科展企劃書

以自動化頻率分析解密以代換密碼加密之密文

11817莊舒涵

1. 科別

電腦與資訊學科

1. 研究動機

在閱讀密碼學相關資料時，了解到古典密碼學的單一字元替代密碼可以用字頻分析（Frequency analysis）來解密，想寫出一個可以自動使用字頻分析解密文章的解碼器。

1. 預計研究方向
2. 撰寫程式，以單個字元(字母)的頻率分析來解碼，計算出正確率
3. 在上述的基礎加上兩個字元為一組的字元組的頻率分析交叉比對，比較正確率
4. 加入更多個字元為一組(三及以上)的頻率分析
5. 探討無空格及標點符號的密文在解碼時如何自動加入空格以分隔單字
6. 探討是否能利用自然語言處理相關技術使解碼正確率更高
7. 預期研究結果

撰寫出一個能自動進行字頻分析並解密密文的程式，比較是否有更多字元的字元組(2個字元一組、3個字元一組等)之頻率分析對於解密的影響，找出能使正確率最高的分析方式，。

1. 文獻探討
2. 單一字元替代密碼

字母一對一代換，但不一定有規律

舉例:

密碼表:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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** |
| T | N | G | S | U | V | W | X | Y | Z | A | B | C | D | E | F | H | I | J | K | L | M | O | P | Q | R |

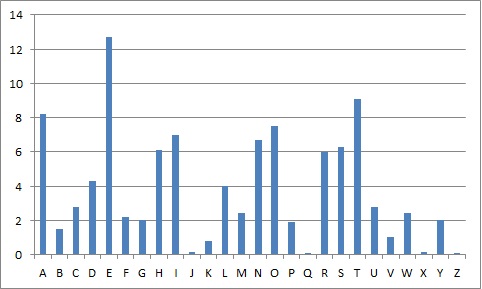
明文:WHERE ARE YOU

密文:OXUIU TIU QEL

1. 字頻分析

在英文文章中，不同字母的使用頻率不同，所以可以利用密文中各個字母出現的頻率來破解所代表的字母。也有一些字元組(由兩個或以上之字元組成)的出現頻率較高，也可以用來分析。

下表為各個字母出現的頻率



1. 自然語言處理(Natural Language Processing，NLP)
2. 自然語言處理是一種透過複雜的數學模型及演算法來讓機器去認知、理解並運用我們的語言的技術。
3. 早期的技術主要基於統計的概念去訓練模型，讓演算法閱讀大量類似字典的文章段落，再讓演算法計算單字、句子出現的機率，然而此種方式無法使系統很好地辨識複雜的文法，同時，這樣子的模型所產生的字句可能生硬且結構錯亂。
4. 隨著深度學習與演算法模型的突破，新的訓練方式已能更好的處理以上所提的問題。
5. 此技術與我的題目的相關:可能可以運用此技術使解碼出的結果更接近自然語言(我們會用的語言)，使結果更接近原本的明文
6. 初步研究結果
7. 加密器

為製造測試資料(以下簡稱測資)，使用此程式將文章使用單一字元代換密碼加密。

1.密碼表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |

2.程式碼(使用C++)

*//向右平移1字母*

#include<iostream>

#include<string>

using namespace std;

int main(){

string s;

int i=0;

getline(cin,s);

while(s[i]!='\0'){

if(s[i]>='a' && s[i]<'z'){

cout<<char(s[i]+1);

}

else if(s[i]>='A' && s[i]<'Z'){

cout<<char(s[i]+1);

}

else if(s[i]=='z'){

cout<<'a';

}

else if(s[i]=='Z'){

cout<<'A';

}

else{

cout<<s[i];

}

i++;

}

return 0;

}

3.產生之測資範例(僅節錄部分，完整文章附於附錄)

|  |  |
| --- | --- |
| 明文 | 密文 |
| THE INFORMATION BELOW IS TAKEN FROM MACTUTOR. ALAN TURING WAS BORN AT PADDINGTON, LONDON. HIS FATHER, JULIU MATHISON TURING, WAS A BRITISH MEMBER OF THE INDIAN CIVIL SERVICE AND HE WAS OFTEN ABROAD. ALAN'S MOTHER, ETHEL SARA STONEY, WAS THE DAUGHTER OF THE CHIEF ENGINEER OF THE MADRAS RAILWAYS AND ALAN'S PARENTS HAD MET AND MARRIED IN INDIA. WHEN ALAN WAS ABOUT ONE YEAR OLD HIS MOTHER REJOINED HER HUSBAND IN INDIA, LEAVING ALAN IN ENGLAND WITH FRIENDSOF THE FAMILY. | UIF JOGPSNBUJPO CFMPX JT UBLFO GSPN NBDUVUPS. BMBO UVSJOH XBT CPSO BU QBEEJOHUPO, MPOEPO. IJT GBUIFS, KVMJV NBUIJTPO UVSJOH, XBT B CSJUJTI NFNCFS PG UIF JOEJBO DJWJM TFSWJDF BOE IF XBT PGUFO BCSPBE. BMBO'T NPUIFS, FUIFM TBSB TUPOFZ, XBT UIF EBVHIUFS PG UIF DIJFG FOHJOFFS PG UIF NBESBT SBJMXBZT BOE BMBO'T QBSFOUT IBE NFU BOE NBSSJFE JO JOEJB. XIFO BMBO XBT BCPVU POF ZFBS PME IJT NPUIFS SFKPJOFE IFS IVTCBOE JO JOEJB, MFBWJOH BMBO JO FOHMBOE XJUI GSJFOETPG UIF GBNJMZ. |

1. 以單個字元(字母)的頻率分析來解碼
2. 英文字母在文章中出現頻率(由低到高): zqxjkvbpygfwmucldrhsnioate
3. 程式碼

#include<iostream>

#include<string>

#include<algorithm>

#include<utility>

using namespace std;

char correspond[30];

int main(){

string s,letterFrequency="zqxjkvbpygfwmucldrhsnioate";

int i;

pair<int,char>letter[30]; *//(字母出現次數,字母)*

for(i=0;i<26;i++){

letter[i]={0,'a'+i};

}

getline(cin,s);

i=0;

while(s[i]!='\0'){

if(s[i]>='a' && s[i]<='z'){ *//在a~z*

letter[s[i]-'a'].first+=1;

}

else if(s[i]>='A' && s[i]<='Z'){ *//在A~Z*

letter[s[i]-'A'].first+=1;

}

i++;

}

sort(letter,letter+26); *//根據字頻排序*

for(i=0;i<26;i++){ *//將密文中的字母依照字頻連結(解碼)*

correspond[letter[i].second-'a']=letterFrequency[i]; *//例如z->a:correspond[25]='a'*

}

i=0;

while(s[i]!='\0'){

if(s[i]>='a' && s[i]<='z'){ *//在a~z*

cout<<correspond[s[i]-'a']; *//輸出解碼後的字母*

}

else if(s[i]>='A' && s[i]<='Z'){ *//在A~Z*

cout<<char(correspond[s[i]-'a'-'A'+'a']-'a'+'A');

}

else{

cout<<s[i]; *//非字母不更動*

}

i++;

}

return 0;

}

1. 測試結果

(測試時以整篇密文測試，此處為節錄，完整解密結果附於附錄)

|  |  |
| --- | --- |
| 明文 | 解密後 |
| THE INFORMATION BELOW IS TAKEN FROM MACTUTOR. ALAN TURING WAS BORN AT PADDINGTON, LONDON. HIS FATHER, JULIU MATHISON TURING, WAS A BRITISH MEMBER OF THE INDIAN CIVIL SERVICE AND HE WAS OFTEN ABROAD. ALAN'S MOTHER, ETHEL SARA STONEY, WAS THE DAUGHTER OF THE CHIEF ENGINEER OF THE MADRAS RAILWAYS AND ALAN'S PARENTS HAD MET AND MARRIED IN INDIA. WHEN ALAN WAS ABOUT ONE YEAR OLD HIS MOTHER REJOINED HER HUSBAND IN INDIA, LEAVING ALAN IN ENGLAND WITH FRIENDSOF THE FAMILY. | THE ONGIRUATOIN YELIM OS TAKEN GRIU UACTWTIR. ALAN TWRONF MAS YIRN AT PADDONFTIN, LINDIN. HOS GATHER, XWLOW UATHOSIN TWRONF, MAS A YROTOSH UEUYER IG THE ONDOAN COVOL SERVOCE AND HE MAS IGTEN AYRIAD. ALAN'S UITHER, ETHEL SARA STINEB, MAS THE DAWFHTER IG THE CHOEG ENFONEER IG THE UADRAS RAOLMABS AND ALAN'S PARENTS HAD UET AND UARROED ON ONDOA. MHEN ALAN MAS AYIWT INE BEAR ILD HOS UITHER REXIONED HER HWSYAND ON ONDOA, LEAVONF ALAN ON ENFLAND MOTH GROENDSIG THE GAUOLB. |

醒目提示為不同處

可以清楚看到，以單一字母字頻分析來解密的正確性仍然很低，所以要加入其他分析方法以增加正確性。

1. 指導教授
2. 參考資料
3. <https://hitcon.org/2018/CMT/slide-files/d1_s2_r4.pdf>
4. <https://crypto.interactive-maths.com/frequency-analysis-breaking-the-code.html>
5. <https://zh.oosga.com/pillars/nlp/>
6. 附錄:測試資料

明文

THE INFORMATION BELOW IS TAKEN FROM MACTUTOR. ALAN TURING WAS BORN AT PADDINGTON, LONDON. HIS FATHER, JULIU MATHISON TURING, WAS A BRITISH MEMBER OF THE INDIAN CIVIL SERVICE AND HE WAS OFTEN ABROAD. ALAN'S MOTHER, ETHEL SARA STONEY, WAS THE DAUGHTER OF THE CHIEF ENGINEER OF THE MADRAS RAILWAYS AND ALAN'S PARENTS HAD MET AND MARRIED IN INDIA. WHEN ALAN WAS ABOUT ONE YEAR OLD HIS MOTHER REJOINED HER HUSBAND IN INDIA, LEAVING ALAN IN ENGLAND WITH FRIENDSOF THE FAMILY. ALAN WAS SENT TO SCHOOL BUT DID NOT SEEM TO BE OBTAINING ANY BENEFIT SO HE WAS REMOVED FROM THE SCHOOL AFTERA FEW MONTHS. NEXT HE WAS SENT TO HAZLEHURST PREPARATORY SCHOOL WHERE HE SEEMED TO BE AN 'AVERAGE TO GOOD' PUPILIN MOST SUBJECTS BUT WAS GREATLW TAKEN UP WITH FOLLOWING HIS OWN IDEAS. HE BECAME INTERESTED IN CHESS WHILE AT THIS SCHOOL AND HE ALSO JOINED THE DEBATING SOCIETY. HE COMPLETED HIS COMMON ENTRANCE EXAMINATION IN 1926 AND THEN WENT TO SHERBORNE SCHOOL. NOW 1926 WAS THE YEAR OF THE GENERAL STRIKE AND WHEN THE STRIKE WASIN PROGRESS TURING CYCLED 60 MILES TO THE SCHOOL FROM HIS HOME, NOT TOO DEMANDING ATASK FOR TURING WHO LATER WAS TO BECOMEA FINE ATHLETE OF ALMOST OLYMPIC STANDARD. HE FOUND IT VERY DIFFICULT TO FIT INTO WHAT WAS EXPECTED AT THIS PUBLIC SCHOOL, YET HIS MOTHER HAD BEEN SO DETERMINED THAT HE SHOULD HAVEA PUBLIC SCHOOL EDUCATION. MANW OF THE MOST ORIGINAL THINKERS HAVE FOUND CONVENT IONAL SCHOOLING AN ALMOST INCOMPREHENSIBLE PROCESS AND THIS SEEMS TO HAVE BEEN THE CASE FOR TURING. HIS GENIUS DROVE HIM IN HIS OWN DIRECTIONS RATHER THAN THOSE REQUIRED BY HIS TEACHERS. HE WAS CRITICISED FOR HIS HANDWRITING, STRUGGLED AT ENGLISH, AND EVEN IN MATHEMATICS HE WAS TOO INTERESTED WITH HIS OWN IDEAS TO PRODUCE SOLUTIONS TO PROBLEMS USING THE METHODS TAUGHT BY HIS TEACHERS. DESPITE PRODUCING UNCONVENT IONAL ANSWERS, TURING DID WIN ALMOST EVERY POSSIBLE MATHEMATICS PRIZE WHILE AT SHERBORNE. IN CHEMISTRW. A SUBJECT WHICH HAD INTERESTED HIM FROMA VERY EARLW AGE. HE CARRIED OUT EXPERIMENTS FOLLOWING HIS OWN AGENDA WHICH DID NOT PLEASE HIS TEACHER. TURING'S HEADMASTER WROTE.IF HE IS TO STAY AT PUBLIC SCHOOL, HE MUST AIM AT BECOMING EDUCATED. IF HE IS TO BE SOLELW A SCIENTIFIC SPECIALIST, HE IS WASTING HIS TIME AT A PUBLIC SCHOOL. THE FIRST PERSON WHO TELLS THE KEY WORD TO THE OWNER OF THE ZIZAIXUAN TEA SHOP WILL GETA COUPON OF ONE HUNDRED TAIWAN DOLLARS.

密文

UIF JOGPSNBUJPO CFMPX JT UBLFO GSPN NBDUVUPS. BMBO UVSJOH XBT CPSO BU QBEEJOHUPO, MPOEPO. IJT GBUIFS, KVMJV NBUIJTPO UVSJOH, XBT B CSJUJTI NFNCFS PG UIF JOEJBO DJWJM TFSWJDF BOE IF XBT PGUFO BCSPBE. BMBO'T NPUIFS, FUIFM TBSB TUPOFZ, XBT UIF EBVHIUFS PG UIF DIJFG FOHJOFFS PG UIF NBESBT SBJMXBZT BOE BMBO'T QBSFOUT IBE NFU BOE NBSSJFE JO JOEJB. XIFO BMBO XBT BCPVU POF ZFBS PME IJT NPUIFS SFKPJOFE IFS IVTCBOE JO JOEJB, MFBWJOH BMBO JO FOHMBOE XJUI GSJFOETPG UIF GBNJMZ. BMBO XBT TFOU UP TDIPPM CVU EJE OPU TFFN UP CF PCUBJOJOH BOZ CFOFGJU TP IF XBT SFNPWFE GSFN UIF TDIPPM BGUFSB GFX NPOUIT. OFYU IF XBT TFOU UP IBAMFIVSTU QSFQBSBUPSZ TDIPPM XIFSF IF TFFNFE UP CF BO 'BWFSBHF UP HPPE' QVQJMJO NPTU TVCKFDUT CVU XBT HSFBUMX UBLFO VQ XJUI GPMMPXJOH IJT PXO JEFBT. IF CFDBNF JOUFSFTUFE JO DIFTT XIJMF BU UIJT TDIPPM BOE IF BMTP KPJOFE UIF EFCBUJOH TPDJFUZ. IF DPNQMFUFE IJT DPNNPO FOUSBODF FYBNJOBUJPO JO 1926 BOE UIFO XFOU UP TIFSCPSOF TDIPPM. OPX 1926 XBT UIF ZFBS PG UIF HFOFSBM TUSJLF BOE XIFO UIF TUSJLF XBTJO QSPHSFTT UVSJOH DZDMFE 60 NJMFT UP UIF TDIPPM GSPN IJT IPNF, OPU UPP EFNBOEJOH BUBTL GPS UVSJOH XIP MBUFS XBT UP CFDPNFB GJOF BUIMFUF PG BMNPTU PMZNQJD TUBOEBSE. IF GPVOE JU WFSZ EJGGJDVMU UP GJU JOUP XIBU XBT FYQFDUFE BU UIJT QVCMJD TDIPPM, ZFU IJT NPUIFS IBE CFFO TP EFUFSNJOFE UIBU IF TIPVME IBWFB QVCMJD TDIPPM FEVDBUJPO. NBOX PG UIF NPTU PSJHJOBM UIJOLFST IBWF GPVOE DPOWFOU JPOBM TDIPPMJOH BO BMNPTU JODPNQSFIFOTJCMF QSPDFTT BOE UIJT TFFNT UP IBWF CFFO UIF DBTF GPS UVSJOH. IJT HFOJVT ESPWF IJN JO IJT PXO EJSFDUJPOT SBUIFS UIBO UIPTF SFRVJSFE CZ IJT UFBDIFST. IF XBT DSJUJDJTFE GPS IJT IBOEXSJUJOH, TUSVHHMFE BU FOHMJTI, BOE FWFO JO NBUIFNBUJDT IF XBT UPP JOUFSFTUFE XJUI IJT PXO JEFBT UP QSPEVDF TPMVUJPOT UP QSPCMFNT VTJOH UIF NFUIPET UBVHIU CZ IJT UFBDIFST. EFTQJUF QSPEVDJOH VODPOWFOU JPOBM BOTXFST, UVSJOH EJE XJO BMNPTU FWFSZ QPTTJCMF NBUIFNBUJDT QSJAF XIJMF BU TIFSCPSOF. JO DIFNJTUSX. B TVCKFDU XIJDI IBE JOUFSFTUFE IJN GSPNB WFSZ FBSMX BHF. IF DBSSJFE PVU FYQFSJNFOUT GPMMPXJOH IJT PXO BHFOEB XIJDI EJE OPU QMFBTF IJT UFBDIFS. UVSJOH'T IFBENBTUFS XSPUF.JG IF JT UP TUBZ BU QVCMJD TDIPPM, IF NVTU BJN BU CFDPNJOH FEVDBUFE. JG IF JT UP CF TPMFMX B TDJFOUJGJD TQFDJBMJTU, IF JT XBTUJOH IJT UJNF BU B QVCMJD TDIPPM. UIF GJSTU QFSTPO XIP UFMMT UIF LFZ XPSE UP UIF PXOFS PG UIF AJABJYVBO UFB TIPQ XJMM HFUB DPVQPO PG POF IVOESFE UBJXBO EPMMBST.

解密後明文

THE ONGIRUATOIN YELIM OS TAKEN GRIU UACTWTIR. ALAN TWRONF MAS YIRN AT PADDONFTIN, LINDIN. HOS GATHER, XWLOW UATHOSIN TWRONF, MAS A YROTOSH UEUYER IG THE ONDOAN COVOL SERVOCE AND HE MAS IGTEN AYRIAD. ALAN'S UITHER, ETHEL SARA STINEB, MAS THE DAWFHTER IG THE CHOEG ENFONEER IG THE UADRAS RAOLMABS AND ALAN'S PARENTS HAD UET AND UARROED ON ONDOA. MHEN ALAN MAS AYIWT INE BEAR ILD HOS UITHER REXIONED HER HWSYAND ON ONDOA, LEAVONF ALAN ON ENFLAND MOTH GROENDSIG THE GAUOLB. ALAN MAS SENT TI SCHIIL YWT DOD NIT SEEU TI YE IYTAONONF ANB YENEGOT SI HE MAS REUIVED GROU THE SCHIIL AGTERA GEM UINTHS. NEJT HE MAS SENT TI HAQLEHWRST PREPARATIRB SCHIIL MHERE HE SEEUED TI YE AN 'AVERAFE TI FIID' PWPOLON UIST SWYXECTS YWT MAS FREATLM TAKEN WP MOTH GILLIMONF HOS IMN ODEAS. HE YECAUE ONTERESTED ON CHESS MHOLE AT THOS SCHIIL AND HE ALSI XIONED THE DEYATONF SICOETB. HE CIUPLETED HOS CIUUIN ENTRANCE EJAUONATOIN ON 1926 AND THEN MENT TI SHERYIRNE SCHIIL. NIM 1926 MAS THE BEAR IG THE FENERAL STROKE AND MHEN THE STROKE MASON PRIFRESS TWRONF CBCLED 60 UOLES TI THE SCHIIL GRIU HOS HIUE, NIT TII DEUANDONF ATASK GIR TWRONF MHI LATER MAS TI YECIUEA GONE ATHLETE IG ALUIST ILBUPOC STANDARD. HE GIWND OT VERB DOGGOCWLT TI GOT ONTI MHAT MAS EJPECTED AT THOS PWYLOC SCHIIL, BET HOS UITHER HAD YEEN SI DETERUONED THAT HE SHIWLD HAVEA PWYLOC SCHIIL EDWCATOIN. UANM IG THE UIST IROFONAL THONKERS HAVE GIWND CINVENT OINAL SCHIILONF AN ALUIST ONCIUPREHENSOYLE PRICESS AND THOS SEEUS TI HAVE YEEN THE CASE GIR TWRONF. HOS FENOWS DRIVE HOU ON HOS IMN DORECTOINS RATHER THAN THISE REZWORED YB HOS TEACHERS. HE MAS CROTOCOSED GIR HOS HANDMROTONF, STRWFFLED AT ENFLOSH, AND EVEN ON UATHEUATOCS HE MAS TII ONTERESTED MOTH HOS IMN ODEAS TI PRIDWCE SILWTOINS TI PRIYLEUS WSONF THE UETHIDS TAWFHT YB HOS TEACHERS. DESPOTE PRIDWCONF WNCINVENT OINAL ANSMERS, TWRONF DOD MON ALUIST EVERB PISSOYLE UATHEUATOCS PROQE MHOLE AT SHERYIRNE. ON CHEUOSTRM. A SWYXECT MHOCH HAD ONTERESTED HOU GRIUA VERB EARLM AFE. HE CARROED IWT EJPEROUENTS GILLIMONF HOS IMN AFENDA MHOCH DOD NIT PLEASE HOS TEACHER. TWRONF'S HEADUASTER MRITE.OG HE OS TI STAB AT PWYLOC SCHIIL, HE UWST AOU AT YECIUONF EDWCATED. OG HE OS TI YE SILELM A SCOENTOGOC SPECOALOST, HE OS MASTONF HOS TOUE AT A PWYLOC SCHIIL. THE GORST PERSIN MHI TELLS THE KEB MIRD TI THE IMNER IG THE QOQAOJWAN TEA SHIP MOLL FETA CIWPIN IG INE HWNDRED TAOMAN DILLARS.