中的单词, 应该是经过词法分析的, 应该把变形词还原成原形词;

● 文本最好是经过局部句法分析或者依存关系分析, 能够反映出题元角色关系或者其他语法关系。

经过这样的初始化处理, 输入文本要进一步提炼为包含相关信息的特征的集合。主要步骤是:

● 选择相关的语言学特征;

● 根据学习算法的要求对这些特征进行形式化描述(或者编码)。

大多数的学习系统使用简单的特征向量(feature vector), 这些特征向量采用数字或者词类标记

来编码。

用来训练词义排歧系统的语言学特征可以粗略地分为两类:

● 搭配特征(collocation feature);

● 共现特征(co-occurrence feature)。

搭配特征对目标词左右的上下文进行编码, 要求指出特定的、能反映这些单词的语法性质的位置特征。典型的特征是单词、词根形式、词类等。这样的特征往往能把目标词特定的含义孤立起来以便处理。

例如, An electric guitar and bass player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps. (电吉他和低音乐器演奏者站在一旁, 他并不是站在舞台的一部分, 大概只是为了等待外国佬的到来。)

我们取特征词 bass(低音乐器)的左右两个词以及它们的词类标记为特征向量, 表示如下:

[guitar, NN1, and, CJC, player, NN1, stand, VVB]

共现特征不考虑相邻词的精确的位置信息, 单词本身就可以作为特征。特征的值就是单词在围绕目标词的环境中出现的次数。目标词的环境一般定义为以目标词为中心的一个固定窗口, 要计算出在这个窗口中实词的出现频度。

例如, 对于目标词 bass, 我们从语料库中选出它的 12 个共现词。然后标出它们在特定窗口中的出现频度。

这 12 个共现词是 fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band.

在上面句子中选取 guitar and bass player stand 作为窗口, 则其特征向量为:

[0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0]

根据这样的特征向量, 由于第四个共现词 player 和第十一个共现词 guitar 在特征向量中的值都是 1, 因此可以确定这个 bass 的词义是“低音乐器”。

在鲁棒的词义排歧系统中, 一般都把搭配特征与共现特征结合起来使用。

5. 有指导的学习方法

有指导的学习方法(Supervised Learning Approaches)又可以再分为朴素的 Bayes 分类法和决策表分类法两种。

(1) 朴素 Bayes 分类法

使用朴素 Bayes 分类法 (naive Bayes classifier) 时,不是去寻找某个特定的特征,而是在综合考虑多个特征的基础上进行词义排歧。这种方法实际上是在给定的上下文环境下,计算一个多义词的各个义项中概率最大的义项。计算公式如下:

$$s = \operatorname{argmax}_{s \in S} P(s|V)$$

$$s \in S$$

其中 S 是词义的集合, s 表示 S 中的每一个可能的义项, V 表示输入上下文中的向量。

直接根据向量的计算公式是:

$$s = \operatorname{argmax}_{s \in S} P(s) \prod_{j=1}^n P(v_j|s)$$

1992 年, Gale 等使用这个方法试验了 6 个英语的多义词 (duty, drug, land, language, position, sentence) 的词义排歧, 正确率达到 90% 左右。

(2) 决策表分类法

决策表分类法 (Decision list classifiers) 根据共现词的等价类的不同制定决策表, 然后利用这个决策表于输入向量, 确定最佳的词义。

例如, 可以制定如下的决策表来确定 bass 的词义:

规则	词义
窗口中出现 fish	→ bass1
striped bass	→ bass1
窗口中出现 guitar	→ bass2
bass player	→ bass2
窗口中出现 piano	→ bass2
窗口中出现 tenor	→ bass2
sea bass	→ bass1
play/V bass	→ bass2
窗口中出现 river	→ bass1
窗口中出现 violin	→ bass2
窗口中出现 salmon	→ bass1
on bass	→ bass2
bass are	→ bass1

其中, bass1 表示 fish 的含义, bass2 表示 music 的含义。如果检测成功, 就选择相应的词义, 如果检测失败, 那就进入下一个检测。这样一直检测到决策表的末尾, 其缺省值就是最大可能的词义。

这个决策表可用于从 bass 的 music 含义中消

除 fish 的含义。第一项检测说明, 如果在输入中出现 fish, 那么, 就选择 bass1 为正确的答案。如果不是这样, 那么, 就检测下一项一直到返回值为 True, 在决策表末尾的缺省值的检测, 其返回值为 True。

决策表中项目的排列根据训练语料的特征来决定。1994 年, Yarowsky 提出一种方法来计算决策表中的每个特征值偶对的对数似然比值 (log-likelihood ratio), 根据计算所得的比值调整 Sense1 和 Sense2 在决策表的顺序, 从而确定整个决策表中特征值的排列顺序。计算公式如下:

$$\operatorname{Abs} \left(\operatorname{Log} \left(- \frac{P(\text{Sense}_1|f_i=v_j)}{P(\text{Sense}_2|f_i=v_j)} \right) \right)$$

其中, v 表示 Sense 的特征向量, f 表示该 Sense 的绝对频度。

根据这个公式来比较个特征值偶对, 便可以获得一个排列最佳的决策表。Yarowsky (1996) 采用这样的方法, 得到了 95% 的正确率。

6. 自力更生的词义排歧方法

有指导的学习方法的问题是需要训练大量的标注语料。Hearst 在 1991 年, Yarowsky 在 1995 年, 都分别提出自力更生方法 (Bootstrapping Approaches)。这种方法不需要训练大量的语料, 而只需要依靠数量相对少的实例, 每一个词目的每一个义项都依靠少量的标记好的实例来判别。以这些实例作为种子 (seed), 采用有指导的学习方法来训练语料从而得到初始的分类。然后, 利用这些初始的分类, 从未训练的语料中抽取大量的训练语料, 反复进行这个过程一直到得到较满意的精确度和覆盖率为止。

这个方法的关键是从较小的种子集合出发, 创造出大量的训练语料。然后在利用这些得出的大量的训练语料来创造出新的、更加精确的分类。每重复一次这样的过程, 所得到的训练语料越来越大。而未标注的语料越来越少。

自力更生法的初始种子可以使用不同的方法来产生。

1991 年, Hearst 用简单的手工标记方法从初始语料中获得一个小的实例集合。他的方法具有如下 3 个优点:

● 种子实例可靠, 保证了机器学习有正确的立足点;

●分析程序选出的实例不仅是正确的,而且可以作为每个义项的意义原型;

●训练简单可行。

1995年,Yarowsky提出“一个搭配一个义项”(One Sense per Collocation)的原则,效果良好。他的方法是为每一个义项选择一个合理的标示词(indicator)作为种子。例如,选择fish作为识别bass1这个义项的种子,选择play作为识别bass2这个义项的种子。

下面是例子:

fish--bass1

The researchers said the worms spend part of their life cycle in such fish as Pacific salmon and striped bass and pacific rockfish or snapper. (研究人员说,蠕虫在它们生命周期的一部分生活在太平洋大马哈鱼和有斑纹的鲈鱼以及太平洋的岩鱼或者甲鱼体内。)

Saturday morning I arise at 8:30 and click on "America's best known fisherman," giving advice on catching bass in cold weather from the seat of a bass boat in Louisiana. (星期六早晨我8:30起床,询问“美国最有名的渔人”,怎样在大冷天从Louisiana的鲈鱼船的座位上捕捉鲈鱼。

play--bass2

We need more good teachers - right now,

there are only a half a dozen who can play the free bass with ease.(我们现在需要很多好老师,只要有五六个能够熟练地演奏低音乐器的就行了。)

An electric guitar and bass player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectation perhaps.(电吉他和低音演奏者站在一旁,他并不是站在舞台的一部分,大概只是为了等待外国佬的到来。)

Yarowsky选择种子的途径有两条:一是机器可读词典;二是利用统计方法根据搭配关系来选择。他对12个多义词的歧义消解正确率为96.5%。

7.无指导的词义排歧方法

无指导的方法(Unsupervised Methods)避免使用通过训练得出义项标注(sense tagging)的语料,只使用无标记的语料作为输入,这些语料根据它们的相似度进行类聚。这样的类聚可以作为成分的特征向量的代表。根据相似度得出的类聚再经过人工的词义标注后,就可以用来给没有特征编码的实例进行分类。

例如,英语多义词bank的义项分别为bank1和bank2,在没有经过训练的语料中,在一个上下文中出现了money,在第二个上下文中出现了loan,在第三个上下文中出现了water,它们在不同上下文中与其他词的共现次数也就是它们的关联向量,如表1所示:

表1

	bank	building	loan	money	mortgage	river	water
loan	150	20	70	100	50	10	40
money	600	500	100	400	50	30	70
water	15	400	40	70	1	400	500

从共现次数的分布(关联向量)可以看出这三个词的相似度的接近程度,water与loan或者money的相似度远远小于money与loan的相似度。也就是说,money和loan的关联向量大于money与water的关联向量,也大于loan与water的关联向量。这样,我们就可以把money与loan类聚在一起,这个类聚是bank1的标示,bank1的含义显然应该是“银行”;把water单独算为一个类聚,这个类聚bank2的标示,bank2的含义显然应该是“岸边”。

经常采用的方法是凝聚法(agglomerative clustering)。N个训练实例中的每一个实例都被指

派给一个类聚,然后用自底向上的方式陆续地把两个最相似的类聚结合成一个新的类聚,直到达到预期的指标为止。

由于无指导的方法不使用人工标注的数据,它存在如下的不足:①在训练语料中,无法知道什么是正确的义项;②所得到的类聚往往与训练实例的义项在性质上差别很大,各不相谋;③类聚的数量几乎总是与需要消解歧义的目标词的义项的数量不一致。

Schuetze在1992年和1998年,先后使用无指导的方法来进行多义词的歧义消解,其结果与有指导的方法和自力更生的方法很接近,达到了

90%的正确率。不过,这种方法所试验的多义词的数量规模都很小。

Schuetze 在 1992 年还使用向量类聚的方法进

行词义排歧,比较了向量类聚的词义排歧与只选择最常见义项的歧义消解结果见表 2。

表 2

词	义项数目	向量类聚词义排歧正确率	选择最常见义项正确率
tank/s	8	95	80
plant/s	13	92	66
interest/s	3	93	68
capital/s	2	95	66
suit/s	2	95	54
motion/s	2	92	54
ruling	2	90	60
vessel/s	7	92	58
space	10	90	59
train/s	10	89	76

由此可见,向量类聚的效果比早期机器翻译系统使用的选择最常见义项的方法好得多。

8. 基于词典的词义排歧方法

上述方法的最大问题是语料的规模问题。许多词义排歧试验的规模只涉及 2 到 12 个词,最大规模的词义排歧试验也只涉及 121 个名词和 70 个动词 (Ng, Lee, 1996)。因此,学者们想到了使用机器可读词典(machine readable dictionary),采用基于词典的词义排歧方法 (Dictionary-Based Approaches)。这时,机器可读词典可以给词义排歧提供义项以及相应义项的定义上下文。

1986 年, M. Lesk 首先使用词典中的定义来进行词义排歧。他把多义词的各个义项的定义进行比较,选择具有最大覆盖上下文的义项为正确的义项。例如,在词组 pine cone(松球)中, cone 是多义词,我们把词典中 pine 的定义与 cone 的定义进行比较如下:

pine 1 kinds of evergreen tree with needle-shaped leaves (一种具有针状树叶的长绿树)

2 waste away through sorrow or illness (因为悲哀或者疾病而憔悴)

cone 1 solid body which narrows to a point(圆锥体)

2 something of this shape whether solid or hollow(硬的东西或者空的东西)

3 fruit of certain evergreen trees (某些长绿树的果实)

我们选择 cone3 作为 pine cone 中多义词

cone 的正确义项,因为在 cone3 的定义中, evergreen 和 tree 两个词与 pine1 定义中的词 evergreen 和 tree 相重合。Lesk 从《傲慢与偏见》(Pride and Prejudice)和 AP newswire 的文章中选取部分语料进行试验,正确率达 50-70%。

K. Jensen 和 J. L. Binot 利用联机词典中的单词的定义来消解英语介词的功能歧义。

例如,英语的 with 这个介词,其功能可以表示 INSTRUMENT(工具),又可以表示 PART-OF(部分-全体)关系,这就出现了功能上的歧义(case ambiguity)。在英语句子“ I ate a fish with a fork ”中, fork(叉子)的定义为“ an instrument for eating food ”,其中的 instrument 与 with 的功能 INSTRUMENT(工具)相同,故可判断 with 在这个句子中的功能应该是 INSTRUMENT(工具),故此句的含义应该为“ 我用叉子吃鱼 ”。

在英语句子“ I ate a fish with bones ”中, bone 在机器可读词典中的定义是“ a part of animal ”,在 fish 的定义中,有“ a kind of animal ”,这与 with 的功能 PART-OF(部分-全体)关系相同,故可判断 with 在这个句子中的功能是 PART-OF 关系,这样,这个句子的含义应该是“ 我吃带骨头的鱼 ”。

这个方法的主要困难是词典中的定义往往太短,不足以为词义排歧提供足够的上下文材料。例如,在 American heritage Dictionary 中, bank(银行)的定义里没有 deposit(存款)这个词,在 deposit(存款)的定义中,没有 bank(银行)这个词,而这两个词有很密切的联系。现在一些词典中有主题分类代码(subject codes),似乎可以弥补这方面的缺

陷,因为 bank 和 deposit 都可以划为 EC(Economics)这个主题。1991年,Guthrie 报告,他使用了 LDOCE(Longman's Dictionary of Contemporary English, 1978)的主题代码来消解歧义,把正确率由 47%提高到 72%。

40 多年来,自然语言处理在词义排歧方面虽然取得了很大的成绩,但是,学者们的各种方法似乎都很难判定 Bar-Hillel 在 1959 年提出的在“the box was in the pen”中 pen 的词义应该是“游戏的围栏”。可见,词义排歧确实是非常困难的问题。要真正解决词义排歧问题,还需要我们做出不懈的努力。过去的成果使我们看到了解决这个问题的一线曙光,尽管这一线曙光还是很微弱的,但它毕竟是黎明前的曙光,还是很鼓舞人心的,因为它预

示了自然语言处理事业光辉的未来。

参 考 文 献

- [1] 冯志伟.自然语言的计算机处理.上海:上海外语教育出版社,1996
- [2] Brown, Peter F., et al, Word-sense disambiguation using statistical methods, In proceedings of ACL 29, 1991
- [3] Gale, W., Church, K., and Yarowsky, D., A Method for Disambiguating Word Senses in a Large Corpus, In Computer and Humanities, 26. 1992
- [4] Lesk, Michael, Automatic sense disambiguation: How to tell a pine cone from an ice cream cone, In Proceedings of the 1986 SIGDOC Conference, New York, Association for Computing Machinery, 1986
- [5] Yarowsky, David, Unsupervised word sense disambiguation rivaling supervised methods, on Proceedings ACL 33, 1995

一部系统化的教材和生动的知识读物 ——《组织机构代码系统工程》

组织机构代码实际上就是单位的身份证号码,它是由国家授权的权威管理机构对我国境内依法注册、依法登记的企业事业单位、机关、社会团体及其他组织颁布的一个在全国范围内统一的、始终不变的代码标识。它同我国公民身份证在性质上是相同的,即都是一种“代码标识”。不过前者对组织机构,后者是对自然人。自然人的注册和标识工作比较简单,新出生的自然人只要在公安部门登记就可以随之获得身份证号,而组织机构则要由机构编制部门、工商部门和民政部门统一赋码。另外,组织机构具有自然人不具备的层次性的特点,即它可以拥有自己的分支机构和内设机构,如总公司和分公司,几个组织机构按照一定的关系可以组成联合体,如法人的联营;很多组织机构还可以跨地域甚至跨国界成立联合机构。所以,组织机构代码的赋码工作及其日常管理比起公民身份证来要复杂得多了。

究竟组织机构代码工作的来龙去脉、国内外情况,它在国民经济建设中的意义和作用,以及此项管理工作的具体内容、运行方式及其推行的法律依据等等是如何,在新近由中国计量出版社出版的《组织机构代码系统工程》一书中作了全面的介绍。该书以通俗而流畅的语言层层递进的清晰条理,通过对国内外产生背景和发展过程的介绍,以及运用系统工作的思路对组织机构代码的重要作用、实践方法和工作特点等的详细分析和说明,给读者提供了多方面的宝贵知识。

全书内容由绪论部分、三大章和若干附件组成。绪论部分介绍了“组织机构代码”的基本概念,国外一些管理水平较高的国家推行对组织机构统一赋码制度的概况(以法国“SIRENE 名录库”和美国“邓氏编码”为典型实例),以及我国组织机构代码工作自 1989 年起步至今的发展历程,并

阐述了系统工程方法对此项工作的指导意义。

第一章是对组织机构代码在国民经济和社会发展中作用的综合介绍。首先阐述了国民经济信息化的内涵,指出电子政务是当前我国信息化工作的重中之重,进而阐明了组织机构代码在国民经济信息化中的基础性地位:它是建立社会监管制度的必要条件,是建设社会保障体系的有效手段,并有助于整顿规范市场经济秩序和建立以实名制为基础的社会信用体系。

第二章全面而系统地讲解了组织机构代码系统工程的实践方法,其内容包括:组织机构的分类、组织机构代码的赋码对象、代码与组织机构代码的基本知识以及作为组织机构代码系统工作依据的相关法律法规、行政规章和标准规范等;组织机构代码的工作制度、工作程序和质量保证工作;组织机构代码的载体(代码证书和 IC 卡)、网络与数据库工作;推广应用、协调与宣传工作;经济运行方式;科研工作和人员系统;电子身份认证(CA)工作;组织机构代码各子系统和要素的相互关系以及对系统的评价方法等。

第三章对组织机构代码系统的特点通过梳理,归纳出系统的整体性、层次性、开放性和稳定性这几大特点,并进行了系统表述。

附录部分收入国务院、原国家技术监督局和其他有关主管部门关于建立代码标识制度的五个文件。

该书由全国组织机构代码管理中心编,主编顾迎建。全书近 30 万字。作者从实践出发,既提供翔实的背景资料,又融入科学和理性的思考。熔过程、方法与理论于一炉,既是一本组织机构代码工作人员的系统化的教材,也是一部适合大众需要的知识读物。

(海萌)