간소화된 DTW방식을 이용한 한국어 숫자음 인식기 구현에 관한 연구

°안 병수, 정 광우, *홍 장석, 박 병철

성균관대학교 전자공학과, *서울보건전문대학 전산정보처리과

A Study on the Realization of Korean Digits Recognition System Using the Simplified DTW Method

°Byung Su An, Kwang Woo Chung, *Kwang Saug Hong, Byung Chul Park

Dept. of Electronics, Sung Kyun Kwan Univ., *Dept. of Computer Information Process, Seoul Health Junior College.

ABSTRACT

This paper describes the simplified DTW algorithm for real time korean digit recognition and construct the digit recognition system using that algorithm. The DTW algorithm which is used nowadays have problems on real time recognition because of its massive computation. But, simplified DTW algorithm, which is proposed in this paper, solved these problems. In the case of single syllable, we use the characteristic of uniform distribution of expansion and contraction on time axis, compare distance of input pattern and reference pattern using constrainedly restricted path. As a result, we can reduce a great deal of computation and achieved that the real time korean digit recognition system.

제 1 장. 서 론

가장 자연스러운 정보 전달 수단인 음성을 이용하여 모든 기계를 말로써 작동시키는, 즉 음성 인식(Speech Recognition)을 기계에 어떻게 구현시키는가 하는 문제가 오래전 부터 연구의 대상이 되어왔다. 이러한 음성 인식(Speech Recognition) 기술은 인식 대상에 따라 고립음 음성 인식(Isolated Word Recognition)과 연속음 음성 인식(Connected Word Recognition)으로, 또 개인차의 취급 방법에 따라 특정 화자 인식과 불특정 화자 인식으로 분류된다. 이들 인식 대상 및 개인차의 취급 방법은 인식 대상 어휘량의 크고 작음과 더불어 실제 음성 인식 구현에 크게 영향을 미친다.

음성 인식에 있어서 가장 문제시 되는것은 같은 음성에 있어서도 발성속도가 일정하지 않아 같은 사람이 같은 말을 하여도 음성의 길이가 다르며, 또한 같은 말을 여러사람이 발음을 하여도 길이의 변동이 크게될 뿐만 아니라, 발성기관의 크기가 개인에 따라 다르므로 발성기관의 형태를 동일하게 하여 발성하여도 포먼트 주파수(Formant Frequency)가 개인에 따라 차이가 생겨 화자마다 독특한 개인성이 나타나게 되어 화자 독립적인 음성 인식은 아직까지 그리 만족스럽지 못한 결과를 나타내고 있다.

이제까지 음성 인식을 위한 여러가지 인식 알고리즘들이 제시되었는데, 패턴매칭(Pattern Matching)을 이용한 DP-매칭(Dynamic Programming)방식, 벡터 양자화(Vector Quantization) 방식, 그리고 통계적 방법을 이용한 HMM(Hidden Marcov Model)이 그것이다. 최근에는 신경회로망과 퍼지논리를 이용한 알고리즘이 대두되어 음성인식을 위한 좋은 방법을 제시하고 있다. 위에서 열거된 가운데 DP-매칭 방법은 Dynamic Programming을 통하여 발성속도에 따른 시간변동을 가장 효율적으로 제거할 수있는 방법으로서 음성인식에 많이 이용된다. 하지만 이는 계산량의 방대함으로 인해 실시간 처리가 되지 않는다는 단점이 있다.

본 논문에서는 이러한 계산량의 증가와 시간적 변동을 효율적으로 흡수하기 위하여 간소화된 DP Matching법을 사용하여 한국어 숫자음 인식 실험을 하여, 실시간처리의 가능성을 제안하며 또한 범용 마이크로 프로세서인 모토롤라사의 MC 68020을 이용하여 한국어 숫자음 인식 시스템을 제작하여 인식 실험을 함으로서 음성 다이얼링 시스템의 실현기술을 향상시킴을 목적으로 한다.

제 2 장. DTW(Dynamic Time Warping)

음성의 발성변화에 따른 음성 패턴의 시간적 변동을 비선형적으로 정규화시키는 패턴 정합방식을 이용한 알고리즘으로서 비선형 Warping함수에 의해 비교되는 두 음성 패턴의 시간적 차이를 분석하고, 두 패턴들 사이의 오차거리를 계산하여 누적된 전체거리 계산값이 최소화되는 경로를 찾아내는 방법이다. 음성의 특징추출에 의한 특징벡터는

$R = [R_1, R_2, \ldots R_i, \ldots R_I]$: Reference Pattern (2-1)

$T = [T_1, T_2, \ldots T_j, \ldots T_J]$: Test Pattern (2-2)

로 나타내어질때, 두 특징 벡터의 Warping 함수는 다음과 같다.

$$F = C(1), C(2), \ldots, C(k), \ldots C(K) \qquad (2\text{-}3)$$

여기서 C(k) = (i(k), j(k))으로 표현된다. F가 Warping 함수이며 입력패턴과 표준패턴의 시간축을 투영하는 함수이다. 만약 두 패턴 사이에 시간차가 없다면 Warping함수는 i=j인 대각선상에 근접하게 되며, 그렇지 않으면 시간적 차이에 의해 대각선에서 벗어나게 된다. 따라서, DTW는 Warping 함수를 이용하여 최소화 거리가 되는 최적경로 i=W(j)를 찾게되는 것이다. 이식에서 i와 j사이의 관계를 함수적으로 표현하기 위하여 공통시간축 k를 만들고, 시간축의 함수 i, j를 k의 함수로 표현하면,

$$i = i(k), \quad k=1, 2, \ldots, K \qquad (2\text{-}4)$$

$$j = j(k), \quad k=1, 2, \ldots, K \qquad (2\text{-}5)$$

여기서 K는 공통시간축의 길이이다.

위에서 언급한 최적 경로를 얻기위한 Warping함수는 다음과 같은 조건을 만족해야 한다.

◆ 소구간 경로 제약

(1) 단조 증가 조건

$$i(k+1) \geq i(k) \qquad (2\text{-}6)$$

$$j(k+1) \geq j(k) \qquad (2\text{-}7)$$

(2) 연속 조건

$$i(k+1)-i(k) \leq 1 \qquad (2\text{-}8)$$

$$j(k+1)-j(k) \leq 1 \qquad (2\text{-}9)$$

위 두식으로 부터 연속되는 두점간에는 다음과 같은 관계를 갖는다. 즉 DP(Dynamic Programming) 방정식은 다음과같다.

$$\text{초기상태} : G(1,1) = 2*D(1,1) \qquad (2\text{-}10)$$

$$\text{그외}: G(i,j) = \text{Min} \begin{bmatrix} G(i,j-1) + D(i,j) \\ G(i-1,j-1)+2*D(i,j) \\ G(i-1,j) + D(i,j) \end{bmatrix} \qquad (2\text{-}11)$$

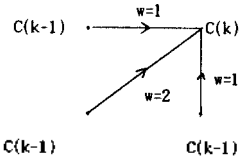

그림.2-1 소구간 경로 제약
Fig.2-1 The constraint of short path

◆ 전체 경로 제약

(3) 경계 조건

$$i(1)=1, j(1)=1 : \text{시작점 제약} \qquad (2\text{-}12)$$

$$i(K)=I, j(K)=J : \text{끝점 제약} \qquad (2\text{-}13)$$

(4) window 조건(그림 2-2)

정상적인 발성상태에서는 과도한 시간적 차이가 생기지 않으므로, 본 논문에서는 DTW의 계산시에 불필요한 계산을 줄이기위해 Window를 사용하는데 다음과 같이 나타내어진다.

$$R = | I - J | + 3 \qquad (2\text{-}14)$$

여기서 R은 Window의 크기이며 I는 표준패턴의 프레임 수이고 J는 시험패턴의 프레임수이다. 또한 시간축 정규화 거리는 다음식과 같다.

(5) 전체 경로 제약

두개의 패턴벡터 a,b의 거리척도 D(a,b)가 음성인식에 있어서 유효하기 위해서는 다음의 조건을 만족해야 한다.

$$\text{대칭성} : D(a,b) = D(b,a) \qquad (2\text{-}15)$$

$$\text{정칙성} : D(a,b) > 0 : a \neq b \qquad (2\text{-}16.a)$$

$$D(a,b) = 0 : a = b \qquad (2\text{-}16.b)$$

두 패턴 사이의 왜곡 정도의 측정은

$$D(c)=D(i(k),j(k))= \| R_i - T_j \| \qquad (2\text{-}17)$$

의 식으로 표시된다.

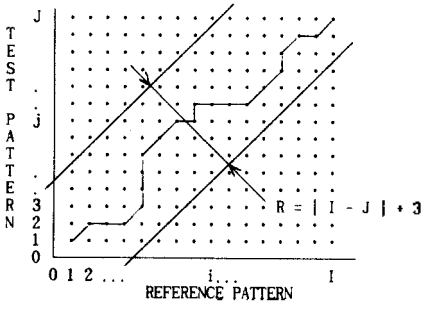

그림. 2-2 음성 패턴 R과 T에 대한 DP매칭 경로제약
Fig. 2-2 DP-matching path constraint for speech pattern R & T

제 3 장. 간소화된 DTW 알고리즘 [15]

음성의 발성은 같은 음성이라도 개인의 발성속도가 다르고 또한 같은 사람이라도 발성할 때마다 조금씩 발성속도가 다르다. 그러나 한 음성을 여러번 발음할 경우에 음성데이타의 시간축상의 신축성이 거의 균일하게 나타남을 알수있는데 이러한 성질을 이용하면 DTW알고리즘의 단점인 계산량의 방대함을 줄일 수 있어 높은 인식률과 더불어 실시간처리의 잇점도 가질수있다. 따라서 DTW의 계산시간을 줄이기 위하여 다음 그림.3-1과 같이 경로제약을 설정한다.

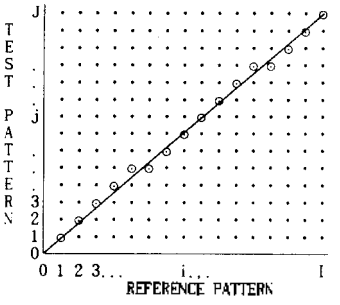

그림. 3-1 간소화된 DP 매칭 경로
Fig . 3-1 Simplified DP matching Path

경로선택을 위한 식은 다음과 같다.

$$j = INT \left| \frac{J-1}{I-1} (i-1) + 1 + 0.5 \right| \qquad (3\text{-}1)$$

위의 식에서는 시작점으로 부터 끝점 까지의 경로를 설정할 때 시작점에서 끝점까지 연결한 직선에 인접한 프레임중 식 (3-1)에 기준하여 프레임을 선택하여 DTW의 경로를 진행함으로서 소구간 경로제약 및 전체 경로제약이 필요없이 위의 식으로 부터 DTW의 진행경로가 강제적으로 얻어지게 된다.

제 4 장. 숫자음 인식기의 구조

4-1. 하드웨어 구성

본 논문에 사용되는 한국어 숫자음 인식기의 전체 블록 다이어그램 은 다음과 같다.

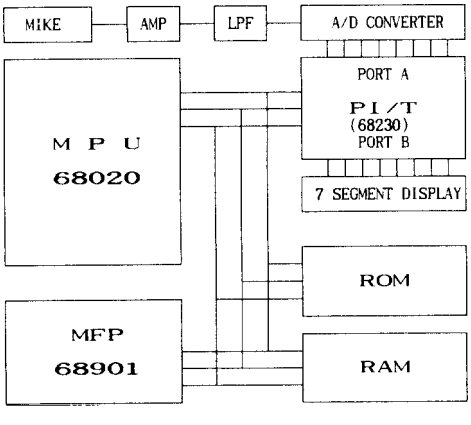

그림.4-1 전체적인 하드웨어 구성
Fig .4-1 Overall Hardware Configuration

■ MC68020 의 특징

MC68020은 모토롤라사의 68000계열중 첫 32비트 마이크로프로세서이다.32비트 레지스터들과 데이타버스 그리고 32비트 어드레스 버스, 풍부한 명령집합들, 다양한 어드레싱 모드를 갖추고있다. MC68020은 초기의 MC68000의 CPU들과 호환적이며 IEEE의 부동 소수점 연산을 완전히 지원하는 Floating Point Coprocessor(MC68881)를 인터페이스 할수있는 기능을 갖고있다.

□ MC68020 CPU의 특징을 요약하면 다음과 같다.

- 4G 바이트의 직접 번지 지정 능력
- 18가지의 어드레싱 방식
- 메모리 맵 입출력 방식
- Co-processor 인터페이싱 능력
- On-Chip Cache 명령
- MC68881 FPC 인터페이싱 능력

일반적으로 숫자음 인식과 같은 일반적인 소규모의 음성 인식 시스템에서는 8비트나 16비트 프로세서로는 충분한 메모리 공간을 확보하기 어렵고 또한 실시간 처리를 할수 없게 된다.장래의 확장성을 고려하며 여유있는 메모리를 확보하고(32Bit의어드레스버스) 또한 실시간 처리를 위해서는 고속도이며 대용량의 메모리를 억세스 할수있는 모토롤라사의 MC68020 프로세서를 사용하였다.

■ PI/T 68230 (Parallel Interface & Timer)

PI/T 68230은 68000 MPU의 주변장치로 개발되었으며 입출력 포트 및 타이머를 내장하고 있다. 병렬 입출력 포트는 단일 방향 혹은 양방향으로 동작 시킬수 있으며, 타이머는 24비트 다운 카운터와 5비트 프리스케일러를 갖고 있으며 시스템클록 및 내부클록을 카운팅 할수있다. PI/T는 벡터 혹은 오토벡터 방식의 인터럽트 어느쪽이나 가능하다. 여기서는 A/D 컨버터로 부터 디지타이즈된 음성신호를 얻기 위하여 그리고 인식된 결과를 7세그먼트로 출력하기 위한 소자이다.

■ A/D Converter (National ADC 0800 PCD)

일반적으로 음성인식을 위해 사용되는 A/D컨버터의 해상도는 12비트의 것이 많이 사용된다. 그러나 이는 하드웨어 구성시에 메모리를 효율적으로 사용하지 못하며, 또한 인식률에도 크게 영향을 미치지 않는다. 이러한 이유로 본 논문에서는 내쇼날사의 8비트 ADC 0800 A/D컨버터를 사용하였다.

4-2. 음성입력부의 설계

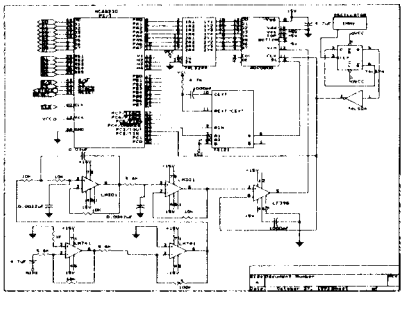

그림.4-2 A/D컨버터와 PI/T 와의 인터페이스 회로도.
Fig .4-2 A/D Converter & PI/T Interface Circuits

제 5 장 실험및 고찰

5-1. 음성 분석 조건

본 논문에서 사용된 음성신호의 분석조건은 표.5-1과 같다. 사용하는 분석 파라미터인 선형예측계수는 프레임길이 10ms에 대해서 프레임 사이에 중첩없이 10차의 선형예측계수를 사용하였다.

표 5-1. 분석조건

샘플링 주파수	10KHz
A/D해상도	8 bit
L P F	4KHz
Frame 길이	10ms
분석 파라미터	LPC계수
분석 차수	10 차
화 자 수	남성 1 인 (화자종속)
발성회수	각 숫자음 40회
인식단위	음절 단위
인식 방법	DTW
	간소화된 DTW

실험에 사용된 음성데이타는 화자종속 실험에서 남성화자 1인을 대상으로 한국어 숫자음 /공/,/일/,/이/,/삼/,/사/,/오/,/육/,/칠/,/팔/,/구/를 각각 40회 발성한 데이타를 실험에 사용하였다. 저장된 표준패턴의 구조는 인식대상인 10개의 숫자음에 대하여 각각 10개씩 총 100개의 표준패턴이 구성되어 있다. 나머지 30개의 단어는 인식실험에 사용한 데이타이다.

5-2. 실험 방법

전처리가 끝난 음성신호는 다음 그림.5-1과 같은 후처리 단계를 거쳐 인식된다. 전처리 단계에서 들어온 음성은 특징벡터 추출후 미리 저장된 표준패턴과 비교된다. 후보단어중 인식단어를 선택하기 위하여 DTW에 의해 계산된 거리값들에 대하여 KNN(K-nearest neighbour)규칙을 적용하여 계산하였다.

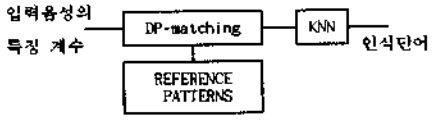

그림.5-1 인식 실험 절차
Fig .5-1 Procedure for recognition experiment.

5-3. 인식 실험 결과

본 논문에서는 한국어 숫자음 인식 실험을 위해 기존의 DTW 알고리즘과 인식시스템의 실시간 처리를 위해 고안된 새로운 방법인 간소화된 DTW알고리즘을 이용해서 실제로 한국어 숫자음 인식기를 구현하여 실험실내의 잡음환경에서 실험하였다. 먼저 컴퓨터 시뮬레이션 결과 기존의 방법에 비해 제안한 방법이 0.5% - 1%정도 인식률이 떨어진다[15]. 이러한 인식률의 저하는 알고리즘의 특성상 정확한 유음구간 검출알고리즘이 요구되는데 이러한 조건을 만족시킬만한 좋은 유음구간의 검출알고리즘과 함께 적용된다면 인식률의 저하는 막을수 있으리라 생각된다. 이에반해 인식시간의 급격한 저하는 실시간 인식 시스템의 구성에 문제가 되어왔던 인식시간을 현저히 줄임으로써 음성인식 시스템의 실시간 인식을 가능하게하는 좋은 알고리즘이라 생각된다.

각각의 인식시간에 대한 비교를 그림.5-2에 나타내었다. 그림.5-2에 나타난 바와 같이 두 알고리즘간에는 작게는 14배(40프레임의 경우)에서 크게는 약25배(30프레임의경우)에 이르는 현저한 차이가 있다. 이러한 현저한 인식시간의 차이는 음성인식 시스템의 실시간처리를 충분히 지원하며 표준패턴의 수를 늘인다면 약간의 인식률의 저하는 충분히 막을수있으리라 생각된다. 두 알고리즘 각각에 대한 DTW경로의 비교를 그림.5-3에 나타내었다. 보는 바와 같이 기존의 방법에 있어서도 입력음성과 표준음성이 일치하는경우 경로가 대각선에 근접하여 진행되므로 정확한 유음구간 검출이 이루어진다면 단음절 발성인경우 오히려 인식률의 상승이 기대된다. 인식 실험 시스템은 MC68020마이크로 프로세서를 사용하여 숫자음 인식기를 구성하여 실험하였으며 인식결과는 표5-2에 나타내었다. 전체인식시간은 음성입력이 끝나고부터 유음구간검출 평균시간이 0.18초 그리고 선형예측계수 추출시간이 평균 0.5696초 DTW거리 계산시간이 제안한 방법의 경우에 0.6419초로서 전체 인식시간은 평균1.39초가 소요된다. 이에반해 기존의 방법은 DTW시간이 평균 10.03초이며 전체 인식 시간은 10.775초가 소요된다.

표 5-2 화자종속의 경우 인식률

I O	일	이	삼	사	오	육	칠	팔	구	공	인식률
일	30										100%
이		30									100%
삼			28					2			93.3%
사			2	27				1			93.3%
오					30						100%
육		1				28			1		93.3%
칠							30				100%
팔								30			100%
구									30		100%
공					1					29	96.7%
평 균 인 식 률											97.3%

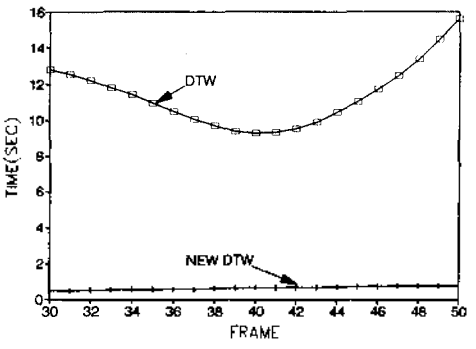

그림 5-2 인식 시간의 비교

Fig 5-2 Comparison of recognition time

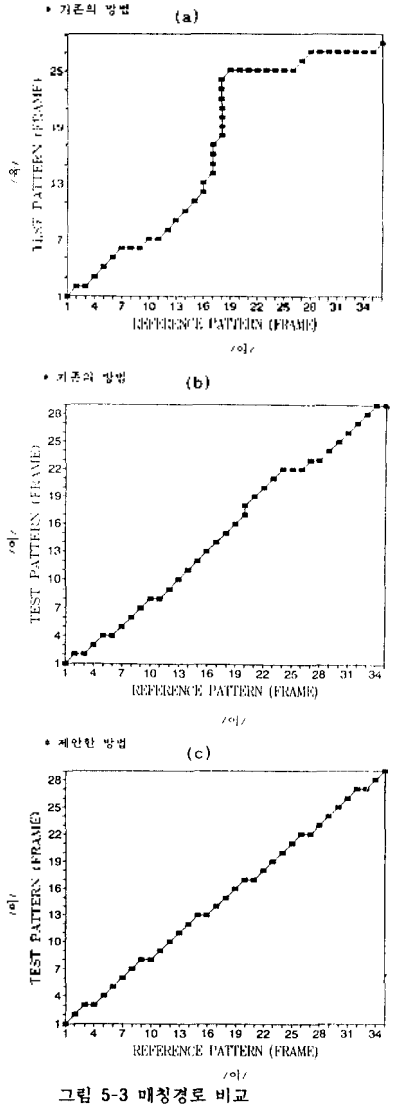

그림 5-3 매칭경로 비교

Fig 5-3 Comparison of Matching Path

제 6 장. 結 論

지금까지 한국어 숫자음 인식기의 구현과 또 이를 실시간으로 빠르게 수행할수있는 방법인 간소화된 DTW알고리즘에 관해 기술하였다. 위에서 언급한 여러사실에서 보듯이 구성된 숫자음 인식시스템은 실시간 인식과 더불어 비교적 높은 인식률을 나타냄으로서 제시한 알고리즘의 타당성을 입증하게 되었다. 그러나 약간의 인식률의 저하를 가져오게 되는데, 이는 표준패턴의 갯수를 늘린다던지 또는 정확한 유음구간검출 알고리즘을 적용한다면 충분히 보상되리라 생각된다.그러나 인식대상이 단음절인 경우에 국한된다면 오히려 인식률의 향상을 기대할수 있을것이다. 결국 이러한 실시간 인식기의 구성은 소규모의 단어인식시스템을 구현하여 실제 제품에의 적용가능성을 제시하였으며 더나아가 연속숫자음 인식기의 실시간 실현을 보장할수 있으리라 생각된다.

參 考 文 獻

1. "MC68020 User's Manual",3ed,PRENTICE HALL 1990
2. Shuzo Saito, Kazuo Nakata," Fundamentals of Speech Signal Processing ", Academic Press, 1985
3. Sadaoki Furui, M.Mohan Sondhi, " Advances in Speech Signal Processing " ,1992
4. T.Stvendson,F.K.Soong,"On the Automatic Segmentation of Speech Signals" ,Proc.ICASSP 87, 1987.
5. Laurence. R. Rabiner, M. R. Sambur," An Algorithm for Determining the Endpoints of Isolated Utterances ",The Bell System Technical Journal, Vol.54, No.2, February, 1975.
6. 홍광석,"규칙을 이용한 포먼트 추정과 한국어 연속단어음성인식 ",성균관대학교 박사학위논문, 1991
7. A. H. Gray, J. D. Markel, " Distance Measure for Speech Processing",IECE Trans.,Acoustics, speech and Signal processing, Vol.ASSP-24,No.5, 1976.
8. H. Saoke, S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Trans, Acuust., Speech and Signal Processing", Vol.ASSP-26, 1978.
9. Laurence. R.Rabiner, "Isolated and Connected Word Recognition-Theory and Selected Applications",IEEE Trans. on Communication,May, 1981.
10. "音聲 タイプライタ の 設計" ,November,CQ出版社,1983.
11. "MC68000 Course Notes", Motorola CO.
12. "MC68000 Educational Computer Board Manual", Motorola CO.
13. Alan D. Wilcox, " 68000 Micro Computer Systems Design and Troubleshooting ", PRENTICE-HALL,INC.,1987.
14. L.A.Leventhal, D.Hawkins, G.Kane and W.D.Cramer, "68000 Assembly Language Progamming",McGRAW-HILL,1986.
15. 정광우,홍광석,박병철,"한국어 숫자음 인식을 위한 간소화된 DTW에 관한 연구",대한전자공학회 하계학술대회 논문집, 1992