

Chinese text Compression using Chinese Language Information Processing

Jun Gao

Xixian Chen

Beijing University of Posts and Telecommunications 103 Box
No. 10, Xi Tu Cheng street, Hai Dian district, Beijing, 100088, China
e-mail: b9507311@bupt.edu.cn

Abstract

To transport the large scale authentic Chinese text, especially in the personal telecommunication system, it is necessary to establish a highly efficient coding method. A good coding system can reduce transportation time. There are two basic ways to compress information. One way is to get rid of the perplexity in the relativity of information source. The other is to remove the perplexity of not equal probabilistic distribution. In this paper, a novel information compressing method is presented. It utilizes the relativity of the information source and catches the information of different probabilistic distribution according to the definition of Chinese words. The relativity and not equal probabilistic distribution are connected by a optimum searching method. The aim of high compressing ratio is reached. And. Some experimental results are also covered.

Keywords: Chinese corpus, optimum searching, Huffman coding, Discrete Stable Information Source, Maximum likelihood, relativity, perplexity, entropy, average code length.

中文信息处理中的信息压缩

高军 陈锡先

北京邮电大学信息技术实验室
北京市海淀区西土城路10号北京邮电大学103信箱
北京, 100088, 中国
e-mail: b9507311@bupt.edu.cn

摘要

在传送大规模真实中文信息中, 特别是在个人通信系统中, 准确高效的信息压缩方法是极为必要的。一个好的信息压缩方法, 可以节约大量的传输时间和存储空间。信息压缩有两种基本途径, 第一种是, 去处寓于信源的相关性之中的冗余度, 另一种是, 去处寓于概率的非等分布之中的冗余度, 改变信源的概率分布, 以期尽可能达到等概率分布的目的。本文提出了一种全新的信息压缩方法, 这种方法在理论上充分利用了信源的相关性, 同时又根据中文信息处理中对词的界定, 把握了其不等概率分布特性, 运用最优搜索方法, 把信源相关性与信源的不等概率分布有机地结合起来, 从而达到了高效压缩信息的目的。在文中, 列举了实例结果。

关键词 汉语语料 最优搜索 Huffman 编码 最大似然度 相关性 冗余度 熵 平均码长

1. 引言

在通信系统中, 采取准确高效的信息压缩技术, 对于解决存储容量的限制、信道带宽的不足, 具有重大意义。尤其是在传输大量中文字符信息的过程中, 信息在压缩后应能够完全恢复, 这对于压缩技术提出了更高的要求。

数据的信息压缩有两种基本途径, 第一种是, 对于联合信源, 其冗余度寓于信源的相关性之中, 去处他们的相关性, 使之成为或差不多成为不相关的信源。另一种是, 对于离散无记忆信源, 其冗余度则寓于概率的非等分布之中, 改变信源的概率分布, 以其尽可能达到等概率分布的目的。

具体到信息压缩编码技术上则表现为两种不同的处理方式: 一种是信息被压缩, 在一定程度上消除了部分冗余信息, 但恢复后得不到信息的完全复原, 即部分有用信息也被压缩掉了, 是有损压缩; 如声码器(Vocoder)对语音信息的编码压缩, 还有差分脉冲编码调制(DPCM)对图像信息的编码等。它们都是从模拟信号转变成数字信号的过程中对信息进行压缩的。这两种信息压缩编码得到的恢复语音和图像信号, 虽然主要信息成分被保留了, 但一些有用的信息也失去了, 从而导致信息的失真。另一种是

缩后的还原信息应能够百分之百地得到恢复。即信息压缩和还原不应带来任何的损失。中文信息处理中的信息压缩就属于这种字符信息的无损压缩。

为了节省信息存储空间和提高通信效率，应尽可能把这些信息的冗余成分去掉。在文件和通信信息中，冗余信息主要有：字符分布冗余，即，少数字符的使用频度明显高于其他多数字符；高频组合冗余，既有些组合字符比其它的字符更常见，这些冗余信息的存在，为字符信息的压缩提供了可能。

本文以信息理论为基础，充分利用以上提到的各种可能的冗余性，对中文信息进行高效压缩。

2. 原理及算法

2.1 理论基础

我们把中文信息作为信源来研究。根据由信息理论关于信源的论述，有以下的结论：

对于离散信源 X ，若其输出的随机序列 $x_1, x_2, x_3, \dots, x_i, \dots$ 各维联合概率分布均与时间起点无关，即，当 $t = i, t = j, (i, j \text{ 为任意的整数, 且 } i \neq j)$ 时，有：

$$P(x_i) = P(x_j)$$

$$P(x_i, x_{i+1}) = P(x_j, x_{j+1})$$

⋮

$$P(x_i, x_{i+1}, \dots, x_{i+N}) = P(x_j, x_{j+1}, \dots, x_{j+N})$$

那么，具有这样性质的离散信源是完全平稳的离散信源，简称为平稳信源。

而长为 N 的信源符号序列中平均每个信源符号所携带的信息量，即平均符号熵为

$$H_N(X) = \frac{1}{N} H(X_1 X_2 \dots X_N) \quad \dots \dots (1)$$

平稳信源 X 的信息熵 $H(X)$ 与平均符号熵 $H_N(X)$ 的关系为

$$H(X) = H_x = \lim_{N \rightarrow \infty} H_N(X) \quad \dots \dots (2)$$

H_x 为平稳信源的极限熵。

$H(X)$ 可由 $H_N(X)$ 来近似。 N 的值越大， $H_N(X)$ 越能更好的描述 $H(X)$ 。

所以，我们可以用 M 组最大长度为 N 的信源符号序列来描述 $H(X)$ 。即，

$$H(X) \approx \frac{1}{\sum_{i=1}^M N_i} H\{(X_{1,1}, X_{1,2}, \dots, X_{1,N_1}), \\ (X_{2,1}, X_{2,2}, \dots, X_{2,N_2}), \dots, \\ (X_{M,1}, X_{M,2}, \dots, X_{M,N_M})\} \dots\dots (3)$$

(3)式中，在M个组之间，信源符号彼此是相互独立的；而在M个组的内部， N_i 个信源符号之间是相互依赖的，其中 $1 \leq N_i \leq N$ ， $i = 1, \dots, M$ 。

我们根据以上分析，把中文信息模型化为特定汉字字符串内部是彼此相关联的，而字符串之间则是相互独立的。

这样模型化的目的是既能够利用汉字字符之间的相关信息，又能够利用汉字字符串的频率信息。如何把这些相关信息及频率信息有机的结合于信息压缩中，是提高压缩效率的关键。在通常的对中文进行编码过程中，都采用已知码本对中文信息进行穷搜索、穷匹配，但这样所得的结果不是全局最优的，甚至不是局部最优的。

2.2 算法

针对上一节中的分析，本文提出了一种最优搜索算法，它利用汉字之间的相关性和字符串的频率特性，使得在给定的范围内，匹配的结果是最佳的。

为了方便描述，我们预先作如下约定：

W：表示一个字符串。W的长度为N。

S：表示对W的一个划分。即：

$$S = s_1 s_2, \dots, s_k, \dots, s_q \quad \dots\dots (4).$$

共q个段。其中 s_k 为W中的一个子列。

$$s_k = [\omega_{i_k}, \omega_{i_k+1}, \dots, \omega_{i_{(k+1)}-1}] \quad \dots\dots (5)$$

I_s ：表示对W中的q个段的每段段首的索引。即：

$$I_s = \{i_1, i_2, \dots, i_q, i_{q+1}\} \quad \dots\dots (6)$$

其中， $i_1 = 1$ ， $i_{q+1} = N + 1$ ，

且 $i_1 < i_2 < \dots < i_{q+1}$ 。

Len：表示每一子列的最大长度。

则由(4),(5),(6)式及其限制条件可推出：

$$1 \leq k \leq q, \text{ 且 } 1 \leq i_{k+1} - i_k \leq \text{Len}.$$

$\{I_s\}$ ：表示满足条件的所有 I_s 的集合。

如前所述, 我们假设各个子列之间是相互统计独立的。而子列内部是相关的, 则最优搜索模型可定义为:

在一个句子中的所有可能的划分中, 子列概率的乘积最大的那个划分——具有最大似然度的那个划分为: $I_{s_{\max}} = \arg \Gamma(W)$, 其中,

$$\begin{aligned} \Gamma(W) &= \max_{\{s\}} \prod_{k=1}^q P(s_k) \\ &= \max_{\{s\}} \prod_{k=1}^q P(\omega_{i_k}, \omega_{i_k+1}, \dots, \omega_{i_{(k+1)}-1}) \quad \dots\dots (7) \end{aligned}$$

在具体实现中, 最优搜索算法由以下步骤实现:

第一步: 计算所有汉字串的出现概率, 并从中挑选出其概率与其长度的乘积最大的前 $NumTop$ 个字符串, 我们称这个乘积为其相应字符串的权值。

设 $N_{init}(s)$ 代表某长度为 $l (\leq Len)$ 的汉字串 s 在总长为 N 的文本 W 中出现的次数。并设 N_{total} 代表所有在长度上小于最大长度 Len 的汉字串出现次数的总和。

$$\text{则: } P_{init}(s) = \frac{N_{init}(s)}{N_{total}} \quad \dots\dots (9).$$

其中 $N_{total} = \sum_{i=1}^{Len} N_i$, N_i 代表所有长度为 i 的汉字串的个数。经过以上的统计, 可以找出所有在长度上小于 Len 的字符串的概率。

$$SetTop = \{s | s \in Top_{NumTop}(N_{init}(s) \times L(s))\} \quad \dots\dots(10)$$

其中, $SetTop$ 表示满足式(10)的 $NumTop$ 个汉字字符串的集合。

$L(s)$ 则表示汉字字符串 s 的长度。

$Top()$ 为求前 $NumTop$ 个具有较大权值 $N_{init}(s) \times L(s)$ 的汉字字符串。

第二步: 把汉字编码中所有一级和二级汉字(国家标准GB2312-80中, 6763个汉字)以其频率得有先顺序加入到 $SetTop$ 的汉字字符串集合, 使之成为有权值优先次序的码本

第三步: 基于第二步生成的码本, 在长度为 Len 的范围内对中文文本进行最优匹配, 得到分割结果, 以便进行编码。

即: 设对长度为 L 的 W 的子串, 则

$$\begin{aligned} &\Gamma(\omega_1 \omega_2 \dots \omega_L) \\ &= \max_{1 \leq i \leq Len} \{P(\omega_{L-i+1} \omega_{L-i+2} \dots \omega_L) \Gamma(\omega_1 \omega_2 \dots \omega_{L-i})\} \quad \dots\dots (11) \end{aligned}$$

重复第二步的迭代计算，最终可求出一个分段序列 $I_{s_{\max}}$ ，它能为长度是N的中文文本W提供最大的似然度值。

第四步 对中文文本的分割结果依其在码本中的权值 按一定的编码规则进行编码。

3. 试验及结果评价

3.1 中文文本的选择

我们采用名为《中国百家新闻报刊……1994》的电子出版物中的中文文本作为试验用语料来源。其总量为789兆字节。我们节选其中四月份的语料，共6.9兆字节，又从中随机节取若干段作为真实中文文本。本文中所引用的语料有23131个汉字。

3.2 中文文本的预处理

由于在电子出版物中存在大量的乱字符(ASCII码值160的半角字符)、全角的汉字与半角的各类字符混排以及统计句长的需要，我们对节取的文本进行了预处理。在汉语中存在者许多自然的切分标志，如标点符号等，汉字字串不能跨越这些标点而存在。经过预处理，我们把整个语料库切分为各种短句，其平均长度为12。标点符号集为{，。！；、：？}以及长度超过两个以上的空格。这样做不仅提高了分词精度而且减少了计算量。句间的分割符均以在其后的句子的长度所代替。平均句长为12。经统计，未曾发现语料中存在超出我国国家标准GB2312-80规定的一、二级汉字的范围。因此，本方法建立的码本可以适应实际要求。

3.3 字串最大长度的设定

对汉语中词的统计和研究(刘源,1994)表明汉语中二字词出现的数量和频度最高。随着词长的增加，词的频度则越来越小。而词长超出4的词个数微乎其微。根据以上语言学的结论，我们在分割汉字字串的过程中，设定字串的最大长度为4。

3.4 编码方式

以统计为基础的编码大致的分为两大类，即等长编码和变长编码。它符合等概率最大熵条件其编码效率最低。而变长编码则与信源中元素出现概率的不均匀性紧密联系在一起。为此，我们采用最优变长编码——Huffman编码。采用Huffman编码的优点

存 它适应环境的能力较强 只要设计合理 概率分布上的细微变化 不至于严重降低编码效率。在本实验中, 我们采用二元Huffman 编码。

3.5 试验条件及运行速度

用Pentium 586/133 MHz、32M 内存 计算机计算, 操作系统为Windows NT 4.0。Visual C++ 4.0 为编程语言及调试工具。在汉字字串的最大长度及真实文本的大小确定后, 本算法的运行速度主要受其中 *NumTop* 值的变化的影响。程序运行时间随 *NumTop* 的变化如表一所示。

<i>NumTop</i>	200	500	1000	1500	2000	3000	5000	7000
时间	0.6 小时	1.8 小时	3.9 小时	7.7 小时	10.5 小时	14.9 小时	25.4 小时	34.9 小时

表 一

3.6 结果及评价

在前 *NumTop* 个具有较大权值的汉字字串集合中, 不同长度的字串在其中出现的比例是不同的。详见表二。

<i>NumTop</i>	200	500	1000	1500	2000	3000	5000	7000
一字串	57%	47.8%	37.9%	31.2667%	26.25%	19.8667%	13.88%	11.1285%
二字串	20.5%	27.6%	33.5%	32.9333%	34.85%	30.8%	28.8%	24.6429%
三字串	12%	13.6%	14.9%	19.4%	20.6%	24.6%	31.82%	29.3%
四字串	10.5%	11%	13.7%	16.4%	18.3%	24.7333%	25.5%	34.9285%

表 二

由信息论中信息熵的概念可知, 对于具有 *n* 个信息元素的信息集合, 其熵为

$$H = - \sum_{i=1}^n p_i \log p_i \quad \dots\dots(12)$$

式中, p_i 表示 *n* 个状态中第 *i* 个状态的发生概率, H 代表消除系统不确定性所需的信息量。Shannon 已证明, 在没有任何干扰的条件下, 一个熵为 H 的信源总可以找到一种编码方法, 使其编码的平均长度 L 任意接近熵 H 。但实际上, 信息编码的平均码长 L 总是大于或等于其熵值。平均码长 L 为

$$L = \sum_{i=1}^n P(s_i)l_i \quad \text{码符号/信源符号} \quad \dots\dots(13)$$

式中 $P(s_i)$ 为第 i 个汉字字符串 s 出现的概率, l_i 为字符串长度。

而衡量编码是否达到最佳, 则用编码效率, 即 $\eta = \frac{H}{L}$, η 的值越接近于 1, 编码效率就越高。

对于不同的 $NumTop$, 所形成的不同的信源, 其熵值见表三。

<i>NumTop</i>	200	500	1000	1500	2000	3000	5000	7000
熵	9.7087	10.1408	10.6405	10.9833	11.2468	11.2468	12.1705	12.5238

表三

对于不同的信源, 用 Huffman 编码后所得的平均码长见表四。

<i>NumTop</i>	200	500	1000	1500	2000	3000	5000	7000
平均码长	9.7489	10.1631	10.6766	11.0121	11.2733	11.6568	12.1981	12.5644

表四

显然, 对于不同信源, 其各自的编码效率均在 99.6% 以上。

对中文文本——23131 个汉字, 经过分割后, 根据 $NumTop$ 的不同, 得到的字符串个数见表五。

<i>NumTop</i>	200	500	1000	1500	2000	3000	5000	7000
字符串数	20868	20773	20633	16501	15654	14464	12960	12189

表五

对中文文本——23131 个汉字, 经过分割后, 根据 $NumTop$ 的不同, 得到的总体编码长度见表六。

<i>NumTop</i>	200	500	1000	1500	2000	3000	5000	7000
总体码长	213165	201765	197321	187080	179515	169522	157398	150719

表六

从以上所得的结果分析, 对于同样的中文文本--23131个汉字, $NumTop$ 不同, 所得的总体编码长度也随之迥然不同。随着 $NumTop$ 的不断增大, Huffman 编码的平均码长也逐渐增大。按直观的想法, 由此编码所得的最后总编码长度也应随 $NumTop$ 的增大而增大。但事实上, 经过本文提出的最优匹配算法对中文文本的分割, 却使总的编码长度在不断地大幅度下降。如 $NumTop$ 从200增加到5000, 总的编码长度减小了35.43%。从另一个角度比较, 如果用等长编码方法对国家标准GB2312-80规定的6763个一、二级汉字以汉字为单位编码, 则所得的二元码最短码长为 $\log_2 6763 \approx 12.72$ 。用它对23131个汉字进行编码, 用本文的算法与之相比, 如对于 $NumTop$ 为5000, 相当于本算法把23131个汉字压缩为14464个(表五), 同时用平均码长为11.6568(表四)进行编码。最后等长编码方法所得总编码长度是本算法编码总长度的1.87倍。随着不断扩大训练语料的规模, 并相应地增大码本容量, 压缩比就能够进一步提高。

以上的实验都是对于训练文本--23131个汉字进行的。我们也把由训练文本所得的码本用于测试文本中。测试文本仍然从《中国百家新闻报刊……1994》的四月份的69兆字节语料中节选, 字数分别为5213、8225、14097、22322。用 $NumTop$ 为5000码本编码, 其最后总的编码长度见表七。

文本规模	5213 字	8225 字	14097 字	22322 字
总体码长	35331	55859	95947	152231

表七

所得的分割字符串个数见表八

文本规模	5213 字	8225 字	14097 字	22322 字
字符串数	2909	4599	7901	12533

表八

应用方法所得出的 $NumTop$ 个权值较大的汉字字符串有偶尔相互交叠缺点。这在串长度大于四时出现较多, 如“会治安综”、“治安综合”、“安综合治”的相互交叠, “燃油附加”、“油附加费”的相互交叠等。这种情况使得码本中各数据之间的冗余度增加了。如果能够消除这种冗余性, 编码效率还能够进一步提高, 总的编码长度还会更小。

本算法除了对在训练文本中出现的一部分高频汉字, 根据其权值在码本中进行了编码以外(见表二中的“一字串”), 对其他汉字, 则均以1为权值对其编码。但汉字的出现频率也是有很大差别的。如果把汉字之间的频率差别引入编码, 尤其对于测试文本, 总的编码长度还会更小。

$NumTop = 1000$ 时的权值较大汉字字串部分内容见附录一。 $NumTop = 5000$ 对真实文本进行分割的部分结果见附录二。

4. 结论与展望

本文提出的用于中文信息处理中的信息压缩方法是建立在信息理论基础上的。该方法通过对中文文本的模型化, 把它作为离散平稳的信源。根据统计方法所得权值, 形成码本。再利用离散平稳信源的组间独立、组内相关的特性, 从求取汉语语料中各个词联合的最大似然度的角度, 在码本的范围内对汉语语料进行分割。从实验结果可以看出, 由于本方法是基于大量文本的统计基础上的, 它能够细致的刻划文本中信息的分布状况。同时, 又充分考虑汉字之间的相关性, 因此, 基于此方法所得码本建立的 Huffman 编码的编码效率极高, 只要不断扩大训练语料的规模, 并相应地增大码本容量, 就能够进一步提高压缩比。

另外, 由于本方法便于随被处理语料的增加越来越集中体现信息的分布及汉字的关联性, 因此, 它特别适合应用于特定领域中的高效信息传输和压缩。

在编码方面, 以 Huffman 码为代表的某些最优变长编码, 由于在数学上缺乏构造性, 难以运用现代数学理论, 来解决它们编码时的工程实际问题, 严重影响了这些编码方法的广泛应用。本方法为此提供了有效途径。

参考文献

- [1] 刘源、谭强、沈旭昆. 信息处理用现代汉语分词规范及自动分词方法, 清华大学出版社、广西科学技术出版社, 1994, pp. 1-56
- [2] 梁南元. 在论汉语自动分词和切分知识, 1987, ICCIP Beijing.
- [3] 黄新亚, 米央. 信息编码技术及其应用大全, 电子工业出版社, 1994, pp. 8-43.
- [4] 许织新. 数据压缩, 国防工业出版社, 1990, pp. 81-85.

附录一

的企业业一有大要工作发展在中是经济工会发企和作人社会
乡镇企业经了国市省市场为不年中型企业产地建国有大中 有大中型
大中型企展 乡镇企 镇企业 政府这 实力建设全政公司改革个电开行
型企业 济动各 宣传思想 大中型 中型企方上机 社会治安 国有大有大中
工作会议 传思想工 思想工作重 技术加定公们以出同社 部委成乡镇
会议改进我好市场经济问题镇企场多来社会主义合生领导家面对
工业群众思想搞好国有好国有大业的企业的图书馆工作会宣传思传思想体
时前他们现党国有宣传大中综合治理中介机构化到书思想工想工作
生产于有大理乡关资部门我省中型搞好型企把一个我们府主设
市场经场经济车量县起国家重要安建材治高全国加强者自调强报
制区导燃油附加油附加费新工作的搞好国好国有利技日思党的图书
机构司规条革种从而他市政府一项强调大局要进一党政稳定的各级党
依靠群众坚人民群众面的指导党委和政平方公里 条龙亿元 的成战略
一批企业 产品质量产的 家企轴集团工 集团工业 时期 形式

附录二

四川富益奇迹的奥秘 本报记者 吴中福 王实 编者按 年是 产权改革 年
如今的改革已经深化到了这样一个层次 一 产权的改革 如今的也已扩展
到了这样一种境界 一把 产权关系 放开在中 把 产权资源 进行重新配置和
优化组合 这是 改革开放 深化 个 年头 之后的 必由之路 它表明 只有
产权关系 按 市场经济 配置 生产要素 企业 才能 优化 才能 四川富益 电力股份
有限公司 仅仅 两年 的实践 就 给我们 提供了 有益的 启示 它 揭示 出
一个 企业 迈向 新生的 进程中 突破 单位 割据 放弃 自己的 部份 产权和 原
有 利益 固然 痛苦 但 只要 从 大处 着眼 按 市 经济 规律 产权关系 便能 换
回 大 效益 真正 解放 生产力 四川 电力 眼下 最紧 缺素 有 千水 之省 美
誉 的四川 水资源 总量 为亿 立方 米 然而 尽管 经过 多年 开发 却 仍有
水资源 在 白白 地 流淌 沱江 穿流 的 富顺 县 也不 例外 据 勘查 滔滔 沱江
流经 的富顺 段 可建 个中 小型 水电站 但 水电 投资 大 地方 投资 力不 从
心 在 富顺 地方 发电 的记录 上 两年 前 还 挂着 零 改革开放 风起 云涌
富顺 县委 县府 不甘 心 守着 流银 的 江河 再 受 穷 几年 前 他们 好
不容易 靠 借贷 来 办电 随着 一声 声 轰轰 隆隆 的 巨响 一座 投资 万元
装机 容量 为万 千瓦 的 水电站 动工 开始 了 富顺 电力 史上 零 的突破 谁
拥有 产权 和 代销 权 这是 电力 行业 的核 心 问题 以往 计划 经济 体制 一切
权 务 归 各级 政府 部门 由此 便 有 几十年 一 贯 制 的 大小 电网 割据 局 面
两年 前 富顺 人 却是 地方 政府 把 配置 权 交给 市场 由 企业 在 市场 上 自我
配置 生产 要素 于是 川 电 便 有了 大小 电网 协调 统一 的 新格局 希望 产生 了
忧虑 也 伴着 希望 产生 早在 川 富电 改组 初 县委书记 王濯 根 就曾 带着 一
班 骨干 人 马 考察 了 乐山 等 地 多 座 水电站 学习 兄弟 电站 的 先进 经验
并 总结 其 面临 的 矛盾 和 困境 他们 发现 在 电站 的 管理 中 大小 电网 的 矛盾
将是 今后 要 面对 的 主要 矛盾 所谓 大小 电网 的 矛盾 据 记者 了解 就是 以

各地电力系统为代表的国家大电网与各地水电系统为代表的地方电网的矛盾矛盾的焦点是争夺供区电是瞬间产品卖不出去就白白浪费于是就有了难以计数的电被低价收购又有多到更难以计数的电白白流失这在电能本身就十分紧张的中国真正是触目惊心的浪费省水电厅分管地方电力的部门向记者介绍他们所辖的地方小水电共万千瓦地方年发电量为亿千瓦时占省总发电量的而地方电站供电地盘则占全省这些地方电站都不同程度地与国家大电网产生着矛盾和磨擦这是不可回避的事实邹永福富顺县政协副主席大小电网的矛盾就是我们常说的条条块块的矛盾国家电力部门与地方水电部门企业以及地方集资兴办的小水电各拥有各的产权各拥有各人的利益这些产权和利益历史遗留下来的使生产力不能解放罗旭良富顺县县长怎样解决大小电网矛盾我们认为要从根子上从人手通过搞股份制企业把产权部分放开和转移使大小电网间形成产权共同体和利益共同体王耀根富顺县委书记搞股份制的深层次问题是解放生产权力和生产资源的配置问题谁来配置以往计划经济靠政府各级政府分管部门不同利益不同矛盾由此而生如今我们用的眼光看产权从自身做起打破地方割据把政府的配置功能放给市场按市场重新组合产权关系矛盾就迎刃而解了年月注册资本万元的四川富益电力股份有限公司正是在这种形势下诞生了公司由代表地方电网的国营富顺电力总公司和两家代表国家电网的自贡电业局电力实业总公司川南电力调度电力技术服务部共同发起空前的产权实体后形成的股份制企业其实质是一种混合经济在这种新的共生体中国家法人股东的产权达到了前所未有的明晰三方的资产前所未有的人格化未有地巨大总结川富电改组一年产生惊人效益奥秘时公司总经理罗崇远如是说的混合经济确实有着强大的生命力它不仅仅是解决了大小电网的矛盾问题更重要的是充分发挥了企业发展的内在动力据记者了解全省约个小电网中也曾有采取电力部门供电部门和小水电实行联营方式来解决大小电网矛盾可是这种联营方式只能缓解而由的利益矛盾对电力长远发展没有保证没有后劲而由发电供电输电三位一体组成的川富电却较好地解决了上述问题这在全省是独一无二的创新模式请看作为法人股两家单位——自贡电力局电力实业总公司及川南电力调度局技术服务部他们除投资万元股本外因为利益一致风险共担积极协同做好电力的发输配工作适时增加销量由此而刺激川富电所属水电站努力多发电再看电站多职工人人均认购了数量可观的股份每个人都成了企业的股东企业的利益再也不是他们的身外之物而与他们息息相关了于是一系列根本性的变化也就由此产生——以往吃惯大锅饭的企业职工名义上是企业的主人而实质上除按月领取相对固定的工资和奖金外很少关心企业的经营改组后不同了企业每个员工都拥有企业的股份相互命运与利益紧紧地捆在一起职工主人翁精神真正