

# 計算をしない計算機\*

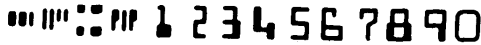
和田 弘\*\*

本日は電子部がしているこの方面の仕事の中で、普通の電子計算機とそれを使う数値計算の話には触れないで、その他のことをご紹介しようと思います。

## 1. 文字を読む機械

電子計算機が符号 (code) だけしか読めないのに、人間に読める文字が読めない現状は不便です。そこで機械に人間の文字を読ませようとする努力は、いくつかの方法になって発表されています。

その一つは Stanford 研究所の開発した磁性インクで書かれた文字を、磁気ヘッドで走査する方法で、これはアメリカで実用になり出しました (第1図)。しかしこの一次元的な方法で多くの種類の文字を読ませようとするのは難しく、文字を改造することも必要になります。

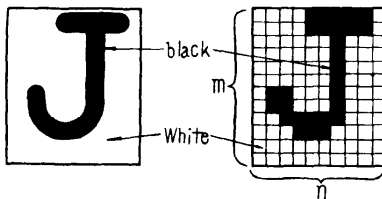


第1図 Stanford が開発した数字

文字を本来の姿のまま、二次元的に走査する方法にも若干の提案があります。ここではその一つとしてわれわれの開発した機械の要点を述べます。

この機械は普通の英文タイプにある文字や記号の合計73字を読むことを目的としたものです。

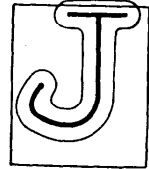
いま第2図のように文字を枠に入れて、これを第3図のような網目の白黒で置き換えます。この際、網目の大きさが問題になります。つまり、網目が大きいとcとoのような小さい相違を区別できませんし、小さすぎると装置が複雑になります。また一つの網目の



第2図 Letter in a tablet  
第3図 Symbolic letter

中が完全に白か黒でない時に、これをどう判定するかにも問題があれば、文字の僅かなズレとか小さいきずのために、同じ文字でもある網目が時によって白になったり黒になることもあります。

そこで第4図のように、文字の周辺の部分を白でもなく黒でもなく、仮に灰色とみなすことにしました。この部分は識別に関係させないのです。この灰色の概念を網目にも採用して、第3図に代えて第5図のようなものを考えました。

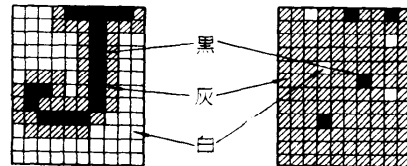


第4図

Colored letter  
つまり第2図の文字と第5図の表示とが一致するか否かで、文字を読ませようとしたわけです。したがって、73の文字について作った第5図のような表示が相互に違うことが必要条件です。

ここまでの考えを実現するには、網目の大きさと字の太さ、ズレても支障を少なくする条件、一つの網目の中でどの位の部分が黒ければ、黒い網目とみなしてよいか、などの検討しなければなりません。それらは相互に密接な関係があります。それを数式で表わし、それを解いて、縦を10、横を12に分けて網目の数を120にしました。

こうして作った第5図では、例えば右下の隅の網目はどの文字でも白です。一々白いことを確認しても無駄です。文字を相互に区別するには、どの網目が急所であるかを検討しました。逐次急所を求めて行ったら51個の網目を採用すれば、73の文字が識別できることが判りました。残った網目は白黒にかかわらず灰色にし



第5図 Letter block  
第6図 Modified letter block

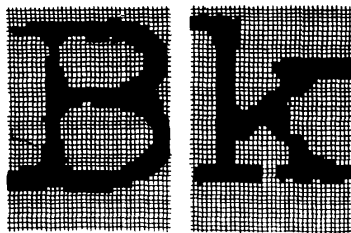
\* Computers for non-numerical use, by Hiroshi Wada

\*\* 工業技術院電気試験所電子部長  
1960年4月22日創立総会の記念講演

ました。こうしてでき上がったのが第6図です。

この工夫によって装置が簡単なもののできただけでなく、いろいろの雑音に対して強い装置となったのです。具体的には、白黒を表示させるゲルマニウム・ダイオードの数はもし簡素化しなければ、約5,600個必要であったのが、このお蔭で800個ですみました。

こうして作ったマトリックスを採り入れた実際の機械は、*flying spot scanner* からの縦方向のラスタ40本で文字を走査させ、1本のラスタを48個のバルスに直して、計1,920個のバルスで文字を表示させています。そして1個の網目は $4 \times 4 = 16$ 個のバルスを総合して、黒のバルスが3個以上あれば、その網目を黒としています(第7図)。



第7図 1,920ビットに分解された文字

タイプライタで用紙一杯(約6,000字)文章を打ったものをこの機械にかけて、約16%の文字が「読めない」記号を打ちました。これは第2図で、文字が枠中の適切な位置に取まらないためです。そこで読めない時は枠を左右上下にそれぞれ網目の半分ばかり動かした上、走査を繰返えして、どれかの位置で読ませるようにしたところ「読めない」率は半減しましたが、こんどは誤読が起るようになりました。残念ながら、まだ実用になるとはいえません。

元来第5図から第6図への簡素化を行なうことと、このように揺ぶることとは両立しないのですから、揺ぶることを前提としたものを改めて設計しています。またカナ文字についても設計しています。

さて、この機械はこうして設計された *matrix* を中心とした特定の用途のための金物 (*hardware*) です。文字にはタイプによるものの他にいろいろありますから、もっと融通性のあるものを考案しなければなりません。それには、*matrix* を作る時知恵を絞ったような操作を機械が自ら行って欲しいものです。つまり、いろいろな形の文字について、機械が自らそれを整理して各字についての本質だけを記憶していて、文字が示された時はそれに基づいて、認識できるよう

になって欲しいわけです。それには、われわれがさきの装置について行なったことを、プログラムの形で電子計算機に与えることから始めねばなりません。

こういう考え方を、一般的に機械の学習 (*machine learning*) といいます。この方針で文字を読むことはすでにイギリスの Manchester 大学、アメリカの M. I.T. や Bell 研究所で始められており、われわれも着手しています。

同じことが音声についてもいえます。京都大学、東北大学、名古屋大学、電気試験所などで研究されています。そして音声、文字、絵などをまとめて、欧米では *pattern* といい、これらを対象とする研究を *pattern 認識 (recognition)* と呼んでいます。

## 2. 外国語の翻訳

外国語を機械で翻訳させる研究は逐語訳 (*word-for-word*) から始まりました。欧米語はソ連を含めて、同じ系統から発しているもので、一つの文章の中で単語の相互の位置には、あまり大きな違いがありません。例えば、

I love you. → Ich liebe dich.

したがって、機械的な単語辞典を作り、原語を逐次機械に入れて、その訳語を打ち出せば一応の翻訳はできます。ところが一つの単語にはいろいろな意味があります。例えば *spring* を名詞に限っても、源泉、泉、バネ、弾性などの他に、季節の春もあります。そこで適当な訳語をどうして求めるかという意味論 (*semantic*) の研究が問題になるわけです。

ところが、英語を日本語に訳す場合は、

I give you some money.

ワタシ アゲル アナタ イクラカノ カネ

では訳せたといいいくいです。彼に与えるのか、彼がくれるのか大事な点が不明確です。少なくともテニオハを入れて、動詞を文章の末尾に廻すくらいことは望まれます。

そこでわれわれは *semantic* な研究よりも、まづ構文的 (*syntactic*) な研究が必要だと判断しました。そしてその方針として、英文の中の動詞に着目して、自動詞であるか他動詞であるか、完全自動詞であるか不完全自動詞であるかなどに応じて、いくつかの標準的な文形を定めました。与えられた文章がそのどれに属するかを判断させ、それによって動詞に続くものが目的語か補語であるかをきめて、テニオハを挿入させることにしました。そしてその決定に当っては、形容詞



われわれが取りあえず採用した方法は、英文の単語を意味のある単語と意味のない単語とに分けて、意味のあるのは名詞だけとし、その他は捨てます。そして同じ単語が出てくる頻度を計算し、選挙のようにある数以上の頻度を持った単語を選び出します。次に再び論文を機械に読ませて、この当選した単語がある数以上含まれている文章だけを印字させます。

「猫の首に鈴をつける鼠の相談」の寓話をこの方針で要約させると、

The cat killed some of the mice.

Put a bell on the cat.

But who is going to put the bell on the cat?

の三文が印字されます。注目した単語が *cat*, *mouse*, *bell* の三語であって、上の文はそれらの語が二つ以上入っていたためです。

1,000 語程度の記事であれば、「やまと」を使うと、約17分で要約されます。この程度の要約で満足してはいませんので、今後は syntactic な考え、単語のつながり、三段論法などを考慮に入れて改善するつもりですが、ここにも machine learning の問題が潜んでいます。

#### 4. 検 索 機

論文を読んで、今後の参考にしたいと感じると、われわれは著者名、雑誌名、その巻号とページ、発行年月、論文の標題、および内容の要約とか備考などを記録、保存しています。これを機械化して、能率を高めることを企てました。これらの事項をテープに打って機械に入れると、磁気テープに記録、保存されます。必要な時に、それらのどれかの項をこの機械に入れると、該当する論文全部について、上記の事項を自動的に打ち出してくれる検索機 (indexer) を最近完成しました。いよいよこれを使うことを考える段階に達しました。

アメリカでのこの方面の状況を見ますと、大部分のところではすべての項目を電話番号のように numerical code になおしたうえ使用しているようです。こういう方法によると、電話帳のような辞典がいるだけでなく、文献に現用されている UDC のように分類が樹枝状 (tree-like) になります。そうすると、例えば chemical physics と physical chemistry とは紙一重の違いであるにもかかわらず、離れて分類されるので、必要な時に一方は出て来なくなる虞があります。

論文の内容は連続的な性質を持っており、いろいろ

な観点から眺められ、引用されるべきものでありますから、それを yes か no かで割り切って行く樹枝状分類には一脈の危険を感じます。何とか合理的な整理法が必要と考えています。そこで取りあえず、符号化すべきものは雑誌名、巻号とページ、発行年月に限ることとし、標題や内容は単語をそのまま使つておいて、書き込む内容は使用する人の自由とし、試用の成果によって、改良法を考えることにしております。

これだけのことを申し合わせて使う場合を考えますと、機械に持ち込まれる文献は、順を追って磁気テープに収められますから、例えばトランジスタの文献がすでに入つていても、いま A 氏が入れるトランジスタの文献はこれと関係のない個所に記録されます。もし B 氏が同じ文献を読んで感心したとすると、内容の記述では少し違うかも知れませんが、同じ文献が二つ以上入ることになります。これはわれわれが知識を記憶している状況と趣が違ふように思われます。

何とかして、同じものは一カ所にまとめ、古いものは忘れて(整理して)、また一つことがいろいろの角度から取り出せるように、例えば、cross-reference というような感念を導入して、要するに合理的な整理方法を案出したいと念じております。

現在の電子計算機でも情報の蒐集 (collection)、貯蔵 (storage)、取り出し (retrieval) は行なわれております。例えば取り出すにはアドレスを指定すればよいのですが、隣アドレスにはこれと全く関係のない情報が入っております。学問として collection, storage, retrieval という場合は、そういうことを意味するのではなく、ここで文献の合理的整理法をどうしたらばよいかと悩んだような内容を指すのです。

これが解決されないと、情報センターのような組織は、物量を整備するだけでも莫大な費用と労力があるだけでなく、活用に当っては不便であって、あまり効果的に使われないことになる心配があります。

#### 5. む す び

以上述べてきたところの対象はどれも数値ではなくて文字や言葉でした。数学は広い意味での論理学の一部門ですが、現在のところ計算機が利用されているのは、演繹的 (deductive) に書かれた数値計算に限られています。定理を証明するような帰納的 (inductive) なことはまだできません。

また、さき程から述べてきたところの諸例は、いずれも多量のデータを整理して共通な本質ともいべき

ものを求めたうえ、判断する性質のこと、すなわち帰納的なことでして、人間の学習課程に似た内容のことです。machine learning と呼ばれる所以です。

本日誕生したこの学会が壮年になる頃のことを想うと、意見なり思想を表わす文章も、この計算機にかかっているでしょう。司法から始まって、行政から政治までも、やがてこれらの機械によって処理されることも夢とばかり片づけられません。

それが実現されるためにはまづ述語論理について、帰納的な理論が確立されることが望まれます。そしてその際にも、言葉（単語）は element として残りま

しょう。単語の意味というものは数値と違ってききに述べたように曖昧な点がありますが、数学、ことに代数学はわれわれの知っているような単純な数や文字を対象とすることから離れて、複雑な集合のような抽象的な対象を操作する方向に進んでいますから、単語や文章の解明に大いに参考になることと思われま

ことに、日本語を科学的に考察し改善することは必要で、これを怠ると機械文明の進歩にもかかわらず、日本人の生活には向上しない面が残されることになりはしないかと案ぜられます。