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：4296

参赛队员　(签名)　：

队员 1：

队员 2：

队员 3：

参赛队教练员 (签名)：

参赛队伍组别：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）：

4296

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）：

第八届数学中国数学建模网络挑战赛
地址：数学中国数学建模网络挑战赛组委会　　网址：www.tzmcm.cn
电话：0471-4969085　　邮编：010021　　Email：2015@tzmcm.cn

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　针对信道干扰的破译算法

关 键 词　破译 明码 筛选 频率分析 数据分析

摘　　要：

随着科学技术的发展，密码在信息传输过程中，应用越来越广泛。而对于每个空格和标点符号都保留的密文而言，是很容易破译的，不利于重要信息的传递与保密。现在，应用较多的是无空格无标点且有信息干扰的密文。

对于这类密码的破译，我们在第一阶段的基础上建立并加以改进了一个新的算法。

首先，我们对密文涉及到的每个单词都进行频率分析，得到其频率，再与常用单字母频率表进行一一对比，将高频字母进行替换。接着，找出含有高频字母的双字母组合，与常用双字母组合频率表对照，得出明文；在此基础上，我们进行断句，查找单字母频率中的次高频率。在达到某个阈值时，既 5%，就无法直接利用频率分析进行替换，我们想到把密文破译较多的字段作为突破口，依据英语造词结构，查询英语词库，找出符合条件的单词，并一一筛选，以确定最后的明文和密文的组合。经过这三阶段，大量密文已经被我们破译，此时我们可以根据上下文之间的关系，对现有的密文和明文进行尝试性断句，再结合英语词库，就可以破译出该段密文。

但如果密码在传输过程中，发生缺失，添加，篡改，此算法仍然是适用的。可假设一个由 n 个字母组成的单词中有一个字母发生缺失或添加，则我们在破解密文时就会在 n 个字母的单词词库中寻找，若搜寻不到，则去 n-1 或 n+1 个词库中查找。假设 n 个字母的词库有q_1个单词,n-1 个字母的词库有q_2个单词，n+1 个字母的词库有q_3个单词，那么正确的概率是$\frac{1}{q_1 \times q_2}$或$\frac{1}{q_1 \times q_3}$，显然该密文正确的概率不高。若密码在传输过程中发生篡改。假设该字母由 n 个字长组成，其中有 i 个字母发生篡改，降低了准确率，最小的错误率为$\frac{n-i}{n}$，100%。

参赛队号：4296

所选题目：B 题

参赛密码 _____
（由组委会填写）

Abstract :

With the development of science and technology, the password in the process of information transmission has been applied more and more widely. Reserved for each spaces and punctuation, cipher text would be cracked easily and not conducive to the transmission of important information. Now, it's widely applied without spaces, punctuation and cipher text information with interference.

For this kind of password cracking, a new algorithm is proposed when we established and improved with the first level results.

First of all, we did frequency analysis on cipher which every word involved, got its frequency, then compared with common single letter frequency table one by one, replaced the high frequency letters. And, identify the high frequency double letter combinations, and compared with common double letter combinations frequency table, reached the result; Find the pausing, and on this basis, we discussed single letters which is the second highest frequency single letters. When reached a certain threshold, 5% had been reached, could not compare the use of frequency analysis for replacement directly, we thought that the letters which had been deciphered a lot, could be a breakthrough point, on the base of English words' structure, then search English word library, to find suitable words, and screening one by one, to determine the final combination of plaintext and cipher text. After the three steps, a large number of cipher text has been deciphered, at this time we could refer to the context, the relationship between the trial was carried out on the existing cipher text and cut out long letter combinations, combining English word library, we would decode the segment cipher text.

But if the password missed, added, tampered in the process of transmission, the algorithm is still applicable. Assuming one of n letters of the word with letters is absent or added, then we decode the segment cipher text to search the n words thesaurus, if we searched nothing, lookup in the library, compared to n - 1 or n + 1 word thesaurus. Suppose that there are n letters of the word library has q_1 word, n - 1 letters thesaurus has q_2 words, n + 1 letters of the word library has q_3, so the correct answer probability is $\frac{1}{q_1 \times q_2}$ or $\frac{1}{q_1 \times q_3}$, obviously the cipher text correct probability is not high. In case of the password will be transmitted with tampering. Assuming that the letter is composed of n letters length, including tampered with i letters, reduces the accuracy, the smallest error rate is $\frac{n-i}{n}, 100\%$.

一.问题重述

1.1 问题背景

在众多的编制密码中，比如摩斯密码，凯撒密码，乘法密码，多表代替密码，比较发现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[1]。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。

在加密后，单词之间的间隔和标点符号全部被删去。

1.2 问题提出

1.根据我们第一阶段的算法，我们对其改进使其能够在没有间隔和标点符号完成破译工作。

2．我们对其进行改进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。

二．问题分析

问题一的分析:

在密码学中，有很多传统的破译密码的经典方法，例如凯撒密码，RSA 算法等，但传统的破译方法花费时间长，安全性低，存在很大的弊端，且传统的算法都是基于密文中含有标点符号的情况下进行解密的[2-3]。因此我们在第一阶段算法的基础上进行改进，通过频率分析先一一对应，破译一部分密文后，发现一些明文字母会连在一起，通过对英语造词的规律我们依此进行部分断句。在此基础上，再通过英语语法，词缀等独有的特征来进行断句，判别。使其能够在没有符号，传输过程中存在丢失，篡改，添加等干扰情况下进行高效率，高精度的破译。

问题二的分析：

密文通过一个带有噪声干扰的信道传输，比如密码在传输过程中发生被篡改，添加，缺失，我们基于问题一的基础上，对问题二进行分析。

可假设一个由 n 个字母组成的单词中有一个字母发生缺失或添加，则我们在破解密文时就会在 n 个字母的单词词库中寻找，若搜寻不到，则去 n-1 或 n+1 个词库中查找。假设 n 个字母的词库有q_1个单词,n-1 个字母的词库有q_2个单词，n+1 个字母的词库有q_3个单词，那么正确的概率是$\frac{1}{q_1 \times q_2}$或$\frac{1}{q_1 \times q_3}$，显然该密文正确的概率不高。若密码在传输过程中发生篡改。假设该字母由 n 个字长组成，其中有 i 个字母发生篡改，降低了准确率，最小的错误率为$\frac{n-i}{n}$，100%。

三．模型假设

1.假设明文是由现代通常使用的英语写成的。

2.假设密码表仅是针对 26 个字母的，每个单词之间的空格，以及标点符号全部删除。

四．算法说明与求解

4.1 算法说明

在第一阶段，密文在传输过程中所有的标点，空格都是保留的，因此通过对单个字母，双字母，三个字母及多个字母的频率分析以及英语语法，造句，词缀的了解就可以破译部分密文。但此方法并不适用于没有标点，空格且传输过程中会发生篡改，缺失，添加的破译过程。下面给出算法：

Step1.频率分析，判断密文字母出现的频率是否大于等于 5% ，若是，则转入 step2，否则转入 step7.

Step2.将 step1 得到的结果与正常频率表中的字母一一对应，替换输出

Step3.找出含有 step2 的字段并计算双字长的组合

Step4.将 step3 算出的概率的组合与常用的双字母组合概率进行比较对应并替换，输出明文

Step5.判断能否断句。若能，则转入 step6，否则转入 step1

Step6.根据英语语法，用词习惯，前后文进行断句，输出

Step7.根据英语语法，构词规则，短语搭配等进行破译

Step8.输出

Step9.退出

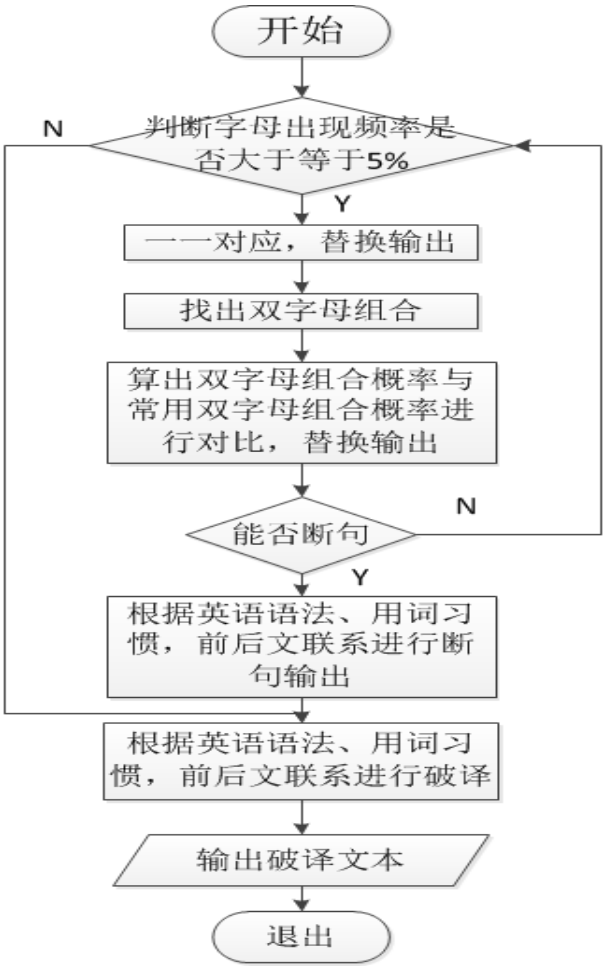

图 1. 算法流程图

比如，在经过频率分析破译部分密文后，我们发现一些明文字母会连在一起。此时，我们可以根据英语单词的构造法判断出一些字母是不可能连在一起作为词尾的，比如**sss**，在英语单词中没有以 **sss** 结尾或开头的单词所以**|sss** 是不 可以截断的，同样常用单词表中也没有以 **ss** 开头的单词，所以 **ss|s** 也不可以，由此我们可以判定截断处在 **ss|s**。依次类推，我们可以在部分字母处进行截断。由此假设我们得到了 n 个截断|，现在我们分析第 i-1 个截断|与第 i 个截断|之间的明密文。以**|ABsCDDEss|**为例。首先在词库表中筛选词尾是 **ss** 且词中含有 **s** 的单词，发现不存在该种形式的单词，则进行下一步判断。将|向后平移一格，判断| **BesCDDEss|**，我们发现依旧没有这种类型的单词，则继续进行该步骤，最终可以得到**|success|**，将此单词保留，继续循环上述步骤。

4.2 算法求解

1. 找出 26 个字母的个数，算出频率，如图二，然后对照各字母频率分布表，如图三，得出 L----e,将密文中的 L 全部替换成 e。

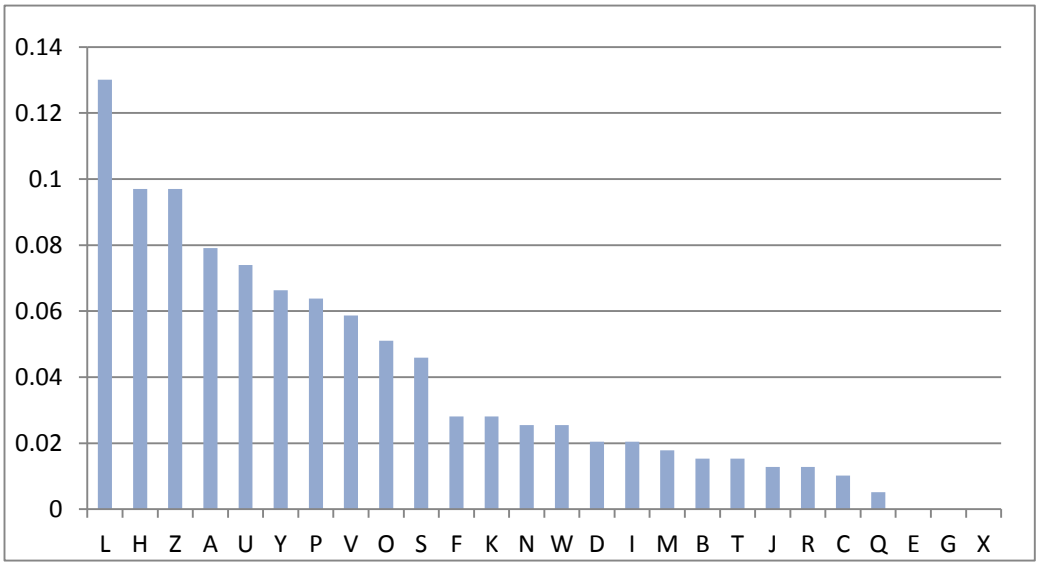

图 2. 密文中各字母的出现频率

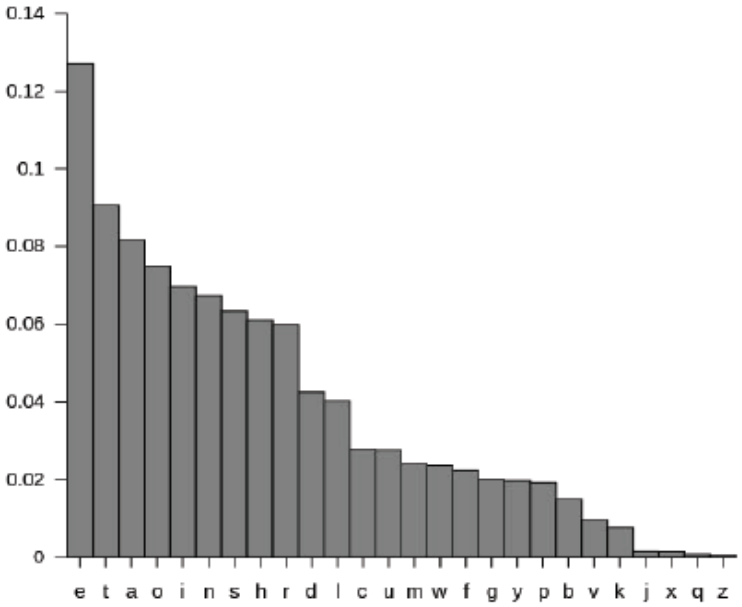

图 3. 单字母出现频率表

TFHBUAQeUUPMeYPZHUHJAYeZZZOeTBZAIeHASeHZAAOPYAFMPCeFeHYVSKP
UZWPAeVMAOPZZOeVMAeUHWWeHYZVUAOeZAHNeHZHFVBUNNPYSQeUUPMe
YDPSSOHCeAVAHReWHYAPUHUeDWSHFZVVUAOPZAPTeZOeDPSSIeHNPYSVMZ
eCeUAeeUPUAOeWSHFZOeTBZAHWWeHYPUHIYPNOAYeKKYeZZHUKSVUNISHJR
ZAVJRPUNZSHZAFeHYPUHUVAOeYWSHFZOeOHKAVDeHYZOVYAZVJRZHUKHI
YPNOAVYHUNeJVSVBYeKKYeZZPMHUFVUeeCeUHZReKOeYOVDVSKZOePZZOe
HSDHFZHUZDeYZTFKeHZPATZAIeAeYYPISeAVIeNYVDUBW

2. 在第一步中，我们已经知道了密文 L 对应的明文是 e，下面我们查找出含有 e 的词段，即如表一。

表一：含 e 的字段

QeU	MeY	YeZ	CeFeH	IeH	SeH	OeT
AeV	OeV	AeU	QeU	OeZ	NeH	WeH
MeY	CeA	ReW	IeH	TeZ	OeD	UeD
OeW	OeT	WeH	ZeCeU	YeZ	FeH	YeK
OeY	OeO	DeH	AeeU	YeK	YeZ	NeJ
CeU	ReK	OeY	UeeC	OeH	DeY	OeP
KeH	SeA	IeN	IeAeY			

得出‘*e’、‘e*’的各种频率如图 4.

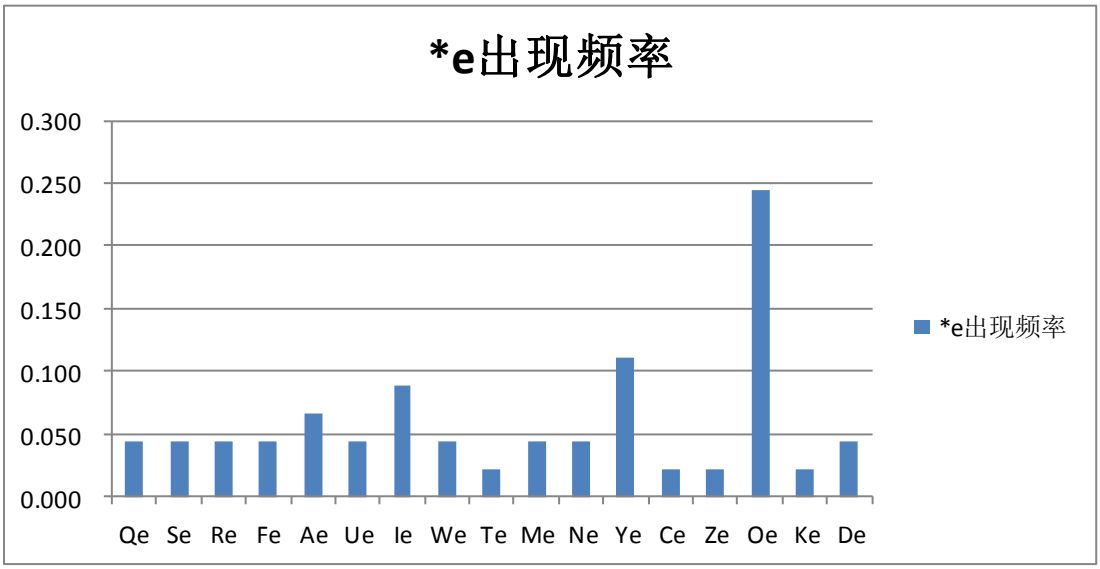

图 4. *e 出现的频率

然后对照双字母组合频率前十的图五

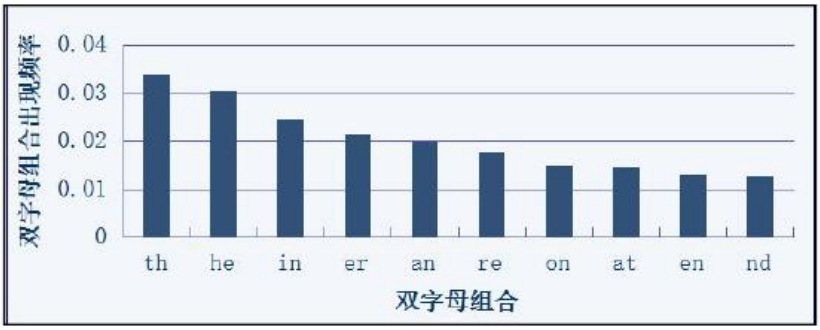

图 5. 双字母组合频率表

可知 Oe 对应的明文应该是 he，Ye 对应的明文是 re。，即 **O----h、Y----r**。

查找出含有 h 的词段，如表二。

表二：含 h 的字段

FHB	ZHUHJ	**e**HA	**e**HZ	**eHr**	UHW
AHN	**e**HZHF	hHC	AHR	WH**r**	UHU
eHN	SHF	AHW	**eHr**	UHI	ZHU
SHZ	**eHr**	UHU	SHF	hHK	**eHr**
KHI	**r**HU	MHU	UHZ	**e**HS	DHF
eHZ	**eHr**	SHF	SHJ	ZHU	ZHU

得到*h 的频率分布图，图 6

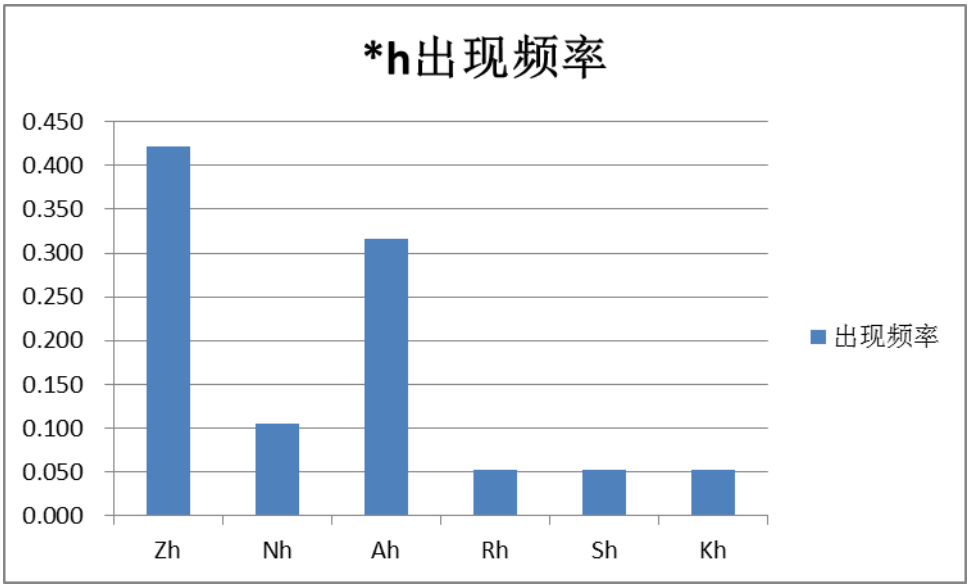

图 6. *h 出现频率

再结合之前的密文中单个字母的频率，和高频单个字母表对应，进行分析可以知道 **Z----s、A----t**。

现在将 O、Y、Z、A 进行替换，可以得到部分明文，即

TFHBUtQeUUPMerPsHUHJtressheTBstIeHtSeHstthPrtFMPCeFeHrVSKPUsWPteVMthPs
sheVMteUHWWeHrsVUthestHNeHsHFVBUNNPrSQeUUPMerDPSShHCetVtHReWHrtPU
HUeDWSHFsVVUthPstPTeshеDPSSIeHNPrSVMseCeUteeUPUtheWSHFsheTBstHWWeHr
PUHIrPNhtreKKressHUKSVUNISHJRstVJRPUNsSHstFeHrPUHUVtherWSHFshehHKtVD
eHrshVrtsVJRsHUKHIrPNhtVrHUNeJVSVBreKKressPMHUFVUeeCeUHsReKherhVDVS
KshePssheHSDHFsHUsDersTFKeHsPtTstIeterrPISetVIeNrVDUBW

我们对现在的密文分析可以发现一个字段“shePsshe”，根据已知的英语语法的知识，我们可以对其进行部分断句，即“she|Ps|she”，再查询常用词汇表，可以知道 Ps 为 as、is，即 p 对应 a、s。

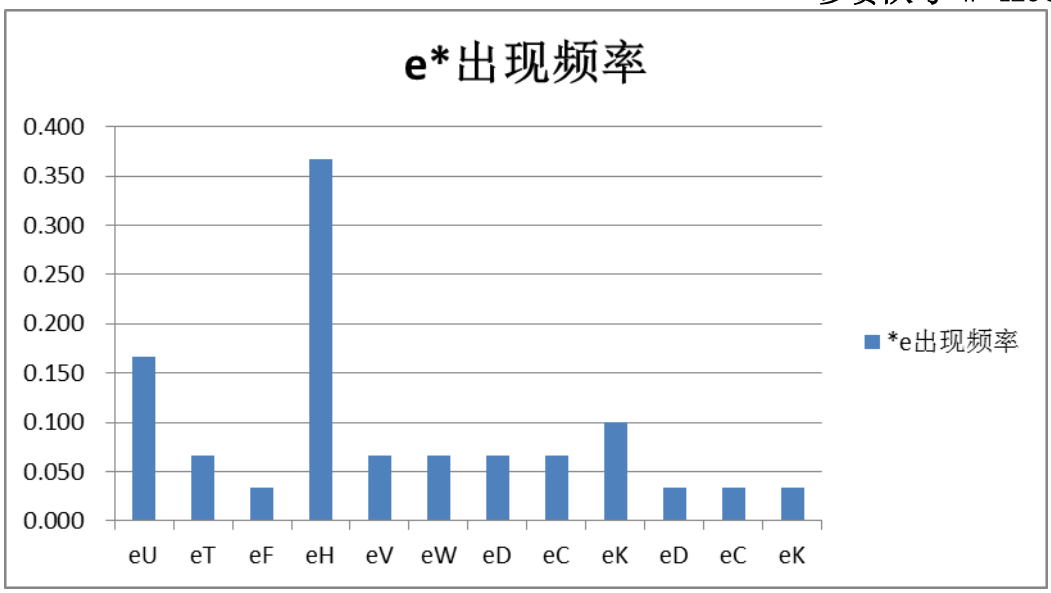

图 7.　e*出现的频率

根据 e*出现的频率次数，如图，再结合单个字母出现的频率，　　可知，H、U----a、n，但是不能一一对应。所以，要结合之前的密文 P，进行二次判断。

对密文进行分析，我们发现字段“sHUHJtressshe”分别带进去 ana*tress、nan*tress，由已知的语法知识及相关的构词法可以知道这两种都是不符合的，换言之，我们可以知道，**U----n、H----a、P----i**。

将 U、H、P 这三个密文进行替换可以得到：

TFaBntQenniMerisanaJtressstheTBstIeatSeastthirtFMiCeFearVSKinsWiteVMthissheVMtenaWWearsVnthestaNeasaFVBnNNirSQenniMerDiSShaCetVtaReWartinaneDWSaFsVVnthistiTesheDiSSIeaNirSVMseCenteenintheWSaFsheTBstaWWearinaIriNhtreKKressanKSVnNISaJRstVJRinNsSastFearinanVtherWSaFshehaKtVDearshVrtsVJRsanKaIriNhtVranNeJVSVBreKKressiManFVneeCenasReKherhVDVSKsheissheaSDaFsansDersTFKeasitTstIeterriISetVIeNrVDnBW

3. 在之前破译的基础上，我们注意到有部分字段的密文已经被破译的很多了，只有中间的 1 到 2 个密文字母没有被破译，而我们就可以将其作为破译的关键所在，即对其进行语法结构、语境、单词构造等方面进行逐层分析，最后得到密文所对应的明码。

对之前的密文进行分析，我们应注意到“**erisanaJtress**”，这段字母中只有“J”是密文，其余都已经全部转换成明码了，所以，我们现在就要依据已有的知识对其进行尝试性的断句，根据英语构词规则，知道“tress”为单词的后缀，所以我们可以尝试一下几种情况：“**erisana，Jtress**”、“**erisan，aJtress**”、“**erisa，naJtress**”、“**eris，anaJtress**”，接着去常用单词表中进行比对查询，我们可以确定单词为“actress”,即可以知道 J----c，而“**erisanactress”，**就应该断为**“er，is，an，actress”。**

我们还发现了字段“**seCenteeninthe”，**和之前的那个字符串类似，这个字符串中也只有一个字母“C”没有被破译，据我们的语感和对语法的熟悉程度，以及知道“teen”通常是作为后缀出现的，可以对其进行断句，可得以下几种情况：“**se，Centeen，in，the**”、“**s，eCenteen，in，the**”、“**seCenteen，in，the**”，再次查询常用词汇表，我们可以进行分析得出此字段为“**seventeeninthe**”，

对其进行断句即为“**seventeen，in，the**”。即 C----v。

对于字段“**thistiTeshe**”，我们可以采取同样的方法对其进行尝试性断句，再根据已知的词汇查询表，得出明码，“**this，tiTe，she**”、“**thi，stiTe，she**”、“**th，istiTeshe**”、“**t，histiTeshe**”等情况，而根据词汇表的对照，我们可以轻易地知道密文 T 对应的明文是

m,即 T----m，断句为“**this，time,she**”。

对于密文“**neevenas**R**edherh**”，同样的中间只有一个是未知的，在学习这么多年的英语基础上，我们一眼就可以判定“**even**”是一个单词，“**her**”应该也是一个单词，而“**ed**”通常是作为后缀的，也差不多可以判定“**as**R**ed**”是一个单词，经过这样的分析，我们可以快速的得到密文 R 对应的明文是 k，即 R----k,断句为“**ne,even,asked,her,h**”。

与之前的那个字段类似，对于这个字段“**have**t**V**t**ake**”，我们一眼看出“**have**”、“**take**”，这两个单词，那么意味着“**t**V”必为一个单词，而查询常用字母表，我们发现符合条件的只有“**to**”,即密文 V 对应的是 o，V----o，断句为“**have,to,take**”。

还有字段“**shem**B**st**”，和之前的理论一样，我们可以得到 B 对应的明码是 u，即 B----u，“**she，must**”。

同理可以得到字段“**m**F**aunt**”中，F 对应的应该是 m，即 F----y，“**my,aunt**”。

之前的都是对一段中只有一个位置密文，现在是两个未知密文，即字段“**re**KK**ress**”，我们发现这两个密文是同样的，所以我们可以对其进行断句“**re**KK**ress**”、“**re**KK，**ress**”、“**re**K，K**ress**”，查询常用单词，我们发现不存在“**re**KK**ress**”这样的单词或字段，又“reKK”此字段只能是“ress”,又 s 已经被破译出，所以此种情况也不可能，只剩下最后一种情况，而对其进行查询，发现只有将“d”代入才符合，所以，得到结论，K----d。 我们发现了这样的字段“**ayshemusta**WW**earina**”，虽然对于这个我们不好判断，但是我们还知道一个条件就是 W 是在文章的最末尾，而在之前我们就已经知道了 j、q、v、z，不能放在句末，再结合已经破译出来的密文，我们发现 W 的可能性为 b、g、l、p、w、x，现在就是将这几个字母一一代入进行判断，发现符合条件的只有“**ayshemustappearina**”，即 W 对应的是 p,W----p,“**ay，she，must，appear，ina**”。

将 J----c、C----v、K----d、T----m、R----k、V----o、B----u、F----y、W----p 带入密文中，可以得到以下的密文：

myauntQenniMerisanactressshemustIeatSeastthirtyMiveyearoSdinspiteoMthissheoMtenappearsonthestaNeasayounNNirSQenniMerDiSShavetotakepartinaneDpSaysoonthistimesheDiSSIeaNirSoMseventeeninthepSayshemustappearinaIriNhtreddressandSonNISackstockinNsSastyearinanotherpSayshehadtoDearshortsocksandaIriNhtoranNecoSoureddressiManyoneevenaskedherhoDoSdsheissheaSDaysansDersmydeasitmstIeterriISetoIeNroDnup

4. 现在我们已经破译了大部分的密文，好多长的字段也都全部被解码，因此，可以对此时的密文进行断句，将已知的单词可根据语法、单词构造、前后文联系之类的对余下的密文再次分析破解。

密文可以断分如下：

My aunt QenniMer is an actress she must IeatSeast thirty MiveyearoSd in spite oMthis she oMten appears on the staNea say ounNNirS QenniMer DiSS have to take part in aneDpSay soon this time she DiSSIeaNirSoM seventeen in the pSay she must appear in aIriNht red dress and SonNISackstockinNsSast year in another pSay she had to Dear short socks and aIriNht or anNecoSou red dress iManyone even asked her hoDoSd she is she aSDaysansDers my deas it must IeterriISe to IeNroDnup

从上段短文中，我们可以发现有些常用的单词短语的，我们就可以从这方面入手，对其进行分析，破译密文，进而进行全文的破译。

字段“in spite oMthis”，依照固定搭配，我们立刻想到短语“in spite of”,也没有其他的符合这个语境，所以可得到 M----f，断句即为“in spite of this”。我们还发现，“pSay”，根据前后文的意思基本可以判定这是一个单词，查询常用词汇表，发现这是 S 只可能是 l 和 r，又因为 r 已经破译出来了，所以 S----l，即为“play”。再者就是“Ie”，查询之后，我们发现 I 对应的是 i 或者 b，又因为 i 被破译出，那么 I 只能对应 b，I----b，单词为“be”。对于字段“abriNht”不能像之前那样一次性就看出结果，此刻我们要对其进行断分“a，

briNht”、“ab，riNht”、“abr，iNht”等几种可能性，而查询常用词汇表，对比，发现符合条件的只有一种可能性“a bright”，即 N----g。对于字段“aneD”，对其断分，可以知道“a，neD”、“an，eD”、“ane，D”，进行对比分析，再结合之前我们已经判断出来的结果，可知 D----w，即“a new”。

将 M----f、S----l、I----b、N----g、D----w 代入密文进行破译，并且对此再次进行断分，得到：

my aunt Qennifer is an actress she must be at least thirty five year old in spite of this she often appears on the stage as a young girl Qennifer will have to take part in a new play soon this time she will be a girl of seventeen in the play she must appear in a bright red dress and long black stockings last year in another play she had to wear short socks and a bright orange coloured dress if any one even asked her how old she is she always answers my dear it must be terrible to be grown up

5. 在前三步中，对于原始的密文中的 23 个单词，我们已经破译出了 22 个，还有一个为破译出，而我们根据已有的明文，猜测字段“Qennifer”是人名，不是单词或者单词短语，所以，我们很难判断出 Q 所对应的明码是什么，而我们确实也没有办法对其做准确的判断，因为不管哪一个字母代入都是有理由成立的。因此，对于此篇密文，我们只有 Q 无法得出正确明码。

最终的密文可以破译为：

my aunt Qennifer（人名） is an actress she must be at least thirty five year old in spite of this she often appears on the stage as a young girl Qennifer（人名） will have to take part in a new play soon this time she will be a girl of seventeen in the play she must appear in a bright red dress and long black stockings last year in another play she had to wear short socks and a bright orange coloured dress if any one even asked her how old she is she always answers my dear it must be terrible to be grown up.

但是在实际密码传输过程中会出现**篡改**，**丢失**，**添加**的问题。

1. 发生篡改的情况

密文在传输过程中发生篡改，即某些字母会发生替换，这样整篇文章篡改的字母都会发生改变。下面以 boy 为例。

假设正确的明文对应是表一.

表一：正确的明文对应表

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A

明文 boy 在加密后变为 CPZ，现在 C 发生篡改后，密文变为 UPZ.假设 PZ 已经破译是 oy。那么此时我们查询三字库，发现符合条件的单词有 boy,joy,toy，即 U-- $\begin{cases} b \\ j \\ t \end{cases}$。而原文加密是 C--b，最小错误率为 1.89%。以此类推，n 个字长的单词，篡改 i 个单词后，错误率即为 $\frac{n-i}{n}$，100% 。

2. 发生缺失的情况

缺失，就是原来 n 个字长的字母在传输过程中变为 n-i 个字长的字母。（i 表示缺失的字长）以 sing 为例。

明文 sing 加密后变为 TJOH，现在 g 缺失，密文变为 TJO。此时是三个字长，我们通过查找三字库再根据英语语法结构，单词构造破译出该密文。但此时破译出的明文定

与正确的不符，错误率是 11.54%。假设 n 个字母的词库有q_1个单词,n-1 个字母的词库有q_2个单词，n+1 个字母的词库有q_3个单词，正确的概率可表示为$\frac{1}{q_1 \times q_2}$或$\frac{1}{q_1 \times q_3}$。

五．模型优缺点

优点：

1、本算法脱离了传统方法，大胆尝试，结合多种算法，将主体较大的检索目标转化为一级一级较小的分目标，然后对分目标进行排序，告别了以往破译软件对加密文档内容无选择的排列匹配的冗杂过程。

2、本算法考虑到了英语语法、单词构造，短语搭配等的影响，有效排除了明显错误的结果，大大提高了准确率，完成了破译算法的根本目的，具有一定实用性。

3、本题中考虑到去除标点空格的情况，应此得到的算法对复杂难解的密文有一定的破解性。

缺点：

1、未设置电脑程序，采用人工现实法演绎的实例比较简单，对普遍性缺乏说服力。

2、题目中的前提假设有局限性，在现实生活中已渐渐被淘汰，因此，本算法还需要完善的地方。

3、对于题目中的干扰设置，我们在算例中模拟出丢失、篡改、添加的情况，并与最后结果做了比较。但是无法确定干扰对最后结果的影响。

六．参考文献

[1]陈华友　周礼刚 刘金培，《数学模型与数学建模》[M]北京：科学出版社，2014
[2]刘来福 曾文艺，《数学模型与数学建模》[M]北京师范大学出版社，1997
[3]范九伦 刘宏月，《密码学原理》[M]西安电子科技大学出版社,2008