Codis correctors d'errors: de les travesses de futbol als viatges espacials

Com es pot aconseguir una transmissió de dades, per exemple en telecomunicacions, amb la màxima fidelitat? Com es poden emmagatzemar per molt de temps dades fixes en un ordinador sense que les minúscules traces del material radioactiu de la màquina alterin el contingut de les cel·les de memòria? Incorporant codis correctors d'errors.

FRANÇOIS SIGRIST

La majoria dels jugadors empedernits de travesses de futbol es coneixen aquesta taula 1 de memòria:

x	x	x	1	1	1	2	2	2
x	1	2	x	1	2	x	1	2
x	1	2	1	2	x	2	x	1
x	1	2	2	x	1	1	2	x

Taula 1: Travessa amb 4 partits.

Aquesta llista de 9 pronòstics per a 4 partits és coneguda perquè garanteix almenys 3 encerts. En altres paraules: sigui quin sigui el resultat dels 4 partits, hi ha una columna de la llista amb 3 o 4 signes correctes.

El codi de les travesses

Una manera carregosa de comprovar aquesta remarcable propietat consisteix a mirar un per un els $3^4 = 81$ possibles resultats que es poden donar en 4 partits. També podem procedir a l'inrevés: cada columna de la nostra llista té un domini d'influència sobre 9 diferents resultats: la pròpia columna (4 encerts) i les 8 possibilitats en què la columna té 3 encerts. Com que $9 \times 9 = 81$, cal verificar que aquests dominis d'influència són tots disjunts. Ara, si dos d'aquests dominis d'influència tinguessin intersecció, hi hauria dues columnes de la nostra llista amb dos o més signes en comú, i aquest no és el cas, com es comprova ràpidament.

Il·lustrem a continuació alguns termes tècnics amb l'exemple de la taula de les travesses: les 9 columnes (composta cadascuna de 4 elements) són les paraules d'un codi de longitud 4. Els signes {x,1,2} constitueixen l'alfabet del codi. La distància entre dues paraules és el nombre de posicions en què difereixen les paraules. La distància del codi, la definim com la mínima distància entre dues paraules diferents. En el codi de les travesses la distància és 3.

Veiem ara com el nostre codi de les travesses pot corregir errors: si escrivint una paraula del codi cometem 1 error (per exemple, xxx1 en comptes de xxxx), aleshores la paraula falsa és a distància 1 de la correcta i a distància almenys 2 de totes les altres paraules del codi. Per tant, la paraula correcta és la paraula del codi que està més a prop de la paraula escrita. Aquest mètode es pot aplicar a qualsevol codi de distància 3. Aquests codis corregeixen automàticament 1 error: si una paraula té un signe equivocat no pertany al codi i la corregim considerant la paraula del codi que és més a prop seu!

Deixem ara el futbol de banda i apliquem la propietat correctora d'un codi a les telecomunicacions. Imaginem un lèxic format per 9 paraules diferents, cadascuna integrada per dues xifres de l'alfabet {x,1,2}. Per exemple, podem triar el principi de les 9 paraules de 4 xifres del codi de les travesses: {xx,x1,x2,1x,11,12,2x,21,22}. (Si substituïm x per 0 ens adonem que tenim els números

del 0 al 8 escrits en base 3. Així doncs, les columnes de la taula 1 vénen numerades per les seves dues primeres xifres.) Suposem que aquestes paraules han de ser trameses de manera no fidel, amb una probabilitat p d'error en la transmissió de cada signe; per exemple, $p = 0,01$ (l'1 %). Per terme mitjà, doncs, 1 signe de cada 100 serà incorrectament tramès. Tot i així, reduïm la taxa de transmissió a la meitat, és a dir, disposem de quatre xifres per a la transmissió de cada paraula de dues xifres, amb el propòsit d'utilitzar les dues xifres addicionals per reduir encara més la possibilitat d'error en la transmissió.

Una primera idea és repetir cada paraula; per exemple, enviar x1x1 en comptes de x1. Si en la recepció ens arriba una paraula amb dues meitats iguals és molt improbable que s'hagi comès un error. Si les dues meitats de la paraula són diferents, hem estat de pega, i necessitem un mètode per a reconstruir la paraula correcta. Es pot provar que tots els mètodes són igual de bons (i igual de dolents) que el d'escollir la primera meitat de la paraula rebuda. La probabilitat d'una descodificació correcta amb aquest procediment és per tant $(1-p)^2 = 0,9801$; és a dir, 1 paraula de cada 50, per terme mitjà, serà incorrectament interpretada.

Hi ha una altra idea molt més enginyosa: les nostres paraules són el principi de les 9 paraules de 4 xifres del codi de les travesses. Doncs bé, enviem cada vegada la paraula sencera del codi; per exemple, x111 en comptes de x1. Per a la recepció apliquem la següent regla: busquem a la taula una paraula a distància mínima de la paraula rebuda i prenem les seves dues primeres xifres. Si s'han comès 0 o 1 errors en la transmissió, el resultat que obtenim és el correcte ja que el codi corregeix 1 error! La probabilitat d'una descodificació correcta amb aquest mètode, ens la dóna la fórmula: $(1-p)^4 + 4p(1-p)^3 = 0,9994$. Ara, per tant, només 1 de cada 1.689 paraules, de mitjana seran incorrectament interpretades. Una millora notable del rendiment!

Mètodes anàlegs a aquest, amb l'alfabet $\{0,1\}$, han comportat canvis radicals en les tècniques de transmissió de dades, en particular en les telecomunicacions (transmissió digital). Una altra aplicació, potser fins i tot més important, fa referència a la protecció de dades fixes dels ordinadors. En cada ordinador cal emmagatzemar una gran quantitat de dades a les cel·les de memòria. Pràcticament tots els materials tenen traces ínfimes d'urani o tori radioactius. La radiació alfa corresponent pot alterar el contingut de les cel·les de memòria d'un ordinador. Suposem, per exemple, que es produeix una alteració de cel·la de memòria cada milió d'anys. Pels ordinadors actuals, proveïts de quantitats ingents de cel·les de memòria, això significa un període de seguretat, com a molt, d'algunes setmanes. Si incorporem codis correctors d'errors, els períodes de seguretat s'allarguen considerablement. Com a conseqüència de la ben coneguda *paradoxa de l'aniversari*, els errors en una sola posició són automàticament corregits. Per més informació sobre aquest punt es pot consultar un excel·lent article de McEliece (*Scientific American*, gener 1985). Avui en dia, pràcticament totes les dades són regularment posades al dia i els ordinadors estan fortament protegits dels errors de magatzematge.

Codis de Hamming

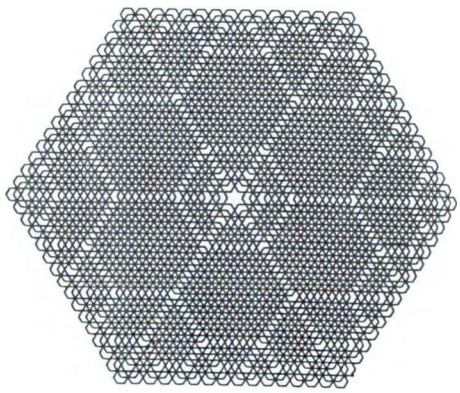
Els codis correctors d'errors van ser descoberts després de la segona guerra mundial. Aquí teniu la història: situem-nos als laboratoris Bell en els anys 1946-1947: en el seu temps lliure, el matemàtic Richard Hamming experimenta amb una d'aquelles famoses computadores monstruoses, farcides de cables connectors i lampadetes, que varen ser desenvolupades durant la segona guerra mundial. Tot i que la màquina no té cap codi corrector d'errors (no n'hi havia encara!), té un sistema molt elemental d'alerta que avisa quan es comet un error encenent una bombeta. Cal desconnectar aleshores la màquina i tornar a carregar les dades. El cap de setmana, el temps que Hamming pot utilitzar la màquina, els tècnics que fan aquesta feina no hi són i el nostre matemàtic ha d'interrompre l'experiment cada cop que té una pana, i esperar fins que tornin els tècnics. Es proposa aleshores la tasca de desenvolupar un codi automàtic per corregir errors. Amb una mica d'àlgebra lineal aconsegueix desenvolupar un sistema de *travesses* sobre l'alfabet $\{0,1\}$ amb 16 columnes de 7 partits (vegeu la taula 2). El sistema garanteix 6 encerts. La comparació amb el codi de les travesses ens fa veure com cal utilitzar el sistema. Les paraules a transmetre són les formades per les primeres 4 xifres de cada columna. Per cert, noteu que representen els números del 0

al 15 en base 2 (anàlogament al codi de les travesses), de manera que també serveixen per a numerar les columnes. Per a la transmissió d'una paraula enviem la columna sencera, i quan es rep es consideren les 4 primeres xifres de la columna del codi que és més a prop de la columna rebuda. Aquest mètode té una taxa de transmissió lleugerament superior a la del codi de les travesses: passem de 4 a 7 xifres en comptes de doblar!

0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	0	0	0	1	1	1	1	0	0	0	0	1	1	1
0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
0	0	1	1	1	1	0	0	1	1	0	0	0	0	1
0	1	0	1	1	0	1	0	1	0	1	0	0	1	0
0	1	1	0	0	1	1	0	1	0	0	1	1	0	0

Taula 2: Codi de Hamming de longitud 7.

El mètode de Hamming permet fabricar molts altres codis. Entre ells, un de longitud 15, que transmet 11 xifres, és utilitzat per l'exèrcit nordamericà. Per a un alfabet de 3 xifres (apte per als partidaris de les travesses) hi ha, fins i tot, un codi de longitud 13 (el nombre de partits de la travessa suïssa) que garanteix 12 encerts. Per desgràcia, aquest codi té $3^{10} = 59.049$ columnes!



Codis de Golay

Els codis de Hamming només poden corregir 1 error per paràmetre. El 1949 l'enginyer suís Marcel Golay, que havia treballat als Estats Units com a expert en radars, va construir dos codis remarcables que permeten corregir uns quants errors per paraula. El primer codi utilitza un alfabet de 3 xifres i no té cap aplicació practica en les telecomunicacions; no obstant això, per als matemàtics és, encara avui dia, un objecte molt interessant. (De fet, la descoberta original del codi és deguda al finès Juhari Virtakallio, que el va descriure, el 1947, en la revista finesa de futbol *Vejkkaja*.) El segon codi de Golay, amb alfabet $\{0,1\}$, és extraordinàriament important, tant a efectes pràctics com teòrics. Consta de $2^{12} = 4.096$ paraules de longitud 23 i pot corregir 3 errors per paraula.

Tots tenim a la memòria la imatge de Júpiter que va poder ser enregistrada amb el sonar del *Voyager*. El sonar va ser programat amb un codi de Golay, per tal d'aconseguir una transmissió extremadament segura de les dades.

Per acabar, volem mencionar encara una altra aplicació del codi de Golay en un àmbit completament diferent. Fa aproximadament quinze anys es van fer enormes progressos en la classificació dels grups finits simples (els grups són estructures molt importants, que apareixen en molts vessants de la matemàtica). A més de les diverses menes de grups simples clàssics, hi ha 26 grups esporàdics. La majoria d'aquests grups es van poder realitzar explícitament per primera vegada amb l'ajuda del codi de Golay, i en connexió amb empaquetaments d'esferes en espais de dimensió 24. La teoria necessària per a la completa classificació dels grups finits simples és el resultat del treball conjunt, ingent i sense parió, de molts matemàtics. Omple uns quants milers de pàgines i es calcula que no hi ha cap especialista capaç de dominar-la totalment.

L'exemple dels codis de Golay, amb les seves aplicacions en diversos dominis, mostra com pot arribar a ser d'inabastable la recerca en matemàtiques.

François Sigrist és professor a la Universitat de Neuchâtel.