September 1997
Features

Decoding a war time diary

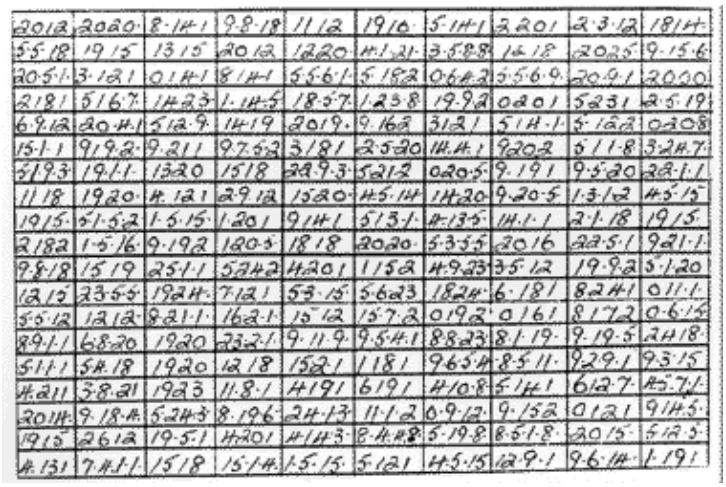
by P.J. Aston

Donald Hill was a young squadron leader attached to the Far East command in Kai Tak, Hong Kong in 1941. The Japanese attacked on 8 December and seventeen days later, on Christmas day, the outnumbered defending forces surrendered and were put into prisoner of war camps in which many died. Donald kept a diary during the battle and for a time during his captivity. He wrote it in a numerical code, disguised as "Russels Mathematical Tables". After the war he brought the diary home with him but his experiences were so traumatic that he did not like to talk about them. The diary was never translated before his death in 1985.

Hong Kong harbour shortly before the Japanese invasion.
(Source: Veterans Affairs Canada.)

Decoding

I stared at the pages of numbers in front of me. How could these numbers be turned into a story? Since this was done during the war, it must be fairly simple I reasoned and with a computer to help, it should not take too much effort to crack the code, should it?



A sample of the coded diary. Note the dot either side of a single digit number.

The first observation was that only the numbers 1−26 were used. That seemed like a good start. I guessed that the method used was a *substitution cipher* in which each number stands for a particular letter and all that is required is to match up numbers with letters and then it would be possible to read off the story.

There is a standard method of cracking such a code which consists of checking frequencies. If a long piece of English text is taken from a book or a newspaper and the number of times each letter occurs is determined, then the letter E will always be the most common letter. The second most common letter is T and this is followed by A. Similarly, the least common letters are J, Q and Z. The whole alphabet can be written down in order of frequency. If long pieces of text are used, then it is found that this ordering will be almost the same every time.

So I went to my computer, typed in the first page of numbers and wrote a little program to check the frequency of each of them. I found that the first few most frequent numbers were 5, 20 and 1 while the least frequent were 10, 26 and 17. Now the 5th letter of the alphabet is E, the 20th is T, the 1st is A and so, comparing with the alphabet rearranged in order of frequency, this suggested that there was a simple translation of numbers into letters given by 1=A, 2=B, 3=C etc. (Note that the last two numbers 26 and 17 are in the wrong order compared with the frequency list but this small change is not important.) However, this translation from numbers to letters did not result in a story, just lots of jumbled letters. This seemed like a step of progress but obviously the method used was not a substitution cipher.

Questions

1. Write a computer program to read in some text, add up the number of times each letter in the alphabet occurs and write the letters out in order of their frequencies with the most frequent first. Run your program on several long pieces of text (such as this article or some other text you can download from the web) and see how much difference there is in the lists in each case. How many differences are there for short pieces of text and for long pieces? Can you explain your findings? (You might like to compare this with tossing a coin many times and counting the number of heads and tails.)

2. The following message is written using a substitution cipher in which each number stands for the letter which is 3 places further on in the alphabet (assuming that the letter after Z is A). Can you translate the message?

10 24 17 5 2 10 24 17 6 26 16
6 16
3 24 16 26 6 11 24 17 6 11 4

3. A similar method is used for the following message except that each number stands for the letter which is *n* places further on in the alphabet for some value *n*. Can you find *n* and translate the message? This can be done either by hand or by writing a computer program to help you.

19 8 17 19 1 21 20
25 10
17 10
2 17 9 10

The next step

Having converted the numbers into letters, I now had to consider *transposition ciphers* in which the letters are rearranged in some way. By looking at some marks on the pages of the diary, I realised that the letters should be written out in rectangular blocks consisting of 33 rows of 34 letters. Also, on the front page of the tables, two names were written. These were

DONALD SAMUEL HILL
PAMELA SEELY KIRRAGE.

These names were his own and that of his fiancée Pamela. I was not sure of the significance of these names until, lying in bed early one morning, it suddenly occurred to me to count the number of letters in these names − 34, the same as the number of columns in the blocks of letters! This suggested that these names were used as a keyword for rearranging the columns in the blocks, another standard method.

Donald and Pamela on their wedding day in 1946.
(Source: Chris Hill.)

Example

To illustrate the method using a keyword, suppose we want to code the message

MATHEMATICS IS VERY INTERESTING.

There are 28 letters in this message so we could write it in a rectangular block consisting of 7 rows of 4 letters each. Writing the message across in rows gives

```
M A T H
```

```
E  M  A  T
I  C  S  I
S  V  E  R
Y  I  N  T
E  R  E  S
T  I  N  G
```

We now use a 4 letter keyword to rearrange the columns of letters. Suppose we use the keyword HIDE. The method of coding consists of writing the keyword over the columns of the block and then moving the columns around to put the letters of the keyword in alphabetical order. In this case, we rearrange the letters of the keyword as DEHI and this corresponds to taking the columns in the order 3, 4, 1, 2. Rearranging the columns in this way gives the following block.

```
T  H  M  A
A  T  E  M
S  I  I  C
E  R  S  V
N  T  Y  I
E  S  E  R
N  G  T  I
```

The final step is then to write out the columns of the block as the coded message which is

TASENEN HTIRTSG MEISYET AMCVIRI

I think you will agree that it is not at all obvious what the original message was looking at these jumbled letters! One advantage of this method is that it is very difficult to decode the message if you do not know the keyword which was used in the coding.

Question

4. Can you reverse the process described above to decode the following message using the keyword CIPHER?

CAAIP SCDOR EEECE ONCTU DBKHT CRWMS

The solution

Having written out all the letters from the coded diary in rectangular blocks, I wrote a program to use the keyword method in reverse, taking the names as the keyword, to reorder the columns and then, there on the screen in front of me, was a story that I could read! The 12 pages of numbers in the diary turned into approximately 11 pages of text once it was all translated. The diary told a fascinating, first hand account of the battle for Hong Kong, the confusion after the surrender and then some details about life in a prisoner of war camp.

British Prisoners of war marching to captivity in Hong Kong 28th December 1941.

What did it say?

Some extracts from the diary are as follows:

December 23rd. Up early, lucky for me, as a bomb lands on my bed just as I leave the room wrecking everything including my kit.

December 25th. What a Christmas day, empty stomachs, tired out, and heaven knows what is going on. At ten am a message arrives saying their is a truce until midday. This news is immediately followed by a terrific bombardment of our positions. Not my idea of a truce.

December 26th. Several (Japanese) officers started arguing and kept pointing at me and looking aggressive. Suddenly one of the officers whipped out his sword and I thought they had decided to bump me off but to my amazement he produced a bottle of beer, nipped the top off with his sword, and handed me the bottle. I was then given a loaf of bread. Two officers decide to drive me back in a Ford Ten. They don't use any lights and we have several narrow escapes from hitting lamp posts. Suddenly I see we are heading for one of the islands in the middle of the road and shout a warning. Too late and there's a terrific crash and we finish up on our backs. By now I am fed up so, bowing politely, I leave them and walk the two miles to China Command.

December 30th. It would appear that we are going to Sham Shui Po. The whole camp has been stripped of every useful article by looters and had also been bombed. All doors, windows, furniture, and fittings had been taken leaving just hulks of buildings. Even in peace time it was an awful dump, but now it looked as if a typhoon had hit it.

December 31st. There are over six thousand men in the camp with no sanitation and rotten food. We have no lights and go to bed soon after dusk. We have one meal at nine and another at five consisting of soggy rice and are permanently hungry. And so ended nineteen forty one.

February 6th. Spend hours these days thinking of home and family, especially Pam. They probably think I am dead and I pray to God that the Japs will get news through. Thank God for you Pammy darling, your memory is ever with me. I still have your photograph, signet ring and cigarette case. I will never lose them.

Further reading

A decoded diary reveals a war time story

Translation of "Russels Mathematical Tables"
Full text of the diary

Canadians in Asia
A Canadian account of the Japanese invasion of Hong Kong

P.J. Aston, Dept of Mathematics and Statistics, University of Surrey, Guildford, Surrey GU2 5XH

Solutions to questions

2. MATHEMATICS IS FASCINATING

3. $n = 10$ and the message is CRACKED IT AT LAST

4. CODES CAN BE CRACKED WITH COMPUTERS