第八届“认证杯”数学中国

数学建模网络挑战赛
承　诺　书

我们仔细阅读了第八届“认证杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们接受相应处理结果。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：

参赛队员 (签名) ：

队员 1：余万玉

队员 2：王　震

队员 3：朱翠艳

参赛队教练员 (签名)： 无

参赛队伍组别（例如本科组）： 本科组

第八届“认证杯”数学中国

数学建模网络挑战赛
编 号 专 用 页

参赛队伍的参赛队号:（请各个参赛队提前填写好）:#4824

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

2015 年第八届“认证杯”数学中国数学建模网络挑战赛第一阶段论文

题　　目　带有噪声干扰信道传输替换式密码破译

关 键 词　多目标优化模型 频率分析算法 模拟仿真 检验函数

摘　　要：

本文研究的是带有噪声干扰信道传输的替换式密码破译问题，主要根据字符频率建立起多个目标函数，运用频率分析算法求解出密文与明文字符之间对应的关系。

针对带有噪声干扰信道传输替换式密码破译问题，首先对噪声干扰概率进行处理，字符在传输过程中，某个字符的重复频率越高，受到的干扰概率越大，则用字符在密文中出现的频率代替字符传输干扰的概率，并把概率平分给三个干扰项，通过 MATLAB 编程仿真得出密文。

其次建立目标函数，通过密文与正常英文字符之间频率差最小值，明文单词与词料库单词相似比最大值，以及 1 减去明文和正常英文的 26 个字符频率之差的和最小值三者作为目标函数；以字符频率，密文传输概率等因素建立 4 个约束条件，最终建立多目标优化模型。

利用字符在正常英文的频率，根据单个字符，双字符对，单个字符作为首字母的高频统计表，设计一种频率分析算法。再按照英文构词与语法的统计规则入手，人工干预 37 次，得到密文与明文的字符对应法则如下。并且把结果从定性和定量两个方面进行了分析。得出模型和算法的正确性。

明文	a	b	c	d	e	f	g	h	i	j	k	l	m
暗文	t	l	o	u	d	e	b	w	v	y	k	n	h
明文	n	o	p	q	r	s	t	u	v	w	x	y	z
暗文	x	s	p	v	j	z	g	c	l	m	q	f	r

将文本进行 10 次加密得到 10 种不同密文，运用所建模型和设计的算法，将密文进行转化成明文，运用检验函数，求出不同密文相对应的明文的最佳检验值。

次 数	1	2	3	4	5	6	7	8	9	10
检验值	0.855	0.849	0.855	0.857	0.861	0.845	0.851	0.857	0.849	0.853

最后用检验函数，得到检验函数值均值为 $Fitness = 0.8532$，检验最后得到明文的正确性。

参赛队号： #4824

所选题目： B　　题

参赛密码 _____

（由组委会填写）

Abstract

This study is substitution cipher deciphering problems with noise channel transmission, the main character to establish multiple objective functions according to frequency, using frequency analysis algorithm for solving the correspondence between cipher text and plaintext character relationships.

For channel transmission substitution cipher deciphering problems with noise, the first of the noise probability processing, character during transmission, disruption frequency increases, the probability that the frequency of characters appear in place of the cipher text character transmission interference, and the probability of interference equally to the three items, the cipher text obtained by MATLAB simulation program.

Second objective function, through between the cipher text and the difference between the minimum frequency of normal English characters plaintext word and the word library material than the difference between the maximum word similarity, and one minus the plaintext and the normal English 26 characters and the minimum frequency three as the objective function; the character frequency, the cipher text transmission probability and other factors to establish four constraints, the eventual establishment of multi-objective optimization model.

English characters in normal use frequency, based on a single character, two characters right, a single character as a high frequency statistics initials design a frequency analysis algorithm. Then in accordance with the rules of the English word formation and grammar statistics start, manual intervention 37 times to obtain Character corresponding cipher text and plaintext rule follows

Plaintext	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**
Cipher text	**t**	**l**	**o**	**u**	**d**	**e**	**b**	**w**	**v**	**y**	**k**	**n**	**h**
Plaintext	**n**	**o**	**p**	**q**	**r**	**s**	**t**	**u**	**v**	**w**	**x**	**y**	**z**
Cipher text	**x**	**s**	**p**	**v**	**j**	**z**	**g**	**c**	**l**	**m**	**q**	**f**	**r**

The encrypted text 10 times to get 10 different cipher text, using the model and design of algorithms, will be converted into plaintext cipher text, using test function, obtaining different cipher text corresponding plaintext best test values.

次 数	1	2	3	4	5	6	7	8	9	10
检验值	0.855	0.849	0.855	0.857	0.861	0.845	0.851	0.857	0.849	0.853

Finally, test function, Be tested function value is $Fitness = 0.865312$, testing the validity of the final plaintext。

Key words: Multi-objective optimization model Best test function Simulation Frequency Analysis- Algorithm

1. 问题重述

1.1 问题背景

历史上有许多密码的编制方法。较为简单的是替换式密码，也就是将文中出现的字符一对一地替换成其它的符号。对拼音文字而言，最简单的形式是单字母替换加密，也就是以每个字母为一个单位，将每个字母替换成另外的字母或者另外的符号。较为复杂的形式是以多个字母为一个单元，整体替换成其它的字符。这个映射方法被称为密码表，拿到密码表的人就能够将密文破译成明文。单字母替换加密是在古代就使用过的一种加密方法，但由于其容易被破解，所以在现代需要加密的场合已经很少使用。

单字母替换加密的破译方法有频率分析等。这种密码和破译方法在小说中也经常提到，例如爱伦·坡的《金甲虫》和柯南·道尔在福尔摩斯系列故事《归来记》中的“跳舞小人”。但当获取的密文篇幅不是很大时，光凭借频率分析是不足以破译全部密文的。往往还要熟知该种语言的人，经过对可能出现的词汇及字母组合进行分析，才能完整地破译密码。

1.2 问题相关信息

在密码的实际使用中，我们获取的密文往往是经过各种干扰的。例如密文不完整，或者在通信传输的过程中，密文存在被丢失、添加和篡改字符的可能。为简单起见，我们假设密文是通过一个带有噪声干扰的信道传输的。在通信过程中，每个字符经过信道传输的结果都属于如下四种情形之一：

1. 该字符在传输过程中被丢失，其概率为 p1；
2. 该字符本身正常传输，在其之后添加了一个随机字符，其概率为 p2；
3. 该字符在传输过程中被篡改为一个随机字符，其概率为 p3；
4. 该字符正常传输，其概率为 1 - p1 - p2 - p3。

1.3 需要解决的问题

在加密后，单词之间的间隔和标点符号全部被删去。请你改进之前的破译算法，使其能够在这种带有干扰的条件下完成破译工作，并推广对破译能力的评价指标，使其能应用于这种带有干扰的条件。

2. 问题假设与符号说明

2.1 问题假设

假设一：文章字母之间不存在连接符“一”；

假设二：题中所给出的字母丢失、添加、篡改三种出错概率大小相等；

假设三：在简单的替换密码中，明文中的每一个字母都被另一个字母替换，而且明文中相同的字母在转换为密文时总是被同一个字母所替换。比如，所有的 E 都会被替换成 X. 一个含有大量 X 的密文消息会向密码破译者暗示 X 替换 E。

2.2 符号说明

符号	含义
P_1	表示字符在传输过程中被丢失的概率
P_2	表示字符本身正常传输，在其之后添加了一个随机字符的概率
P_3	表示字符在传输过程中被篡改为一个随机字符的概率
P_4	表示字符正常传输的概率
$SF(\mathrm{i})$	表示第 i 个字符在密文中出现的频率
$DF(\mathrm{i})$	表示第 i 个字符在正常英文中出现的频率

$CF(i)$	表示第 i 个字符在明文中出现的频率
ω_1	表示字符丢失概率权重
ω_2	表示字符本身正常传输，在其之后添加了一个随机字符的概率权重
ω_3	表示字符在传输过程中被篡改为一个随机字符的概率权重
S_n	表示两个单词对比有 n 个字符相同
M	表示明文集合
S	表示暗文集合
k	表示 26 个字母所有排列
$M=(M_1,M_2\cdots M_i)$ 且$(M_i\in Z_{26},1\le j\le i)$ $S=(S_1,S_2\cdots S_i)$ 且$(C_j\in Z_{26},1\le j\le i)$ $k=(k_1,k_2\cdots k_{26})$ 且$(k_i=Z_{26})$	

3. 问题分析

针对第二阶段问题，首先是替换式密码在传输过程中受到传输信道的干扰，即字符丢失，字符正常传输的情况下，其后又增加一个随机字符以，及字符传输过程中被篡改成另外的字符，并且传输信道对字符干扰的概率未知，所以对大量的英文统计分析得到，每一个字母英文中出现的频率都有相对应的一个稳定的频率，由于每个字母在密文中出现的频率不一样，所以对于每一个字符在通信传输过程受到的干扰也是不一样的，可以很容易想到每个字母在密文中出现的频率越高，则该字符受到的干扰的频率也会相应的增大，又因为密文在传输过程中受到的三种干扰是由于通信信道影响，所以把每一种干扰传输的概率取为该种字符在密文中出现频率的均值，根据概率约束，通过计算机模拟将明文转成密文。

然后由于在加密后，单词之间的间隔和标点符号全部被删去，就不知道了密文转成明文时的单词的长度，明文所转成的密文就是一段由 26 个字母 A□ Z 组成的字符串，通过 *wiki* 找到单个字母和双字符组在英文中出现的频率， e 是最常见的单字母， *th* 是最为常见的双字母组，而 the 则为最为常见的三字母组，先将密文中出现最高频率字符做一个统计，然后将密文中的最高频率字符频率减去单个字符在英文出现的频率，并取绝对值，如果绝对值最小的，就认为得到两个字符是对应的。然后找到该字符，根据双字母组在英文出现的频率，运用与单字母相似的方法，就可以找到双字母对应的密钥，根据单字符作为首字母的频率，就可以来确定单词的首字母。由于字符传输过程中，受到干扰，所以有的单词有可能找不到与词料库相对应的单词来匹配，所以定义一个相似度函数，如果两个单词相似度较大，就认为这两个单词是对应的。

最后定义一个最佳检验函数，通过密钥将密文转化成的明文进行检验，通过将明文中字符出现的频率统计出来，再将明文中对应的字符频率和正常英文所字符对应频率进行求差，并对 $A\sim Z$ 26 个字符频率差进行求和并取绝对值，最后用 1 减去求和的值。就得到了最佳匹配函数。

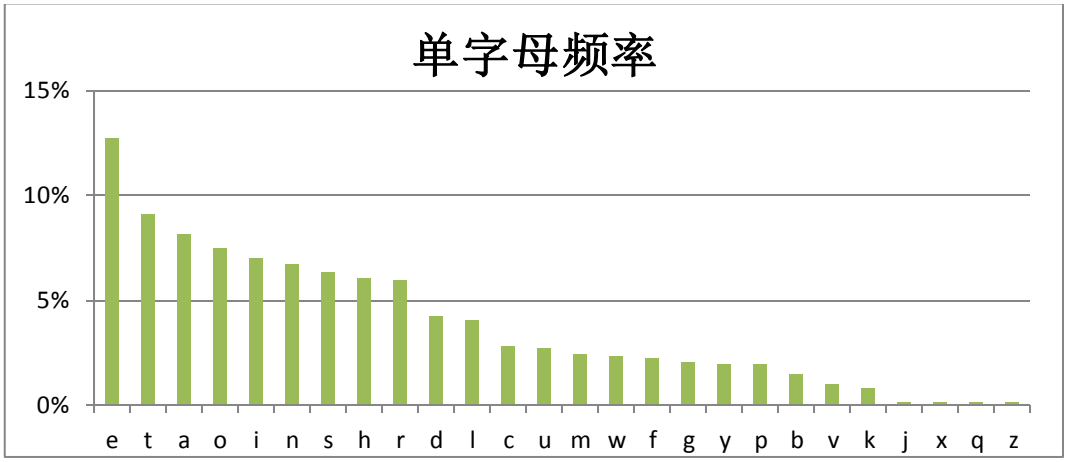
4. 数据处理

4.1 单字母频率统计

4.1.1 字母频率

字母频率，指的是各个字母在文本材料中出现的频率,常被应用于密码学，尤其是可破解古典密码的频率分析，在英语中最常见的字母是 e。引用 Wiki 百科单字母数据[1]，通过数据排列得到单字母频率统计[6]（图 1）如下：

图 1：单字母频率统计排序

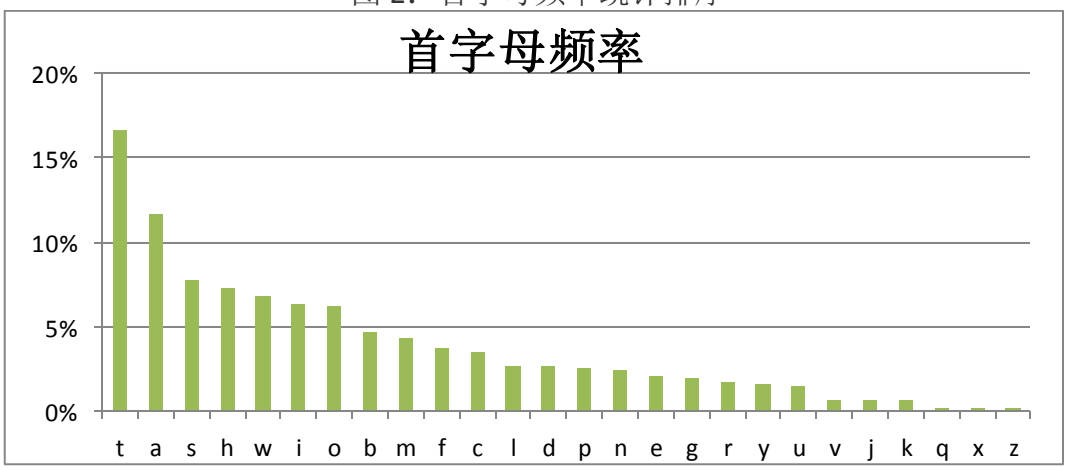


在任何一种书面语言中，不同的字母或字母组合出现的频率各不相同。而且，对于以语言书写的任意一段文本，都具有大致相同的特征字母分布。根据上图可知，字母“E”出现的频率最高，而“Z”则出现得最少，差异显示为 180 倍。另外，发现排列前五的字母中元音占了四个，可见元音字母使用频率之高。

4.1.2 首字母频率

引用 Wiki 百科首字母数据，通过数据排列得到首字母频率统计（图 2）如下：

图 2：首字母频率统计排序



由上图可得，排名第一的首字母为“T”，根据英语单词大全不难发现原因；次之为“A”，出现频率达到 11.6%，结合图 1 可得元音字母“A”无论是以单字母的形式出现还是以首字母出现，其频率都较高；除“Q”“X”“Z”出现频率极低外，其他字母为首字母的频率都相对较稳定徘徊在 8%-1%之间。

4.2 双字母频率统计

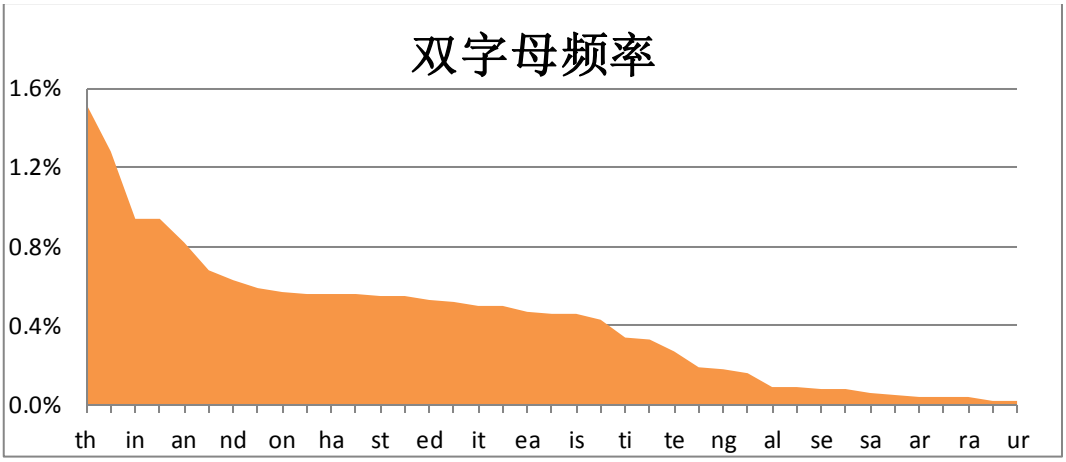
双字母组或称二元语法，作为统计分析文本使用非常广泛，它是由两个字母，或者两个音节，或者两个词构成的双字母组。这种组被用在最成功的一种 语音识别的语言模型中。引用 Wiki 百科双字母数据[2]，下面列出了英语里最常见的双字母组（图 3）如下：

字母对	频率	字母对	频率	字母对	频率	字母对	频率
th	1.52%	en	0.55%	ng	0.18%	ar	0.04%
he	1.28%	ed	0.53%	al	0.09%	ra	0.04%

……	……	….	….	…	…	….	…
st	0.55%	ti	0.34%	sa	0.06%	ur	0.02%
es	0.56%	as	0.33%	si	0.05%	ld	0.02%

做出数按字母频率的图像，可以直观观察双字母对的变化规律。图如下：

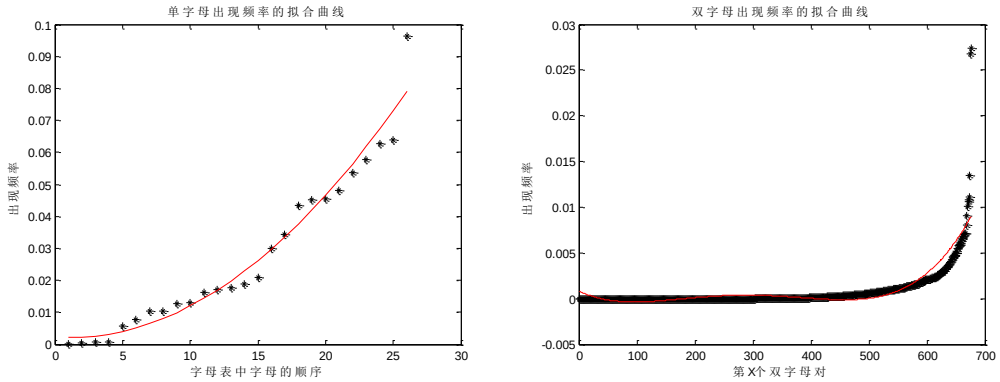
图 3：双字母频率统计排序[7]



根据上图可观察出双字母的频率分布情况，“TH”频率最高，“UR”最低。将分布图划分为高频率、中频率、低频率三份，其中高频率出现的双字母为：“TH”“HE”“IN”“ER”“AN”、中频率出现的双字为：”RE””VD””AT””ON””NT””HA””ES”等、其次为低频率出现的双字母。

4.3 对字符频率的拟合

通过把单字符在正常英文中的频率和双字母对在英文的频率，在直角坐标系中描点，再用曲线拟合的方式，拟合得到单字母频率，和双字母频率分布的曲线如下：



5. 问题的求解

5.1 模型的建立

5.1.1 密文的生成

密文的生成，将明文转成密文，由于在信道传输过程中，会出现字母的丢失，在字母正常传输的过程中，会在后面加上随机干扰字母，篡改字符以及正常传输，分别对应的概率为 p_1, p_2, p_3, p_4（其中 $p_4 = 1 - p_1 - p_2 - p_3$），在通信过程中，每个字符经过信道传输的结果都属于如上四种情形之一。每个字符受到干扰的都是不一样的，由于每一个字母的在英文出现的频率是不一样，密文在通信过程中，字符出现频率越高的，在传输过程中，受到的干扰的频率也就越高，所以将字符在英文出现的频率，作为该字符在传输过程受到干扰的概率，根据假设有，三种干

扰的使密文出错的概率相同，所以就有：

$$\begin{cases} p_1=p_2=p_3=\sum_{j=1}^{3}\omega_j p_j \\ \sum_{j=1}^{3}\omega_j p_j + p_4 = 1 \\ \omega_1+\omega_2+\omega_3=1 \end{cases}$$

p_4 表示字符正常传输的概率， $\omega_1,\omega_2,\omega_3$ 分别表示字符丢失，字符正常传输而后面被加上新的字符以及字符被篡改的概率的权重。

最后通过 *MATLAB* 模拟仿真就可以按照题目要求将明文装成密文。

5.1.2 约束条件建立

1. 约束条件一

由于每一个字符在正常英文中都有一个稳定频率，所以密文转成明文的的统计得出字符的频率大于等于零，且不能超过字符在正常英文中的频率加上 5%, 即得出约束条件为：

$$0 \le SF_i \le DF_i + 0.05$$

2. 约束条件二

因为统计密文和明文各字符重复频率都应该满足统计的基本规则，密文和明文各字符重复频率之和应该等于 100%，

$$\sum_{i=1}^{26} SF_i = 100\%$$

3. 约束条件三

因为第二阶段的密文没有间隔与标点符号，每一篇密文中的字符都应该是 $A\sim Z$ 的 26 个字符中的，不应该有其它的字符出现。传输的干扰只是字符丢失，篡改，和增加随机英文字符。所以也不会引进其它的字符。所有一下约束条件。

$$0 \le i \le 26 \quad SF_i \in A\sim Z$$

4. 约束条件四

每个字符经过信道传输的结果都属于字符丢失，篡改，增加随机英文字符以及正常传输四种情况中的一种，且每种情况都各对应一个概率，四种情况的概率之和应该等于 1。且正常传输的概率应该大于其余三种受到干扰概率之和。

$$\begin{cases} \sum_{j=1}^{4} p_j = 1 \\ \sum_{j=1}^{3} p_j \le p_4 \end{cases}$$

综上所述，得到约束条件为：

$$\begin{cases} \sum_{i=1}^{26} SF_i = 100\% \\ \sum_{j=1}^{4} p_j = 1 \\ 0 \le SF_i \le DF_i + 0.05 \\ 0 \le i \le 26 \quad SF_i \in A \sim Z \\ \sum_{j=1}^{3} p_j \le p_4 \end{cases}$$

5.1.3 目标函数建立

首先将密文中字符出现的频率统计出来，并将密文中字符出现频率与英文中单字符出现的频率做差，两个字符之间的频率之差的绝对值最小，则就认为两个字符是相匹配的。目标函数为：

$$\min Match = |SF(i) - DF(i)|$$

式中：$SF(i)$表示第i个字符在密文中出现的频率，$DF(i)$表示第i个字符在正常英文中出现的频率。

然后，由于题目要考虑字符传输过程有可能丢失，在正常传输的情况下，会在后面增加一个随机字符以及字符被篡改成其他的字符。所以从密文替换出来的明文，就与最开始的明文加密成密文时不一致，所对应出来的单词与词料库中对比，有可能找不到。所以定义一个最大的相似度函数，让两个单词相似度最大，就表示这两个单词可以对应。

$$\max Similarity = \frac{S_n}{D_m} \times 100\% \quad (n \le m)$$

式中，S_n表示两个单词对比有n个字符相同，D_m表示单词的总长度。

最后定义以最佳检验函数，通过密钥将密文得到的明文检验，通过将明文中字符出现的频率统计出来，再将明文中对应的字符频率和正常英文所字符对应频率进行求差，并对$A \sim Z$ 26 个字符频率差进行求和并取绝对值，最后用 1 减去求和的值。就得到了最佳匹配函数。

$$\max \ Fitness = 1 - \sum_{i=1}^{26} |DF(i) - CF(i)|$$

式中：$DF(i)$表示第i个字符在密文中在正常英文出现的频率，$CF(i)$表示第i个字符在明文中出现的频率。最佳检验函数值越大，就表示密文转成明文的正确性就越高。

综上所诉，得到目标函数为：

$$\begin{cases} \max Similarity = \frac{S_n}{D_m} \times 100\% \\ \min Match = |SF(i) - DF(i)| \quad (n \le m) \\ \max \ Fitness = 1 - \sum_{i=1}^{26} |DF(i) - CF(i)| \end{cases}$$

综合以上对约束条件和目标函数的建立，得到多目标优化模型：

$$\begin{cases} \sum_{i=1}^{26} SF_i = 100\% \\ \sum_{j=1}^{4} p_j = 1 \\ 0 \le SF_i \le DF_i + 0.05 \\ 0 \le i \le 26 \quad SF_i \in A \sim Z \\ \sum_{j=1}^{3} p_j \le p_4 \end{cases}$$

$$\begin{cases} \max\ Similarity = \dfrac{S_n}{D_m} \times 100\% \\ \min\ Match = \left| SF(i) - DF(i) \right| \qquad (n \le m) \\ \max\ Fitness = 1 - \sum_{i=1}^{26} \left| DF(i) - CF(i) \right| \end{cases}$$

5.1.4 破译替换式密码的频率分析算法

*Step*1：首先找出密文中重复字符的频率，然后将密文中的字符频率与正常英文字符频率求差并取绝对值，最后确定由两个字符频率之差的绝对值最小来确定为两个字符为相对应关系，通过这种方式找到密文中最大频率字符对应的关系。并替换密文中相应的字符。例如密文字符 t 与正常密文字符 a 。即密文的 t 用 a 替换。

*Step*2：在 *Step* 1中得到密文最大频率字母对应的关系，然后根据双字母对找出密文中最大频率字符对应的两个明文双字母对，并取双字符组频率较大的双字符组。再对密文进行替换。例如双字母对 th 的频率为 1.52%，而 ht 的频率为 0.55%。

*Step*3：在 *Step* 2中，然后结合单个字符作为首字母的频率，把单词的首字母确定下来，就把得出单词与词料库的单词相比较，例如，字符 t 和 a 作为首字符的频率最高，可以先将 t 作为首字符。并算出其两个单词对应相似比，如果较大，则该单词与词料库中单词相对应，则就可以将得到另外的字符对应情况，并保留此单词，并把单词两边加上空格，否则，就继续重新找对应密钥。

*Step*4：在 *Step* 3中，得到密文与明文的对应关系后，根据多目标优化模型定义的最佳检验函数，观察结果的正确性，根据算出检验函数值，检验函数值越大，说明密文转化成明文的的正确性越高。

*Step*5：最后根据英语语法，将密文对应出得到的明文进行断句。

*Step*6：进行多次，模拟仿真，每次都会得到一个最佳检验函数值，看最佳检验函数值是否符合要求。

5.2 模型的求解

本文采用频率分析法，利用 Matlab 编程统计密文中字符出现的频率，并将密文中字符出现频率与英文中单字符出现的频率做差，得出频率之差为最小平均值，证明两个字符是相匹配的；其次定义一个最大的相似度函数，使对应出来的单词与词料库中单词可相对应；以最佳检验函数，通过密钥将密文得到的明文检验，得出最佳匹配函数；最佳检验函数值越大，就表示密文转成明文的正确性就越高。计算求解得到明文与暗文替换的字符如下（图 4）：

图 4：明文与暗文字符替换表

明文	a	b	c	d	e	f	g	h	i	j	k	l	m
暗文	t	l	o	u	d	e	b	w	v	y	k	n	h
明文	n	o	p	q	r	s	t	u	v	w	x	y	z
暗文	x	s	p	v	j	z	g	c	l	m	q	f	r

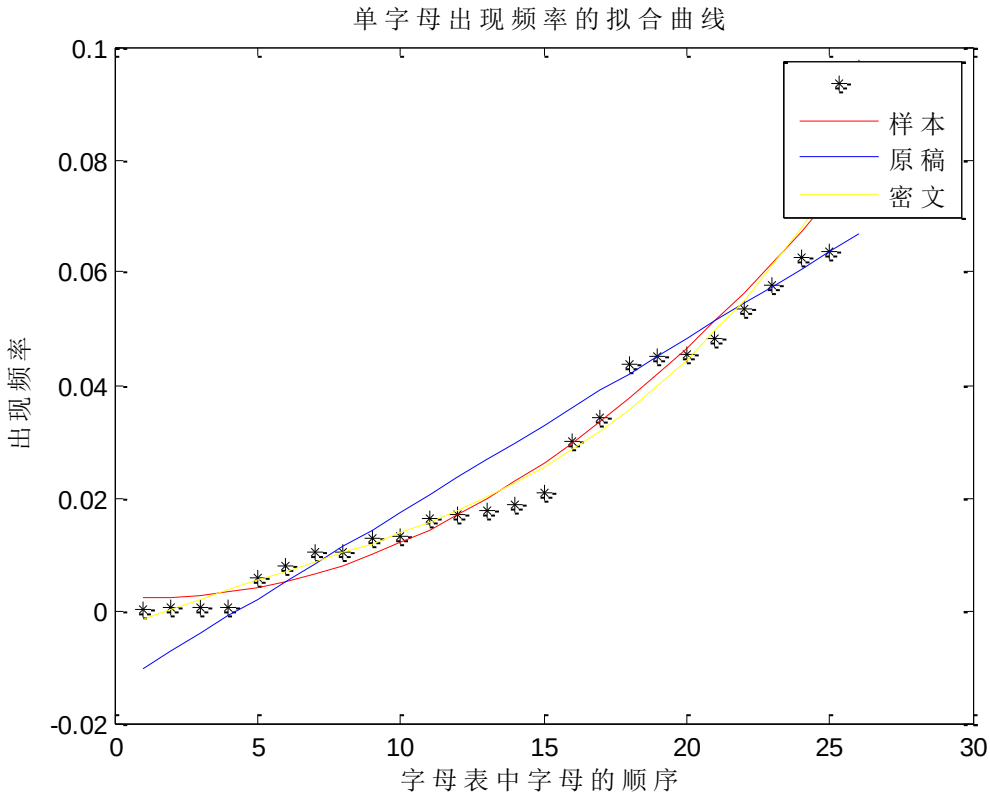
并计算得到检验函数值为 $Fitness = 0.865312$，由于最佳检验函数值越大，密文转成明文的正确性就越高，所以可判定本文建立的多目标优化模型的正确性。

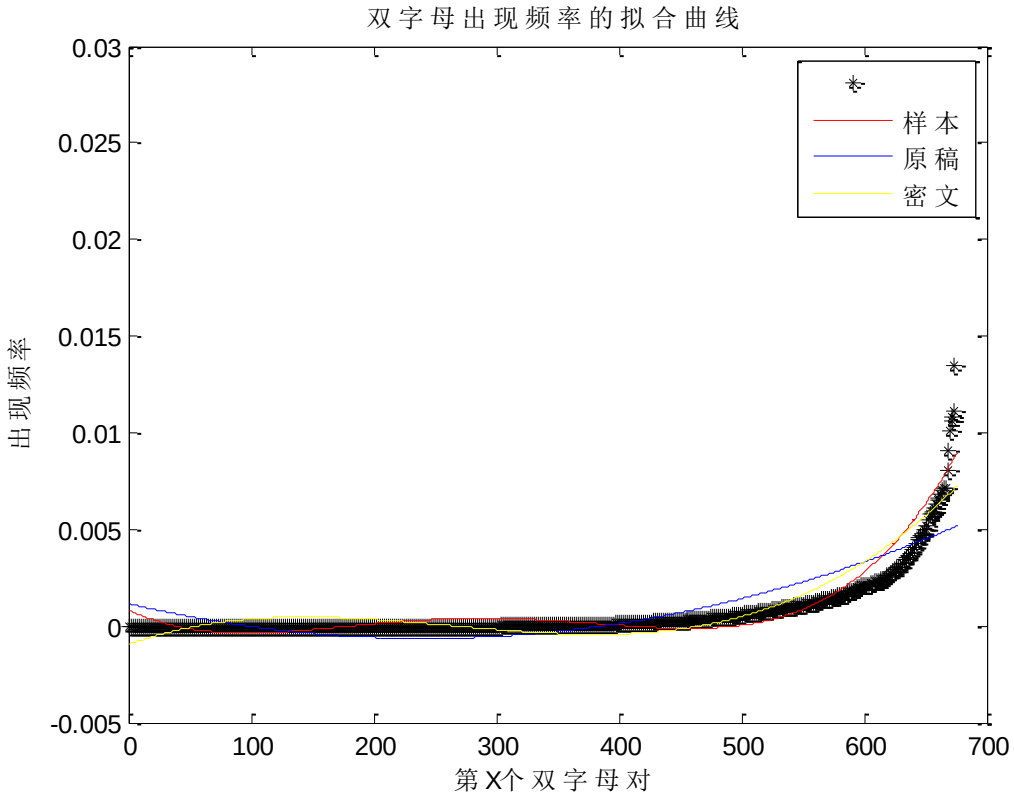
5.3 结果分析

5.3.1 结果的定性分析

通过对密文字符频率，双字符对频率的拟合，明文字符频率，双字符对频率的拟合及正常英文字符频率，双字母对频率的拟合，得出拟合曲线如下（图 5）：

图 5：拟合程度





通过对密文单字母，双字母对的频率曲线拟合、明文单字母，双字母对频率的曲线拟合及正常单字母和双英文字母对频率的曲线拟合，得到拟合值R1=0.9991，说明拟合效果很好。从图中可看出，曲线走势差异甚微，呈缓慢上升状态。因此说明本文所建模型对密文的适用性及对结果准确性的定性分析。

5.3.2 结果定量分析

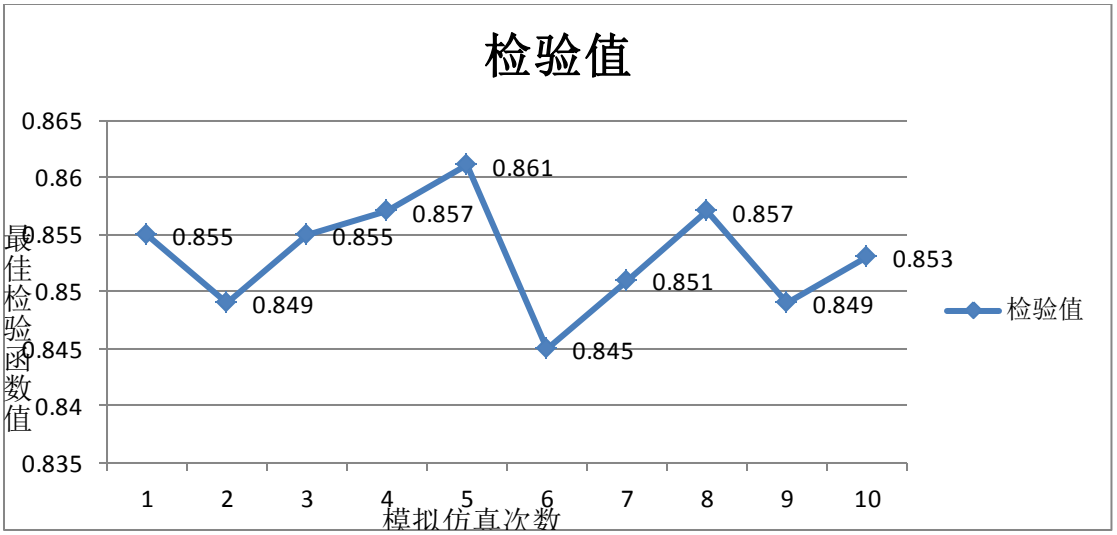
将得到的明文和正常的英文，把字符的频率代入到最佳检验函数中，进行最佳检验计算，计算得到检验函数值为 $Fitness = 0.865312$，由于最佳检验函数值越大，密文转成明文的正确性就越高，所以可判定本文建立的多目标优化模型的正确性。

5.4 模拟仿真分析

通过字符在信道传输过程中，受到干扰的概率的约束，运用 MATLAB 编程模拟仿真文本加密的方式，生成不同的密文。然后运用论文中建立的模型和设计的频率分析算法。将密文进行一一破译，得到明文。最后运用最佳检验函数得到检验函数值如下：

次 数	1	2	3	4	5	6	7	8	9	10
检验值	0.855	0.849	0.855	0.857	0.861	0.845	0.851	0.857	0.849	0.853

画出模拟仿真的，经过最佳检验函数计算的检验值的折线图如下：



从上面的表和图中可以看出，经过 10 次模拟仿真得到明文，用最佳检验函数检验的数值都在平均值 0.8532 附近波动。并且振幅不大，所以通过模拟仿真进一步证明了模型和算法的正确性和实用性。

6. 模型的评价

6.1 优点

（1）本文利用 Excel 软件对数据精心处理并作出个汇总图表，简便、直观、快捷；利用多种数学软件取长补短，使计算结果更加精确；

（2）在破译算法中，建立的模型与实际精密相连，充分考虑各字母出现频率的高低，从而使模型更贴近实际，通用性强；

6.2 缺点

（1）为使所得结果更加理想，忽略了一些次要影响因素。其中包括不同作者或写作主题的作品中往往各不相同，导致统计的字母频率存在偏差，对明文加密也产生影响；

（2）题中给出字符丢失、添加、篡改三种出错概率，而本文假设三种出错概率均相同，对所计算的结果有误差。

7. 模型的改进及推广

7.1 模型的改进

在破译算法中,要使其能够在带有干扰的条件下完成破译工作，存在一定的不精确性。有分析显示字母频率就像词频，不同作者或写作主题的作品中往往各不相同。当为 x 射线（x-rays）撰文时，文章中就会有大量的字母 X。而撰写用 x 射线治疗卡塔尔（Qatar）的斑马（zebras）时，一般很少出现的字母 X、Q 和 Z 就会充斥文中。可从作者的字母使用频率中看出他的某些写作习惯。例如，海明威的写作风格明显不同于福克纳。字母、双字母组、三字母组、单词频率、单词长度和句子长度，这些都可以经统计后用以证明或反驳某一作品是某作者所写，甚至待鉴别作品与作者的写作风格相近也可用这一方法。只能靠分析大量有代表性的文本才可得出准确的字母平均频率，所以要使破译算法计算结果更加精确，还需分析大量代表作文本。

7.2 模型的推广

作为一种高效的智能全局搜索的优化破译算法，正日益应用到科学研究和工程实践的众多领域，越来越多的人开始关注、研究这一技术。本文建立模型计算的结果表明,系统能够很好地完成密码破译工作,并且具有较高的性能。而密码技

术作为一种保密技术对于一个国家来说，无论在军用上还是民用上，其地位都是无可替代的，而且有变得越来越重要的趋势。

8. 参考文献

【1】何晓琴，一种新式 Vigenere 密码的破译和研究[J]. 计算机科学 第 40 卷,第 12 期,2013

【2】巩延文，马尔可夫链蒙特卡洛方法在密码学中的应用[J]，哈尔滨工业大学,2013 年 6 月

【3】王彩霞，密码分析中几种方法的研究及其设计与实现[J]，西北大学,2004 年

【4】裴治捷，密码学中布尔函数性质研究[J]，上海交通大学,2012 年 9 月

【5】龚洁中，遗传算法在密码分析中的应用研究[J]，上海交通大学, 2007 .12

【6】字母频率：
http://zh.wikipedia.org/wiki/%E5%AD%97%E6%AF%8D%E9%A2%91%E7%8E%87

【7】双字母组：
http://zh.wikipedia.org/wiki/%E5%8F%8C%E5%AD%97%E6%AF%8D%E7%BB%84

9．附录

附录一：

```
clear;clc;
%% 导入样本圣经 txt 文件，用作词库
A=fileread('Bible.txt');
%% 计算每个字母出现的频率 P，并存入 gailv.xls
for i=1:26
    B(i)=char(i+96);
    p{i}=strfind(A,B(i));
    P(i)=length(p{i})/length(A);
end
%% 将字母按出现的频率进行排序
y1=B(:);
Y1(:,1)=P(:);
[X1,a1]=sort(Y1);
for i=1:length(Y1)
    x1{i,1}=y1(a1(i));
end
%% 对单字母出现频率进行拟合
% %% 计算出多项式拟合的阶数
z1=(1:length(X1))';
for j=1:10
    y2=polyfit(z1,X1,j);
    Z1=polyval(y2,z1);
    if sum((Z1-X1).^2)<0.001
        c1=j;
        R1=1-sum((Z1-X1).^2)
        break;
    end
end
% %% 绘制拟合曲线
figure(1)
y2=polyfit(z1,X1,c1);          %c 是多项式的次数，拟合出来的结果 y2 是系数向量
y1=polyval(y2,z1);             %计算出拟合的 y 值
plot(z1,X1,'k*',z1,y1,'r-');  %画出数据对比图，黑点是原始数据，红线是拟合曲线
% %% 得到拟合多项式的表达式
syms t f(t)
disp('得到单字母的拟合多项式为：')
y2=vpa(y2);
f(t)=poly2sym(y2,t)
title('单字母出现频率的拟合曲线');
xlabel('单字母');
ylabel('出现频率');
xlswrite('gailv.xls',P,'单字符');
```

```
%% 计算双字母出现的频率 Q，并存入 gailv.xls
B={};
for i=1:26
    for j=1:26
        B{i,j}=strcat(char(i+96),char(j+96));
        q{i,j}=strfind(A,B{i,j});
        Q(i,j)=length(q{i,j})/length(A);
    end
end
xlswrite('gailv.xls',Q,'双字母');
%% 将双字母按出现的频率进行排序
y=B(:);
Y(:,1)=Q(:);
[X,a]=sort(Y);
for i=1:length(Y)
    x{i,1}=y(a(i));
end
%% 对双字母出现频率进行拟合
% %% 计算出多项式拟合的阶数
z=(1:length(X))';
for j=1:10
    y2=polyfit(z,X,j);
    Z=polyval(y2,z);
    if sum((Z-X).^2)<0.001
        c=j;
        R=1-sum((Z-X).^2)
        break;
    end
end
% %% 绘制拟合曲线
figure(2)
y2=polyfit(z,X,c);          %c 是多项式的次数，拟合出来的结果 y2 是系数向量
y1=polyval(y2,z);           %计算出拟合的 y 值
plot(z,X,'k*',z,y1,'r-');  %画出数据对比图，黑点是原始数据，红线是拟合曲线
% %% 得到拟合多项式的表达式
syms t f(t)
disp('得到双字母的拟合多项式为：')
y2=vpa(y2);
f(t)=poly2sym(y2,t)
title('双字母出现频率的拟合曲线');
xlabel('字母对');
ylabel('出现频率');
```

附录 2：

```
clear;clc;
%% 去除原文 A 中的标点和空格及改变大小写所得的明文 C
A=fileread('Obama.txt');
B=A;
for i=1:length(B)
    if B(i)<65||B(i)>122||B(i)>90&&B(i)<97
        B(i)='?';
    end
end
B(find(B=='?'))=[];
C=B;
for j=1:length(C)
    if C(j)>=65&&C(j)<=90
        C(j)=char(C(j)+32);
    end
end
%% 明文 C 随机编码得到密文 D，并将得到的密文写入 miwen.txt 中
D=C;
global M;
global N;
M=char(randperm(26)+96);
N=char(randperm(26)+96);
for k=1:length(C)
    p=find(M==C(k));
    D(k)=N(p);
end
fprintf('原文是：\n %s\n', A)
fprintf('密文是：\n %s\n', D)
fid=fopen('miwen.txt','wt');
fprintf(fid,'%s\n',D);
fclose(fid);
fprintf('\n 结束:已完成对一段文字的随机加密\n')
```

附录 3：

```
clear;clc;
%% 导入密文的 txt 文件
C=fileread('miwen.txt');
D=C;
%% 模拟密文传输过程中出现的干扰
p=xlsread('gailv.xls');        % 字母出现的频率
N=char(97:122);                % 字母的顺序排列
%%% 密文中的字符在传输过程中被篡改为一个随机字符
for i=1:length(D)
    for j=1:length(N)
```

```
        if D(i)==N(j)
            s1=[1,0];g1=[1-p(j),p(j)];
            D(i)=C(i)*randsrc(1,1,[s1;g1]);
            if D(i)==0
                a=char(randperm(26,1)+96);   % 字符在传输过程中被篡改为一个随机字符
                b='!';                       % 字符本身正常传输，在其之后添加了一个随机字符
                c='?';                       % 字符在传输过程中被丢失
                s2=[a,b,c];g2=[1/3,1/3,1/3];
                D(i)=randsrc(1,1,[abs(s2);g2]);
            end
        end
    end
end
D=char(D);
%%% 字符在传输过程中被丢失
D(find(D=='?'))=[];
E=C(find(D=='!'));
e=find(D=='!');
n=1;
F=[D,E];
%%% 字符本身正常传输，但在其之后添加了一个随机字符
for k=1:length(F)
    if F(k)=='!'
        x=[];
        x=D(k-n+2:end);
        F(k)=E(n);
        F(k+1)=char(randperm(26,1)+96);
        F(k+2:end-length(E)+n)=x;
        n=n+1;
    end
end
%% 将干扰处理后的密文存入 ganrao.txt 文件中
fid=fopen('ganrao.txt','wt');
fprintf(fid,'%s\n',F);
fclose(fid);
fprintf('密文传输干扰完成\n')
```

附录 4：

```
clear;clc;
%% 导入经过干扰处理后的密文 ganrao.txt 文件
F=fileread('ganrao.txt');
%% 计算每个字母在密文中出现的频率 p
```

```
G=char(97:122);       % 字母的顺序排列
for i=1:26
    B=G(i);
    P{i}=strfind(F,B);
    p(i)=length(P{i})/length(F);
end
% for k=1:length(N)
%     fprintf('字母%s 在密文中出现的概率是%f:\n', N(k),p(k))
% end
%% 计算双字母在密文中出现的频率 q
B={};
for i=1:26
    for j=1:26
        B{i,j}=strcat(char(i+96),char(j+96));
        Q{i,j}=strfind(F,B{i,j});
        q(i,j)=length(Q{i,j})/length(F);
    end
end
%% 将双字母按出现的频率进行排序
y=B(:);
Y(:,1)=q(:);
[X,a]=sort(Y);
for i=1:length(Y)
    x{i,1}=y(a(i));
end
%% 将密文的双字母与统计频率进行破译替换
W=F;
global M;
global N;
for k=1:-5+length(F)
    r=find(N==F(k));
    W(k)=M(r);
end
C={};
for i=1:26
    for j=1:26
        C{i,j}=strcat(char(i+96),char(j+96));
        T{i,j}=strfind(W,C{i,j});
        t(i,j)=length(T{i,j})/length(W);
    end
end
s=xlsread('gailv.xls',2);
for i=1:length(t)
    for j=1:length(s)
```

```
            z(i,j)=3*length(P)*(s(i,j)-t(i,j))^2;
        end
    end
Z=1-sum(sum(z));
fprintf('得到破译的正确率：%f\n',Z);
```

附录 5：

helloeverybodythankyouthankyouthankyoueverybrodyallrighteverybodygoaheadandhaveseathowiseverybodydoingtodayapplausehowabouttimspicerapplauseiamherewithstfudentsatwakefieldhighpschoolinarlingtonvirginiandwevegotstudentssxuninginfromallacrossamericafromkindergartenthroughthgradeandiamjustsogladthatallcouldjoinustodayandiwanttothankwakefieldforrzeingsuchanoutstaniinghostgiveyourselvesabigroundfapplauseapplauseiknowthatformanyofyoutumytyisthefirstdayofschoolandforthoseofyouinkindergartenorstkrtingmiddleorhighschoolitsyourfirstdayinanewschoolsoitsunderstandableifyourealittlenervousiimagnethereareasomeseniorsouttherewhfareageelingprettygoodrightnowapplauswithjustonemoreyeartogoandnomatterwhatgradeyoureinsomeofyouareprobablywishingitwerestilrsummerandyoucouldvestxyedinbedjstalittlebitlongerthismorningiknowthatfeelttngwheniwasyoungmyfamilylivedoverseasilivedinindonesiaforafewyarsandmymothersoiedidnthavethemoneytosendmewhereallttheamericasnkidswenttoschoolbutshethoughtitwasimportantformetokeeupwithanamericaneducationsoshedecidedtoteachmeextramnessonsherselsfmondaytroughyfridaybutbecauseshehadtogotoworktheonlytimeshecoutddoitwasatinthemorningnowasyoumightimuegineiwasnttoohappyaboutgettngpthatearlyandaolotoftmesidfallasleeprighttherеatthekitchentablebutwheneveridcomplainmiamootheriouldjustgivemeoneofthoselooksandshedsaythisisnopicnicformeitherbusterlaughtersoiknowthatsomeofyuttrestilladjustingtobeingackatshoolbutimheretodaybecauseihvesomethingiportanttodiscusswithyouimherebecauseiwanttotalkwithyoigboutyoureducationandwhatsexpectedofallofyouinthisnewschoolyearnowivegivenalotofspeechesabouteducationandivetalkedaboutresponsbiltyalotivetalkedabouteachersrespnsibilityforinspiringstudentsandpushingyutolearnivetalkedubоutyurarentsresponsiafiotityformakingsureyoustayontrackndyougetyourhomeworkdoneanddontspendeverywakinghourinfronteqfthetvorwiththexboxivetalkedlotaboutyourgbdvernmentsresponsibilityforsettinghighstandardsandsupportisngteachersandprincipalsanduurningaroundschoolsthatarentworkngwhedgestudentsarentgettingtheopportunitiesthattheydeservebutttheenofthedaywecanhavethemostdedicatedteachersthemostsupportiveparentsthebestschoolsinthewoldandnoneofitwillmakeadifferencenoneofitwillmaterunlessallofyoufulfillyourresponsibilitiesunlessyvushowuptothosescholsunlessyoupayattentiontothoseteachersunlessyoueistentoyourparentsandgrandparentsandotheradultsandputinthehardworkittakestosucceedthatswhatiwanttofoabusontodaytheresponsibilityeachofyouhasforyoureducationiwanttostartwiththeresponsibilityyouhavetoyourselfeverysingleoneofyouhassomufthingthatysauregoodateverysingleoneofyoaehassomethingtoofferwndyouhavearesponsibilitytkyourselftodiscoverwhtthatisthatstheoppftrtunityanеducetioncanprovihpemaybeyoucouldbeagreatwritermkybeevengoodenoughtowriteabookorarticlesinanewspiperbutyoumightnotknowituntilyouwritethatenglhoshpapertilatenglishclasspaperthatsassignedtoyoumaybeyoucouldbeaninnoviqtororaninventormaybeevengoodenoughtocomilupwiththenextiphoneorthenewmedicineorvaccinebutyoumightnotknowituntilyoudoyourprojmectforyourscienceclassmaybeyoucouldbeayqayororosenatororasupremecourtjusticebutyoumightnotknowthatuntilyoujoinstudentgvernmentorthedebateteamandnomatterwhatyouwanttodowityourlifeiguaranteethatyoullnedaneducationtodoityuwanttasbeadoctororateacherorapoliceofficeryouwanttobeanurseoranarchitectalawyeroraemberofourmilitaryyouregoingtoneedagoodeducationforeverysingleoneofthosecarersyoucannotdropoutofschoolandjustdropintoagoazdjotyouvegottotrainforitandworkfwritandlearnforitandthisisntjustiwaportantforyourownlifeandyourwnfuturewhatyoumakeoeyoureducationwilldecifenothinglessthanthefutureofthnpscountrythefutureofaericadependsonyouwhatyourlearninteinschomfltodaywpjlldetnmrminewhetherweasanationcanmeetourgreatestchallengesinthefutureyoullneedtheknowledgeandproblemsolvingskillsyoulearninscienceandmathtocurediseaseslikecancerandaidsandtodevelopnewepnergytechnologiesandprotectourenvironmentyullneedtheinsightsanrrcriticalthinkngskillsyougaininhis

toryansocirolstudiestofightpovertyandhoiyelessnesscrimemnddiscriminationandmakeournwtion moreairkndorefreyorvllneedthecreativityandigenuityyoudevelopinallyourclassestrlbuldnewcomp aniesthatwillcreatenewjobsandboostoureconomywenee deverysngleoneofyoutodevelopyburtalen tsndyourskillsandyourintellectsoyoucanhelpusoldfvlkssolveuiurmostdifficultptxoblemsiyoudontd othatifyouquitonschoolyourenotjustquittingonyourselfyourequittingonyourcountrynowiknowitsn otalwayseasytodowellinscholiknowclotofyouhavechallengesinyourlivesrightnowthatianmpkeithar dtofocusonyourschogworkigetitiknowwhatitslikemyfotherdzeftmyfamilywheniwastwoyearsoldan diwasraisedbyasinglmomwhohadtoworkandwhostruggledattimestopaythebillsandwasntalwaysab letogiveusthethingsthatotherkidshadtverewereretimeswhenimissedilavintwafatherinmylifetherewe retimeswheniwaslonelyandiejeltlikeididntfitinsoiwasntalwaysasfobusnldasishouldhavebeenonsch oolandididsomethingsimnotproudofinsmiagotinmoretroublethanishouldhaveandmylifecouldhave easilytakenaturnfogrtheworsebutiwasiwasluckyigotalotofsecondlphancesandihadtheopportunityt ogtocollegeandlwschoolandfolowmydreamsmywifeourfirstladymichelleobamashehasasimilarstor yneitherofherparentshadgonetocollegeandtheydidnthavealtofmoneybuttheyworkedhardandshe workedhardsothatshecouldgotothebestschoolsinthiscountrysmeofyoumightnothavethoseadvant agesmaybeyoudonthaveadultsinyourlifewhogiveyouthesupportthatyouneedmaybesomeoneinyou rfamilyhaslosttheirjobandtheresnotenoughmoneytogoarounzmaybeyouliveinaneighborhoodwhe ryoudontfeelsafeorhatefriendswhoarepressuringyoutodothingsyouknowarentrightbutattheendof thedaythecircumstancesofyourlifewhatyoulooklikewhereyoucomefromhowmuchmoneyyouhave whatyouvegotgotdngonwthvmenoneofthatisanexmiuseforneglectingyourhomeworkorhavingaba dattitudeinscholthatsnoexcusefortalkingbacktyourteacherorcuttingclassordroppingotofschoolthe reisnoexcusefornottryingwhereouarerightnowdoesnthavetodeterminewhereyoullendupnoonesw rittenyourdestinyforyoubecausehereinamericayoewwriteyourowndestinyyoumakeyourownfuture thatswhatoungpeoplelikeyouredoingeverydayallacrossamericayoungpeoplelikejazminperzfrvmro matexasjazmgyndidntspeakenglishwhensheirststartedschoolneitherofherparentshadgonetoollege exutseworkedhardearnedgoodradesanocgotascholarshiptobrownuniversityisnowingrhqduatesch oolstudyingpublichealthonherwaytobecomingdrjazminperezimthinkingaboutandonischutzfromlo saltoscaifforniawhosfoughtbrrpincancersincehewasthreeheshaptendureallsroisoftreatmentsandsu rgerisoneofwhchaczfecbtedhismemorysoittookhimmuchlongerhundredsofextrekhourstodohisscho oelworkbuthe neverfellbehindhesheadedtocollegethisftllandthentheresshntellstevefrommyhom etownvffchicagmillinehisevenwhenbouninghsromfosterhometofosterhomeinthetoughestnemzgh borhoodsinthecityshemanagedtogetajobatalofalhealthtuarecenterstartaprogramtokeepyoungpeo pleoutofgangsandshevsontracktograduatehighschzlwithhonorsandgoontocollegeandwazminljndo niandshannsellarentanydifferentfromanyofyoutheyfacechallengesintheirlivesjustlikeyoudonsome casestheyvegotitalotworseoffthanmanyofyoubuttheyresxusehctogiveuptheychosetotakeresponsi biityvaortheirlivesfortheireducatowonandsetgoalsforthemselvesandiexpectallkmfyoutodothesam ethatswhytodcyimcallingoneachofyoutosetyourowngoalsoryoureducationanddoeverythingyouca ntomeettheyiyourgcalcanbesomethingassimpleasdoanngfallyoureoomeworkpayingattentionincla ssorspendingsometimeeachdayreadingabookmaybeyolldecidetogetinvolvedinanextracurriularacti vityorvolunteeriny hnurcommunitymaybeyoulldecidetostandupforkidswhoarebeingteasedorbllied becauselpfwhotheyareorhowtheylookbecauseyoubelioevlikeufdothatalyoungpeopledeserveasafe environenttostudyandlearnmaybeyoulldecidetotakebetterckqreofyourselfsoyoucanbemorereadyt olearnandalongthoselinesbythewayihopeallofyouarewashingyourhandsalotandthatyoustayhomef romschorxnkwhenyoudontfeelwlszsowecankeeppeoplefromgettingthfluthisfallandwinterbutwha evryouresovetdoiwantyoutocommittoitiwantyoutoreallyworkatitiknowthatsometimesyugetthtse nsefromtvthatyoucanberichandsuccessfulitithoutanyhardworkthatyourtckettosuccessisthroughra ppingorszasketbalorbeingarealittvstarchancesareyourenotgoingobwianyofthosethingsthetruthisb eingsuccessulishardyouwontloveeverysubjectthatyoustudyyouwontclickwitheveryteacherthatyou havenoteveryhomeworkassignmentwillseeoqcompletelyrelevetnttoyourlscferigovtatthisminutea ndyouwontnecessarilysucceedateverythingthefirsttimeyotrythatsokaysomeoftheeostsuccessfulpe opleintheugrldaretheoneswhovehadthemostfailuresjkriywlingswhowroteharrypotterherfirstha rry potterbookwasrejecljedtimesbeforeitwasfinallypublishedmichaeljordanwascutfromhishighschool basketalltevmhelosthundredsofgamesandmissedthousandsfshotsduringhiscareerbutheoncnxsaidi havefiledoverandoverandoveragaininmylifeandthatswhyisucceedthesepeoplesucceoxdedbecause

theyunderstodthatyoucantletyourfailuresdefineyouyouhavtoletyourfailuresteachyuyouhavetolett
hemshowyouwhattododifferentlythenexttimesoifyougetintotroublethatdoesntmeanyoureatroubl
emakeritmeansyouneedtotryhardertoactrightifyougetabdgradethatdoesntmeanyourestupiditjusa
ltzeansyuneedtospendmoretimestudyingnoonesbornbeinggoodatallthingsyoubecomegoodrltthin
gsthrougmohardworkyourenotararsityathletethefirsttimeyouplayanewsportyoudonthiteverynotet
hefirsttimeyousingasongyouvegottopracticetcesameprincipleappliestoyourschoolworkyoumighth
avetodoamathproblemafewtimesbeforeyougetitrightyoumighthavetoreadsomethingafewtimesbe
foreyouunderstandityoudefinitelyhavetodiafewdraftsofapaperbeforozitsgoodenoughtohandindon
tbefraidtoaskquestionsdontbeaymraidtoaskforhelpwheny ouneeditidothateverydayaskngforhelpis
ntasignofweuiknessitsasignofstrengthbecauseit showsyouhavethecouragetoadmiwhenyoudontkn
owsomethingandthatthenallowsyoutolearnsomekphingnewsoindanadultthatyoutrustaparentagra
ndparentortecheracoaenhoracounselorandaskthemto helpyoustayontracktoaueetyourgoabsetnde
venwhenyourstruggingevenwhenyourediscouragedandyoufeellikeotherpeoplehavegknvenuponyo
udontevergiveuponyourselfbecausewhenyougiveuponyurselyougiveuponyourcountrythestoryofa
mericifisntaboutpeoplewhoquitwhenthingsgottoughitsaboutpeoplewhokeptgoingwhotriedharde
rwholovedthircountrytuoomuchtodoanythinglessthntheirbpystitsthestoryofstudentswhosatwher
eyousityearsagozndwentontowagearevolutonandtdteyfoundedthisnationyoungpeoplestudentsw
hosatwhereyousityearsagowhooverameadepressionandwonworldwarwhofoughtforcivirightsandp
utamanonthemoonstudentswhosatwheryousityearsagowhofoundedgoogleandtwitterandfegcebo
okandchangedthewaywecommunicatewitheachothersoufoaywanttoaskallofyouwhaysyourcontrb
utiongoingtobewhatproblemsareyougoingtosolvewhatdiscoverieswillyoumakewhatwilapresident
whocomeshereinororyearssayaboutwhatallofyoudidforhiscountrynowyourfamiliesyourteachersa
ndiaremoingeverythingwecantomkesureyouhavetheeducatonyouneedtoanswerthesequestionsim
workinghardtofixupyourclassroomsandgetyouthebooksandtheequipmentandthecomputersyoune
edtolhwarntyutyouvegotodoyourparttoosoiexpectallofyoutogetserioustisleariexpectyoutoputyour
besteffortintoeverythingyoudoiexpectgreatthingsfromeachofyousodontltdtusdowndontletyoodra
mpzlydownoryourcountrydownmostofalldontletyueurselfdownmakeusallproudthankyouverymuc
heverybodygodblessyougodblessamericathankyou