Information Theory와 音聲의 識別問題

姜 麟 求
(海軍士官學校)

정보이론은 비교적 연륜이 짧은 새로운 학문이다. 이를 처음 體系化한 것은 1942년에 MIT 교수인 노먼·위이너(N.Wiener)가 필터의 설계에 대한 이론을 확립한 뒤부터다. 그러나 戰爭中이었기 때문에 서로 모르기는 했지만, 동시에 소련의 킨치너(Khintchine) 박사가 거의 같은 이론을 편 것으로 알려져 요즘은 '위이너-킨치너의 이론'이라고 부른다. 그 후 1948년에 섀논(C.E. Shannon : 現 Illinois 大學敎授)이 당시 벨 전화 연구소에서 근본적인 통일된 정보이론을 세워서, "The Mathematical Theory of Communication"이란 통신의 수학적 이론을 발표했던 것이다. (Shannon and Weaver, 1963)

그 이론은 근본적으로 두 가지의 전제 조건에 근거를 둔 것이다. 첫째, 통신 즉 정보를 전달하는 과정은 유한(有限)의 가능한 선택중에서 하나씩 선택하는 과정이다. 그러니까 어휘가 오십만 혹은 백만이라면 그것으로 한정되는 것이고, 그 한정된 어휘에서 하나씩 하나씩 뽑아내서 어떤 것을 구성하는 그런 계통이라는 전제 조건이다. 둘째, 모든 통신과정중의 선택의 운용에 있어선 필연적으로 확률(probability)의 개념이 근본적인 역할을 한다. 그 구조에 있어서 어떤 어휘나 문장을 선택했을 때 다음에 선택하는 것은 임의의 것이 아니고 어떤 구조에 따라 선택을 하는데 그 과정에서는 확률적인 요소가 있다는 통계적 전제하에 이론이 확립된 것이다. 그러므로 어떤 정보의 전달을 통계적으로 다루는 학문은 모두 정보이론의 테두리 속에 있다고 볼 수도 있다. 그리고 통신 분야만으로 국한하면 이론이 간결하게 되어 있지만, 언어학에까지 관계를 지으면 과연 이렇게 이론화되는지 하는 의혹도 나올 수 있다.

그러나 대개 요사이 정보이론의 중요 분야는 위이너 이론의 필터 부분 즉 잡음이 섞인 신호를 받아서 어떻게 처리하면 가장 원형에 가까운 것을 얻겠느냐는 필터 분야와 섀논이 발표한 이론을 중심으로 해서 부호화(coding), 즉 어떤 메시지(언어나 음악)를 어떻게 가장 적절한 다른 신호로 바꾸면 가장 탈없이 송신할 수 있느냐는 부호화의 분야가 있다. 또 현

재 큰 발전을 이룬 전자계산기를 이용하여 연구되는 뒤에 말할 패턴인식(pattern recognition)이 최근 정보이론 분야에서 가장 활발히 논의되고 있다. 패턴인식이란 어떤 현상에서 가장 특출한 특징이 무엇이고 그것을 어떻게 골라내서 어떻게 처리하는가 하는 것이다.

여기서는 우선 정보이론의 기초 개념과 패턴인식중 언어의 기계적 인식에 관하여 간단히 진술하겠다.

1. 정보이론의 기초 개념

정보이론의 기초 개념을 설명하기 위하여 그림 1과 같은 일반적인 통신계통을 생각해 보자.

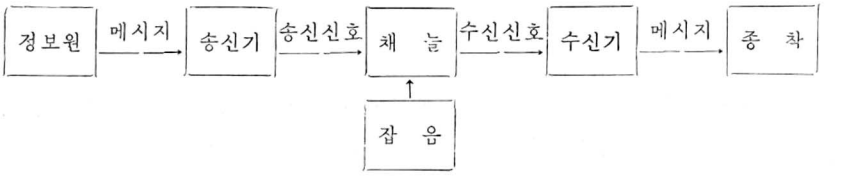

그림 1

우선 정보원(情報源, source)에서 나온 가능한 많은 선택중의 하나인 메시지를 종착(destination)으로 보내는 과정은 첫째 송신기를 통과시키면서 메시지를 어떤 신호로 바꾸는 것이다. 이것을 채늘을 통해서 수신기에 수신신호로서 전달하는 과정이 따르고, 이 때 대부분은 잡음이 반드시 개입된다. 가령 전화를 쓸 때 여러분이 정보원이라면 전화기는 송신기며 이것이 메시지를 전기적 신호로 바꿔서 전깃줄을 통해 수신처로 가게 되는 것이다. 이 때 전화기가 깨끗하고 틀림없다면 '누구세요'란 말조차 필요없을 것이다. 왜냐하면 상대방 말소리가 그대로 나온다면 즉시 누구인지 알 것인 까닭이다. 그러나 의례히 '누구세요'라 묻는 것은 도중에 잡음이 들어가는 때문이라고 생각할 수 있다.

이렇게 수신기에서 나오는 소리가 종착되는 피전달자나 전달하고자 하는 기관에 닿게 되면 다시 메시지로 재현되는 것이다. 이 때 이 메시지를 섀논과 같이 전기공학에 입각한 수학적 처리만을 한다거나 또는 더 확장하여 의미론에까지 관련을 묻는 학자도 있다. 후자의 입장은 다시 말하자면 이 각 요소 하나하나 사이에 의미론적인 변환(transformation)이 개재되고 이 경우 반드시 잡음이 개입되게 마련이니까 종착직전에 의미론적인 수신기가 다시 변환을 해서 종착에 도달한다고 생각하는 것이다. 또 이렇게 변환이 있을 때마다 정보는 반드시 감소하게 된다고 한다.

섀논의 이론의 장점의 하나는 얼마나 많은 정보가 정보원에 있는가를 수학적으로 측정할 척도를 마련한 것이다. 그 측도를 소위 '앤트로피(entropy)'라 하며 이것을 수식으로 정의하면,

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \text{ bit}$$

우선 정보원에 있는 유한의 선택을 1, 2, 3……i로 번호를 매기고 그 각 번호를 정보원의 각단위에 대한 (즉 해당되는 가능한 사건의) 발생확률로 나타내게 된 p_i에 $\log_2 p_i$를 곱하면 평균화되는 것으로 이것이 바로 '앤트로피'며 그 단위는 보통 bit로 나타낸다. 이 때 대수함수의 기수(基數)가 2인 것은 역시 전기공학에서 가장 단순한 구분인 on 이냐 off 냐는 것을 나타낸다. 전자계산기가 아무리 복잡해도 근본 원리는 그 어느 회로에 전기가 통하느냐 안 통하느냐는 구분, 혹은 전압이 어느 기준선 이상이냐 이하냐는 구분밖에 없는 것이다. 이렇게 작용하는 한 계산 단위체를 bit 라 하고, $\log_2$에 2를 넣으면 1 이 되니 on-off로서 bit 단위를 삼아 정보를 평균해서 나타낼 수 있게 한 것이다. 그밖에 앤트로피에는 수학적으로 여러가지 장점이 있으나 생략한다. 이것이 또 무슨 척도인가를 알기 위해 상대적 앤트로피(Hr)를 설정한다. 이것은 앤트로피의 maximum possible 한 것 나누기 실제 앤트로피를 말한다.

$$H_r = \mathrm{H} / \mathrm{H\ max.\ poss.}$$

즉 최대로 가능한 값과 실제 앤트로피의 비율이며, 예를 들어 부호가 여덟 개가 있다면,

$$\log_2 8 = 3$$

이고 그 가능한 최대치는 여덟 개가 골고루 퍼져 있을 때, 즉 확률이 1/8일 때가 가장 높다. 따라서

$$\sum_{i=1}^{8} \frac{1}{8} = 1$$

이며, 결과적으로 다음과 같이 3 bit 가 된다.

$$H = -\sum_{i=1}^{8} \frac{1}{8} \log_2 8 = 3$$

즉 부호가 8 개면 3 bit가 최대로 가능한 수다. 또 16개일 때는 4개의 부호로 그것을 분류 표시할 수 있다.

다음 올장도(冗長度 : redundancy)란 최대치 1에서 상대적 앤트로피를 뺀 값으로 결국 그만큼은 없어도 좋을 잉여(剩餘)의 것이라는 뜻이다.

$$R = 1 - Hr$$

여기서 한글 단모음에 대한 앤트로피를 계산한 결과를 보자. 그런데 한글의 경우에선 각 자모의 발생확률 Pi를 계산하는 것이 매우 복잡한 것으로, 필자는 문교부 자료에 의거했지만 다른 자료(공병우·김동훈씨에 의한)와는 오차 범위가 클 뿐만 아니라 서열에까지 영향을 주어 뒤바뀌는 경우도 있다.

한글 모음의 앤트로피

	p_i	$p_i \log_2 p_i$		p_i	$p_i \log_2 p_i$
ㅏ	0.2605	0.506	ㅗ	0.131	0.380
ㅣ	0.1875	0.451	ㅜ	0.077	0.285
ㅡ	0.144	0.403	ㅟ	0.049	0.213
ㅓ	0.142	0.400	ㅐ	0.029	0.148

'아, 이, 으, 어, 오, 우, 에, 애'의 8 기본모음에 대해서 조사하면 그 발생확률(p_i)과 $p_i \log_2 p_i$는 '아'인 경우 각각 0.2605와 0.506이 나오고, 결국 총 합계는 2.786이 나온다. 이 반면 8 개의 단모음의 발생확률이 균일하다면 $H=3$이 최대치로 되는데 위 합계와의 차이만큼 정보의 양이 작아진 것이라고 볼 수 있다. 여기서 H_r을 계산해 보면,

$$2.786/3 \fallingdotseq 0.919$$

정도가 되어, 한국어 모음의 올장도 (R)는 약 8%가 된다는 결과다. 그런데 실제로 자음까지 다 생각하면 그 올장도는 증가할 것이며 음과 음, 단어와 단어간의 상호관계를 참작하면 더욱 증가할 것이다. 영어에서는 R이 50% 정도다.

이 R은 어느 언어에나 있는 것이며 특히 고대 언어 및 사어(死語)를 해독하는 일과 현대에선 암호 해독에 필요한 기본 요건이다. 또 암호 작성에 있어서도 되도록 R을 줄여 쉽게 풀리지 않도록 하는 것이 근본 원칙이다. 그래도 암호가 풀리는 것은 1차적 단계에서 더 나아가 2차, 3차의 체계화된 구조를 언어가 드러내고 있기 때문에 가능한 것이다.

채늘의 능력(C)는 bit/second를 단위로 한다. 이 계산 과정은 생략하나, 쉬운 예를 들어 무전수(無電手)들이 1분간에 치는 글자 수는 한정되어 있는 것 같은 일정한 능력을 말한다.

그러면 정보이론 중에 정보원과 채늘 사이에 관해 섀논이 세운 두 개의 근본 정리를 소개하겠다. 그 첫째는 잡음이 없는 채늘에서는 정보원의 출력을 부호화하면 평균해서 C/H에 임의로 근사한 속도로서 채늘을 통해 전달시킬 수 있다는 것이다. 즉 C/H에 꼭 맞게는 못해도 거의 비슷하게 하는 것이 가능하다. C라는 채늘 능력이 있고 정보원이 H인 체제에서 그 채늘에 잡음이 없으면 C/H에 임의로 가까운 속도로, 즉 수학에서 +e를 붙이는, 아주 적은 숫자만 더 하면 그 정도의 속도로 전달이 가능하다. 그러나 반면 C/H 이상의 속도로는 절대로 전달할 수 없다는 일종의 존재의 정리다.

섀논의 제2의 정리는, 잡음이 있는 채늘에서는 정보원의 앤트로피가 채늘 능력보다 적을 경우에는 즉 $H \leq C$인 경우에는 임의의 작은 오차(천만 개 중에 하나의 오차 정도)로 정보원의 출력을 채늘을 통해 전달하는 부호화 방법이 있는데 채늘 능력이 앤트로피보다 적을 때 즉 $H > C$이면 $H-C+e$ 보다 작은 오차를 일으키는 부호화방법은 가능하나 $H-C$ 보다 작은 오차를 일으키는 부호화방법은 없다는 것이다.

앞서 말한 세 큰 분야, 즉 필터, 부호화방법(coding), 패턴 인식에서 이 섀논의 정리가 나온 이후 부호화방법이 어떻게 가능하며 그것이 무엇인가를 완전히 체계화하려는 노력이 이루어진 것이다. 앤트로피의 정의와 이 두 정리가 완성되자 이와 관련하여 거기에 존재한다는 가장 적합한 정체를 찾고 있으나 20년이 지난 오늘날 아직도 확실한 해답은 얻지 못하고 있는 실정이다. 그러나 언어의 정화한 앤트로피가 무엇인가를 알기 위하여 음과 음, 알파벳과 알파벳, 단어와 단어 사이의 상호 생기(生起) 확률에 대한 연구가 많이 추진되어 있다.

이 정보이론은 통신의 심볼(알파벳, Morse 신호 등)이 얼마나 정확하겠는가라는 문제에 초점을 둔 것이라 말할 수 있다. 그러나 전달된 심볼이 얼마나 똑바로 원하는 '뜻'을 전하는가 하는 차원에 대해서도 언어학, 특히 수리언어학(Mathematical Linguistics)의 입장에서 이용할 수 있는 적지 않은 계시를 준 이론이라고 본다.

다음 위이너의 정리를 열거하면 그 근본이 시계열(時系列) 이론으로 역시 전기적인 신호를 주로 쓰고 있으나 결국 Markov 과정인데 이것은 어떤 확률과정에 있어서 어떤 현상이 시간적으로 과거에서 현재로 계속되고 있을 때 그것이 유한의 몇 사상(事象)으로 나타나면 그 다음 현상은 확률적으로 추측할 수 있다는 것이다.

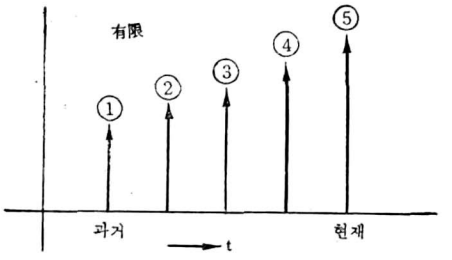

그림2. 時 系 列

즉 현재에서 과거까지의 몇 개의 사상을 관측하면 그 다음 것은 확률적으로 추측 가능하여, 가령 ① 다음 ②가 일어나는 확률이 결정되는 식으로 나간다. 그런데 현재란 하나의 현상만으로 미래의 일을 결정하는 것을 1차 마르코프과정(Markov process)라 하고, 바로 전 것과 그 다음 것 두 개로 결정되는 것을 2차, 이렇게 하여 3, 4, 5 차로 발전되어 가는 것이다. 시계열이란 이런 과정 전체중 ①,②,③을 알면 ④,⑤를 알아 낼 수 있는 것을 말하며, 시간적으로 봐서 각 관측에 확률적인 함수가 결정되어 있는 것을 마르코프 과정이라고 하여 시계열이론의 연구대상이 된다. 이 이론은 요즘 전기공학 뿐만 아니라 경제분야에서도 경기변동론중 경기를 측정하는데 이용하고 있다. 언어학에서는 언어를 기계적으로 인식할 때는 마르코프 과정에 대한 구조가 수량적으로 완연히 확립되기 전에는 많은 난점이 따를 것 같다.

다음 위이너-킨치너(Wiener-Khintchine)의 정리에 대해 한마디로 언급하면, 자기 상관함수와 전력 스펙트럼과의 관계를 하나의 수리적 변환으로 정의한 것으로 두 함수가 매우 밀접한 관계가 있음을 입증한 것이다.

2. 언어의 기계적 인식

언어의 기계적 인식의 필요성에 대해서 간단히 논하겠다. 이 문제는 전자계산기의 발달과 더불어 크게 등장된 것으로, 전자계산기 자체의 계산은 10^{-6} 초 정도로 굉장히 빠르므로 1 초 간에 100만 단위의 계산을 할 수 있으나 입력으로 집어 넣는 과정이나 출력으로 나오는 과정은 3000 단위 정도로 상당히 늦다. 또 현재로서는 여자 타자수가 펀치·카드나 테이프를 찍어서 입력으로 집어 넣어야 하기 때문에 그 찍는 속도가 문제되고 또 확인하는 것등의 과정이 시간과 경비를 많이 소요하게 된다.

그래서 이 과정을 간소화하여 궁극적으로는 입력이나 출력에서 글이나 말을 쓰려는 움직임이 활발해지고 있다. 이것을 위해서는 언어 가운데에서 특징을 뽑아내서 가령 자음에는 파열음등등이 있다는 것을 밝혀야 한다. 그러나 기계가 그것을 인식해서 알아 내는가 혹은 그것이 꼭 중요한 요인인가는 또 별개의 문제인 것이다.

여기서 특징의 추출은 여러가지를 생각할 수 있겠으며 그 특징의 정보량이 어떻게 되겠는가 하는 것이 더 큰 문제일 것으로 이것은 앞서의 정보이론 분야와도 밀접한 관계가 있다. 즉 정보이론의 한 분야인 패턴 인식에 관한 이론은 어떤 패턴에 있어서 그 특징이 추출되면 그 특징의 정보량 평가를 먼저 하게 된다. 그 다음 추출된 특징을 이용하는 문제로 어떻게 하면 가장 짧은 시간 내에 가장 효과적으로 그 특징의 패턴을 인식할 수 있는가는 방법론을 제시해야 한다.

이 때 특징의 추출에 있어서 대분하여 글씨와 말이 있겠는데, 글씨는 영어의 예와 한자, 일본글자(가다가나·히라가나)는 많이 연구되어져서 일부에서 실현하는 단계나 한글은 아직 손대지 못하고 있다. 한글은 다른 말과 달리 합자(合字)로 한 글자를 만들기 때문에, 분류하기 나름으로 42 가지의 단위를 나누기도 하지만 필자는 18 가지로 분류하여 위에서 말한 앤트로피인 개념을 적용하면, 평균하여 균일하게 있을 경우 $\log_2 18 = 4.2$ 다. 그러나 확률적으로 계산하면 2.7 정도다. 다시 말하면 이론상 세번만 yes, no를 거듭하면 답이 나오게 되어 있으나 실제로는 반드시 그런 결과가 나오지도 않는다. 필자의 실험에 따르면 기껏해야 평균 3.5 bit 정도, 즉 긴 것은 8 번 짧은 것은 2 번정도에 끝나는 구조밖에 체계화시킬 수 없었다. 이 구조는 좀더 연구해 봐야 할 과제다.

음소(phoneme)에 대하여 고찰하면 말은 음소라는 심볼의 연결로서 메시지가 부호화되어 있으며 따라서 말을 인식하는 첫 과제는 어떻게 사람이 말한 신호를 특정한 음소와 관련시키는가 하는 문제다. 음소를 특징지우는 패러미터에는 여러가지가 있으나 우선 제 1·제 2·제 3 포오먼트 주파수가 가장 큰 특징을 나타낸다고 보며 다음 기본 음성 주파주(fundamental frequeney), 즉 말의 기본 주파수로서 남자는 약 120-130, 여자는 그 두배 정도다. 다음 제 1·제 2 의 반공진(反共振) 주파수(anti-resonance frequency)로서 자음에서 특히 중요시되는

것이 있다. 또 음의 길이, 음의 순간적 강도와 평균의 강도 등을 들 수 있으나 이 특징의 상호관계가 어떻고 그 중 어느 패러미터가 가장 중요한가는 그 때의 대상이 되는 언어에 따라 달라질 것이다.

영어의 10개 모음에 대해서는 근래 Lindgren(1965)은 다음과 같이 4개의 중요한 특징으로 구분하고 있다.

1. Acute/Grave (제2 포오먼트의 High/Low)
2. Compact/Diffuse(제1 포오먼트의 High/Low)
3. Flat/Plain (F_1+F_2 Threshold F_1+F_2 Threshold)
4. Tense/Lax (장음/단음)

여기서 이들을 과연 어떤 순서로 비교해야만 가장 적합하겠는가 하는 문제가 자연히 제기된다. 즉 특징이 구분되면 그것을 이용해서 어떻게 패턴을 인식하는가를 생각해 보자. 위의 글에서는 1, 2, 3, 4 에 따라 각각의 모음들을 구분하는데, 영어에서 그 발생 확률이 가장 많은 i와 그 다음인 a가 (한글에서는 먼저 ㅏ와 다음 ㅣ로 뒤바뀜) 그림 3에서 보는 바와 같이 이 네 구분을 다 거쳐야 한다는 사실이 정보이론상 타당한 것이냐에 대해 많은 의문이 남게 된다.

위의 구분은 양자중 택일이니 그가 가진 최대의 앤트로피 $H=-\frac{1}{2}\log\frac{1}{2}=1$ bit다. 그런데 10개의 모음이 똑 같이 발생한다면 $H=-\frac{1}{10}\log\frac{1}{10}=3.32$ 다.

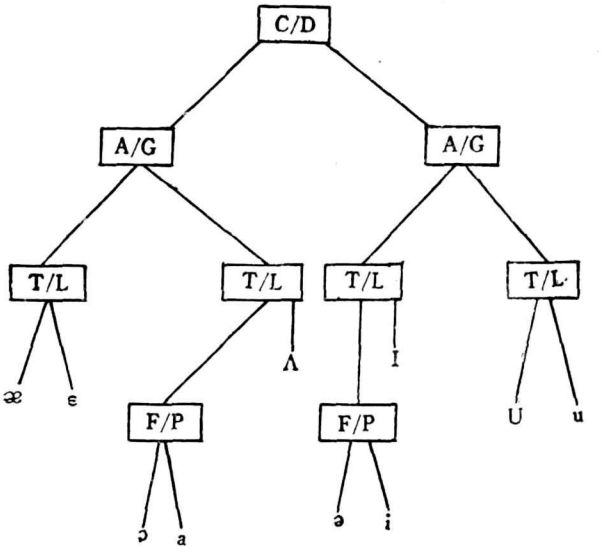

그림 3

따라서 4 가지의 구분은 꼭 필요하다는 계산이 나오고, 정확히 말하면 평균 3 번에 조금 더하는 정도로 구분될 것임을 가리킨다. 다음의 순서로는 제일 처음에 가장 균등히 양분하는 구분부터 채택해 나가야 한다.

그림 2는 이런 방식에 의한 결정을 표시한 것이다. 이외에 Neuron을 이용한 인식방법도 연구되고 있다. 아직도 음소의 식별을 위해서는 개인차가 별로 없는 불변의 특징을 얻든지 혹은 개인에 적응할 수 있는 기구가 마련되어야 한다.

언어의 인식에서 가장 절실한 것은 마르코프 과정으로 언어를 수리화하는 문제로서 음소와 음소간의 천이 확률에 의해서 음절이 구성되고 음절과 음절간의 천이확률에 의해서 단어가 구성되는 과정을 *Shannon diagram*으로 나타내는 문제다. 이 과정이 수량적으로 확립되고 그 확률에 의한 결정 방법이 세워져야만 완전한 기계인식이 이루어지리라고 본다. 이것은 인간이 글을 읽을 때 일어나리라고 추측되는 계층적(hierarchical)인 구조의 인식이 글자 자체의 인식과 더불어 매우 필요한 것이라는 공통된 결론에서 나온 것이다.

예를 들면 한글 ㅎ은 자음 확률중 하위나 '히'는 순위가 제 3위 정도니 ㅎ이 나오면 십중팔구 '히'가 되리라는 것을 암시한다. 이런 구조가 기계인식에서 매우 필요한 것이며, 이것이 앞으로 언어학자가 정보이론을 전공한 전자공학자와 공동 연구의 과제로서 추진되어야 할 점이라고 믿는다.

참 고 문 헌

(1) Shannon, C.E. and Weaver, W., *The Mathematical Theory of Communication*, Urbana: Univ. of Illinois Press, 1963

(2) Wiener, N., *Cybernetics*, Cambridge: MIT Press, 1961

(3) Reza, F.M., *An Introduction to Information Theory*, McGraw-Hill, New York, 1961

(4) Lindgren, N., *Machine Recognition of Human Language*, Part I-III, IEEE Spectrum, March-May, 1965.