



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2025년04월23일
(11) 등록번호 10-2799088
(24) 등록일자 2025년04월17일

(51) 국제특허분류(Int. Cl.)
G06F 16/906 (2019.01) G06F 16/28 (2019.01)
G06F 18/00 (2023.01)
(52) CPC특허분류
G06F 16/906 (2019.01)
G06F 16/285 (2019.01)
(21) 출원번호 10-2021-7019429
(22) 출원일자(국제) 2019년02월06일
심사청구일자 2021년12월06일
(85) 번역문제출일자 2021년06월23일
(65) 공개번호 10-2021-0118067
(43) 공개일자 2021년09월29일
(86) 국제출원번호 PCT/JP2019/004315
(87) 국제공개번호 WO 2020/161845
국제공개일자 2020년08월13일
(56) 선행기술조사문헌
JP05205058 A
US20080058668 A1

(73) 특허권자
도호쿠 다이가쿠
일본 미야기켄 센다이시 아오바쿠 가타히라 2쵸메 1반 1고
(72) 발명자
엔도오 테츠오
일본 9808577 미야기켄 센다이시 아오바쿠 가타히라 2쵸메 1반 1고 도호쿠 다이가쿠 내
센 후이
일본 9808577 미야기켄 센다이시 아오바쿠 가타히라 2쵸메 1반 1고 도호쿠 다이가쿠 내
마 이타오
일본 9808577 미야기켄 센다이시 아오바쿠 가타히라 2쵸메 1반 1고 도호쿠 다이가쿠 내
(74) 대리인
장수길, 이증희, 허용

전체 청구항 수 : 총 13 항

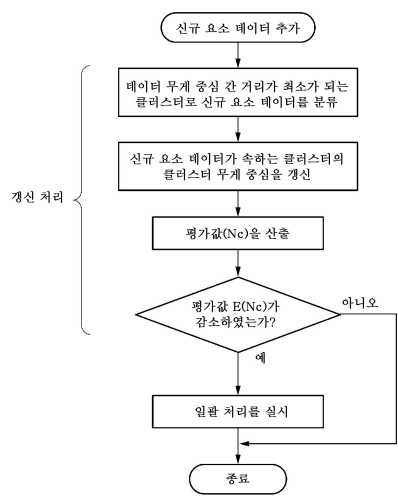
심사관 : 안지현

(54) 발명의 명칭 클러스터링 장치 및 클러스터링 방법

(57) 요약

효율적으로 고정밀도의 클러스터링에 이바지하는 클러스터링 장치 및 클러스터링 방법을 제공한다. 클러스터수를 변화시키면서, 모든 요소 데이터를 사용한 클러스터링을 행함으로써 얻어지는 각 클러스터수에 대한 평가값으로부터 최적의 클러스터수를 구한다. 구한 최적의 클러스터수로 요소 데이터를 분류한 클러스터링 결과를 얻는다. 이 일괄 처리 후에, 신규 요소 데이터가 추가된 경우에는, 신규 요소 데이터를 가장 가까운 기존의 클러스터로 분류한다. 그 후에, 평가값을 구하여 클러스터링 결과의 타당성을 판단한다.

대표도 - 도15



(52) CPC특허분류
G06F 18/2321 (2023.01)

명세서

청구범위

청구항 1

복수의 요소 데이터를 분류하는 클러스터링 장치에 있어서,
 상기 요소 데이터를 기억하는 데이터 기억부와,
 분류 결과를 평가하는 평가값을 산출하는 평가값 산출부와,
 설정되는 클러스터수의 어느 클러스터로 상기 복수의 요소 데이터를 분류함으로써 얻어지는 상기 평가값에 기초하여, 최적의 클러스터수의 상기 클러스터로 상기 복수의 요소 데이터를 분류하는 일괄 처리부와,
 추가되는 신규의 요소 데이터를, 상기 일괄 처리부에 의해 상기 복수의 요소 데이터가 분류되어 있는 복수의 상기 클러스터 중 상기 신규의 요소 데이터에 가장 거리가 가까운 상기 클러스터로 분류하는 갱신 처리부와,
 상기 갱신 처리부가 상기 신규의 요소 데이터를 분류함으로써 얻어지는 상기 평가값과, 당해 신규의 요소 데이터를 분류하기 전의 상기 평가값을 비교함으로써, 상기 신규의 요소 데이터의 분류 후의 분류 결과의 타당성을 판단하는 판단부를 구비하는 것을 특징으로 하는, 클러스터링 장치.

청구항 2

제1항에 있어서,
 상기 판단부는, 분류 결과가 타당하지 않은 경우에, 상기 일괄 처리부에 의한 상기 신규의 요소 데이터를 포함하는 모든 요소 데이터의 분류를 실행시키는 것을 특징으로 하는, 클러스터링 장치.

청구항 3

제1항 또는 제2항에 있어서,
 상기 데이터 기억부는, 불휘발성 메모리로 구성되는 것을 특징으로 하는, 클러스터링 장치.

청구항 4

제1항 또는 제2항에 있어서,
 상기 일괄 처리부는, 상기 요소 데이터가 분류되는 클러스터 정보를 보유하는 레지스터부와 데이터 계산을 행하는 계산 회로를 포함하고,
 상기 평가값 산출부는, 상기 일괄 처리부가 상기 복수의 요소 데이터를 분류할 때에 얻어지는 상기 평가값을 계산하기 위한 데이터를 보유하는 평가 레지스터부와 상기 평가값을 계산하는 로직부를 갖고,
 전원 공급이 독립적으로 제어되는 복수의 파워드 메인이 형성되고, 적어도 상기 데이터 기억부와, 상기 계산 회로와, 상기 갱신 처리부와, 상기 로직부가 다른 상기 파워드 메인으로 나뉘고 동시에, 이들 상기 파워드 메인과는 다른 상기 파워드 메인으로 상기 레지스터부 및 평가 레지스터부가 나뉘어 있는 것을 특징으로 하는, 클러스터링 장치.

청구항 5

제1항 또는 제2항에 있어서,
 상기 평가값 산출부는, 상기 클러스터 내의 상기 요소 데이터의 분산의 정도를 나타내는 각각의 상기 클러스터에 대한 제1 지표값을 당해 클러스터의 데이터수에 기초하는 제1 값으로 규격화한 값의 각 상기 클러스터의 제1 총합값인 내적 결합도와, 상기 클러스터 간의 거리의 지표가 되는 각각의 상기 클러스터에 대한 제2 지표값의 제2 총합값을 상기 클러스터수에 기초하는 제2 값으로 규격화한 외적 분리도를 각각 구하여, 상기 내적 결합도와 상기 외적 분리도를 변수로 하는 소정의 연산식으로부터 상기 평가값을 산출하는 것을 특징으로 하는, 클러스터링 장치.

청구항 6

제5항에 있어서,

상기 일괄 처리부는, 상기 요소 데이터를 분류할 때에 각각의 상기 클러스터에 대해서, 상기 클러스터 내의 요소 데이터를 가산한 데이터 가산값과 상기 요소 데이터의 개수를 구하여, 상기 데이터 가산값과 상기 요소 데이터의 개수로부터 당해 클러스터의 클러스터 무게 중심을 산출하고,

상기 평가값 산출부는, 각각의 상기 클러스터에 대해서, 상기 일괄 처리부에서 구해진 상기 데이터 가산값과 상기 요소 데이터의 개수를 취득하고, 취득한 상기 데이터 가산값과 상기 요소 데이터의 개수를 사용하여 당해 클러스터의 상기 제2 지표값을 구하고, 각 상기 클러스터에 대한 상기 요소 데이터의 개수 및 상기 제2 지표값을 보유하고, 보유하고 있는 상기 제2 지표값을 사용하여 상기 평가값을 산출하는 것을 특징으로 하는, 클러스터링 장치.

청구항 7

제6항에 있어서,

상기 평가값 산출부는, 상기 신규의 요소 데이터의 분류 후의 클러스터 무게 중심을, 상기 신규의 요소 데이터의 분류 전의 당해 클러스터에 대한 요소 데이터의 개수, 클러스터 무게 중심 및 상기 신규의 요소 데이터를 사용하여 구하고, 구한 분류 후의 상기 클러스터 무게 중심을 사용하여, 당해 클러스터의 상기 제2 지표값을 갱신하고, 갱신된 상기 제2 지표값을 사용하여 분류 후의 상기 평가값을 산출하는 것을 특징으로 하는, 클러스터링 장치.

청구항 8

복수의 어느 클러스터로 복수의 요소 데이터를 분류하는 클러스터링부와,

상기 복수의 요소 데이터를 분류하는 상기 클러스터링부의 계산 과정의 계산값을 사용하여, 상기 클러스터 내의 상기 요소 데이터의 분산의 정도를 나타내는 각각의 상기 클러스터에 대한 제1 지표값을 당해 클러스터의 데이터수에 기초하는 제1 값으로 규격화한 값의 각 상기 클러스터의 제1 총합값인 내적 결합도와, 상기 클러스터 간의 거리의 지표가 되는 각각의 상기 클러스터에 대한 제2 지표값의 제2 총합값을 클러스터수에 기초하는 제2 값으로 규격화한 외적 분리도를 각각 구하여, 상기 내적 결합도와 상기 외적 분리도를 변수로 하는 소정의 연산식으로부터, 상기 클러스터링부에 의한 분류 결과를 평가하는 평가값을 산출하는 평가값 산출부를 구비하는 것을 특징으로 하는, 클러스터링 장치.

청구항 9

제8항에 있어서,

상기 클러스터링부는, 1개 또는 복수의 신규의 요소 데이터의 추가에 응답하여, 적어도 상기 신규의 요소 데이터를 어느 상기 클러스터로 분류하는 것을 특징으로 하는, 클러스터링 장치.

청구항 10

제8항 또는 제9항에 있어서,

상기 클러스터링부는, 상기 요소 데이터를 분류할 때에 각각의 상기 클러스터에 대하여 클러스터 무게 중심을 산출하는 무게 중심 산출 회로를 갖고,

상기 무게 중심 산출 회로는,

상기 클러스터 내의 상기 요소 데이터를 가산한 데이터 가산값을 산출하는 데이터 가산기와,

상기 클러스터 내의 상기 요소 데이터에 대응한 신호수로부터 상기 클러스터 내의 상기 요소 데이터의 개수를 산출하는 개수 산출기와,

상기 데이터 가산기로 얻어지는 상기 데이터 가산값을 보유하는 제1 레지스터와,

상기 개수 산출기로 얻어지는 상기 요소 데이터의 개수를 보유하는 제2 레지스터와,

상기 제1 레지스터에 보유되어 있는 상기 데이터 가산값을 상기 제2 레지스터에 보유되어 있는 상기 요소 데이터의 개수로 제산함으로써 상기 클러스터 무게 중심을 산출하는 제산기를 갖고,

상기 평가값 산출부는, 각각의 상기 클러스터에 대해서, 상기 제1 레지스터로부터 취득한 상기 데이터 가산값과 상기 제2 레지스터로부터 취득한 상기 요소 데이터의 개수를 사용하여 상기 제2 지표값을 구하고, 각 상기 클러스터에 대한 상기 제2 지표값을 사용하여 상기 평가값을 산출하는 것을 특징으로 하는, 클러스터링 장치.

청구항 11

클러스터수를 변화시키면서, 모든 요소 데이터를 사용한 분류를 행함으로써 얻어지는 각 클러스터수에 대한 분류 결과를 평가하는 평가값으로부터 최적의 클러스터수를 구하고, 구한 상기 최적의 클러스터수로 상기 요소 데이터를 분류한 분류 결과를 얻는 일괄 처리 공정과,

상기 일괄 처리 공정 후에 신규의 요소 데이터가 추가되었을 때에, 상기 신규의 요소 데이터를 기존의 클러스터의 어느 것으로 분류하고, 분류 후에 상기 평가값을 구하는 갱신 처리 공정과,

상기 갱신 처리 공정에서 상기 신규의 요소 데이터의 분류 후에 얻어지는 상기 평가값과, 당해 신규의 요소 데이터의 분류 전의 상기 평가값을 비교함으로써, 상기 갱신 처리 공정에 의한 상기 신규의 요소 데이터의 분류 후의 분류 결과의 타당성을 판단하는 판단 공정을 갖는 것을 특징으로 하는, 클러스터링 방법.

청구항 12

제11항에 있어서,

상기 판단 공정에서 상기 분류 결과가 타당하지 않다는 판단이 이루어진 경우에, 상기 신규의 요소 데이터를 포함하는 모든 요소 데이터를 사용하여 상기 일괄 처리 공정을 실시하는 것을 특징으로 하는, 클러스터링 방법.

청구항 13

제11항 또는 제12항에 있어서,

상기 평가값은, 상기 클러스터 내의 상기 요소 데이터의 분산의 정도를 나타내는 각각의 상기 클러스터에 대한 제1 지표값을 당해 클러스터의 데이터수에 기초하는 제1 값으로 규격화한 값의 각 상기 클러스터의 제1 총합값인 내적 결합도와, 상기 클러스터 간의 거리의 지표가 되는 각각의 상기 클러스터에 대한 제2 지표값의 제2 총합값을 상기 클러스터수에 기초하는 제2 값으로 규격화한 외적 분리도를 각각 구하여, 상기 내적 결합도와 상기 외적 분리도를 변수로 하는 소정의 연산식으로부터 산출되는 것을 특징으로 하는, 클러스터링 방법.

발명의 설명

기술 분야

[0001] 본 발명은 클러스터링 장치 및 클러스터링 방법에 관한 것이다.

배경 기술

[0002] 분류 대상의 복수의 요소 데이터의 집합을 복수의 클러스터(데이터군)로 분류하는 클러스터링(클러스터 분석)이 알려져 있고, 근년, 화상 해석, 데이터 마이닝, 빅 데이터의 해석 등에서 빈번히 이용되고 있다. 또한, 이러한 클러스터링은, 기계 학습 중 비지도 학습이 된다. 클러스터링에서는, 유사한 요소 데이터끼리를 동일한 클러스터로 분류하고, 클러스터 간에서는 요소 데이터가 가능한 한 유사하지 않도록 분류함으로써, 그 분류 결과로부터 요소 데이터의 경향이나 특징의 추출 등을 행하는 것이다.

[0003] 복수의 요소 데이터를 클러스터로 분류하는 클러스터링 방법으로서, 다양한 방법이 알려져 있고, 대표적인 클러스터링의 방법의 하나로서 k-means법이 알려져 있다. k-means법에서는, 미리 클러스터수 k가 설정되고, 예를 들어 N개의 전체 요소 데이터 중으로부터 임의의 k개의 요소 데이터가 선택되어, 그들을 k개의 클러스터의 무게 중심의 초깃값으로 한다(수순 1). 이어서, 각 요소 데이터 각각에 대해서, 당해 요소 데이터와의 거리가 가장 짧은 무게 중심의 클러스터로 분류한다(수순 2). 각 클러스터에 대해서, 클러스터 내의 요소 데이터의 평균을 새로운 클러스터의 무게 중심으로 설정한다(수순 3). 그리고, 각 클러스터의 무게 중심이 변화하지 않게 될 때까지, 수순 2와 수순 3을 반복하여 행한다.

[0004] 상기와 같은 클러스터링 방법은, 미리 설정된 클러스터수의 클러스터로 요소 데이터를 분류하는 것으로서, 클러스터수를 미리 설정할 필요가 있다. 통상, 최적의 클러스터수를 결정하기 위해서, 클러스터수를 변화시켜서 클러스터링한 각 결과 각각에 관한 평가값을 구하고, 평가값이 극값, 혹은 최댓값 내지 최솟값의 클러스터수를 최적의 클러스터수로 하고 있다.

[0005] 분류해야 할 신규 요소 데이터가 추가된 경우에, 신규 요소 데이터를 포함하는 전체 요소 데이터를 사용하여, 상기와 같이 최적의 클러스터수를 결정하고, 최적의 클러스터수로 클러스터링을 행하는 방법이 알려져 있다(비특허문헌 1을 참조). 또한, 신규 요소 데이터가 추가된 경우에, 신규 요소 데이터와의 거리가 최소가 되는 클러스터 무게 중심을 특정하고, 그 특정한 클러스터 무게 중심의 클러스터로 신규 요소 데이터를 분류하는 방법이 알려져 있다(비특허문헌 2를 참조).

선행기술문헌

비특허문헌

[0006] (비특허문헌 0001) Z. Hou et.al., "Real-Time Very Large-Scale Integration Recognition System with an On-Chip Adaptive K-Means Learning Algorithm", JJAP, Vol. 52,04CE11
 (비특허문헌 0002) Chen, Tse-Wei, and Makoto Ikeda. "Design and implementation of low-power hardware architecture with single-cycle divider for on-line clustering algorithm."IEEE Transactions on Circuits and Systems I: Regular Papers60.8(2013): 2165-2176.

발명의 내용

해결하려는 과제

[0007] 그런데, 신규 요소 데이터가 추가된 경우에, 신규 요소 데이터를 포함하는 전체 요소 데이터를 사용하여, 최적의 클러스터수를 구하여 클러스터링을 행하는 방법은, 고정밀도의 분류 결과를 기대할 수 있는 반면, 연산 부하가 크기 때문에, 효율이 나뉘었다. 또한, 클러스터링 결과에 대하여 고속의 응답이 요구되는 용도에는 부적합하다. 한편, 신규 요소 데이터가 추가된 경우에, 신규 요소 데이터와의 거리가 최소가 되는 클러스터 무게 중심의 클러스터로 신규 요소 데이터를 분류하는 방법은, 클러스터링 결과에 대하여 고속의 응답성이 얻어지지만, 분류 결과의 정밀도가 낮아지는 경우가 있다. 이것은, 요소 데이터의 추가 전에 있어서의 클러스터수를 요소 데이터의 추가 후에 있어서도 적정으로 간주하고 있기 때문이다.

[0008] 본 발명은 상기 사정을 감안하여 이루어진 것이며, 효율적으로 고정밀도의 클러스터링에 도움을 주는 클러스터링 장치 및 클러스터링 방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0009] 본 발명의 클러스터링 장치는, 복수의 요소 데이터를 분류하는 클러스터링 장치에 있어서, 상기 요소 데이터를 기억하는 데이터 기억부와, 분류 결과를 평가하는 평가값을 산출하는 평가값 산출부와, 설정되는 클러스터수의 어느 클러스터로 상기 복수의 요소 데이터를 분류함으로써 얻어지는 상기 평가값에 기초하여, 최적의 클러스터수의 상기 클러스터로 상기 복수의 요소 데이터를 분류하는 일괄 처리부와, 추가되는 신규의 요소 데이터를, 상기 일괄 처리부에 의해 상기 복수의 요소 데이터가 분류되어 있는 복수의 상기 클러스터 중 상기 신규의 요소 데이터에 가장 가까운 상기 클러스터로 분류하는 갱신 처리부와, 상기 갱신 처리부가 상기 신규의 요소 데이터를 분류함으로써 얻어지는 상기 평가값에 기초하여, 상기 신규의 요소 데이터의 분류 후의 분류 결과의 타당성을 판단하는 판단부를 구비하는 것이다.

[0010] 본 발명의 클러스터링 장치는, 복수의 어느 클러스터로 복수의 요소 데이터를 분류하는 클러스터링부와, 상기 복수의 요소 데이터를 분류하는 상기 클러스터링부의 계산 과정의 계산값을 사용하여, 상기 클러스터 내의 상기 요소 데이터의 분산의 정도를 나타내는 각각의 상기 클러스터에 대한 제1 지표값을 당해 클러스터의 데이터수에 기초하는 제1 값으로 규격화한 값의 각 상기 클러스터의 제1 총합값인 내적 결합도와, 상기 클러스터 간의 거리의 지표가 되는 각각의 상기 클러스터에 대한 제2 지표값의 제2 총합값을 상기 클러스터수에 기초하는 제2 값으로 규격화한 외적 분리도를 각각 구하여, 상기 내적 결합도와 상기 외적 분리도를 변수로 하는 소정의 연산식을

로부터, 상기 클러스터링부에 의한 분류 결과를 평가하는 평가값을 산출하는 상기 평가값 산출부를 구비하는 것이다.

[0011] 또한, 본 발명의 클러스터링 방법은, 클러스터수를 변화시키면서, 모든 요소 데이터를 사용한 분류를 행함으로써 얻어지는 각 클러스터수에 대한 분류 결과를 평가하는 평가값으로부터 최적의 클러스터수를 구하고, 구한 상기 최적의 클러스터수로 상기 요소 데이터를 분류한 분류 결과를 얻는 일괄 처리 공정과, 상기 일괄 처리 공정 후에 신규의 요소 데이터가 추가되었을 때에, 상기 신규의 요소 데이터를 기존의 상기 클러스터의 어느 것으로 분류하고, 분류 후에 상기 평가값을 구하는 갱신 처리 공정과, 상기 갱신 처리 공정에서 얻어지는 상기 평가값에 기초하여, 상기 갱신 처리 공정에 의한 분류 결과의 타당성을 판단하는 판단 공정을 갖는 것이다.

발명의 효과

[0012] 본 발명에 따르면, 모든 요소 데이터를 사용하여 구해지는 최적의 클러스터수로 요소 데이터를 분류한 후에, 신규 요소 데이터가 추가되었을 때에는, 신규 요소 데이터를 기존의 복수의 클러스터 중 가장 가까운 클러스터로 분류하고, 이 분류 후에 얻어지는 평가값에 기초하여 분류 결과의 타당성을 판단하므로, 효율적으로 고정밀도의 클러스터링을 행할 수 있게 된다.

[0013] 또한, 본 발명에 따르면, 상기 복수의 요소 데이터를 분류하는 상기 클러스터링부의 계산 과정의 계산값을 사용하여, 클러스터 내의 상기 요소 데이터의 분산의 정도를 나타내는 내적 결합도와, 클러스터 간의 분리의 정도를 나타내는 외적 분리도에 기초하는 평가값을 산출하므로, 과잉의 분류를 억제한 고정밀도의 평가값을 효율적으로 구할 수 있고, 그 결과 효율적으로 고정밀도의 클러스터링을 행할 수 있게 된다.

도면의 간단한 설명

- [0014] 도 1은 본 발명을 실시한 클러스터링 장치의 구성을 도시하는 블록도이다.
- 도 2는 메인 메모리 및 지연 회로의 구성을 도시하는 설명도이다.
- 도 3은 거리 산출 회로의 셀의 구성을 도시하는 회로도이다.
- 도 4는 최댓값 검출 회로의 구성을 도시하는 회로도이다.
- 도 5는 CID 마스크 회로의 구성을 도시하는 회로도이다.
- 도 6은 무게 중심 산출 회로의 구성을 도시하는 회로도이다.
- 도 7은 근린 탐색 회로의 구성을 도시하는 회로도이다.
- 도 8은 평가값 산출 회로의 구성을 도시하는 블록도이다.
- 도 9는 거리 레지스터부의 셀에 접속된 인에이블 신호 회로를 도시하는 회로도이다.
- 도 10은 MID 레지스터로 접속된 인에이블 신호 회로를 도시하는 회로도이다.
- 도 11은 연산 유닛의 파워드 메인을 도시하는 설명도이다.
- 도 12는 일괄 처리에 있어서의 각 파워드 메인에 대한 전력 공급의 타이밍을 도시하는 타이밍 차트이다.
- 도 13은 갱신 처리에 있어서의 각 파워드 메인에 대한 전력 공급을 도시하는 타이밍 차트이다.
- 도 14는 일괄 처리의 수순의 개략을 도시하는 흐름도이다.
- 도 15는 갱신 처리의 수순의 개략을 도시하는 흐름도이다.
- 도 16은 검증에 있어서의 갱신 처리 전후에 있어서의 일례를 도시하는 설명도이다.
- 도 17은 검증에 있어서의 갱신 처리 전후에 있어서의 다른 예를 도시하는 설명도이다.
- 도 18은 검증에 있어서의 갱신 처리 전후에 있어서의 또다른 예를 도시하는 설명도이다.

발명을 실시하기 위한 구체적인 내용

[0015] 도 1에 있어서, 클러스터링 장치(10)는 복수의 요소 데이터를 대상으로 클러스터링을 행한다. 이 클러스터링 장치(10)는 일괄 처리(배치 처리), 갱신 처리(온라인 처리)를 행한다.

[0016] 일괄 처리는, 클러스터수를 변화시키면서, 모든 요소 데이터를 사용한 클러스터링을 행함으로써 얻어지는 각 클러스터수에 대한 평가값으로부터 최적의 클러스터수를 구하고, 구한 최적의 클러스터수로 요소 데이터를 분류한 클러스터링 결과(분류 결과)를 얻는 처리이다. 이 예에서는, 일괄 처리에 있어서의 클러스터링 방법으로서 k-means법을 사용하고, 또한 평가값이 최대(극대)가 되는 클러스터수를 최적의 클러스터수로 하고 있다.

[0017] 갱신 처리는, 일괄 처리 후에 신규 요소 데이터가 추가된 경우에, 그 추가된 신규 요소 데이터를 기존의 클러스터로 효율적이고 또한 고속으로 분류하는 처리이다. 또한, 이 갱신 처리에서는, 추가된 신규 요소 데이터의 클러스터로의 분류 후에 평가값을 구하고, 그 평가값을 사용하여 갱신 처리의 타당성 판단을 행한다. 이 갱신 처리에서 구해지는 평가값은, 일괄 처리로 구하는 평가값과 동등한 것이다. 타당성 판단은, 갱신 처리에 의한 클러스터링 결과가 타당한지 여부를 판단하여, 타당한 경우에는 갱신 처리에 의한 결과를 최종 결과로 하고, 타당하지 않은 경우에는 일괄 처리를 실행시킨다.

[0018] 일괄 처리 및 갱신 처리에서 구하는 클러스터수 N_c 에 관한 평가값을 $E(N_c)$ 라 했을 때, 이 예에서 구하는 평가값 $E(N_c)$ 는 식 (1)과 같이 표현된다. 또한, i 는, 1, 2... N_c 이며, 이 예에서는 클러스터 ID로 하고 있다.

$$E(N_c) = \frac{SBS/N_c}{SWD} \quad \dots (1)$$

$$\text{단, } SWD = \sum_{i=1}^{N_c} \left(\frac{1}{n_i} \sum_{X \in C_i} d(X, V_i) \right)$$

$$SBS = \sum_{i=1}^{N_c} (n_i \cdot d(V_i, GG))$$

[0019]

[0020] 상기 식 (1) 중의 각 값은, 다음과 같다.

[0021] SWD: 제1 총합값인 내적 결합도(규격화 완료)

[0022] SBS/ N_c : 외적 분리도(규격화 완료)

[0023] SBS: 제2 총합값

[0024] X: 요소 데이터

[0025] GG: 전체 요소 데이터의 무게 중심인 데이터 무게 중심

[0026] C_i : 클러스터 ID가 「i」인 클러스터

[0027] V_i : 클러스터 C_i 의 무게 중심인 클러스터 무게 중심

[0028] n_i : 클러스터 C_i 의 요소 데이터의 개수인 데이터수

[0029] $d(V_i, GG)$: 클러스터 무게 중심 V_i 와 데이터 무게 중심 GG 간의 거리인 무게 중심 거리

[0030] $d(X, V_i)$: 요소 데이터 X와 클러스터 무게 중심 V_i 간의 거리인 데이터 무게 중심 간 거리

[0031] 또한, 이하의 설명에서는, 무게 중심 거리 $d(V_i, GG)$ 를 무게 중심 거리 DGV_i , 데이터 무게 중심 간 거리 $d(X, V_i)$ 를 데이터 무게 중심 간 거리 DXV_i 라고 칭한다. 데이터 무게 중심 간 거리 DXV_i 중, 클러스터 C_i 내의 요소 데이터 X와 당해 클러스터 C_i 의 클러스터 무게 중심 V_i 사이의 거리를 특별히 구별할 때에는 클러스터 내 거리 DXV_i 라고 칭하는 경우가 있다. 또한, 개개의 요소 데이터 X를 특별히 구별하는 경우에는, 요소 데이터 X_1, X_2, \dots 등이라고 칭하여 설명한다.

[0032] 요소 데이터 X는, q (q 는, 1 이상의 정수)차원 벡터이며, 예를 들어 화상의 색이나 농담, 색의 분포 등의 각 특장량을 나타낸다. 요소 데이터 X의 각 차원은, N비트(예를 들어 8비트)로 표현된다. 데이터 무게 중심 GG는, 모든 요소 데이터 X의 상가평균으로서 구해지고, 클러스터 무게 중심 V_i 는, 클러스터 내의 각 요소 데이터 X의 상가평균으로서 구해진다. 데이터 무게 중심 GG 및 클러스터 무게 중심 V_i 는, 요소 데이터 X와 동일한 q 차원 벡터이다. 무게 중심 거리 DGV_i 및 데이터 무게 중심 간 거리 DXV_i 는, 이 예에서는 맨해튼 거리로서 구해진다.

[0033] 식 (1)의 우변의 분모의 값 SWD는, 클러스터 C_i 내에 있어서의 요소 데이터 X의 분산(요소 데이터끼리의 유사)

의 정도를 전체 클러스터에 대하여 나타내는 내적 결합도로 이루어져 있다. 식 (1)에 있어서의 내적 결합도는, 클러스터 C_i 마다의 제1 지표값 SD_i 를 당해 클러스터 내의 데이터수 n_i 로 제산함으로써 규격화한 결합 지표값 SWD_i 의 각 클러스터 C_i 에 관한 총합인 제1 총합값으로서 구해진다. 제1 지표값 SD_i 는, 클러스터 C_i 내의 각 요소 데이터 X 에 관한 클러스터 내 거리 DXV_i 의 총합으로 하고 있다.

[0034] 또한, 식 (1)의 우변의 분자는, 클러스터 C_i 의 분리의 정도를 전체 클러스터에 대하여 나타내는 외적 분리도로 이루어져 있다. 식 (1)에 있어서의 외적 분리도는, 클러스터 지표값 SBS를 클러스터수 N_c 로 제산하여 규격화한 것으로 되어 있다. 이 예에 있어서의, 클러스터 지표값 SBS는, 클러스터 간의 거리의 지표가 되는 각각의 클러스터 C_i 에 관한 제2 지표값 SBS_i 의 총합인 제2 총합값으로서 구해진다. 제2 지표값 SBS_i 는, 클러스터 C_i 의 데이터수 n_i 로 가중치 부여한 무게 중심 거리 DGV_i 로 하고 있다. 이와 같이, 제2 지표값 SBS_i 로서, 무게 중심 거리 DGV_i 를 사용하는 것은, 클러스터 상호 간의 거리를 사용하는 것보다도 계산수를 적게 함에 있어서 유리하다. 또한, 무게 중심 거리 DGV_i 에 가중치 부여를 하는 경우, 가중치 부여하는 값은, 데이터수 n_i 에 한정되지 않고, 데이터수 n_i 에 기초하는 값으로 가중치 부여해도 된다.

[0035] 내적 결합도 SWD는, 결합 지표값 SWD_i 를 사용하여, 식 (2)와 같이 표현되고, 클러스터 지표값 SBS는, 제2 지표값 SBS_i 를 사용하여, 식 (3)과 같이 표현된다. 또한, 클러스터 C_i 에 있어서의 결합 지표값 SWD_i 및 제2 지표값 SBS_i 는, 각각 식 (4), 식 (5)와 같이 표현된다.

$$SWD = SWD_1 + SWD_2 + \dots \quad \dots (2)$$

$$SBS = SBS_1 + SBS_2 + \dots \quad \dots (3)$$

$$SWD_i = \frac{1}{n_i} \cdot SD_i = \frac{1}{n_i} \cdot \sum_{X \in C_i} d(X, V_i) \quad \dots (4)$$

$$SBS_i = n_i \cdot d(V_i, GG) \quad \dots (5)$$

[0036]

[0037] 상기한 바와 같이 평가값 $E(N_c)$ 에서는, 클러스터 지표값 SBS를 클러스터수 N_c 로 규격화한 값을 외적 분리도로 하고, 클러스터 C_i 의 데이터수 n_i 로 규격화한 결합 지표값 SWD_i 의 제1 총합값을 내적 결합도로 하고 있다. 이 때문에, 전체 요소 데이터 X 에 있어서, 다른 요소 데이터 X 의 집합에 대하여, 범위가 큰 요소 데이터 X 의 집합이 있는 경우나, 요소 데이터 X 의 분포 밀도가 높은 요소 데이터 X 의 집합이 있는 경우 동일지라도, 개개의 클러스터 C_i 내에 있어서의 요소 데이터 X 의 분산의 정도가 내적 결합도에 적절하게 반영된다. 즉, 과잉의 분류가 된 경우에, 평가값 $E(N_c)$ 가 보다 커질 일이 없어, 과잉의 분류가 억제된다.

[0038] 클러스터링 장치(10)는 시스템 컨트롤러(11)와, 연산 유닛(12)을 구비하고 있다. 연산 유닛(12)은 메인 메모리(14), 무게 중심 메모리(15), 클러스터링 연산부(16), 근린 탐색 회로부(17), 평가값 산출 회로(18)를 구비하고 있다.

[0039] 시스템 컨트롤러(11)는 연산 유닛(12)에 대한 요소 데이터 X 의 입력, 일괄 처리의 클러스터링 및 갱신 처리의 실행 지시, 클러스터수 N_c 의 설정이나, 연산 유닛(12)으로부터 취득한 평가값 $E(N_c)$ 에 의한 최적의 클러스터수 N_c 의 판단, 상술한 타당성 판단, 연산 유닛(12)에 대한 파워 게이팅의 제어 등을 행한다. 파워 게이팅에 대해서는 후술한다. 또한, 시스템 컨트롤러(11)는 클러스터링 결과로서 연산 유닛(12)으로부터 각 요소 데이터 X 에 할당되는 클러스터 ID, 추가된 신규 요소 데이터 X_{new} 가 분류된 클러스터 C_i 의 클러스터 ID의 취득 등을 행한다. 이 예에서는, 시스템 컨트롤러(11)가 판단부가 된다.

[0040] 또한, 시스템 컨트롤러(11)는 연산 유닛(12)에 의한 클러스터링의 실행 시에 무게 중심 메모리(15)의 내용 즉 각 클러스터 무게 중심 V_i 를 감시하고, 각 클러스터 무게 중심 V_i 의 변동이 없어졌을 때, 즉 클러스터 무게 중심 V_i 가 수렴했을 때에 클러스터링을 종료시킨다. 또한, 클러스터 무게 중심 V_i 가 수렴하는 것 대신에, 미리 설정한 횟수의 후술하는 분류 계산을 행했을 때에 클러스터링을 종료시켜도 된다.

[0041] 시스템 컨트롤러(11)는 일괄 처리 시의 클러스터수 N_c 마다의 복원용 데이터를 기억한다. 복원용 데이터는, 연산 유닛(12)에 보유되는 클러스터 ID, 클러스터 내 거리 DXV_i , 클러스터 무게 중심 V_i , 데이터수 n_i 등을 최적의

클러스터수 N_c 로 클러스터링한 상태로 복원하기 위한 데이터이다. 이 예에서는, 복원용 데이터로서, 각 요소 데이터 X 에 할당되는 클러스터 ID를 기억한다. 복원용 데이터로서는, 이 밖에, 각 클러스터 무게 중심 V_i 등으로 해도 되고, 클러스터 ID와 클러스터 무게 중심 V_i 로 해도 된다.

[0042] 최적의 클러스터수 N_c 에 의한 클러스터링을 다시 실행함으로써, 연산 유닛(12)에 보유되는 데이터를 복원할 수 있지만, 클러스터 ID나 클러스터 무게 중심 V_i 를 사용함으로써, 적은 계산량으로 고속으로 복원할 수 있다. k-means법과 같은 클러스터링에서는, 그 클러스터링에 요하는 계산 시간의 대부분이 클러스터 무게 중심을 수렴시키기 위한 반복 계산의 시간이다. 계산의 반복 횟수는, 요소 데이터 X 의 총 데이터수에 따라 다르지만, 십수 내지 수백회, 많은 경우에는 1000회 정도에도 달한다. 그러나, 상기한 바와 같이 한번 수렴한 클러스터 무게 중심 V_i 나 그에 의하여 결정된 클러스터 ID를 사용하면, 반복 계산을 하지 않고, 고속(단시간)으로 또한 고정밀도로 클러스터링한 상태로 할 수 있다.

[0043] 연산 유닛(12)은 상기 일괄 처리, 갱신 처리를 행하는 ASIC(Application Specific Integrated Circuit)로서 제작되고, 연산 유닛(12)의 각 부는 클럭 발생기(도시 생략)로부터의 클럭에 기초하여 동기 동작하도록 구성되어 있다. 이 연산 유닛(12)은 시스템 컨트롤러(11)에 의한 일괄 처리의 클러스터링 및 갱신 처리의 지시를 트리거로 하여 작동한다.

[0044] 클러스터링 연산부(16)는 k-means법에 의한 클러스터링을 행하는 회로이며, 일괄 처리의 클러스터링의 각종 계산이나 각 요소 데이터 X 의 클러스터 C_i 로의 분류 등을 행한다. 이 클러스터링 연산부(16)는 지연 회로(21), 거리 산출 회로(22), 거리 레지스터부(24)와 CID(클러스터 ID) 레지스터부(25)로 이루어지는 메인 레지스터 유닛(26), 최댓값 검출 회로(27), CID(클러스터 ID) 마스크 회로(28), 무게 중심 산출 회로(29)를 갖고 있다. 클러스터링 연산부(16)는 시스템 컨트롤러(11)와 함께 일괄 처리부를 구성한다.

[0045] 데이터 기억부로서의 메인 메모리(14)는 시스템 컨트롤러(11)에 의해 기입되는 복수의 요소 데이터 X 를 기억한다. 도 2에 일례를 도시하는 바와 같이, 메인 메모리(14)는 1개의 요소 데이터 X 의 각 차원의 성분(이하, 벡터 성분이라고 함)과 동일한 N 비트의 용량을 갖는 단위 블록(14a)을 $q \times M$ 의 행렬형으로 배열한 것이며, 각 열에 요소 데이터 X 의 차원과 동일한 q 개의 단위 블록(14a)이 배열되어 있다. 행방향으로 배열하는 단위 블록(14a)의 개수 M 은, 분류하는 요소 데이터 X 의 최대 개수 이상이다. 메인 메모리(14)는 1개의 단위 블록(14a)에 1개의 벡터 성분을 기억하고 1개의 요소 데이터 X 가 동일 열의 각 단위 블록(14a)에 기억되도록 하여, 각 요소 데이터 X 를 기억한다. 도 2에서는, 요소 데이터 X_p 의 각 벡터 성분 $X_{p1}, X_{p2} \dots X_{pq}$ 가 기입된 상태를 도시하고 있다.

[0046] 메인 메모리(14)는 요소 데이터 X 를 판독하는 때에는, 행방향으로 배열하는 M 개의 단위 블록(14a)을 1단위로 하여 순차적으로 판독한다. 이에 의해, 각 요소 데이터 X 의 벡터 성분이 1차원분씩 병렬적으로 메인 메모리(14)로부터 출력된다. 또한, 요소 데이터 X 가 쓰여져 있지 않은 열로부터도 마찬가지로 판독이 행하여져, 이 경우에는 예를 들어 「0」의 성분 벡터가 판독된다. 또한, 메인 메모리(14)는 요소 데이터 X 의 기입에 대해서는, 예를 들어 단위 블록(14a)마다와 같이 소용량씩 기입을 행하는 것이어도 된다. 또한, 무게 중심 메모리(15)에 대해서도 마찬가지이다.

[0047] 무게 중심 메모리(15)는 각 클러스터의 클러스터 무게 중심 V_i 를 기억하는 것이며, 도시를 생략하지만, 메인 메모리(14)와 마찬가지로의 구성이다. 즉, 무게 중심 메모리(15)는 N 비트의 단위 블록을 열방향으로 q 개가 배열하도록 하여 단위 블록을 행렬형으로 마련한 것이며, 각 열에 q 차원의 클러스터 무게 중심 V_i 를 기억한다. 무게 중심 메모리(15)는 메인 메모리(14)보다도 소용량이다. 무게 중심 메모리(15)로부터는, 클러스터 무게 중심 V_i 마다 판독되고, 1개의 클러스터 무게 중심 V_i 에 대해서는, 1열분의 벡터 성분이 순차적으로 판독된다.

[0048] 이 예에서는, 메인 메모리(14) 및 무게 중심 메모리(15)에는, 불휘발성 메모리가 사용되고 있다. 이에 의해, 요소 데이터 X , 클러스터 무게 중심 V_i 의 판독, 기입을 행하지 않을 때에는, 이들에의 전력 공급을 정지할 수 있어, 보다 한층의 전력 절약화를 도모할 수 있다. 메인 메모리(14) 및 무게 중심 메모리(15)에 사용하는 불휘발성 메모리로서는, MTJ 소자 등을 기억 소자로 한 것이 바람직하게 사용된다.

[0049] 지연 회로(21)는 메인 메모리(14)와 거리 산출 회로(22) 사이에 마련되어 있다. 이 지연 회로(21)는 메인 메모리(14)로부터 판독되어서 거리 산출 회로(22)에 입력되는 요소 데이터 X 의 입력 타이밍과, 당해 요소 데이터 X 로부터 무게 중심 산출 회로(29)에 의해 산출되어서 거리 산출 회로(22)에 입력되는 클러스터 무게 중심 V_i 의

입력 타이밍을 동기시킨다. 도 2에 도시된 바와 같이, 지연 회로(21)는 복수단에 접속된 레지스터부(31)로 구성된다. 각 레지스터부(31)는 각각 M개의 셀(31a)로 구성되어 있다. 각 셀(31a)은 각각 N비트의 용량을 갖는 레지스터이다. 1개의 레지스터부(31)의 각 셀(31a)로부터 다음단의 레지스터부(31)의 각 셀(31a)에 순차적으로 벡터 성분을 보냄으로써, 요소 데이터 X의 거리 산출 회로(22)에의 입력을 지연시킨다. 지연 회로(21)에 의한 지연 시간 즉 레지스터부(31)의 단 수는, 데이터 무게 중심 GG, 클러스터 무게 중심 V_i 의 계산에 필요한 클럭수 등으로부터 미리 정해져 있다.

[0050] 거리 산출 회로(22)는 일괄 처리 시에는, 지연 회로(21)를 통하여 메인 메모리(14)로부터 입력되는 각 요소 데이터 X와 무게 중심 산출 회로(29)로 산출되는 클러스터 무게 중심 V_i 가 입력된다. 또한, 갱신 처리에서는, 클러스터 무게 중심 V_i 로서, 무게 중심 산출 회로(29)로부터의 것 대신에 무게 중심 메모리(15)로부터의 것이 입력된다. 이 거리 산출 회로(22)는 입력되는 각 요소 데이터 X에 관한 데이터 무게 중심 간 거리 DXV_i 를 병렬적으로 산출한다. 또한, 요소 데이터 X 및 클러스터 무게 중심 V_i 등의 벡터 데이터의 회로에 대한 입력은, 벡터 성분이 순차적으로 입력되는 것을 의미한다.

[0051] 거리 레지스터부(24)는 거리 산출 회로(22)로 산출되는 각 데이터 무게 중심 간 거리 DXV_i 를 보유하고, CID 레지스터부(25)는 클러스터 ID(클러스터 정보)를 보유한다. 거리 레지스터부(24)의 내용은, 대응하는 거리 산출 회로(22)로 산출되는 새로운 데이터 무게 중심 간 거리 DXV_i 가 그 시점에서 기억되어 있는 데이터 무게 중심 간 거리 DXV_i 보다도 작은 경우에, 그 새로운 데이터 무게 중심 간 거리 DXV_i 로 갱신된다. 클러스터링 시에 최종적으로 CID 레지스터부(25)에 보유하고 있는 각 클러스터 ID가, 각 요소 데이터가 분류되는 클러스터를 나타내고, 거리 레지스터부(24)에 보유하고 있는 데이터 무게 중심 간 거리 DXV_i 가 클러스터 내 거리를 나타낸다.

[0052] 도 3의 (A)에 도시하는 바와 같이, 거리 산출 회로(22)는 데이터 무게 중심 간 거리 DXV_i 를 계산하는 M개의 셀(22a)을 갖고 있다. 마찬가지로, 거리 레지스터부(24)는 도 3의 (C)에 도시하는 바와 같이, 데이터 무게 중심 간 거리 DXV_i 를 보유하는 M개의 셀(24a)을, CID 레지스터부(25)는 클러스터 ID를 보유하는 M개의 셀(25a)을 각각 갖고 있다. 셀(24a, 25a)은, 각각 복수 비트의 용량을 갖는 레지스터이다. 상술한 바와 같이, 요소 데이터 X가 q차원 벡터이며, 각 차원이 N비트일 경우에는, 데이터 무게 중심 간 거리 DXV_i 가 $(N+q)$ 비트의 데이터이므로, 셀(24a)은 적어도 $(N+q)$ 비트의 용량을 갖는다. 메인 메모리(14)의 각 열에 각각 1개의 셀(22a, 24a, 25a)이 대응하고 있다. 후술하는 CID 마스크 회로(28)의 셀(28a)(도 5 참조)에 대해서도 마찬가지이다. 이에 의해, 메인 메모리(14)에 기억되어 있는 각각의 요소 데이터 X는, 각각 1개의 셀(22a, 24a, 25a, 28a)에 대응하고, 요소 데이터 X, 셀(22a, 24a, 25a, 28a)의 각 내용이 서로 대응지어진다. 또한, 도 3 및 후술하는 도 4 내지 도 7에서는, 주요부의 구성 및 신호만을 도시하고 있다.

[0053] 거리 산출 회로(22)의 셀(22a)은 도 3의 (B)에 일례가 도시되는 바와 같이, 전가산기(32), XOR(배타적 논리합) 회로(33), 셀렉터(34), 전가산기(35), 계산 레지스터(36)로 구성되어 있다. 전가산기(32)는 그 한쪽의 입력단에 지연 회로(21)로부터의 요소 데이터 X가, 다른 쪽의 입력단에 무게 중심 산출 회로(29) 또는 무게 중심 메모리(15)로부터의 클러스터 무게 중심 V_i 가 반전 입력되고, 그 계산 결과와 자릿수 올림 신호(부논리)를 XOR 회로(33)에 입력하고 있다. 이에 의해, 요소 데이터 X와 클러스터 무게 중심 V_i 의 차원마다의 거리가 XOR 회로(33)로부터 순차적으로 출력된다.

[0054] XOR 회로(33)로부터의 각 차원의 거리는, 셀렉터(34)를 통하여 전가산기(35)의 한쪽의 입력단에 순차적으로 입력된다. XOR 회로(33)로부터의 1차원분의 거리가 전가산기(35)에 입력될 때마다, 이것에 동기하여 계산 레지스터(36)의 내용이 판독되어서 전가산기(35)의 다른 쪽 입력단에 입력되고, 이에 의해 얻어지는 전가산기(35)의 계산 결과로 계산 레지스터(36)의 내용이 갱신된다. 계산 레지스터(36)의 내용은, 초깃값은 「0」이다. 이에 의해, 전가산기(35)에의 q차원분의 거리의 입력에 의해, 계산 레지스터(36)에는, 데이터 무게 중심 간 거리(맨해튼 거리) DXV_i 가 보유된다.

[0055] 셀렉터(34)는 상기와 같이 계산 레지스터(36)가 데이터 무게 중심 간 거리를 기억한 후, 즉 데이터 무게 중심 간 거리 DXV_i 가 계산된 후에, 셀(24a)의 내용을 전가산기(35)의 한쪽의 입력단에 반전 입력한다. 그 후에, 계산 레지스터(36)의 내용과 셀(24a)의 내용이 동기하여 판독되어, 전가산기(35)에 각각 입력된다. 셀(24a)에는, 계산 레지스터(36)의 내용이 입력 데이터로서 부여됨과 함께, 갱신 신호로서 전가산기(35)의 자릿수 올림 신호

가 부여되어 있다. 이에 의해, 전가산기(35)로 자릿수 올림이 발생했을 때에, 셀(24a)의 내용이 계산 레지스터(36)의 내용으로 갱신된다. 이러한 동작이 거리 산출 회로(22)에 입력되는 클러스터 무게 중심 V_i 마다 행하여짐으로써, 각 클러스터 무게 중심 V_i 에 대한 데이터 무게 중심 간 거리 DXV_i 중에서 최소의 것이 셀(24a)에 보유된다. 각 셀(24a)은 독립적으로 그 내용이 갱신되기 때문에, 거리 레지스터부(24)에는, 각 요소 데이터 X의 각각에 관한 데이터 무게 중심 간 거리 DXV_i 가 보유된다.

[0056] CID 레지스터부(25)의 각 셀(25a)은 거리 산출 회로(22)가 데이터 무게 중심 간 거리 DXV_i 의 계산을 행하고 있을 때에, 처리 대상으로 하고 있는 클러스터 C_i 의 클러스터 ID가 지정 CID로서 입력되어 있다. 이 셀(25a)은 셀(24a)과 마찬가지로, 전가산기(35)에서 자릿수 올림이 발생했을 때에, 입력되어 있는 클러스터 ID에 보유되어 있는 내용이 갱신된다. 이에 의해, 셀(25a)에는, 데이터 무게 중심 간 거리 DXV_i 가 최소가 되는 클러스터 C_i 가 보유된다. 즉, 셀(25a)이 보유하는 클러스터 ID에 의해, 대응하는 요소 데이터 X가 분류되어 있는 클러스터의 클러스터 ID가 나타내진다.

[0057] 또한, 연산 유닛(12)에 있어서의 각종 데이터를 전송하는 버스는, 전송하는 데이터에 따른 버스 폭(비트수)의 것이 사용되고 있다. 예를 들어, 전가산기(32)의 각 입력단은, 요소 데이터 X, 클러스터 무게 중심 V_i 의 N비트의 벡터 성분을 패러렐로 입력하므로, 그들 각 입력단에는 각각 버스 폭이 N비트인 버스가 접속되어 있다. 또한, 전가산기(35)와 계산 레지스터(36) 사이에서는, $(N+q)$ 비트가 되는 데이터 무게 중심 간 거리 DXV_i 의 전송을 패러렐로 행하므로, 버스 폭이 $(N+q)$ 비트인 버스가 사용되고 있다. CID 마스크 회로(28)의 셀렉터(37c)와 같이, 요소 데이터 X 및 클러스터 무게 중심 V_i 의 양쪽을 선택적으로 출력하는 버스에서는, 비트수가 큰 것, 이 예에서는 클러스터 무게 중심 V_i 에 따른 $(N+q)$ 비트의 버스 폭으로 되어 있다. 또한, 도 3, 도 5, 도 7에서는, 도면 중에 주요부의 버스 폭을 기재하고 있다.

[0058] 거리 레지스터부(24) 및 CID 레지스터부(25)에는, 쌍이 되는 셀(24a)과 셀(25a)마다, 그들 셀(24a, 25a)의 래치 동작의 유효/무효를 전환하는 인에이블 신호 회로(71)(도 9 참조)가 마련되어 있다. 이 인에이블 신호 회로(71)로부터의 인에이블 신호를 셀(24a, 25a)에 입력함으로써, 셀(24a) 및 셀(25a)의 래치 동작의 유효와 무효가 각종 동작에 따른 타이밍에 전환된다. 이러한 인에이블 신호 회로(71)를 마련함으로써, 시스템 컨트롤러(11)에서의 개별의 셀(24a, 25a)의 제어를 요하지 않게 하고 있다. 또한, 셀(25a)과 대응하는 셀(28a) 사이에서는, 서로 한쪽의 내용을 다른 쪽이 래치 가능하게 되어 있다.

[0059] 최댓값 검출 회로(27)는 클러스터 C_i 의 클러스터 무게 중심 V_i 를 초기 설정할 때에 거리 레지스터부(24)에 보유되어 있는 각 데이터 무게 중심 간 거리 DXV_i 를 비교하여, 최대가 되는 데이터 무게 중심 간 거리 DXV_i 를 검출한다. 최댓값 검출 회로(27)는 M개의 셀(24a)에 대응한 M개의 최대 플래그(1비트)를 출력한다. 도 4에 일례를 도시하는 바와 같이, 최댓값 검출 회로(27)는 M비트 입력의 AND 회로(27a)와, 거리 레지스터부(24)의 M개의 각 셀(24a)에 각각 대응한 OR 회로(27b), NAND 회로(27c) 및 1비트의 레지스터(27d)가 마련되어 있다. OR 회로(27b), NAND 회로(27c), 레지스터(27d)는 셀(24a)에 대응하는 것끼리 접속되어 있다. 셀(24a)은 최댓값 검출 회로(27)에 대하여 보유하고 있는 클러스터 무게 중심 간 거리 DXV_i 의 데이터를 상위 비트로부터 1비트씩 차례로 송출한다.

[0060] OR 회로(27b)의 한쪽의 입력단에, 대응하는 셀(24a)로부터의 1비트 신호가 반전 입력되고, 다른 쪽의 입력단에 레지스터(27d)의 출력이 반전 입력된다. NAND 회로(27c)의 한쪽의 입력단에 OR 회로(27b)의 출력이 입력되고, 다른 쪽의 입력단에 AND 회로(27a)의 출력이 반전 입력된다. 레지스터(27d)는 NAND 회로(27c)의 출력의 논리(「1」 또는 「0」)를 보유하고, 보유하고 있는 논리를 출력한다. 이 구성에 의해, 셀(24a)로부터 클러스터 무게 중심 간 거리 DXV_i 의 전체 비트의 송출 완료 후에는, 거리 레지스터부(24)에 보유되어 있는 클러스터 무게 중심 간 거리 DXV_i 중에서 최대의 클러스터 무게 중심 간 거리 DXV_i 를 보유하고 있는 셀(24a)에 대응한 레지스터(27d)가 보유하는 논리만이 「1」로 된다. 각 레지스터(27d)의 내용은, 클러스터 무게 중심 간 거리 DXV_i 가 최대인지의 여부를 나타내는 최대 플래그로서, CID 마스크 회로(28)가 대응하는 셀(28a)에 각각 출력된다.

[0061] CID 마스크 회로(28)는 메인 메모리(14)로부터 입력되는 각 요소 데이터 X 또는 거리 레지스터부(24)로부터 입력되는 각 데이터 무게 중심 간 거리 DXV_i 중으로부터 처리에 필요한 것만을 출력하도록 구성된다.

[0062] CID 마스크 회로(28)는 도 5의 (A)에 도시하는 바와 같이, M개의 셀(28a)을 갖고 있다. 도 5의 (B)에 일례를 도시하는 바와 같이, 셀(28a)은 셀렉터(37a 내지 37c), MID 레지스터(37d), 비교기(37e) 등으로 구성된다. 각 MID 레지스터(37d)에는, 셀(24a, 25a)과 마찬가지로, 인에이블 신호 회로(75)(도 10 참조)가 각각 마련되어 있고, 시스템 컨트롤러(11)에서의 각 MID 레지스터(37d)의 개별의 제어를 요하지 않게 하고 있다. 인에이블 신호 회로(75)로부터의 인에이블 신호를 MID 레지스터(37d)에 입력함으로써, MID 레지스터(37d)의 래치 동작의 유효와 무효가 각종 동작에 따른 타이밍에 전환된다.

[0063] 셀렉터(37a)는 요소 데이터 X와 데이터 무게 중심 간 거리 DXV_i 의 어느 한쪽을 입력 데이터로서 선택하여, 셀렉터(37c)로 출력한다. 셀렉터(37b)는 CID 레지스터부(25)의 셀(25a)로부터의 클러스터 ID와 시스템 컨트롤러(11)로부터의 외부 설정 ID(클러스터 ID)의 어느 한쪽을 선택하여 MID 레지스터(37d)에 입력한다.

[0064] MID 레지스터(37d)는 셀렉터(37b)로부터의 클러스터 ID를 래치 동작으로 보유하고, 보유하고 있는 클러스터 ID를 비교기(37e)에 입력한다. 비교기(37e)는 지정 CID와 MID 레지스터(37d)로부터의 클러스터 ID를 비교하여, 그 비교 결과를 나타내는 1비트의 비교 플래그(C-flag)를 외부로 출력함과 함께, 셀렉터(37c)에 입력한다. 비교 플래그는, 지정 CID와 MID 레지스터(37d)로부터의 클러스터 ID가 일치하고 있는 경우가 「1」이 되고, 불일치하는 경우에는 「0」이 된다.

[0065] 셀렉터(37c)는 비교기(37e)로부터의 비교 플래그가 「1」인 경우에 입력 데이터(요소 데이터 X 또는 데이터 무게 중심 간 거리 DXV_i)를 출력하고, 「0」인 경우에 전체 비트가 「0」인 널(null) 데이터를 출력한다. 이에 의해, 예를 들어 MID 레지스터(37d)에 대응하는 CID 레지스터부(25)의 셀(25a)의 클러스터 ID가 보유되어 있는 때에는, 지정 CID와 일치하는 클러스터 ID로 분류되어 있는 요소 데이터 X 또는 그 데이터 무게 중심 간 거리 DXV_i 가 셀(28a)로부터 출력되어, 그들 셀(28a)로부터의 비교 플래그가 「1」로 된다.

[0066] 무게 중심 산출 회로(29)는 CID 마스크 회로(28)를 통하여 입력되는 각 요소 데이터 X와 비교 플래그로부터 데이터 무게 중심 GG 및 클러스터 무게 중심 V_i 를 산출한다. 또한, 이 무게 중심 산출 회로(29)는 클러스터 무게 중심 V_i 의 계산 도중에 얻어지는 데이터수 n_i 및 데이터 가산값 SS_i 를 평가값 산출 회로(18)에 입력한다. 데이터수 n_i 및 데이터 가산값 SS_i 는, 평가값 산출 회로(18)에 있어서, 제2 지표값 SBS_i 의 계산 등에 사용된다. 또한, 무게 중심 산출 회로(29)는 CID 마스크 회로(28)를 통하여 입력되는 각 요소 데이터 X와 비교 플래그로부터, 결합 지표값 SWD_i 를 산출하여 평가값 산출 회로(18)로 보낸다. 상술한 바와 같이, 결합 지표값 SWD_i 는, 클러스터 내 거리 DXV_i 의 총합인 제1 지표값 SD_i 를 데이터수 n_i 로 제산한 값으로서 구해진다.

[0067] 상기 데이터 가산값 SS_i 는, 식 (6)에 나타나는 바와 같이, 클러스터 C_i 의 각 요소 데이터 X를 차원마다 가산한 데이터이며, 요소 데이터 X와 동일한 q차원 벡터이다. 클러스터 무게 중심 V_i 는, 식 (7)에 나타나는 바와 같이, 데이터 가산값 SS_i 를 데이터수 n_i 로 제산함으로써 구해진다. 또한, 전체 요소 데이터 X에 관한 데이터 가산값을 전체 데이터수로 제산함으로써 데이터 무게 중심 GG를 구한다. 무게 중심 산출 회로(29)는 클러스터 무게 중심 V_i 를 무게 중심 메모리(15)와 거리 산출 회로(22)로 보낸다.

$$SS_i = \sum_{X \in C_i} X \quad \dots (6)$$

$$V_i = SS_i / n_i \quad \dots (7)$$

[0068]

[0069] 무게 중심 산출 회로(29)는 도 6에 일례를 도시하는 바와 같이, M개의 셀렉터(38a)로 이루어지는 셀렉터부(38)와, 가산기(39)와, 제1 레지스터(41) 및 제2 레지스터(42)와, 제산기(43)를 구비하고 있다. 각 셀렉터(38a)는 CID 마스크 회로(28)로부터의 출력 데이터(요소 데이터 X 또는 데이터 무게 중심 간 거리 DXV_i)와 비교 플래그(C-flag)를 차례로 선택하여 가산기(39)로 출력하도록 동작한다. 가산기(39)는 입력되는 데이터를 가산한다.

[0070] 제1 레지스터(41)는 셀렉터부(38)로부터 가산기(39)에 요소 데이터 X 또는 데이터 무게 중심 간 거리 DXV_i 가 입력되는 경우에, 그 가산기(39)의 계산 결과를 보유하고, 이에 의해, 가산기(39)에 요소 데이터 X가 입력되어 있는 경우에는, 데이터 가산값이 제1 레지스터(41)에 보유하고, 데이터 무게 중심 간 거리 DXV_i 가 입력되어 있는 경우에는, 제1 지표값 SD_i 가 제1 레지스터(41)에 보유된다. 또한, 셀렉터부(38)로부터 비교 플래그가 입력되는 경우, 비교 플래그 즉 1비트 데이터를 가산한 값을 가산기(39)가 계산하고, 이 값이 제2 레지스터(42)에

보유된다. 제2 레지스터(42)에 보유되는 값은, CID 마스크 회로(28)로부터 출력되어 있는 요소 데이터 X 또는 데이터 무게 중심 간 거리 DXV_i 의 개수가 된다. 이에 의해, 데이터수 n_i 또는 전체 데이터수를 얻을 수 있다. 이 예에서는, 가산기(39)는 요소 데이터 X가 입력되어 있는 때에 데이터 가산기로서 기능하고, 비교 플래그가 입력되어 있는 때에 개수 산출기로서 기능한다.

[0071] 계산기(43)는 제1 레지스터(41)에 보유되어 있는 값을 제2 레지스터(42)에 보유되어 있는 값으로 계산한 값을 출력한다. 이 계산기(43)의 출력으로서, 데이터 무게 중심 GG, 클러스터 무게 중심 V_i , 결합 지표값 SWD_i 가 얻어진다.

[0072] 근린 탐색 회로부(17)는 시스템 컨트롤러(11) 및 무게 중심 메모리(15)와 함께, 갱신 처리부를 구성한다. 이 갱신 처리부와 상술한 일괄 처리부에 의해, 클러스터링부를 구성한다. 이 근린 탐색 회로부(17)는 갱신 처리시에, 추가되는 신규 요소 데이터 X_{new} 와의 데이터 무게 중심 간 거리 DXV_i 가 최소가 되는 클러스터 ID를 특정하고, 그 특정한 클러스터 C_i 로 신규 요소 데이터 X_{new} 를 분류한다.

[0073] 이 근린 탐색 회로부(17)는 계산부(17a)와, 근거리 레지스터부(17b)와, 근거리 CID 레지스터부(17c)를 갖는다. 계산부(17a)는 신규 요소 데이터 X_{new} 와, 무게 중심 메모리(15)로부터 순차적으로 판독되는 각 클러스터 무게 중심 V_i 의 데이터 무게 중심 간 거리 DXV_i 를 계산한다. 근거리 레지스터부(17b), 근거리 CID 레지스터부(17c)는 계산부(17a)의 계산 결과에 기초하여, 최소가 되는 데이터 무게 중심 간 거리 DXV_i , 클러스터 ID를 보유한다. 근거리 CID 레지스터부(17c)에 최종적으로 보유되어 있는 클러스터 ID가, 신규 요소 데이터 X_{new} 가 분류되는 클러스터 C_i 의 클러스터 ID가 된다. 근거리 CID 레지스터부(17c)에 최종적으로 보유되어 있는 클러스터 ID는, 신규 요소 데이터 X_{new} 에 대응하는 CID 레지스터부(25)의 셀(25a)에 기입된다. 또한, 근린 탐색 회로부(17)를 구성하는 계산 회로의 일부가, 갱신 처리 시의 평가값 $E(Nc)$ 를 구하기 위하여 사용된다.

[0074] 도 7에 일례를 도시하는 바와 같이, 계산부(17a)는 셀렉터(44), 전가산기(45), XOR 회로(46), 셀렉터(47), 전가산기(48), 계산 레지스터(49), 가산기(61)를 갖고 있다. 이 계산부(17a)에 의해, 신규 요소 데이터 X_{new} 와 각 클러스터 무게 중심 V_i 의 데이터 무게 중심 간 거리 DXV_i 가 순차적으로 계산된다. 또한, 최소가 되는 데이터 무게 중심 간 거리 DXV_i 가 근거리 레지스터부(17b)에 보유되고, 그 데이터 무게 중심 간 거리 DXV_i 에 대응하는 클러스터 ID가 근거리 CID 레지스터부(17c)에 보유된다. 이들 계산부(17a), 근거리 레지스터부(17b) 및 근거리 CID 레지스터부(17c)로 이루어지는 근린 탐색 회로부(17)의 회로 구성은, 거리 산출 회로(22)의 셀(22a)과 거리 레지스터부(24)의 셀(24a) 및 CID 레지스터부(25)의 셀(25a)과 마찬가지로의 구성이므로, 상세한 설명은 생략한다.

[0075] 갱신 처리 시의 평가값 $E(Nc)$ 를 구할 때에는, 전가산기(45), XOR 회로(46) 및 가산기(61)를 사용하여, 신규 요소 데이터 X_{new} 가 분류되는 클러스터 무게 중심 V_i 와 데이터 무게 중심 GG의 무게 중심 거리 DGV_i 를 산출한다. 무게 중심 거리 DGV_i 는, 평가값 산출 회로(18)로 보내진다. 또한, 클러스터 무게 중심 V_i 는, 무게 중심 메모리(15)로부터 입력된다.

[0076] 평가값 산출부로서의 평가값 산출 회로(18)는 일괄 처리에 있어서의 각 클러스터링의 종료 시 및 갱신 처리시에 각각 평가값 $E(Nc)$ 를 산출한다. 도 8에 일례를 도시하는 바와 같이, 평가값 산출 회로(18)는 로직부(18a)와 평가 레지스터부(18b)로 크게 구별된다. 로직부(18a)는 셀렉터(50), 승산기(51), 감산기(52), 적산기(53), 병렬 가산기(54), 승산기(55), 가산기(56, 57), 분모 레지스터(58), 분자 레지스터(59), 계산기(60)로 구성된다. 이 로직부(18a)는 평가 레지스터부(18b)에 보유되어 있는 각종 데이터, 시스템 컨트롤러(11)로부터 입력되는 클러스터수 Nc 등을 사용하여, 상기 식 (1)에 의한 평가값 $E(Nc)$ 를 산출한다. 또한, 이 로직부(18a)는 갱신 처리시에, 신규 요소 데이터 X_{new} 가 분류된 클러스터 무게 중심 V_i 를 계산하고, 이것을 무게 중심 메모리(15)에 기입한다. 또한, 평가값 산출 회로(18)의 상세한 동작은 후술한다.

[0077] 평가 레지스터부(18b)는 GG 레지스터(63), 데이터수 레지스터(64), SBS 레지스터 유닛(66), SWD 레지스터 유닛(67) 및 셀렉터(68)를 포함한다. GG 레지스터(63)는 무게 중심 산출 회로(29)로 산출되는 데이터 무게 중심 GG를 보유한다. 데이터수 레지스터(64)는 무게 중심 산출 회로(29)로 얻어지는 각 클러스터 C_i 의 데이터수 n_i 를 각각 보유한다. 또한, 데이터 무게 중심 GG를 무게 중심 메모리(15)에 기억하는 구성으로서, GG 레지스터(63)를 생략해도 된다. 또한, GG 레지스터(63), 데이터수 레지스터(64), SBS 레지스터 유닛(66), SWD 레지스터 유

닛(67)을 각각 불휘발성 레지스터로 하는 것도 바람직하다.

- [0078] SBS 레지스터 유닛(66)은 제1 SBS 레지스터(66₁), 제2 SBS 레지스터(66₂)...을 갖는다. 제i SBS 레지스터(66_i)는, 데이터 가산값 SS_i로부터 구해지는 제2 지표값 SBS_i를 보유한다. SWD 레지스터 유닛(67)은 제1 SWD 레지스터(67₁), 제2 SWD 레지스터(67₂)...을 갖는다. 제i SWD 레지스터(67_i)는, 결합 지표값 SWD_i를 보유한다. 셀렉터(68)는 SBS 레지스터 유닛(66), SWD 레지스터 유닛(67)의 한쪽을 선택하고, 선택한 레지스터 유닛에 보유되어 있는 데이터를 로직부(18a)로 보낸다.
- [0079] 도 9에, 거리 레지스터부(24)의 셀(24a) 및 CID 레지스터부(25)의 셀(25a)에 접속된 인에이블 신호 회로(71)의 일례를 도시한다. 인에이블 신호 회로(71)는 상술한 바와 같이 쌍을 이루는 셀(24a)과 셀(25a)마다 마련되어 있고, 인에이블 신호 회로(71)는 CID 레지스터부(25)의 셀(25a)에도 접속되어 있지만, 도 9에서는 셀(25a)의 도시를 생략하고 있다.
- [0080] 인에이블 신호 회로(71)는 AND 회로(71a, 71c, 71f), OR 회로(71b, 71d), NAND 회로(71e)로 구성되어 있다. OR 회로(71b)에는, 제1 제어 신호(CIDM flag) 및 제2 제어 신호(Fupdate_preset_N)가 입력되고, NAND 회로(71e)에는 제3 제어 신호(OF) 및 제4 제어 신호(Fset)가 입력되고, AND 회로(71f)에는, 제1 제어 신호 및 제5 제어 신호(Flag_enable)가 입력된다. OR 회로(71d)에는, NAND 회로(71e)의 출력과 AND 회로(71f)의 출력이 입력된다. AND 회로(71c)에는, OR 회로(71d)의 출력과, 제6 제어 신호(Fauto)가 입력된다. AND 회로(71a)에 AND 회로(71c) 및 OR 회로(71b)의 각 출력이 입력되고, 이 AND 회로(71a)의 출력이 인에이블 신호로서, 거리 레지스터부(24)의 레지스터인 셀(24a)의 인에이블 단자에 입력되고 있다.
- [0081] 각 셀(24a)에는, 셀렉터(도시 생략)를 통하여, 초기화용의 데이터, 계산 레지스터(36)의 내용, 셀(24a)이 관독한 데이터(데이터 무게 중심 간 거리) 자체의 어느 것이 입력되도록 되어 있다. 초기화용의 데이터는, 전체 비트가 「1」로 된 데이터 또는 「0」으로 된 데이터이다. 또한, 셀(24a)은 시프트 레지스터와 같이, 보유하고 있는 데이터를 상위 비트측으로 시프트함으로써 최상위 비트로부터 차례로 1비트씩 출력할 수 있다. 이 경우에, 셀(24a)이 관독한 데이터를 입력으로 복귀시키는 것에 의해, 전체 비트의 송출 완료 후에 셀(24a)의 내용이 원래의 상태로 복귀되도록 하고 있다.
- [0082] 제1 내지 제6 제어 신호는, 연산 유닛(12) 내에서 생성되는 것이다. 제1 제어 신호는 비교 플래그이다. 제1 제어 신호는, CID 마스크 회로(28)의 셀(28a)로부터 출력되는 비교 플래그이다. 제3 제어 신호는, 전가산기(35)의 자릿수 올림 신호이다. 제2, 제4, 제5, 제6 제어 신호는, 시스템 컨트롤러(11)로부터의 신호이다. 제2 제어 신호는, 거리 레지스터부(24)의 각 셀(24a)의 내용을 초기화의 제어하기 위한 로우 액티브 신호이다. 제4 제어 신호는, 「1」 또는 「0」으로 세트되는 세트 신호이다. 또한, 제5 제어 신호는, 제1 제어 신호(비교 플래그)를 유효하기 위한 신호이다. 제6 제어 신호는, 분류 계산 시에 거리 레지스터부(24)의 각 셀(24a)의 자동적인 갱신을 서포트하는 신호이다.
- [0083] 예를 들어, 요소 데이터 X에 대응한 각 셀(24a)의 각각에 최댓값(전체 비트가 「1」이 된 데이터)을 기입하고, 기타의 셀(24a)이 최솟값(전체 비트가 「0」이 된 데이터)을 남기도록 하는 거리 레지스터부(24)를 초기화하는 경우에서의, 인에이블 신호 회로(71)의 동작은 다음과 같이 된다. 또한, 이 초기화에서는, 거리 레지스터부(24)의 셀(24a)에 각 비트가 「0」인 데이터가 미리 기입되고, 또한 요소 데이터 X에 대응한 각 셀(28a)의 MID 레지스터(37d)에 「1」을, 기타의 MID 레지스터(37d)에는 「0」을 보유시킨 상태로 하여 CID 마스크 회로(28)가 초기화됨과 함께, 비교기(37e)에는 「1」의 지정 CID를 입력한 상태에서 행하여진다. 또한, CID 레지스터부(25)의 각 셀(25a)의 내용을 「1」로 한 상태에서 행하여진다. 또한, 각 셀(24a)에 전체 비트가 「1」이 된 초기화용의 데이터가 입력된다.
- [0084] 상기 초기화의 타이밍에 제2 제어 신호가 「0」, 제3 제어 신호가 「0」, 제4 제어 신호가 「1」, 제6 제어 신호가 「1」, 제5 제어 신호가 「1(0이어도 된다)」로 된다. 한편, 상기와 같이 설정되어 있는 CID 마스크 회로(28)로부터의 비교 플래그(제1 제어 신호)는 요소 데이터 X에 대응한 것이 「1」로, 그 밖의 것이 「0」으로 된다. 이에 의해, 요소 데이터 X에 대응한 셀(22a)의 인에이블 신호가 「1」로, 기타의 셀(22a)의 인에이블 신호가 「0」으로 되어, 요소 데이터 X에 대응한 셀(24a)만이 전체 비트가 「1」로 된 초기화용의 데이터를 래치하여 보유한다.
- [0085] 분류 계산에서는, 제2 제어 신호가 「1」, 제4 제어 신호가 「1」, 제5 제어 신호가 「0」, 제6 제어 신호가 「1」로 된다. 이에 의해, 제1 제어 신호(비교 플래그)에 상관없이, 전가산기(35)의 자릿수 올림 신호인 제3 제어 신호의 논리와 동일한 인에이블 신호가 입력된다. 이 결과, 전가산기(35)에서 자릿수 올림이 발생하면, 즉

셀(24a)이 보유하는 데이터 무게 중심 간 거리 DXV_i 보다도 계산 레지스터(36)가 보유하는 데이터 무게 중심 간 거리 DXV_i 가 작을 때에 인에이블 신호가 「1」로 되어, 당해 셀(24a)이 계산 레지스터(36)의 데이터 무게 중심 간 거리 DXV_i 를 래치하여 보유한다.

[0086] 또한, 신규 요소 데이터 X_{new} 가 분류된 클러스터 C_i 의 클러스터 ID를 CID 레지스터부(25)의 셀(25a)에 기입하는 경우에는, 먼저 CID 마스크 회로(28)의 신규 요소 데이터 X_{new} 에 대응하는 셀(28a)의 MID 레지스터(37d)의 내용을 분류처의 클러스터 ID로 한다. 또한, 그 클러스터 ID를 지정 CID로 하여, 각 셀(25a), 각 비교기(37e)에 입력한 상태로 한다. 이 상태에서, 제2 제어 신호를 「0」, 제4 제어 신호를 「1」, 제5 제어 신호를 「0」, 제6 제어 신호를 「1」로 한다. 이에 의해, 제1 제어 신호(비교 플래그)가 「1」로 되는 셀(25a)만의 인에이블 신호가 「1」로 된다. 비교 플래그는, 신규 요소 데이터 X_{new} 에 대응하는 셀(28a)로부터의 것만이 「1」로 되어 있으므로, 그 결과, 신규 요소 데이터 X_{new} 에 대응하는 셀(25a)에만 그 분류처의 클러스터 ID가 기입된다.

[0087] 도 10에, CID 마스크 회로의 MID 레지스터(37d)에 마련된 인에이블 신호 회로(75)의 일례를 도시한다. 상술한 바와 같이 인에이블 신호 회로(75)는 MID 레지스터(37d)마다 마련되어 있다. 인에이블 신호 회로(75)는 셀렉터(75a), AND 회로(75b), NAND 회로(75c), NOT 회로(75d)로 구성되어 있다. NOT 회로(75d)에는, 제7 제어 신호(Max Detector)가 입력되어 있다. 이 NOT 회로(75d)의 출력과 제8 제어 신호($i_presetMIDreg_N$)가 NAND 회로(75c)에 입력되어 있다. AND 회로(75b)에는, NAND 회로(75c)의 출력과, 제9 제어 신호(Disable_N)가 입력되어 있다. 셀렉터(75a)는 AND 회로(75b)로부터의 출력과, 제10 제어 신호(Column Decoder)가 입력되고, 어느 한쪽을 인에이블 신호로 하여, MID 레지스터(37d)의 인에이블 단자에 입력한다.

[0088] 제7 제어 신호는, 최댓값 검출 회로(27)로부터의 최대 플래그이며, 새로운 클러스터 무게 중심 V_i 의 초깃값을 설정할 때에는, 제7 제어 신호를 사용하여 MID 레지스터(37d)의 래치 동작을 제어한다. 또한, 제8 제어 신호는, 시스템 컨트롤러(11)로부터의 로우 액티브의 신호이며, 이 제8 제어 신호에 의해, 초기화 시에 MID 레지스터(37d)의 래치 동작을 제어한다. 제9 제어 신호는, 시스템 컨트롤러(11)로부터의 로우 액티브의 신호이며, 이 제9 제어 신호에 의해, 제8 제어 신호의 유효/무효를 전환한다. 제10 제어 신호는, 메인 메모리(14)의 칼럼 디코드 신호이며, 메인 메모리(14)의 열에 대응하는 CID 마스크 회로(28)의 셀(28a)을 제어하기 위한 신호이다. 이 제10 제어 신호는, MID 레지스터(37d)의 내용을 외부 설정 CID로 지정할 때 등에 인에이블 신호로서 사용된다. 구체적으로는, 예를 들어, 장래에 추가되는 신규 요소 데이터 X_{new} 을 위한 미사용 셀(28a)을 지정하는 경우에 사용된다. 제10 제어 신호는, 시스템 컨트롤러(11)로부터의 신호이다.

[0089] 상기와 같이 구성되는 연산 유닛(12)은 도 11에 도시한 바와 같이, 제1 파워드 메인(PD1) 내지 제6 파워드 메인(PD6)으로 나뉘어 있다. 제1 내지 제6 파워드 메인(PD1 내지 PD6)은, 전원(PS)으로부터의 전력 공급이 게이트 회로부(PG)를 통하여 시스템 컨트롤러(11)에 의해 독립적으로 제어된다. 시스템 컨트롤러(11)는 연산에 필요한 회로를 포함하는 파워드 메인에 필요한 타이밍에 전력 공급을 한다.

[0090] 이 예에서는, 메인 메모리(14)가 제1 파워드 메인(PD1)에, 무게 중심 메모리(15)가 제2 파워드 메인(PD2)에, 근린 탐색 회로부(17)가 제3 파워드 메인(PD3)에, 평가값 산출 회로(18)의 로직부(18a)가 제4 파워드 메인(PD4)에 각각 포함된다. 또한, 클러스터링 연산부(16)의 거리 산출 회로(22), 최댓값 검출 회로(27), CID 마스크 회로(28) 및 무게 중심 산출 회로(29)가 제5 파워드 메인(PD5)에 포함된다. 또한, 클러스터링 연산부(16)의 거리 레지스터부(24) 및 CID 레지스터부(25)와, 평가값 산출 회로(18)의 평가 레지스터부(18b)가 제6 파워드 메인(PD6)에 포함된다.

[0091] 메인 메모리(14)에의 요소 데이터 X 의 기입과, 그에 이어지는 일괄 처리에 있어서의 전력 공급의 상태를 도 12에 도시하는 바와 같이, 기간 T1에서는, 각 요소 데이터 X 를 기입하기 위하여 메인 메모리(14)의 제1 파워드 메인(PD1)에 전력 공급된다. 제1 파워드 메인(PD1)은, 클러스터링이 완료될 때까지 전력 공급이 계속된다. 기간 T2는, 클러스터링을 위한 초기 클러스터 무게 중심 V_i 를 초기 설정하는 기간이며, 제2 파워드 메인(PD2), 제5 파워드 메인(PD5), 제6 파워드 메인(PD6)에의 전력 공급이 개시되는데, 이 기간 T2에서는 로직부(18a)에 의한 제2 지표값 SBS_i 를 계산하지 않기 때문에, 제4 파워드 메인(PD4)의 전력 공급은 정지되어 있다.

[0092] 기간 T3이 되면, 제4 파워드 메인(PD4)에의 전력 공급이 개시된다. 이 기간 T3부터 기간 T7까지의 각 기간은, 각각 연산 유닛(12)으로 실질적으로 클러스터링하는 분류 계산하는 기간이다. 이 기간 T3부터 기간 T7까지의 각 기간에는, 갱신 처리에 사용되는 근린 탐색 회로부(17)의 제3 파워드 메인(PD3)을 제외하고, 각 파워드 메인에 전력 공급된다.

- [0093] 최후의 분류 계산이 종료되고 기간 T8이 되면, 제1 파워드 메인(PD1), 제2 파워드 메인(PD2), 제5 파워드 메인(PD5)에의 전력 공급이 정지된다. 이 기간 T8부터 기간 T10까지에서는, 평가값 E(Nc)를 평가값 산출 회로(18)의 로직부(18a)로 계산하기 위해서, 제4 파워드 메인(PD4)과 제6 파워드 메인(PD6)에의 전력 공급이 계속된다. 평가값 E(Nc)를 계산할 때의 기간 T10에서는, 클러스터링 연산부(16)를 사용하여 결합 지표값 SWD_i를 계산하기 위하여 제5 파워드 메인(PD5)에 전력 공급이 행하여진다.
- [0094] 상기 기간 T2부터 기간 T10까지가 1개의 클러스터수 Nc에 대한 처리 기간이며, 일괄 처리에서는, 최적의 클러스터수 Nc를 구하기 위해서, 기간 T2부터 기간 T10까지와 동일한 전력 공급의 제어가 반복하여 행하여진다. 이 예에서는, 분류 계산마다 로직부(18a)로 제2 지표값 SBS_i를 구하여 수시로 갱신하고 있기 때문에, 기간 T3로부터 제4 파워드 메인(PD4)에 전력 공급을 행하고 있다. 이렇게 제2 지표값 SBS_i를 수시로 갱신하는 경우에는, 클러스터 무게 중심 V_i의 수렴과 동시에 최종값으로서의 제2 지표값 SBS_i가 얻어진 상태로 되므로, 클러스터 무게 중심 V_i의 수렴 후에 제2 지표값 SBS_i를 취득하기 때문에, 다시 각 클러스터 무게 중심 V_i를 구하는 계산이 불필요하게 되어, 클러스터링의 고속화에 유리해진다.
- [0095] 또한, 제2 지표값 SBS_i는, 최종값만을 취득할 수 있으면 되므로, 예를 들어 이점쇄선으로 나타낸 바와 같이, 데이터 무게 중심 간 거리 DXV_i가 수렴한 것을 검출하고 나서 다음 클러스터 무게 중심 V_i의 계산 개시 전에 제4 파워드 메인(PD4)에 전력 공급을 행하여 제2 지표값 SBS_i를 계산해도 된다. 이와 같이 하면, 전력 절약화에 유리해진다.
- [0096] 도 13은, 갱신 처리에 있어서의 전력 공급의 상태를 도시하고 있다. 갱신 처리에서는, 신규 요소 데이터 X_{new}와, 무게 중심 메모리(15)로부터 판독한 각 클러스터 무게 중심 V_i를 사용하여, 근린 탐색 회로부(17)에 의해, 최소의 데이터 무게 중심 간 거리 DXV_i가 되는 클러스터 ID를 특정한다. 이 때문에, 최초의 기간 T11에 무게 중심 메모리(15)의 제2 파워드 메인(PD2)과, 근린 탐색 회로부(17)의 제3 파워드 메인(PD3)에 전력 공급한다. 제6 파워드 메인(PD6)은, 기간 T11부터 계속하여 전력 공급한다.
- [0097] 근린 탐색 회로부(17)에서 클러스터 ID를 특정한 후의 기간 T12에서는, 제3 파워드 메인(PD3)에의 전력 공급을 정지한다. 대신에, 로직부(18a)에 의해, 신규 요소 데이터 X_{new}가 분류된 클러스터 C_i의 새로운 클러스터 무게 중심 V_i를 계산하기 위하여 제4 파워드 메인(PD4)에의 전력 공급을 개시한다.
- [0098] 기간 T13에서는, 신규 요소 데이터 X_{new}가 분류된 클러스터 C_i의 새로운 제2 지표값 SBS_i를 로직부(18a)로 구하기 위해서, 제4 파워드 메인(PD4)에의 전력 공급을 계속한다. 이 새로운 제2 지표값 SBS_i의 계산에 필요로 되는 무게 중심 거리 DGV_i를 근린 탐색 회로부(17)로 계산하기 위해서, 기간 T13에서는 제3 파워드 메인(PD3)에의 전력 공급을 행한다. 기간 T14에서는, 제2 파워드 메인(PD2), 제3 파워드 메인(PD3)에의 전력 공급을 정지하고, 클러스터링 연산부(16)를 사용하여 새로운 결합 지표값 SWD_i를 계산하기 위하여 제5 파워드 메인(PD5)에 전력 공급한다. 이 기간 T14에 있어서, 로직부(18a)에 의해 평가값 E(Nc)가 계산된다. 기간 T14의 종료에서 제6 파워드 메인(PD6)을 제외하고 각 파워드 메인에 전력 공급이 정지된다.
- [0099] 제6 파워드 메인(PD6)은, 평가 레지스터부(18b)의 SBS 레지스터 유닛(66) 및 SWD 레지스터 유닛(67)을 불휘발성의 구성으로 한 경우에는, 도 12 및 도 13에 있어서 이점쇄선으로 나타낸 바와 같이, 전력 공급 및 그 정지를 제어할 수 있다. 예를 들어, 일괄 처리의 경우에는, 평가값 E(Nc)의 계산 후에, 제6 파워드 메인(PD6)의 전력 공급을 정지할 수 있다. 또한, 갱신 처리에서는, 기간 T11에 전력 공급을 개시하고, 갱신 처리의 종료 후에 전력 공급을 정지할 수 있다.
- [0100] 또한, 평가 레지스터부(18b)의 SBS 레지스터 유닛(66) 및 SWD 레지스터 유닛(67)의 내용을, 전력 공급의 정지 전에 다른 기억 장치에 퇴피시키고, 또한 전력 공급 개시 후에 퇴피시킨 내용을 복귀시키는 구성으로 하여, 마찬가지로 제6 파워드 메인(PD6)의 전력 공급 및 그 정지를 제어해도 된다. 또한, 제6 파워드 메인(PD6)은, 일괄 처리로부터 갱신 처리를 계속하여 행하는 경우나 갱신 처리 후에 계속하여 일괄 처리를 행하는 경우에는, 그 전력 공급을 계속한다.
- [0101] 다음으로 상기 구성의 작용에 대하여 설명한다. 메인 메모리(14)에는, 각 요소 데이터 X가 기입된 상태로 된다. 메인 메모리(14)의 각 열 중 요소 데이터 X가 기입되어 있지 않은 열의 각 단위 블록에는, 「0」이 기입

된다. 메인 메모리(14)에 각 요소 데이터 X가 기입된 상태에서 일괄 처리가 미실시인 경우에는, 일괄 처리가 행하여진다.

- [0102] 이 예의 일괄 처리에서는, 도 14에 도시하는 바와 같이, 클러스터수 N_c 를 1씩 증가시키면서, 전체 요소 데이터 X를 사용한 클러스터수 N_c 에 의한 클러스터링 처리를 행하여, 클러스터링 처리마다 평가값 $E(N_c)$ 를 산출한다. 그리고, 평가값 $E(N_c)$ 가 증가로부터 감소로 돌아서는 클러스터수 N_c , 즉 전회의 평가값 $E(N_c)$ 보다도 금회의 평가값 $E(N_c+1)$ 가 작아졌을 때의 클러스터수 N_c 를 최적의 값으로 하고, 그 최적의 클러스터수 N_c 로 클러스터링된 상태로 한다. 또한, 설정된 클러스터수 N_c 의 범위 내에서 평가치가 최대로 되는 클러스터수를 최적의 값으로 해도 된다.
- [0103] 일괄 처리에 의한 클러스터링의 실시에 앞서, 거리 레지스터부(24)는 기입된 요소 데이터 X에 대응하는 셀(24a)에 최댓값이, 어느 요소 데이터 X에도 대응하고 있지 않은 셀(24a)에 「0」이 기입되어서 초기화된다. 또한, CID 레지스터부(25)는 기입된 요소 데이터 X에 대응하는 각 셀(25a)에 「1」이, 어느 요소 데이터 X에도 대응하고 있지 않은 셀(25a)에는 「0」이 기입되어서 초기화된다. 그 후에, CID 마스크 회로(28)의 각 셀(28a)의 MID 레지스터(37d)에, 대응하는 셀(25a)의 클러스터 ID가 래치되어서 보유된다.
- [0104] 또한, 이하에 설명하는 클러스터링에서는, 메인 메모리(14)의 모든 열로부터 데이터가 판독되지만, 상기와 같이 거리 레지스터부(24), CID 레지스터부(25), CID 마스크 회로(28)를 초기화함으로써 요소 데이터 X 이외의 데이터는, 각종 계산 및 그 결과에 영향을 주지 않기 때문에, 요소 데이터 X 이외의 데이터에 관한 설명은 생략한다.
- [0105] 클러스터링 장치(10)는 시스템 컨트롤러(11)의 지시에 의해, 클러스터수 N_c 를 「2」로 한 클러스터링 처리를 개시한다. 클러스터링 처리에서는, 먼저 초기 설정을 행한다. 초기 설정에서는, 데이터 무게 중심 GG를 구함과 함께, 클러스터 무게 중심 V_2 의 초깃값이 되는 요소 데이터 X를 특정한다. 이 요소 데이터 X의 특징은, 클러스터 무게 중심 V_2 의 초깃값이 되는 요소 데이터 X를 클러스터 C_2 로 분류함으로써 행하고, 또한 데이터 무게 중심 GG가 잠정적인 클러스터 무게 중심 V_1 로 된다.
- [0106] 먼저, 메인 메모리(14)로부터 요소 데이터 X의 판독이 행하여진다. 판독된 요소 데이터 X는, CID 마스크 회로(28)로 보내짐과 함께, 지연 회로(21)로 보내진다. 이때에, CID 마스크 회로(28)는 지정 CID로서 「1」이 입력되고, 또한 입력 데이터로서 메인 메모리(14)로부터의 요소 데이터 X를 선택한다. 이 때문에, MID 레지스터(37d)에 「1」이 보유되어 있는 셀(28a)만이 요소 데이터 X를 출력하고, 또한 그 셀(28a)로부터의 비교 플래그만이 「1」로 된다. 따라서, 메인 메모리(14)의 요소 데이터 X가 기입되어 있지 않은 열에 대응한 셀(28a)의 비교 플래그는 「1」이 되지 않는다.
- [0107] 무게 중심 산출 회로(29)에서는, 셀렉터부(38)가 초기 상태에서, 예를 들어 CID 마스크 회로(28)로부터의 각 비교 플래그를 선택한다. 이에 의해, 「1」의 비교 플래그의 플래그수(신호수)가 가산기(39)에 의해 구해지며, 그리고, 그 결과가 제2 레지스터(42)에 보유된다. 이어서, 셀렉터부(38)가 요소 데이터 X를 선택하여 가산기(39)에 입력한다. 이에 의해, 입력되는 각 요소 데이터 X를 차원마다 가산한 q차원 벡터인 데이터 가산값이 제1 레지스터(41)에 보유된다. 그 후에, 제산기(43)에 의해, 제1 레지스터(41)에 보유하고 있는 데이터 가산값이 제2 레지스터(42)에 보유하고 있는 플래그수로 제산된다.
- [0108] 초기화에 의해, 요소 데이터 X에 대응하는 모든 셀(25a)의 클러스터 ID가 「1」로 이루어져 있기 때문에, 이때의 데이터 가산값 및 플래그수는, 전체 요소 데이터 X에 대하여 각각 구한 값이 된다. 따라서, 제산기(43)의 제산 결과로서 데이터 무게 중심 GG가 얻어진다.
- [0109] 상기와 같이 제산기(43)로 구해진 데이터 무게 중심 GG가, 평가값 산출 회로(18)의 GG 레지스터(63)에 보유된다. 또한, 시스템 컨트롤러(11)가 제2 레지스터(42)의 내용을 전체 요소 데이터 X의 데이터수로서 취득한다. 또한, 제산기(43)로 구해진 데이터 무게 중심 GG가, 잠정적인 클러스터 무게 중심 V_1 로서 거리 산출 회로(22)에 입력된다.
- [0110] 잠정적인 클러스터 무게 중심 V_1 의 입력에 동기하여, 먼저 메인 메모리(14)로부터 판독된 각 요소 데이터 X가 거리 산출 회로(22)에 입력된다. 거리 산출 회로(22)의 각 셀(22a)에서는, 입력되는 요소 데이터 X와 잠정적인 클러스터 무게 중심 V_1 로부터 데이터 무게 중심 간 거리 DXV_1 이 산출된다. 그리고, 이 산출된 데이터 무게 중심 간 거리 DXV_1 이, 그 시점에서 거리 레지스터부(24)의 셀(24a)에 보유되어 있는 값보다도 작은 경우에는, 셀

(24a)의 내용이 갱신된다.

- [0111] 현시점에서는, 요소 데이터 X에 대응한 각 셀(24a)은 최댓값이 보유되어 있기 때문에, 모든 요소 데이터 X에 대응한 각 셀(24a)의 내용이, 거리 산출 회로(22)에 의해 급회 산출된 데이터 무게 중심 간 거리 DXV_1 로 갱신된다. 동시에, 요소 데이터 X에 대응한 CID 레지스터부(25)의 각 셀(25a)의 내용이 갱신되는데, 이때에는 지정 CID로서 「1」이 부여되어 있으므로, 실질적으로 내용의 변경은 없다. 또한, 요소 데이터 X에 대응하고 있지 않은 셀(24a)에는, 초기화에서 거리 「0」이 부여되어 있기 때문에, 그 셀(24a)의 내용 및 그것에 대응한 셀(25a)의 내용(클러스터 ID=0)은 변경될 일은 없다.
- [0112] 상기와 같이 하여 셀(24a)의 내용이 갱신되어, 모든 셀(24a)에 보유되어 있는 데이터 무게 중심 간 거리 DXV_1 이 최댓값 검출 회로(27)에 입력된다. 이에 의해, 최댓값 검출 회로(27)로부터 출력되고 있는 M개의 최대 플래그 중, 입력되는 데이터 무게 중심 간 거리 DXV_1 중에서 최대의 것에 대응하는 최대 플래그만이 「1」로 된다.
- [0113] CID 마스크 회로(28)에서는, 최대 플래그가 「1」에 대응한 셀(28a)의 MID 레지스터(37d)만의 래치 동작이 허용된 상태로 된다. 또한, MID 레지스터(37d)에는, 「2」의 외부 설정 CID가 입력되어 있다. 이 때문에, 상기와 같이 최대 플래그가 변화한 후의 클럭에 의해, 「1」의 최대 플래그에 대응한 셀(28a)의 MID 레지스터(37d)의 내용만이, 외부 설정 CID의 내용(「2」)으로 갱신된다.
- [0114] 상기와 같이 1개의 MID 레지스터(37d)의 내용을 갱신한 후, 1회째의 분류 계산을 행한다. 1회의 분류 계산은, 분류 계산용으로 초기화, 클러스터 C_1 에 관한 처리, 클러스터 C_2 에 관한 처리를 차례로 행한다.
- [0115] 먼저, 거리 레지스터부(24)의 각 셀(24a), CID 레지스터부(25)의 각 셀(25a)의 내용이 분류 계산용으로 초기화된다. 즉, 기입된 요소 데이터 X에 대응하는 셀(24a)에 최댓값이, 어느 요소 데이터 X에도 대응하고 있지 않은 셀(24a)에 「0」이 기입되어서 초기화된다. 또한, CID 레지스터부(25)는 기입된 요소 데이터 X에 대응하는 각 셀(25a)에 「1」이, 어느 요소 데이터 X에도 대응하고 있지 않은 셀(25a)에는 「0」이 기입되어서 초기화된다. 또한, 요소 데이터 X에 대응하는 각 셀(25a)의 내용은, 계속하여 행하여지는 처리에서 반드시 갱신되기 때문에, 이 초기화에서는, 「1」 이외의 값으로 해도 된다.
- [0116] 분류 계산용의 초기화 후에, 클러스터 C_1 에 관한 처리를 행한다. 메인 메모리(14)로부터 요소 데이터 X가 판독되어, 지연 회로(21)와 CID 마스크 회로(28)에 입력된다. 이때에, CID 마스크 회로(28)에서는, 메인 메모리(14)로부터의 요소 데이터 X가 비교기(37e)에 입력되도록 셀렉터(37a)가 절환됨과 함께, 지정 CID로서 「1」이 비교기(37e)에 부여되어 있다. 이때의 CID 마스크 회로(28)의 각 MID 레지스터(37d)는, 클러스터 무게 중심 V_2 로 된 요소 데이터 X에 대응하는 MID 레지스터(37d)의 내용만이 「2」이며, 다른 요소 데이터 X에 대응하는 MID 레지스터(37d)의 내용은 「1」이다.
- [0117] CID 마스크 회로(28)의 각 셀(28a)에 요소 데이터 X가 각각 입력되면, 입력되는 요소 데이터 X 중, 대응하는 셀(25a)의 내용이 「1」로 되어 있는 셀(28a)에 입력된 요소 데이터 X만이 CID 마스크 회로(28)로부터 출력되어, 그 셀(28a)로부터의 비교 플래그가 「1」로 된다.
- [0118] 상기와 같이 요소 데이터 X와 비교 플래그가 CID 마스크 회로(28)로부터 출력된 상태에서, 무게 중심 산출 회로(29)에서는, 먼저 비교 플래그가 입력된 상태에서 가산기(39)가 가산을 행함으로써, 「1」로 되어 있는 비교 플래그의 플래그수가 계산되어, 그 계산 결과가 제2 레지스터(42)에 보유된다. 즉, 클러스터 C_1 로 분류되어 있는 요소 데이터 X의 데이터수 n_1 이 제2 레지스터(42)에 보유된다. 이어서, 요소 데이터 X가 가산기(39)에 입력되고, 가산기(39)에 의한 가산이 행하여져서, 그 계산 결과가 제1 레지스터(41)에 보유된다. 이에 의해, 제1 레지스터(41)에는, 클러스터 C_1 로 분류되어 있는 각 요소 데이터 X를 가산한 q차원 벡터인 데이터 가산값 SS_1 이 보유된다.
- [0119] 그 후에, 제산기(43)에 의해, 제1 레지스터(41)의 데이터 가산값 SS_1 을 제2 레지스터(42)의 데이터수 n_1 로 제산함으로써 클러스터 무게 중심 V_1 이 계산된다. 클러스터 무게 중심 V_1 은, 무게 중심 메모리(15)에 기입됨과 함께, 거리 산출 회로(22)에 입력된다. 또한, 제1 레지스터(41)의 데이터 가산값 SS_1 및 제2 레지스터(42)의 데이터수 n_1 이 각각 평가값 산출 회로(18)로 보내진다.
- [0120] 무게 중심 산출 회로(29)로부터의 데이터수 n_1 은, 데이터수 레지스터(64)에 보유되고, 또한 데이터 가산값

SS₁은, 감산기(52)에 입력된다. 그 후에, 평가값 산출 회로(18)에서는, 데이터수 레지스터(64)의 데이터수 n₁과 GG 레지스터(63)의 데이터 무게 중심 GG가 판독되어, 이들이 승산기(51)로 승산된다. 이 승산기(51)의 출력값(q차원 벡터)과, 무게 중심 산출 회로(29)로부터의 데이터 가산값 SS₁의 차분(q차원 벡터)을 감산기(52)로 구하고, 그 차분의 각 벡터 성분을 적산기(53)로 적산한다. 이에 의해, 현시점에 있어서의 클러스터 C₁의 제2 지표값 SBS₁이 계산된다. 제2 지표값 SBS₁은, 제1 SBS 레지스터(66₁)에 보유된다.

[0121] 여기서, 클러스터 무게 중심 V_i, 데이터수 n_i 및 데이터 가산값 SS_i는, 상기 식 (7)에 나타내어지는 관계가 있다. 이 때문에, 상술한 식 (5)로 표시되는 제2 지표값 SBS_i는, 다음 식 (8)과 같이 변형할 수 있다. 따라서, 데이터 무게 중심 GG, 데이터 가산값 SS_i 및 데이터수 n_i를 사용한 상기와 같은 평가값 산출 회로(18)에서의 계산에 의해, 제2 지표값 SBS_i를 구할 수 있다.

$$\begin{aligned}
 SBS_i &= n_i \times |V_i - GG| \\
 &= |n_i \times V_i - n_i \times GG| \\
 &= |SS_i - n_i \times GG| \quad \dots (8)
 \end{aligned}$$

[0122] 한편, 거리 산출 회로(22)에는, 무게 중심 산출 회로(29)로부터의 클러스터 무게 중심 V₁과 동기하여 지연 회로(21)로부터의 각 요소 데이터 X가 입력된다. 거리 산출 회로(22)의 각 셀(22a)에서는, 입력되는 각 요소 데이터 X에 대하여 클러스터 무게 중심 V₁과의 데이터 무게 중심 간 거리 DXV₁이 산출된다. 그리고, 셀(22a)로 산출된 데이터 무게 중심 간 거리 DXV₁이, 대응하는 거리 레지스터부(24)의 셀(24a)에 그 시점에서 보유되어 있는 거리보다도 짧으면, 당해 셀(24a)의 내용이 산출된 데이터 무게 중심 간 거리 DXV₁로 갱신되고, 대응하는 각 셀(25a)의 클러스터 ID도 갱신된다. 이때에, 요소 데이터 X에 대응하는 모든 셀(24a)은 최댓값을 보유하고 있기 때문에, 요소 데이터 X에 대응하는 모든 셀(24a, 25a)의 내용이 갱신된다. 각 셀(25a)에 대해서는, 지정 CID로서 「1」이 부여되어 있으므로, 요소 데이터 X에 대응하는 모든 셀(25a)의 클러스터 ID는 「1」로 된다.

[0124] 계속해서, 클러스터 C₂에 관한 처리를 행한다. 클러스터 C₂에 관한 처리는, 지정 CID로서 「2」를 사용하는 위에는 상기 클러스터 C₁에 관한 처리와 마찬가지로이다. 즉, 메인 메모리(14)로부터 판독되어서 CID 마스크 회로(28)에 입력된 요소 데이터 X 중, 대응하는 CID 레지스터부(25)의 셀(25a)의 내용이 「2」로 이루어져 있는 요소 데이터 X만이, 무게 중심 산출 회로(29)에 입력된다. 또한, CID 마스크 회로(28)로부터 출력되는 요소 데이터 X와 동일한 개수의 비교 플래그가 「1」로 된다. CID 마스크 회로(28)에서는, 상기한 바와 같이 클러스터 무게 중심 V₂로 되어 있는 요소 데이터 X에 대응하는 MID 레지스터(37d)의 내용만이 「2」이므로, 클러스터 무게 중심 V₂로 된 요소 데이터 X가 무게 중심 산출 회로(29)에 입력됨과 함께, 1개의 비교 플래그만이 「1」로 된다.

[0125] 그리고, 무게 중심 산출 회로(29)에 의해, CID 마스크 회로(28)로부터 출력되는 요소 데이터 X와 각 비교 플래그로부터, 클러스터 C₂에 관한 데이터수 n₂와, 클러스터 C₂에 관한 데이터 가산값 SS₂가 구해지고, 이들로부터 클러스터 무게 중심 V₂가 계산된다. 또한, 이 1회째의 클러스터 C₂에 관한 처리에서는, 클러스터 무게 중심 V₂로 된 요소 데이터 X만이 무게 중심 산출 회로(29)에 입력되기 때문에, 산출되는 클러스터 무게 중심 V₂는, 그 1개의 요소 데이터 X와 동일하다. 제산기(43)로 얻어지는 클러스터 무게 중심 V₂는, 무게 중심 메모리(15)에 기입됨과 함께, 거리 산출 회로(22)에 입력된다. 또한, 제1 레지스터(41)의 데이터 가산값 SS₂ 및 제2 레지스터(42)의 데이터수 n₂가 각각 평가값 산출 회로(18)로 보내진다.

[0126] 무게 중심 산출 회로(29)로부터의 데이터수 n₂는, 먼저 기입된 데이터수 n₁과는 별도로 데이터수 레지스터(64)에 보유되고, 또한 데이터 가산값 SS₂가 감산기(52)에 입력된다. 그 후에, 데이터수 레지스터(64)의 데이터수 n₂와, GG 레지스터(63)의 데이터 무게 중심 GG가 판독되어, 상기 제2 지표값 SBS₁과 마찬가지로 하여, 제2 지표값 SBS₂가 산출된다. 이 제2 지표값 SBS₂가 제2 SBS 레지스터(66₂)에 보유된다.

- [0127] 무게 중심 산출 회로(29)로부터의 클러스터 무게 중심 V_2 와 동기하여, 지연 회로(21)로부터의 각 요소 데이터 X 가 거리 산출 회로(22)에 입력된다. 거리 산출 회로(22)의 각 셀(22a)에 의해, 입력되는 요소 데이터 X 에 대하여 클러스터 무게 중심 V_2 와의 데이터 무게 중심 간 거리 DXV_2 가 산출된다. 그리고, 산출된 데이터 무게 중심 간 거리 DXV_2 가, 대응하는 거리 레지스터부(24)의 셀(24a)에 보유되어 있는 거리보다도 짧으면, 당해 셀(24a)의 내용이 산출된 데이터 무게 중심 간 거리 DXV_2 로 갱신되고, 대응하는 각 셀(25a)의 클러스터 ID도 갱신된다. 이때에, 각 셀(25a)에 대하여 지정 CID로서 「2」가 부여되어 있기 때문에, 그들 셀(25a)의 클러스터 ID는 「2」로 갱신된다. 따라서, 그때까지, 클러스터 C_1 로 분류되어 있었던 요소 데이터 X 중, 클러스터 무게 중심 V_1 보다도 클러스터 무게 중심 V_2 에 가까운 요소 데이터 X 는, 대응하는 셀(24a)의 내용과 동시에 셀(25a)의 내용이 갱신되어서 클러스터 C_2 로 분류된 상태로 된다.
- [0128] 그 후에, 각 MID 레지스터(37d)의 내용이 각각 대응하는 CID 레지스터부(25)의 셀(25a)의 내용으로 갱신된다. 여기서 1회째의 분류 계산이 종료한다.
- [0129] 1회째의 분류 계산이 종료되면, 2회째의 분류 계산이 행하여진다. 2회째의 분류 계산은, 1회째의 분류 계산과 마찬가지로의 수순으로, 분류 계산용의 초기화 후에 클러스터 C_1 에 관한 처리와 클러스터 C_2 에 관한 처리를 차례로 행한다. 먼저, 분류 계산용의 초기화에서는, 기입된 요소 데이터 X 에 대응하는 셀(24a)에 최댓값이, 각 셀(25a)에 「1」이 기입된다.
- [0130] 다음 클러스터 C_1 에 관한 처리에서는, 무게 중심 산출 회로(29)에 의해, 새로운 데이터수 n_1 , 데이터 가산값 SS_1 , 클러스터 무게 중심 V_1 이 구해진다. 그리고, 무게 중심 메모리(15)에 보유되어 있는 클러스터 무게 중심 V_1 및 데이터수 레지스터(64)에 보유되어 있는 데이터수 n_1 이 새롭게 산출된 것으로 각각 갱신된다. 또한, 평가값 산출 회로(18)에 있어서, 새로운 데이터수 n_1 과 데이터 가산값 SS_1 을 사용하여, 새로운 제2 지표값 SBS_1 이 산출되어, 이것으로 제1 SBS 레지스터(66₁)의 내용이 갱신된다.
- [0131] 또한, 거리 산출 회로(22)에 의해, 각 요소 데이터 X 에 관한 새로운 클러스터 무게 중심 V_1 과의 데이터 무게 중심 간 거리 DXV_1 이 산출된다. 그리고, 이렇게 산출된 새로운 데이터 무게 중심 간 거리 DXV_1 이 거리 레지스터부(24)의 셀(24a)의 내용보다도 짧으면, 그 셀(24a)의 내용이 새로운 데이터 무게 중심 간 거리 DXV_1 로 갱신됨과 함께, 그 셀(24a)에 대응하는 CID 레지스터부(25)의 셀(25a)의 내용이 「1」로 갱신된다.
- [0132] 클러스터 C_1 에 관한 처리에 이어서, 클러스터 C_2 에 관한 처리가 마찬가지로 행하여진다. 이에 의해, 무게 중심 산출 회로(29)에 의해 산출된 새로운 클러스터 무게 중심 V_2 , 데이터수 n_2 에, 무게 중심 메모리(15) 및 데이터수 레지스터(64)의 내용이 갱신된다. 또한, 무게 중심 산출 회로(29)에 의해 산출된 새로운 데이터 가산값 SS_2 및 새로운 데이터수 n_2 를 사용하여, 새로운 제2 지표값 SBS_2 가 평가값 산출 회로(18)로 산출되어, 제2 SBS 레지스터(66₂)의 내용이 갱신된다. 또한, 거리 산출 회로(22)에 의해, 각 요소 데이터 X 에 관한 새로운 클러스터 무게 중심 V_2 와의 데이터 무게 중심 간 거리 DXV_2 가 산출되고, 산출된 새로운 데이터 무게 중심 간 거리 DXV_2 가 거리 레지스터부(24)의 셀(24a)의 내용보다도 짧으면, 그 셀(24a)의 내용이 새로운 데이터 무게 중심 간 거리 DXV_2 로 갱신됨과 함께, 그 셀(24a)에 대응하는 CID 레지스터부(25)의 셀(25a)의 내용이 「2」로 갱신되어, 각 요소 데이터 X 의 각 클러스터로의 분류가 갱신된다. 그 후에, 각 MID 레지스터(37d)의 내용이 각각 대응하는 CID 레지스터부(25)의 셀(25a)의 내용으로 갱신되어, 2회째의 분류 계산이 종료된다.
- [0133] 3회째 이후의 분류 계산도 마찬가지로 행하여져, 클러스터 무게 중심 V_1 , V_2 , 데이터수 n_1 , n_2 , 제2 지표값 SBS_1 , SBS_2 가 갱신된다. 또한, 거리 레지스터부(24)의 각 셀(24a)의 내용 및 CID 레지스터부(25)의 각 셀(25a)의 내용이 갱신되어, 각 요소 데이터 X 의 각 클러스터로의 분류된 상태로 된다.
- [0134] 시스템 컨트롤러(11)는 상기와 같이 분류 계산마다, 무게 중심 메모리(15)의 내용을 감시하고 있다. 그리고, 시스템 컨트롤러(11)는 무게 중심 메모리(15)의 내용 변동이 없으면, 분류 계산을 종료시킨다. 분류 계산의 종료 시에 있어서, 무게 중심 메모리(15), 데이터수 레지스터(64), SBS 레지스터 유닛(66)에 보유되어 있는 클러스터 무게 중심 V_1 , V_2 , 데이터수 n_1 , n_2 , 제2 지표값 SBS_1 , SBS_2 는, 변동하지 않게 된, 즉 수렴한 클러스터 무

게 중심 V_1 , V_2 에 기초하여 각각 산출된 것으로 되어 있다.

[0135] 분류 계산의 종료 후에, 평가값 $E(Nc)$ 의 계산이 행하여진다. 평가값 $E(Nc)$ 의 계산을 위해서, 평가값 산출 회로(18)에서는, 처음에, 예를 들어 셀렉터(68)에 의해 SBS 레지스터 유닛(66)이 선택되고, SBS 레지스터(66₁, 66₂...)의 내용이 병렬적으로 판독되고, 판독된 SBS 레지스터(66₁, 66₂...)의 내용이 병렬 가산기(54)로 가산된다. 급회의 클러스터링에서는 클러스터수 Nc 가 「2」이기 때문에, 실질적으로는, SBS 레지스터 유닛(66)으로부터 제2 지표값 SBS_1 , SBS_2 가 판독되어서 병렬 가산기(54)로 가산된다. 이에 의해, 제2 지표값 SBS_1 의 총합인 클러스터 지표값 SBS 가 구해진다.

[0136] 이어서, 무게 중심 산출 회로(29)를 사용하여 클러스터 C_1 , C_2 에 관한 결합 지표값 SWD_1 , SWD_2 를 산출한다. 거리 레지스터부(24)의 각 셀(24a)로부터 각 클러스터 내 거리 DXV_i 를 판독하여 CID 마스크 회로(28)를 통하여 무게 중심 산출 회로(29)에 입력한다. 이때에, 먼저 CID 마스크 회로(28)의 각 셀(28a)에 지정 CID로서 「1」을 입력함으로써, 클러스터 C_1 로 분류되어 있는 각 요소 데이터 X 에 대응하는 클러스터 내 거리 DXV_1 만을 무게 중심 산출 회로(29)로 출력한다.

[0137] 무게 중심 산출 회로(29)에서는, 「1」로 되어 있는 비교 플래그의 플래그수로부터 클러스터 C_1 로 분류되어 있는 각 요소 데이터 X 의 데이터수 n_1 이 가산기(39)로 구해지고, 이것이 제2 레지스터(42)에 보유된다. 이어서, 각 클러스터 내 거리 DXV_i 를 적산했는데 제1 지표값 SD_1 이 가산기(39)로 구해지고, 이것이 제1 레지스터(41)에 보유된다. 이 후, 제산기(43)에 의해, 제1 레지스터(41)의 제1 지표값 SD_1 을 제2 레지스터(42)의 데이터수 n_1 로 제산함으로써 결합 지표값 SWD_1 이 구해진다. 이 결합 지표값 SWD_1 은, 평가값 산출 회로(18)로 보내져, 제1 SWD 레지스터(67₁)에 보유된다.

[0138] 계속해서, 클러스터 C_2 에 관한 결합 지표값 SWD_2 를 무게 중심 산출 회로(29)로 마찬가지로 하여 구하여, 제2 SWD 레지스터(67₂)에 보유한다. 또한, 이 경우에는, CID 마스크 회로(28)의 각 셀(28a)에 지정 CID로서 「2」를 입력하여, CID 마스크 회로(28)를 통하여 클러스터 C_2 로 분류되어 있는 각 요소 데이터 X 에 대응하는 클러스터 내 거리 DXV_2 만을 거리 레지스터부(24)로부터 무게 중심 산출 회로(29)에 입력한다.

[0139] 상기와 같이 하여 각 결합 지표값 SWD_1 , SWD_2 를 취득한 후, 평가값 산출 회로(18)에서는, 셀렉터(68)에 의해 SWD 레지스터 유닛(67)이 선택되어, SWD 레지스터(67₁, 67₂...)의 내용이 병렬적으로 판독된다. 그리고, 판독된 SWD 레지스터(67₁, 67₂...)의 내용이 병렬 가산기(54)로 가산된다. 제2 지표값 SBS_1 의 경우와 마찬가지로, 실질적으로는 SWD 레지스터 유닛(67)으로부터 결합 지표값 SWD_1 , SWD_2 가 판독되어서 병렬 가산기(54)로 가산된다. 이 병렬 가산기(54)의 가산 결과로서 내적 결합도 SWD 가 얻어진다.

[0140] 병렬 가산기(54)로 얻어진 내적 결합도 SWD 는, 승산기(55)에 입력되고, 이 승산기(55)에 의해 시스템 컨트롤러(11)로부터 입력되는 클러스터수 $Nc(=2)$ 로 승산된다. 승산기(55)의 승산 결과는 분모 레지스터(58)에 보유된다. 또한, 내적 결합도 SWD 에 클러스터수 Nc 를 곱하는 것은, 평가값 $E(Nc)$ 를 다음 제산으로 구할 때에, 클러스터 지표값 SBS 를 클러스터수 Nc 로 규격화하여 외적 분리도(SBS/Nc)로 하기 위해서이다. 계속해서, 제산기(60)에 의해, 분자 레지스터(59)의 내용을 분모 레지스터(58)의 내용으로 제산함으로써, 클러스터수 Nc 가 「2」일 때의 평가값 $E(2)$ 가 산출된다.

[0141] 상기한 바와 같이 평가값 산출 회로(18)는 클러스터링 연산부(16)가 클러스터링을 행하기 위한 계산 과정의 계산값을 이용하여 평가값 $E(2)$ 를 산출한다. 즉, 클러스터 무게 중심 V_1 , V_2 를 계산할 때에 그 계산 도중에 산출되는, 데이터 가산값 SS_1 , SS_2 , 데이터수 n_1 , n_2 를 이용하여 얻어지는 제2 지표값 SBS_1 , SBS_2 를 사용하여 평가값 $E(2)$ 를 산출한다. 따라서, 고속이고 또한 저소비 전력으로 효율적으로 평가값 $E(2)$ 가 구해진다. 그 후에 산출되는 평가값 $E(Nc)$ 에 대해서도 마찬가지이다.

[0142] 시스템 컨트롤러(11)는 상기와 같이 하여 제산기(60)로 얻어지는 평가값 $E(2)$ 와, 이 시점에서 CID 레지스터부(25)에 보유되어 있는 각 셀(25a)의 내용 즉 클러스터수 Nc 가 「2」인 때의 각 요소 데이터 X 에 대응한 클러스터 ID를 취득한다. 시스템 컨트롤러(11)는 취득한 평가값 $E(2)$ 및 각 클러스터 ID를 기억부(도시 생략)에 보유

한다.

- [0143] 평가값 E(2) 및 각 클러스터 ID의 취득 후, 클러스터수 Nc를 「3」으로 한 클러스터링 처리를 행한다. 클러스터수 Nc의 「3」클러스터링 처리에 있어서도, 클러스터수 Nc가 「2」인 경우와 마찬가지로, 초기 설정, 분류 계산, 평가값 산출을 행한다.
- [0144] 클러스터수 Nc가 「3」인 초기 설정에서는, 예를 들어 클러스터수 Nc를 「2」로 한 경우에 얻어진 클러스터 무게 중심 V₁, V₂와 함께, 클러스터 내 거리 DXV₁, DXV₂ 중에서 최대의 클러스터 내 거리 DXV_i에 대응하는 요소 데이터 X인 클러스터 무게 중심 V₃을 초깃값으로 한다. 이에 의해, 클러스터 무게 중심 V_i의 수렴을 빠르게 한다.
- [0145] 클러스터 무게 중심 V₃의 초깃값의 설정은, 상술한 클러스터 무게 중심 V₂의 초깃값을 설정한 경우와 마찬가지로의 수순으로 행하지만, 거리 레지스터부(24)에 클러스터 내 거리 DXV₁, DXV₂가 보유되어 있으므로, 클러스터 무게 중심 V₁, V₂의 계산, 클러스터 내 거리 DXV₁, DXV₂의 계산은 행하지 않는다. 거리 레지스터부(24)의 각 셀(24a)로부터 모든 클러스터 내 거리 DXV₁, DXV₂를 관독하고, 최댓값 검출 회로(27) 및 CID 마스크 회로(28)를 사용하여, 「1」의 최대 플래그 즉 최대의 클러스터 내 거리 DXV_i에 대응한 셀(28a)의 MID 레지스터(37d)만의 래치 동작이 허용된 상태로 한다. 각 MID 레지스터(37d)에는, 「3」의 외부 설정 CID가 입력되어 있으므로, 상기와 같이 최대 플래그가 변화한 후의 클럭에 의해, 「1」의 최대 플래그에 대응한 셀(28a)의 MID 레지스터(37d)의 내용만이, 외부 설정 CID의 내용(「3」)으로 갱신된다.
- [0146] 상기와 같이 초기 설정을 행한 후에, 1회째의 분류 계산을 행한다. 1회째의 분류 계산은, 클러스터수 Nc가 「2」인 경우와 마찬가지로, 분류 계산용의 초기화를 하고 나서, 지정 CID를 「1」로 하여 클러스터 C₁에 관한 처리를 행하고 나서, 지정 CID를 「2」로 하여 클러스터 C₂에 관한 처리를 행한다. 클러스터 C₂에 관한 처리 후에, 지정 CID를 「3」으로 하여, 클러스터 C₃에 관한 처리를, 클러스터 C₁, C₂의 경우와 마찬가지로 행한다.
- [0147] 클러스터 C₃에 관한 처리에서 얻어지는 클러스터 무게 중심 V₃이 무게 중심 메모리(15)에 기입되고, 데이터수 n₃이 데이터수 레지스터(64)에 기입된다. 또한, 데이터수 n₃ 및 데이터 무게 중심 GG와, 데이터 가산값 SS₃으로부터 얻어지는 제2 지표값 SBS₃이 제3 SBS 레지스터(66₃)에 기입된다. 또한, 데이터 무게 중심 GG는, 새롭게 산출해도 되지만, 이 예에서는 클러스터수 Nc가 「2」인 때에 구하여 GG 레지스터(63)에 보유한 것을 그대로 사용하고 있다.
- [0148] 1회째의 분류 계산 후에는, 2회째의 분류 계산을 마찬가지로 행한다. 이후, 분류 계산을 마찬가지로 반복하여 행한다. 분류 계산마다, 거리 레지스터부(24)의 각 셀(24a)의 내용 및 CID 레지스터부(25)의 각 셀(25a)의 내용이 갱신되고, 각 요소 데이터 X의 각 클러스터 C_i로의 분류가 갱신된다. 또한, 무게 중심 메모리(15)의 클러스터 무게 중심 V₁ 내지 V₃이 갱신되고, 데이터수 레지스터(64)의 데이터수 n₁ 내지 n₃, 제1 내지 제3 SBS 레지스터(66₁ 내지 66₃)의 제2 지표값 SBS₁ 내지 SBS₃이 갱신된다.
- [0149] 무게 중심 메모리(15)의 내용(클러스터 무게 중심 V₁ 내지 V₃)이 변동하지 않게 되면, 시스템 컨트롤러(11)에 의해, 분류 계산이 종료된다. 그 후에, 평가값 E(3)가 평가값 산출 회로(18)에 의해 계산된다. 이 평가값 E(3)시에는, 무게 중심 산출 회로(29)를 사용하여 클러스터 C₁ 내지 C₃에 관한 결합 지표값 SWD₁ 내지 SWD₃이 계산된다.
- [0150] 시스템 컨트롤러(11)는 상기와 같이 하여 얻어지는 평가값 E(3)과, 이 시점에서 CID 레지스터부(25)에 보유되어 있는 각 셀(25a)의 내용 즉 클러스터수 Nc가 「3」인 때의 각 요소 데이터에 대응한 클러스터 ID를 취득한다. 시스템 컨트롤러(11)는 취득한 평가값 E(3) 및 각 클러스터 ID를 기억부에 보유한다.
- [0151] 이후, 마찬가지로 하여 클러스터수 Nc를 1씩 증가시키면서, 클러스터링 처리를 행하여, 클러스터수 Nc마다의 평가값 E(Nc)와 각 요소 데이터 X에 대응한 클러스터 ID를 취득하여 기억한다.
- [0152] 시스템 컨트롤러(11)는 전회의 평가값 E(Nc)보다도 금회의 평가값 E(Nc+1)가 작아졌을 때에, 전회의 클러스터수 Nc를 최적의 값으로 한다. 그리고, 그 후에, 연산 유닛(12)의 상태를 최적의 클러스터수 Nc에 의한 클러스터링한 상태로 복원한다.

- [0153] 복원하기 위해서, 예를 들어, 시스템 컨트롤러(11)는 연산 유닛(12)의 메인 메모리(14) 및 GG 레지스터(63)를 제외하고 각 부를 초기화한 상태로 하고 나서, 기억부에 기억하고 있는 최적의 클러스터수 N_c 에 대응한 각 클러스터 ID를 CID 레지스터부(25)의 각 셀(25a)에 재기입한다. 그 후에, 클러스터 $C_1, C_2 \dots C_{N_c}$ 에 관한 분류 계산을 1회 행하고, 그 후에 평가값 $E(N_c)$ 를 산출한다.
- [0154] 이에 의해, 무게 중심 메모리(15)의 클러스터 무게 중심 V_1 내지 V_{N_c} , 거리 레지스터부(24)의 각 셀(24a)의 클러스터 내 거리 DXV_1 내지 DXV_{N_c} , CID 레지스터부(25)의 셀(25a)의 클러스터 ID, SBS 레지스터 유닛(66)의 제2 지표값 SBS_1 내지 SBS_{N_c} , SWD 레지스터 유닛(67)의 결합 지표값 SWD_1 내지 SWD_{N_c} , 데이터수 레지스터(64)의 데이터수 n_1 내지 n_{N_c} 가, 최적의 클러스터수 N_c 로 클러스터링한 최종적인 값으로 각각 복원된다. 또한, CID 레지스터부(25)에 기입된 클러스터 ID는, 분류 계산에 의해 변경될 일은 없다.
- [0155] 또한, 복원의 방법은, 상기의 것에 한정되지 않는다. 예를 들어, 클러스터수 N_c 마다 분류 계산 종료 후의 연산 유닛(12)의 각 부가 보유하고 있는 내용을 기억부에 기억해 두고, 이것 중의 최적의 클러스터수 N_c 의 것을 각 부에 재기입하도록 해도 된다. 또한, 전회의 평가값 $E(N_c)$ 보다도 금회의 평가값 $E(N_c+1)$ 가 작아졌을 때에 전회의 클러스터수 N_c 를 최적의 값으로 하는 경우에는, 복원을 위하여 전회의 클러스터수 N_c 에 의한 CID 레지스터부(25)의 각 클러스터 ID만을 기억부에 기억해도 된다.
- [0156] 이상과 같이, 최적의 클러스터수 N_c 에 의한 클러스터링한 상태로 연산 유닛(12)이 복원되어서, 일괄 처리가 종료된다.
- [0157] 일괄 처리의 종료 후에, 신규 요소 데이터 X_{new} 가 추가되었을 때에는, 갱신 처리를 행한다. 도 15에 도시하는 바와 같이, 갱신 처리에서는, 신규 요소 데이터 X_{new} 에 대하여 데이터 무게 중심 간 거리 DXV_i 가 최소가 되는 클러스터 C_i 로 그 신규 요소 데이터 X_{new} 를 분류한다. 즉, 신규 요소 데이터 X_{new} 를, 그것의 가장 가까운 클러스터 C_i 로 분류한다. 이 분류 후에, 신규 요소 데이터 X_{new} 가 분류된 클러스터 C_i 의 클러스터 무게 중심 V_i 를 갱신하고 나서, 평가값 $E(N_c)$ 를 산출한다. 그리고, 평가값 $E(N_c)$ 에 기초하여 갱신 처리 후의 클러스터링 결과의 타당성을 판단한다.
- [0158] 이하, 갱신 처리에 대하여 구체적으로 설명한다. 이 예에서는, 신규 요소 데이터 X_{new} 의 추가에 응답하여 갱신 처리가 행하여진다. 추가되는 신규 요소 데이터 X_{new} 는, 시스템 컨트롤러(11)에 의해, 먼저 근린 탐색 회로부(17)에 입력되고, 근린 탐색 회로부(17)에는, 클러스터 무게 중심 V_i 가 무게 중심 메모리(15)로부터 차례로 판독되어서 입력된다. 근린 탐색 회로부(17)에서는, 신규 요소 데이터 X_{new} 와, 차례로 입력되는 클러스터 무게 중심 V_i 의 데이터 무게 중심 간 거리 DXV_i 가 계산부(17a)로 순차적으로 계산된다.
- [0159] 계산부(17a)에서 얻어지는 새로운 데이터 무게 중심 간 거리 DXV_i 가 근거리 레지스터부(17b)에 보유되어 있는 내용보다도 작을 때에는, 근거리 레지스터부(17b)의 내용이 새로운 데이터 무게 중심 간 거리 DXV_i 로 갱신된다. 이에 의해, 최종적으로, 신규 요소 데이터 X_{new} 에 대하여 최소가 되는 데이터 무게 중심 간 거리 DXV_i 가 근거리 레지스터부(17b)에 보유된다. 한편, 근거리 CID 레지스터부(17c)에는, 근린 탐색 회로부(17)에 입력되어 있는 클러스터 무게 중심 V_i 에 대응하는 클러스터 ID를 나타내는 지정 CID가 입력된다. 이에 의해, 근거리 CID 레지스터부(17c)에는, 최소가 되는 데이터 무게 중심 간 거리 DXV_i 에 대응한 클러스터 ID가 보유된다. 이와 같이 하여, 신규 요소 데이터 X_{new} 는, 데이터 무게 중심 간 거리 DXV_i 가 최소가 되는 클러스터 C_i 로 분류된다.
- [0160] 이하, 최종적으로 근거리 CID 레지스터부(17c)에 보유되어 있는 클러스터 ID를 「k」로 하여 설명한다. 여기서, k는, 1, 2... N_c 의 어느 값이다.
- [0161] 이어서, 평가값 산출 회로(18)에 의해, 신규 요소 데이터 X_{new} 가 추가된 후의 클러스터 무게 중심 V_k 를 식 (9)에 의해 구한다. 식 (9) 중에 있어서, 값 V_{kOLD} , 값 n_{kOLD} 는, 모두 신규 요소 데이터 X_{new} 가 추가(분류)되기 전의 클러스터 C_k 에 관한 클러스터 무게 중심 및 데이터수이며, 클러스터 무게 중심 V_{kOLD} 는 무게 중심 메모리(15)에, 데이터수 n_{kOLD} 는 데이터수 레지스터(64)에 보유되어 있다.

$$V_k = \frac{n_{kOLD} \cdot V_{kOLD} + X_{new}}{n_{kOLD} + 1} \quad \dots (9)$$

[0162]

[0163]

평가값 산출 회로(18)의 데이터수 레지스터(64)로부터 판독된 데이터수 n_{kOLD} 와, 고정값 「1」이 가산기(56)에 입력되어서 값 「 $n_{kOLD}+1$ 」이 구해지고, 이것이 분모 레지스터(58)에 보유된다. 또한, 무게 중심 메모리(15)로부터 판독된 클러스터 무게 중심 V_{kOLD} 와 데이터수 레지스터(64)로부터 판독된 데이터수 n_{kOLD} 가 승산기(51)에 입력되어서, 값 「 $n_{kOLD} \times V_{kOLD}$ 」가 구해진다. 승산기(51)로부터의 값 「 $n_{kOLD} \times V_{kOLD}$ 」와, 신규 요소 데이터 X_{new} 가 가산기(57)에 입력되고, 값 「 $n_{kOLD} \times V_{kOLD} + X_{new}$ 」가 얻어지고, 이 값이 분자 레지스터(59)에 보유된다. 그리고, 계산기(60)에 의해, 분모 레지스터(58)의 내용으로 분자 레지스터(59)의 내용을 계산함으로써 신규 요소 데이터 X_{new} 가 분류된 후의 클러스터 무게 중심 V_k 가 산출된다. 이렇게 산출된 클러스터 무게 중심 V_k 가 무게 중심 메모리(15)에 기입됨으로써 클러스터 C_k 의 클러스터 무게 중심 V_k 가 갱신된다.

[0164]

클러스터 무게 중심 V_k 의 갱신 후, 제2 지표값 SBS_k 와 결합 지표값 SWD_k 를 갱신한다. 먼저, 새로운 제2 지표값 SBS_k 를 구하기 위해서, 무게 중심 메모리(15)로부터 클러스터 무게 중심 V_k 가, GG 레지스터(63)로부터 데이터 무게 중심 GG가 각각 판독되어, 이들이 근린 탐색 회로부(17)의 전가산기(45)에 각각 입력된다. 이에 의해, 가산기(61)의 출력으로서 클러스터 무게 중심 V_k 와 데이터 무게 중심 GG의 차분 벡터(q 차원 벡터)가 얻어진다. 차분 벡터는, 평가값 산출 회로(18)의 승산기(51)에 셀렉터(50)를 통하여 입력된다. 또한, 이 승산기(51)에 데이터수 레지스터(64)로부터 판독된 데이터수 n_k 가 입력된다. 이에 의해, 데이터수 n_k 를 곱한 차분 벡터가 얻어진다. 이 데이터수 n_k 를 곱한 차분 벡터가, 감산기(52)를 패스스루하여 적산기(53)에 입력됨으로써, 제2 지표값 SBS_k 가 얻어진다. 이렇게 산출된 제2 지표값 SBS_k 에, SBS 레지스터 유닛(66)의 제 k SBS 레지스터의 내용이 갱신된다.

[0165]

계속하여 결합 지표값 SWD_k 를 갱신한다. 새로운 결합 지표값 SWD_k 는, 클러스터링 연산부(16)에서 산출된다. 신규 요소 데이터 X_{new} 가 메인 메모리(14)의 미사용 열에 기입됨과 함께, 그 열에 대응한 CID 레지스터부(25)의 셀(25a)에 클러스터 ID로서 「 k 」가 기입된다. 이 클러스터 ID의 기입에서는, 예를 들어 시스템 컨트롤러(11)가 근거리 CID 레지스터부(17c)에 보유되어 있는 클러스터 ID를 판독하고, 그 클러스터 ID를 지정 CID로서 CID 레지스터부(25)의 각 셀(25a)에 부여한 상태에서, 신규 요소 데이터 X_{new} 에 대응하는 셀(25a)만 래치 동작을 행하게 함으로써 행한다.

[0166]

CID 레지스터부(25)에 신규 요소 데이터 X_{new} 의 클러스터 ID를 기입한 후에, CID 마스크 회로(28)의 각 MID 레지스터(37d)의 내용이, 각각 대응하는 CID 레지스터부(25)의 셀(25a)의 내용으로 갱신된다. 이 후, CID 마스크 회로(28)의 각 셀(28a)의 비교기(37e)에 클러스터 ID로서 「 k 」를 지정하는 지정 CID를 부여한다. 이에 의해, 셀(25a)의 클러스터 ID가 「 k 」로 되어 있는 셀(28a)로부터의 비교 플래그만이 「1」로 된다.

[0167]

계속해서, 메인 메모리(14)로부터 신규 요소 데이터 X_{new} 를 포함하는 모든 요소 데이터 X 가 판독되어서, 지연 회로(21)를 통하여 거리 산출 회로(22)에 입력된다. 또한, 거리 산출 회로(22)에는, 무게 중심 메모리(15)로부터 판독된 클러스터 무게 중심 V_k 가 입력된다. 이에 의해, 각 셀(22a)의 전가산기(32)는 그 한쪽의 입력단에 요소 데이터 X 가 입력되고, 다른 쪽의 입력단에 클러스터 무게 중심 V_k 가 입력된다. 전가산기(32)에 의한 q 차원분의 계산이 완료되면, 각 셀(22a)의 계산 레지스터(36)에는, 각각 대응하는 요소 데이터 X 에 관한 데이터 무게 중심 간 거리 DXV_k 가 보유된 상태로 된다.

[0168]

상기와 같이 하여 전가산기(32)에 의한 계산의 완료 후, 「1」의 비교 플래그에 대응한 셀(24a). 셀(25a), 즉 보유하고 있는 클러스터 ID가 「 k 」인 셀(25a)과, 이것에 대응하는 셀(24a)만이 래치 동작을 행하도록 된다. 이에 의해, 다음 클럭이 발생한 타이밍에, 신규 요소 데이터 X_{new} 를 포함하는 클러스터 C_k 로 분류되어 있는 각 요소 데이터 X_k 에 대응하는 셀(24a), 셀(25a)의 내용이 갱신된다.

[0169]

이 결과, 클러스터 C_k 로 분류되어 있는 각 요소 데이터 X 에 대응하는 셀(24a)에는, 신규 요소 데이터 X_{new} 의 분류 후에 구한 새로운 클러스터 무게 중심 V_k 를 사용하여 산출된 데이터 무게 중심 간 거리(클러스터 내 거리)

DXV_k가 보유된다. 한편, CID 레지스터부(25)에는, 지정 CID로서 「k」를 부여하고 있기 때문에, 클러스터 C_k로 분류되어 있는 각 요소 데이터 X에 대응하는 셀(25a)에는 「k」가 기입되기 때문에, 실제로는 내용은 변경되지 않는다. 이와 같이 하여, 클러스터 C_k로 분류되어 있는 각 요소 데이터 X에 관한 클러스터 내 거리 DXV_k가 새로운 클러스터 무게 중심 V_k에 대응한 값으로 갱신된다.

[0170] 계속하여 거리 레지스터부(24)로부터 각 클러스터 내 거리 DXV_i가 판독되어서, CID 마스크 회로(28)를 통하여 무게 중심 산출 회로(29)에 입력된다. 각 MID 레지스터(37d)의 내용은, 대응하는 CID 레지스터부(25)의 셀(25a)과 동일하게 되어 있고, 지정 CID로서 「k」를 지정하고 있다. 이 때문에, 무게 중심 산출 회로(29)에는, 각 클러스터 내 거리 DXV_k만이 입력된다. 또한, 무게 중심 산출 회로(29)에 입력되어 있는 비교 플래그는, 클러스터 C_k로 분류되어 있는 각 요소 데이터 X의 개수와 동일한 플래그수가 「1」로 되어 있다. 이 때문에, 무게 중심 산출 회로(29)에서는, 분류 계산의 경우와 마찬가지로 하여, 각 클러스터 내 거리 DXV_k를 가산한 제1 지표값 SD_k와, 클러스터 C_k로 분류되어 있는 각 요소 데이터 X의 데이터수 n_k가 구해지고, 이들로부터 새로운 결합 지표값 SWD_k가 계산된다. 새로운 결합 지표값 SWD_k는, 평가값 산출 회로(18)로 보내지고, 이 새로운 결합 지표값 SWD_k으로 SWD 레지스터 유닛(67)의 제k SWD 레지스터의 내용이 갱신된다.

[0171] 이 후, 평가값 산출 회로(18)에 의해, SBS 레지스터 유닛(66)의 각 내용 및 SWD 레지스터 유닛(67)의 각 내용을 사용하여 평가값 E(Nc)가 계산된다. 이때의 평가값 E(Nc)의 계산 수순은, 일괄 처리에 있어서의 클러스터링 후에 행하는 수순과 동일하다.

[0172] 상기 설명으로부터 알 수 있는 바와 같이, 갱신 처리에 있어서는, 신규 요소 데이터 X_{new}가 분류되는 클러스터 C_k의 클러스터 무게 중심 V_k는 갱신하지만, 데이터 무게 중심 GG의 갱신을 행하지 않는다. 통상적으로는, 기존의 요소 데이터 X의 전체 데이터수는 매우 크고, 예를 들어 1개 또는 수개의 신규 요소 데이터 X_{new}의 추가에 의한 데이터 무게 중심 GG의 이동량은 매우 작다. 이 때문에, 무게 중심 거리 DGV_i를 파라미터로서 포함하는 제2 지표값 SBS_i의 변동은 매우 작아, 데이터 무게 중심 GG를 갱신하지 않는 것의 평가값 E(Nc)에 대한 영향은 상당히 작다. 이에 반해, 1개 또는 수개의 신규 요소 데이터 X_{new}의 추가에 의한 클러스터 무게 중심 V_k의 이동은, 데이터수 n_k에 따라 다르지만, 데이터 무게 중심 GG의 이동에 비하여 상당히 크다. 이 때문에, 결합 지표값 SWD_k 및 제2 지표값 SBS_k의 변동이 커져서, 그들 변동에 수반하는 평가값 E(Nc)의 변동이 커지기 때문이다.

[0173] 시스템 컨트롤러(11)는 상기와 같이 하여 산출되는 평가값 E(Nc)를 취득하고, 이 평가값 E(Nc)에 기초하여 갱신 처리에 의한 클러스터링 결과의 타당성을 판단한다. 즉, 상기와 같이 갱신 처리에서 신규 요소 데이터 X_{new}를 분류한 후에도 클러스터링 상태가 적정하게 유지되고 있는지 여부를 판단한다.

[0174] 이 예에서는, 갱신 처리의 직전에 행하여진 일괄 처리로 적정으로 된 클러스터수 Nc에 대하여 얻어진 평가값 E(Nc)를 기준 평가값으로 하고, 이 기준 평가값과 갱신 처리에서 얻어진 평가값 E(Nc)를 비교한다. 이 비교에서, 예를 들어 후자가 전자와 동일하거나 그 이상으로 되어 있는 경우에, 타당하다고 판단하고, 처리를 종료한다. 한편, 후자가 전자보다도 작게 되어 있는 경우에는, 타당하지 않다고 판단하고, 일괄 처리를 실시한다.

[0175] 타당하지 않다고 판단한 경우의 일괄 처리는, 상기와 마찬가지로의 수순에 의해 행한다. 이 경우에, 전회에 일괄 처리로 얻어진 적정한 클러스터수 Nc로 얻어진 각 클러스터 무게 중심 V_i, 각 요소 데이터에 대한 클러스터 ID 혹은 갱신 처리에서 얻어진 각 요소 데이터에 대한 클러스터 ID를 사용하여, 초기 설정을 행하는 것도 바람직하다. 이와 같이 함으로써, k-means법에 의한 클러스터링을 조기에 수렴할 수 있어, 계산 횟수, 계산 시간을 단축할 수 있다.

[0176] 이상과 같이 하여 클러스터링을 행하기 때문에, 신규 요소 데이터 X_{new}가 효율적이고 또한 고속으로 분류된다. 그리고, 이 분류에 의한 클러스터링 결과는, 평가값 E(Nc)에 의해 그 타당성이 판단되어, 클러스터링 결과가 나빠져 있는 경우에는, 일괄 처리로 신규 요소 데이터를 포함하는 전체 요소 데이터 X로 클러스터링이 행하여져, 최적의 클러스터수 Nc로 분류되기 때문에, 클러스터링의 정밀도가 높게 유지된다. 이것을 예를 들어 학습 기능을 갖는 자동 인식 장치에 사용한 경우, 갱신 처리에 의해 고속, 실시간 인식(분류)이 실현되어, 상황에 따라서 일괄 처리에 의한 고정밀도의 학습이 가능해진다.

[0177] 갱신 처리에서 얻어지는 평가값 E(Nc)로부터 클러스터링 상태가 적정하게 유지되고 있는지 여부를 판단하는 방

법은, 상기 방법에 한정되지 않지만, 상기한 바와 같이 갱신 처리보다도 전에 얻어진 평가값 E(Nc)를 기준 평가값으로 하여 비교하는 것은 바람직한 방법이다. 또한, 기준 평가값으로서, 일괄 처리 및 갱신 처리를 막론하고, 급회의 갱신 처리의 직전에 행하여진 클러스터링에 의한 클러스터링 결과를 평가한 평가값을 기준 평가값으로서 사용하는 것도 가능하다. 또한, 갱신 전 평가값에 대한 갱신 후 평가값의 저하된 크기가 소정의 범위 내(예를 들어 갱신 전 평가값의 10% 내지 15% 이하나 미리 정한 값 이하)일 경우에는, 타당하다고 판단하고, 저하값이 그 이상인 경우에는 타당하지 않다고 판단해도 된다. 또한, 이와 같이, 갱신 전 평가값에 대한 갱신 후 평가값의 저하된 크기가 소정의 범위 내인 경우나, 약간의 저하일 경우에, 일괄 처리를 실시할 것인지 여부의 선택 요구를 출력하고, 그에 대한 오퍼레이터로부터의 입력에 응답하여 일괄 처리의 실시 유무가 결정되게 해도 된다.

[0178] 또한, 상기 예에서는, 1개의 신규 요소 데이터가 추가될 때마다 갱신 처리를 행하도록 하고 있지만, 추가되는 신규 요소 데이터의 개수가 2 이상의 일정수가 되는 것에 응답하여 갱신 처리를 행하는 구성으로 해도 된다. 또한, 신규 요소 데이터가 추가된 경우에, 소정의 조건을 충족하는지 여부로 갱신 처리를 행하거나, 갱신 처리를 행하지 않고 당초부터 일괄 처리를 행하도록 구성해도 된다. 예를 들어, 신규 요소 데이터가 미리 설정한 개수 이상인 경우나, 클러스터링 완료의 요소 데이터에 대한 추가되는 신규 요소 데이터의 개수의 비율이 일정값 이상인 경우에는 갱신 처리를 행하지 않고 당초부터 일괄 처리를 행하게 해도 된다.

[0179] 상기 예에서는, 일괄 처리에서의 분류 계산에 있어서의 각 클러스터에 대한 처리의 순번은 임의이다. 또한, 분류 계산에서 클러스터 무게 중심이 얻어질 때마다, 그 클러스터 무게 중심을 무게 중심 메모리에 기입하고 있지만, 클러스터 무게 중심이 수렴하고 나서, 클러스터 무게 중심을 무게 중심 메모리에 기입해도 된다. 이 경우, 무게 중심 산출 회로에서 산출되는 클러스터 무게 중심을 감시함으로써 수렴을 판단하면 된다. 제2 지표값에 대해서도 마찬가지이며, 클러스터 무게 중심이 수렴하고 나서, 제2 지표값의 계산을 행하여, 제2 지표값의 SBS 레지스터 유닛에 기입해도 된다. 데이터수의 데이터수 레지스터에의 기입에 대해서도 마찬가지이다.

[0180] 상기, 클러스터링 장치(10)를 사용한 클러스터링의 검증을 하기의 (1) 내지 (3)의 3 케이스에 대하여 행하였다. 즉, (1) 신규 요소 데이터 Xnew가 기존의 2개의 클러스터 중 한쪽의 내부에 배치되는 경우(도 16), (2) 신규 요소 데이터 Xnew가 기존의 2개의 클러스터의 먼 곳에 배치되고, 기존의 각 클러스터와의 거리에 차가 있는 경우(도 17), 및 (3) 신규 요소 데이터 Xnew가 기존의 2개의 클러스터 중 한쪽의 부근에 배치되는 경우(도 18)이다. 어느 케이스에 있어서든, 기존의 각 클러스터는 각각 50개의 요소 데이터 X가 일괄 처리에 의해 분류되고, 10개의 신규 요소 데이터 Xnew를 추가하여 통합하여 갱신 처리를 행하고 있다. 평가값 E(Nc)는 상술한 식 (1)에 의해 산출하였다.

[0181] (1) 신규 요소 데이터 Xnew가 기존의 2개의 클러스터 중 한쪽의 내부에 배치되는 경우에는, 갱신 처리에 의해 각 신규 요소 데이터 Xnew는, 그들이 내부에 배치된 한쪽의 클러스터로 분류되었다. 신규 요소 데이터 Xnew의 추가 전에 있어서의 전체 요소 데이터 X를 사용한 클러스터링의 각 클러스터수 Nc에 있어서의 평가값 E(Nc)를 표 1의 일괄 처리(추가 전)의 란에 나타내었다. 또한, 신규 요소 데이터 Xnew를 추가하여 갱신 처리를 행한 상태에서의 평가값 E(2)를 표 1의 갱신 처리의 란에 나타내었다. 또한, 참고로서, 신규 요소 데이터 Xnew가 추가된 전체 요소 데이터 X를 사용한 각 클러스터수 Nc에 있어서의 평가값 E(Nc)를 표 1의 일괄 처리(추가 후)의 란에 나타내었다.

표 1

클러스터수 Nc	평가값 E(Nc)		
	일괄 처리 (추가 전)	갱신 처리 후	일괄 처리 (추가 후)
2	217.5	233.0	233.0
3	131.7	-	148.0
4	95.1		101.9
5	85.9		87.4

[0182]

[0183] (2) 신규 요소 데이터 Xnew가 기존의 2개의 클러스터의 먼 곳에 배치되고, 기존의 각 클러스터와의 거리에 차가 있는 경우에는, 갱신 처리에 의해 각 신규 요소 데이터 Xnew는, 기존의 2개의 클러스터 중 거리가 가까운 한쪽

의 클러스터로 분류되었다. 신규 요소 데이터 X_{new} 의 추가 전에 있어서의 전체 요소 데이터 X 를 사용한 클러스터링의 각 클러스터수 N_c 에 있어서의 평가값 $E(N_c)$ 를 표 2의 일괄 처리(추가 전)의 란에, 또한 신규 요소 데이터 X_{new} 를 추가하여 갱신 처리를 행한 상태에서의 평가값 $E(2)$ 를 표 2의 갱신 처리의 란에 나타내었다. 이 경우, 갱신 처리에 의해 평가값 $E(2)$ 가 크게 저하되었기 때문에 타당성 판단 후에 일괄 처리가 행하여졌다. 이때에 각 클러스터수 N_c 에 있어서의 평가값 $E(N_c)$ 를 표 2의 일괄 처리(추가 후)의 란에 나타내었다.

표 2

클러스터수 N_c	평가값 $E(N_c)$		
	일괄 처리 (추가 전)	갱신 처리 후	일괄 처리 (추가 후)
2	485.3	114.1	114.1
3	309.0	-	537.2
4	227.3		397.7
5	185.2		292.8

[0184]

[0185]

(3) 신규 요소 데이터 X_{new} 가 기존의 2개의 클러스터 중 한쪽의 부근에 배치되는 경우에는, 갱신 처리에 의해 각 신규 요소 데이터 X_{new} 는, 그들이 부근에 배치된 한쪽의 클러스터로 분류되었다. 신규 요소 데이터 X_{new} 의 추가 전에 있어서의 전체 요소 데이터 X 를 사용한 클러스터링의 각 클러스터수 N_c 에 있어서의 평가값 $E(N_c)$ 를 표 3의 일괄 처리(추가 전)의 란에, 또한 신규 요소 데이터 X_{new} 를 추가하여 갱신 처리를 행한 상태에서의 평가값 $E(2)$ 를 표 3의 갱신 처리의 란에 나타내었다. 이 경우, 갱신 처리에 의한 평가값 $E(2)$ 의 저하가 10% 정도였다. 참고로서, 갱신 처리 후에, 일괄 처리를 행했을 때의 각 클러스터수 N_c 에 있어서의 평가값 $E(N_c)$ 를 표 3의 일괄 처리(추가 후)의 란에 나타내었다.

표 3

클러스터수 N_c	평가값 $E(N_c)$		
	일괄 처리 (추가 전)	갱신 처리 후	일괄 처리 (추가 후)
2	485.3	435.9	435.9
3	309.0	-	266.4
4	227.3		218.2
5	185.2		150.8

[0186]

[0187]

평가값의 산출식은, 상기 것에 한정되지 않는다. 상기와 같이 구해지는 내적 결합도는, 그 값이 작을수록, 클러스터 내의 데이터 상호의 유사성이 높아지도록 각 클러스터 내의 데이터가 집합된 클러스터링 상태로 되는 것을 의미한다. 한편, 상기와 같이 구해지는 외적 분리도는, 그 값이 클수록, 클러스터 상호 간의 유사성이 낮아지도록 각 클러스터가 서로 떨어진 클러스터링 상태로 되는 것을 의미한다. 이 때문에, 내적 결합도와 외적 분리도로부터 평가값을 구하는 연산은, 내적 결합도가 작아지는 방향으로 변화했을 때, 또한 외적 분리도가 커지는 방향으로 변화했을 때에, 평가값이 증가 또는 감소의 한쪽으로 변화하도록 하면 되고, 그러한 결과가 얻어지는 내적 결합도와 외적 분리도를 변수로 하는 연산식을 사용하면 된다.

[0188]

예를 들어, 식 (1)에 나타내어지는 클러스터 지표값 SBS(제2 총합값) 대신에, 식 (10)에 나타나는 바와 같이, 1의 클러스터 C_i 에 관한 다른 클러스터 C_j 와의 사이의 클러스터 거리($d(V_i, V_j)$) 중 최솟값의 각 클러스터에 대한 총합인 클러스터 지표값 SBS(제2 총합값)를 사용해도 된다. 이 경우, 1의 클러스터 C_i 에 관한 다른 클러스터 C_j 와의 사이의 클러스터 간 거리($d(V_i, V_j)$)의 최솟값이 제2 지표값 SBS_i 이다.

$$SBS = \sum_{i=1}^{Nc} \min_{i \neq j} d(V_i, V_j) \quad \dots (10)$$

[0189]

또한, 데이터 무게 중심 간 거리(클러스터 내 거리), 무게 중심 거리, 클러스터 간 거리로서, 맨해튼 거리 이외의, 예를 들어 유클리드 거리, 민코프스키 거리, Point symmetry distance 등을 사용해도 되고, 각종 거리의 제곱한 값을 사용해도 된다.

[0191]

이 예에서는, 제1 값으로서의 클러스터 내의 데이터수 n_i 로 제1 지표값 SD_i 를 계산함으로써 규격화한 결합 지표값 SWD_i 를 구하고 있지만, 제1 값은, 이것에 한정되는 것은 아니며, 클러스터 C_i 의 데이터수 n_i 에 기초하는 값으로 할 수 있다. 예를 들어, 클러스터 C_i 의 데이터수 n_i 의 거듭제곱(n_i^2 , n_i^3 , $n_i^{1/2}$, n_i^{-2} 등), 데이터수 n_i 를 상수배로 한 값, 데이터수 n_i 에 일정수를 감산 또는 가산한 값, 데이터수 n_i 와 클러스터수 Nc 를 미리 정한 함수 $F(n_i, Nc)$, 예를 들어 「 $F(n_i, Nc)=n_i-Nc$ 」, k 를 상수로 한 「 $F(n_i, Nc)=k \cdot n_i-Nc$ 」 등에 적용한 결과 등을 제1 값으로서 사용할 수 있다.

[0192]

또한, 제2 값으로서의 클러스터수 Nc 로 클러스터 지표값 SBS 를 계산함으로써 규격화한 외적 분리도(SBS/Nc)를 구하고 있지만, 제2 값은, 이것에 한정되는 것은 아니며, 클러스터수 Nc 에 기초하는 값으로 규격화할 수 있다. 클러스터수 Nc 이외의 클러스터수 Nc 에 기초하는 값으로서, 제1 값과 마찬가지로, 예를 들어, 클러스터수 Nc 의 거듭제곱, 클러스터수 Nc 를 상수배로 한 값, 클러스터수 Nc 에 일정수를 감산 또는 가산한 값 등을 제2 값으로서 사용할 수 있다.

[0193]

또한, 외적 분리도로서, 예를 들어 값 「 SBS/Nc 」를 전체 데이터수 Nn 으로 계산한 값(= $SBS/(Nc \times Nn)$)을 사용한 평가값(이하, 이러한 평가값 $E(Nn, Nc)$ 라고 칭한다)도 있지만, 이러한 평가값 $E(Nn, Nc)$ 는 요소 데이터의 증가가 없는 경우에, 각 클러스터수 Nc 의 상호 간에 있어서의 클러스터링의 상태를 평가함에 있어서 유용하다. 그러나, 평가값 $E(Nn, Nc)$ 는 전체 데이터수 Nn 이 증가한 경우에, 클러스터링의 상태가 양호해져도, 요소 데이터의 증가에 따라 평가값이 전체 데이터수 Nn 의 증가 전과 비교하여 저하된다. 이 때문에, 요소 데이터의 추가가 발생하도록 요소 데이터의 전체 데이터수 Nn 이 동적일 경우, 또한 상기 갱신 처리와 같이, 요소 데이터의 증가 전 후에 있어서 평가값을 비교하는 경우에는, 식 (1)에 나타내어지는 평가값 $E(Nc)$ 나 식 (10)에 나타내어지는 클러스터 지표값 SBS 를 사용한 평가값 $E(Nc)$ 와 같이 전체 데이터수 Nn 을 포함하지 않는 것이 바람직하다.

[0194]

클러스터링 상태(분류 결과)를 상기 평가값 $E(Nc)$ 를 사용하여 평가하는 방법, 클러스터링 연산부가 클러스터링을 행하기 위한 계산 과정의 계산값을 이용하여 평가값 $E(Nc)$ 를 산출하는 방법은, 신규 요소 데이터가 추가된 경우로 갱신 처리를 행하는 경우에 한정되지 않고, 예를 들어 신규 요소 데이터가 추가된 경우에 일괄 처리를 행하는 구성에도 적용할 수 있다.

[0195]

또한, 상기 예에서는, 클러스터 내 거리의 기점인 제1 대표점으로서, 클러스터 무게 중심을 사용하여, 클러스터 간의 거리의 지표가 되는 무게 중심 거리 및 클러스터 간 거리의 각 클러스터의 기점인 제2 대표점을 클러스터 무게 중심으로 하고 있지만, 제1 대표점, 제2 대표점은 이들에 한정되지 않는다. 예를 들어, 제1 대표점, 제2 대표점을 각 클러스터 내에 있어서, 클러스터 무게 중심에 가장 가까운 요소 데이터로 해도 된다. 또한, 제2 대표점은, 클러스터끼리 또는 후술하는 기준점과 클러스터의 이격된 상태(거리)를 측정함에 있어서의 기점이기 때문에, 인접한 클러스터 혹은 기준점에 가장 가깝거나 또는 가장 떨어진 클러스터 내의 요소 데이터 등으로 해도 된다. 또한, 제1 대표점 및 제2 대표점을, 별도 정한 기준으로 의해 결정되는 클러스터 내의 점 또는 데이터나 클러스터 내의 임의의 점 또는 요소 데이터로 해도 된다.

[0196]

또한, 상기 예에서는, 데이터 무게 중심을 기준점으로 하고 있지만, 기준점은, 임의의 점이나 요소 데이터로 설정할 수 있고, 상기와 같은 데이터 무게 중심 이외에, 데이터 무게 중심에 가장 가까운 요소 데이터, 어느 클러스터의 제2 대표점, 외부로부터 지정하는 점이나 요소 데이터, 외부로부터 지정된 메인 메모리의 어드레스에 저장되어 있는 요소 데이터, 처음에 입력된 요소 데이터 등과 같이 별도 정한 기준에 의해 결정되는 점이나 요소 데이터로 해도 된다.

[0197]

평가값 $E(Nc)$ 는 식 (1)의 역수로 해도 된다. 이 경우에는, 평가값 $E(Nc)$ 는 최소(극소)로 되는 것이 최적의 클러스터수 Nc 가 된다. 또한, 평가값을 구하는 연산으로서, 내적 결합도와 외적 분리도의 비로 하는 것 이외에, 예를 들어, 다음 각 식에 나타내는 바와 같이, 내적 결합도와 외적 분리도의 한쪽의 역수와 다른 쪽을 각각 가중치 부여하여 가산함으로써 평가값 $E(Nc)$ 를 구하는 것, 내적 결합도와 외적 분리도의 한쪽으로부터 다

른 쪽을 각각 가중치 부여하여 감산함으로써 평가값 E(Nc)를 구하는 것 등을 사용할 수 있다. 또한, 식 중의 값 Wa, Wb는 가중치 부여의 상수이며, Wa, Wb ≠ 0이다.

$$E(Nc) = Wb \cdot \left(\frac{SBS}{Nc} \right) + Wa \cdot \left(\frac{1}{SWD} \right)$$

$$E(Nc) = Wb \cdot \left(\frac{Nc}{SBS} \right) + Wa \cdot SWD$$

$$E(Nc) = Wb \cdot \left(\frac{SBS}{Nc} \right) - Wa \cdot SWD$$

$$E(Nc) = Wa \cdot SWD - Wb \cdot \left(\frac{SBS}{Nc} \right)$$

[0198]

[0199]

상기에서는 k-means법도 사용하여 클러스터링을 행하고 있지만, k-means법에 한하지 않고, 각 요소 데이터를 1개의 클러스터에 속하도록 분류하는 하드 클러스터링에 적용할 수 있다. 이러한 하드 클러스터링으로서는, 「k-means++」법, 스펙트럴 클러스터링, 단연결법, 워드법 등이 있다.

부호의 설명

[0200]

10: 클러스터링 장치

11: 시스템 컨트롤러

14: 메인 메모리

15: 무게 중심 메모리

16: 클러스터링 연산부

17: 근린 탐색 회로부

18: 평가값 산출 회로

18a: 로직부

18b: 평가 레지스터부

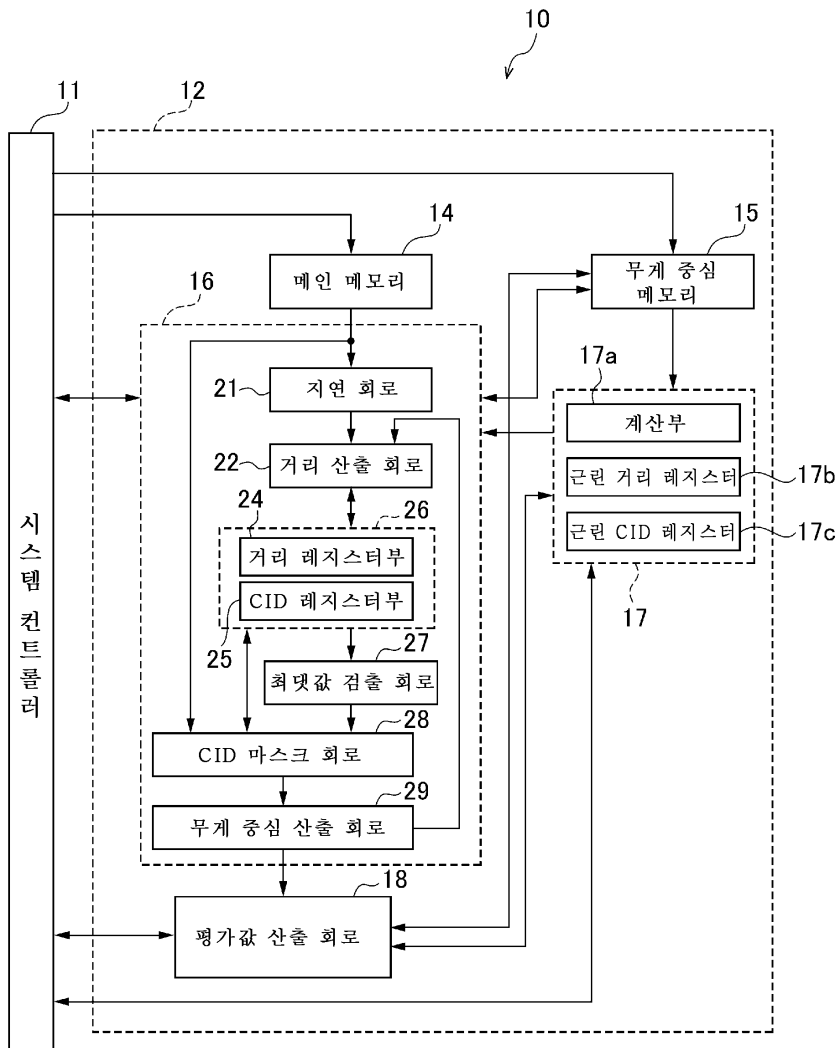
25: CID 레지스터부

29: 무게 중심 산출 회로

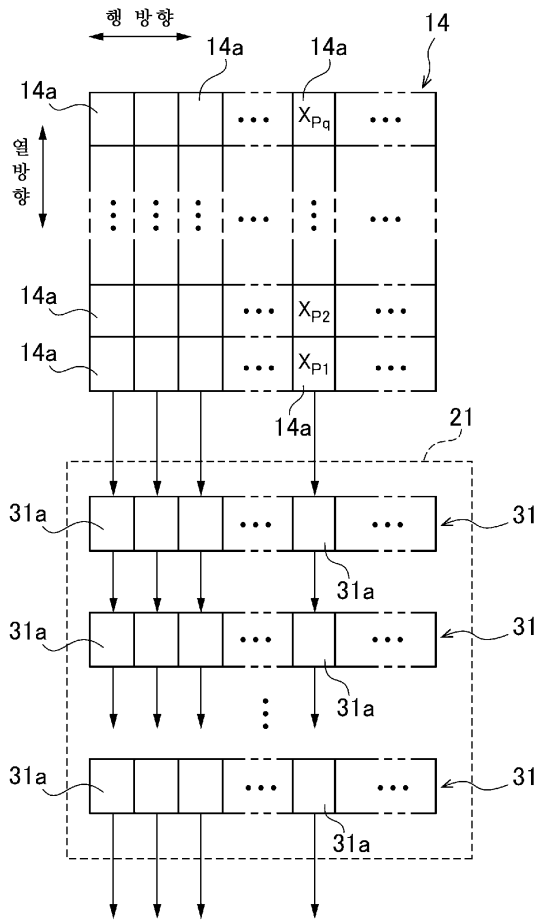
PD1 내지 PD6: 파워드 메인

도면

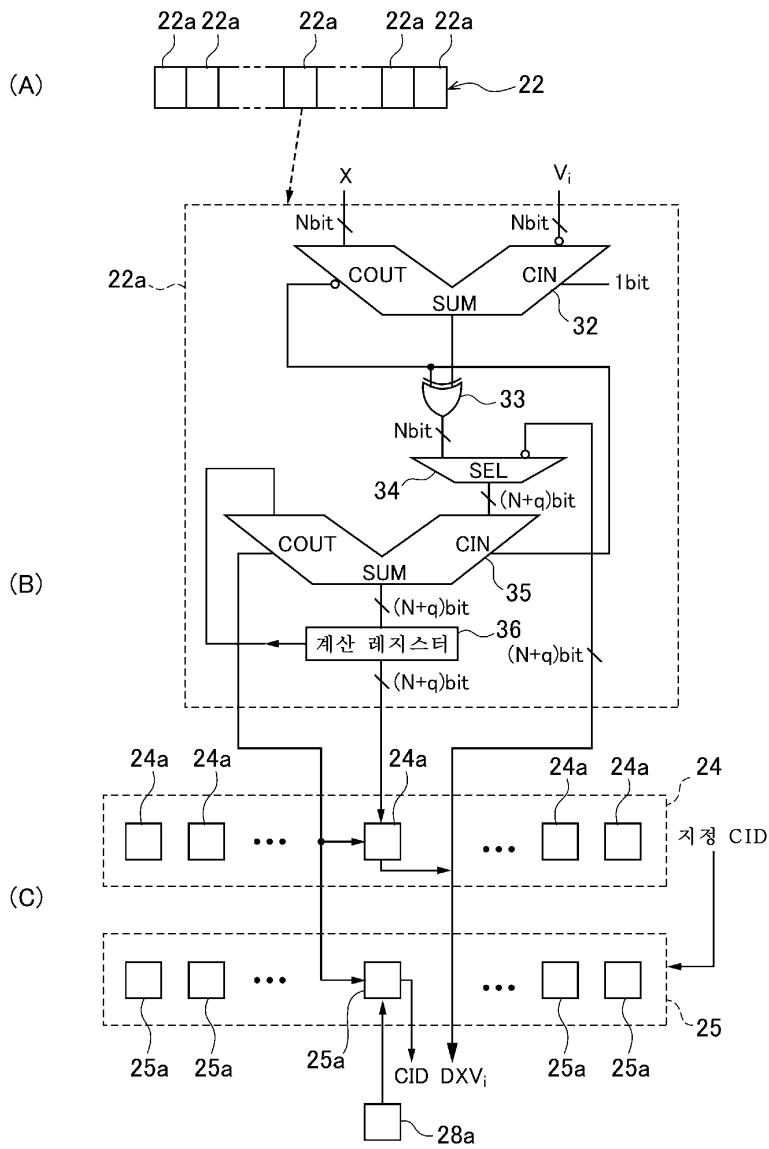
도면1



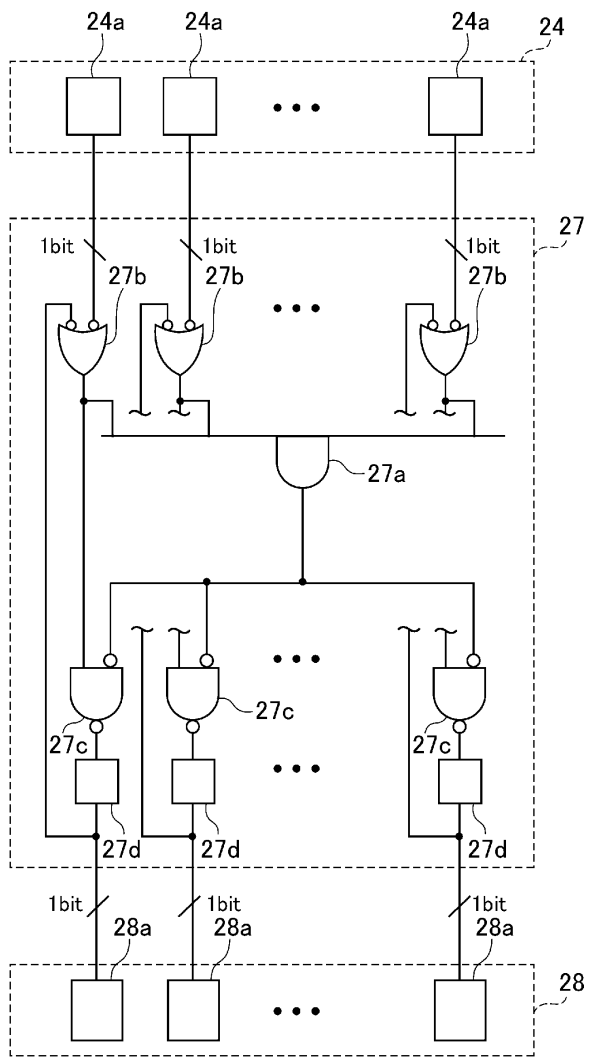
도면2



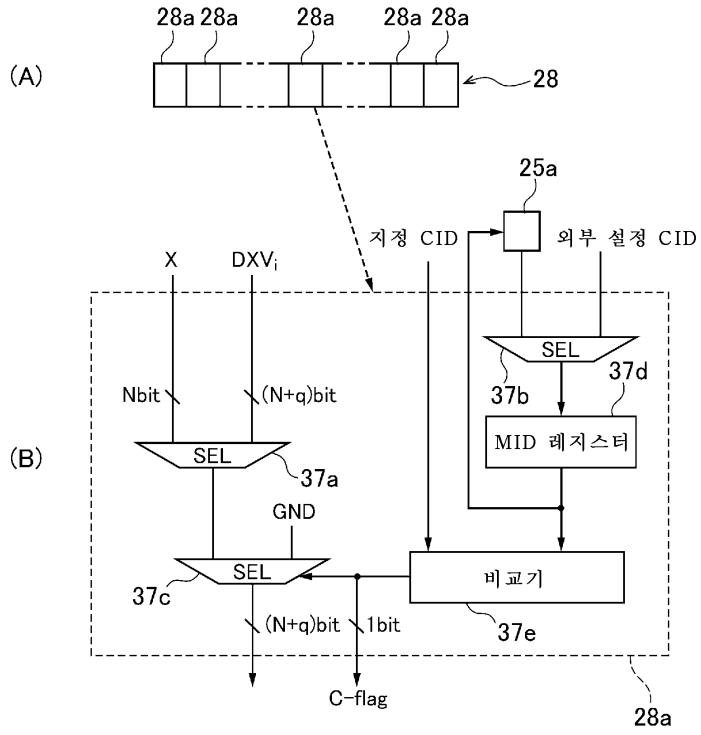
도면3



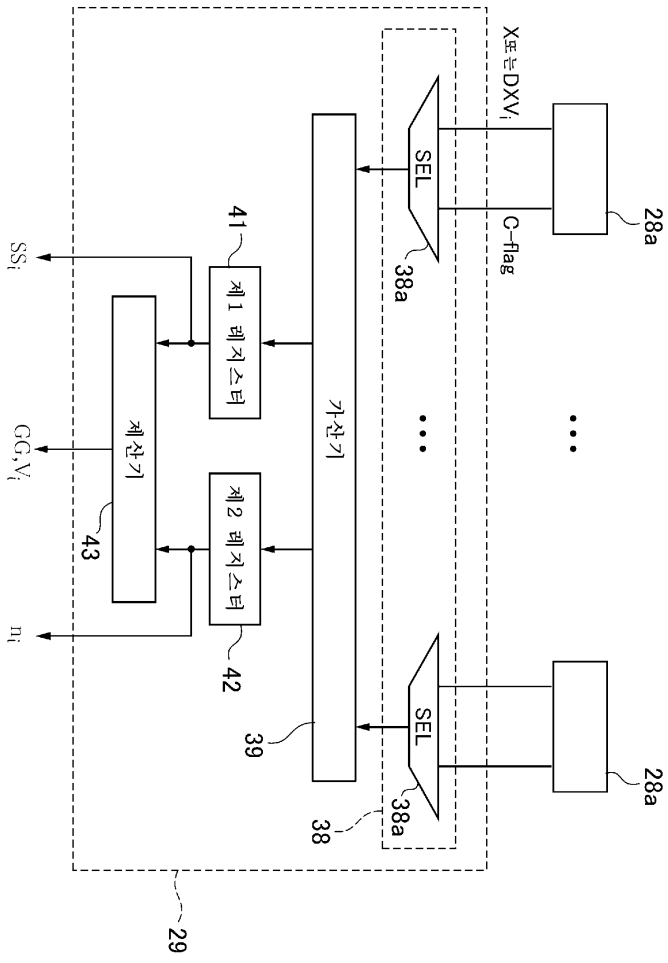
도면4



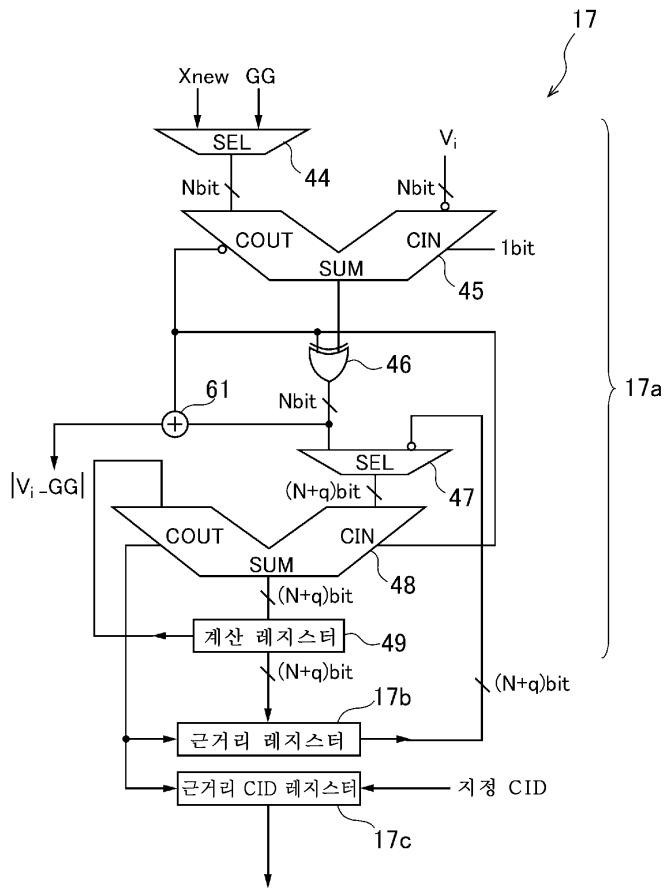
도면5



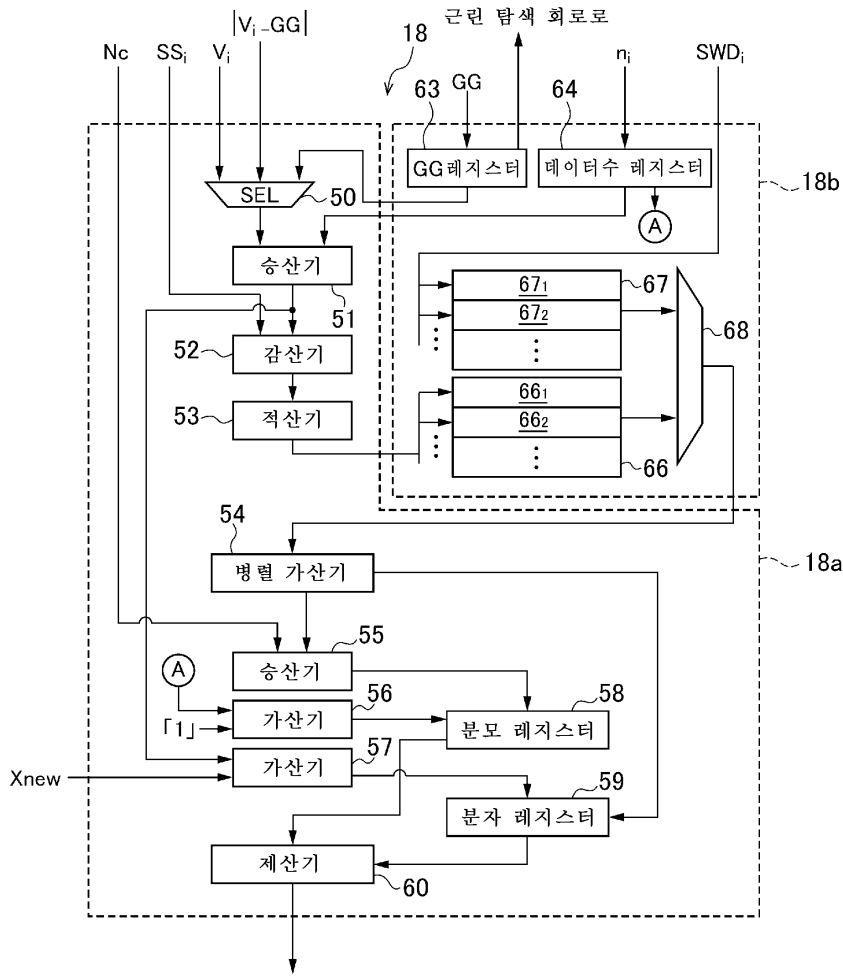
도면6



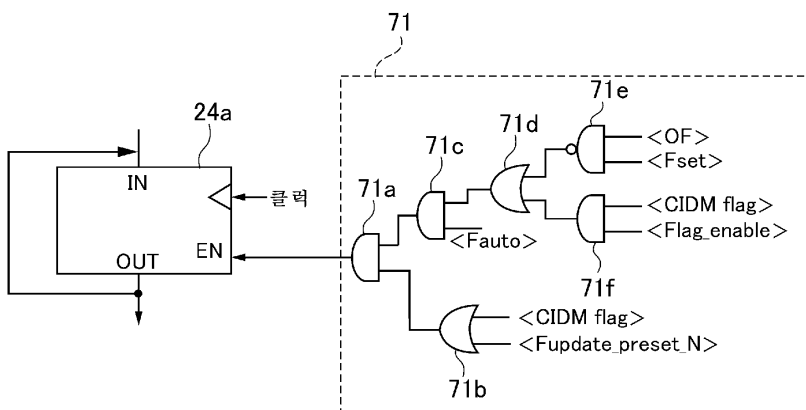
도면7



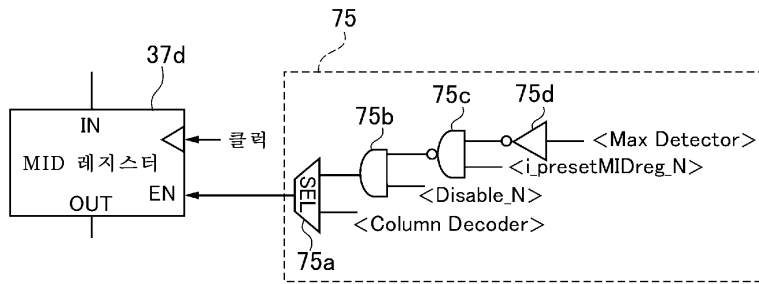
도면8



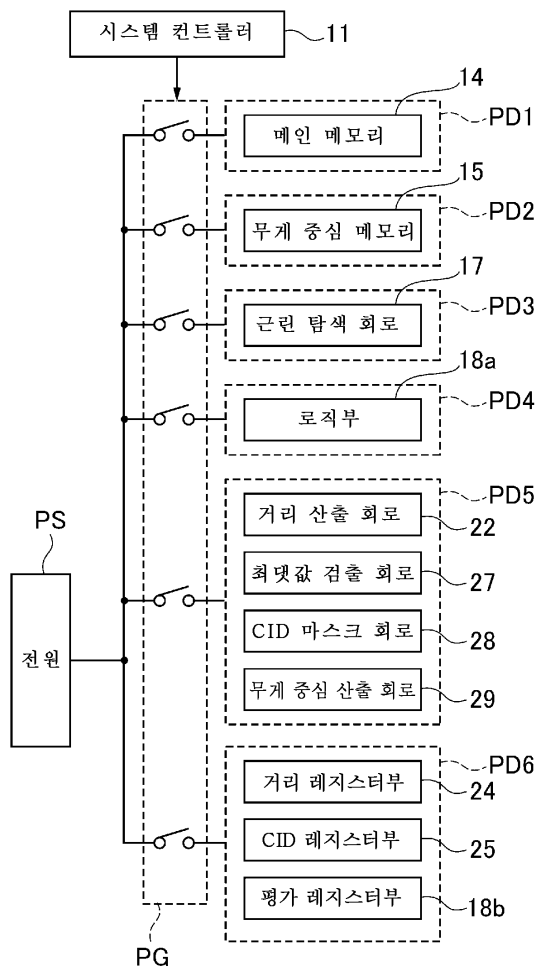
도면9



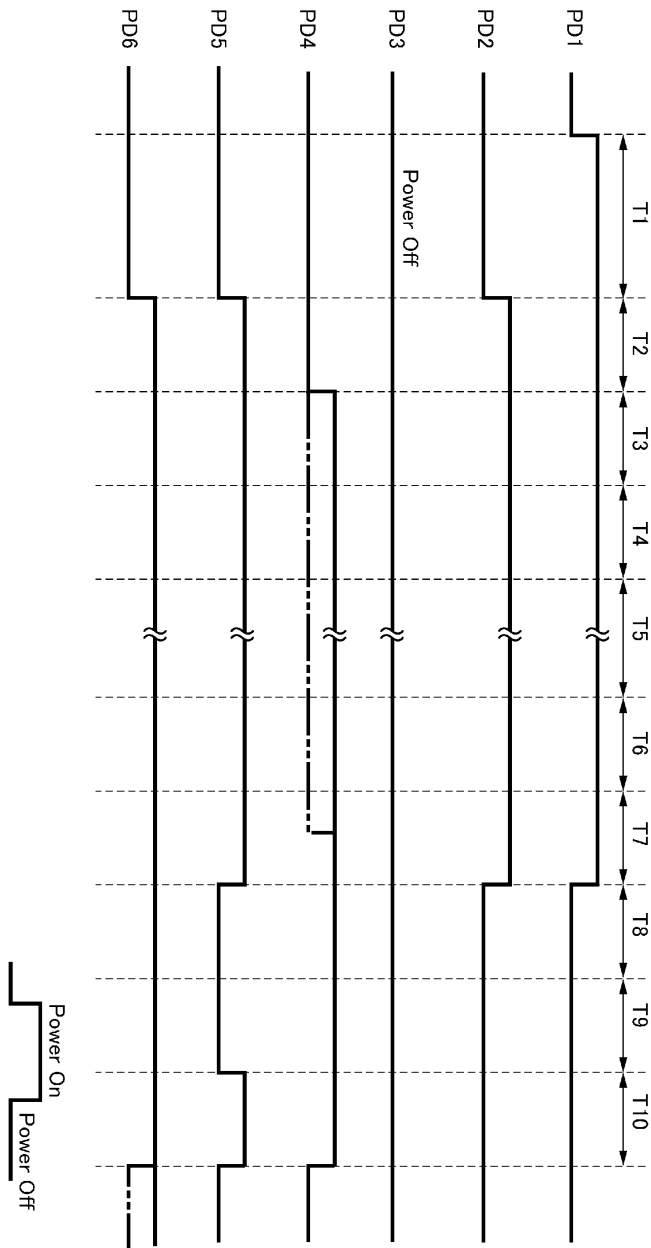
도면10



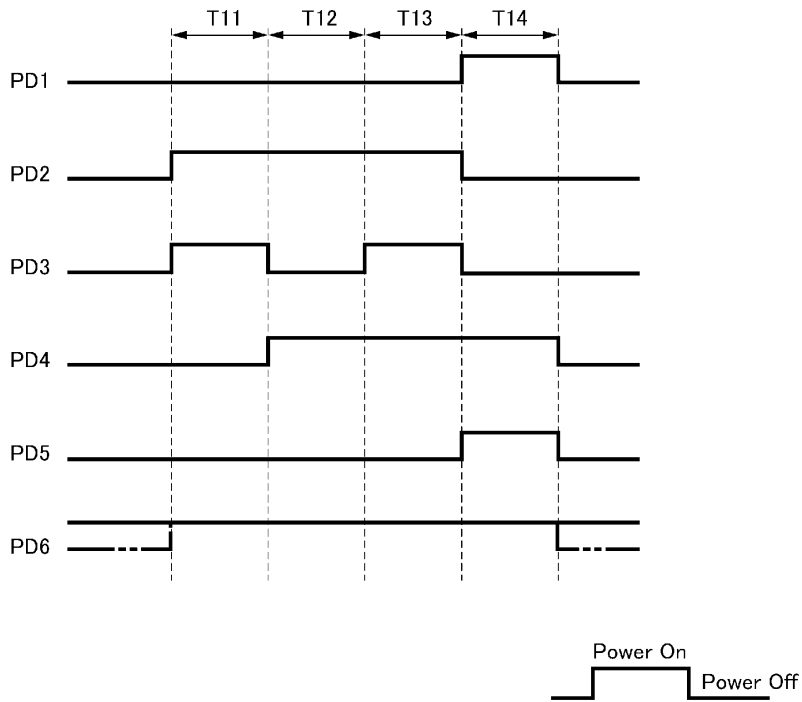
도면11



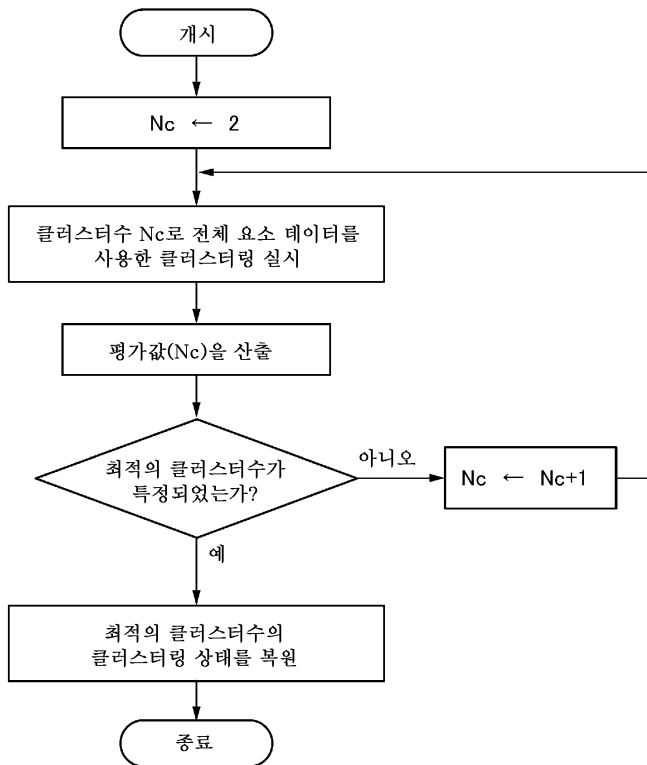
도면12



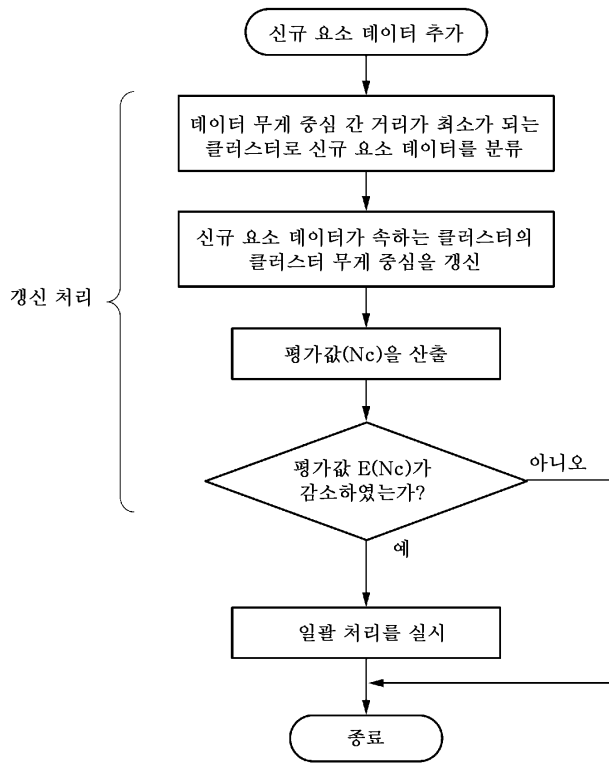
도면13



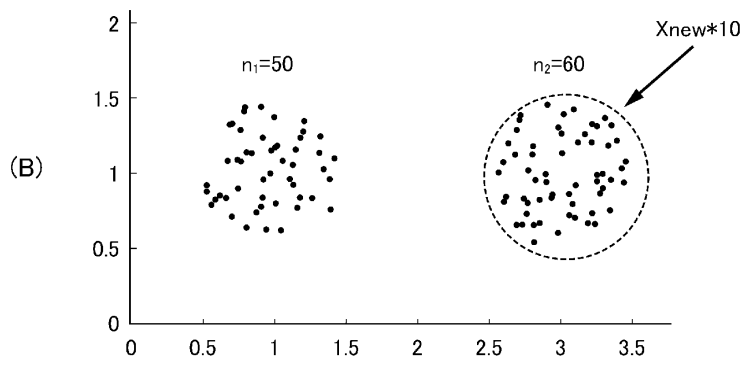
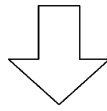
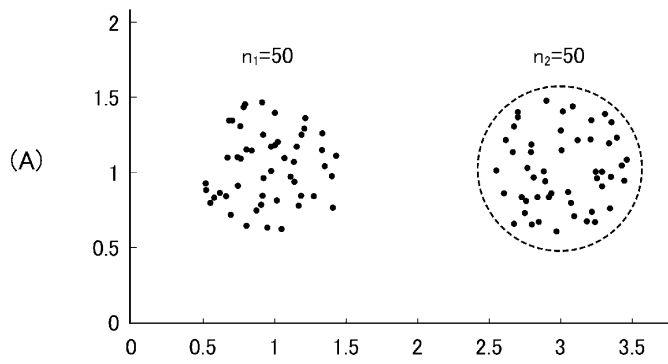
도면14



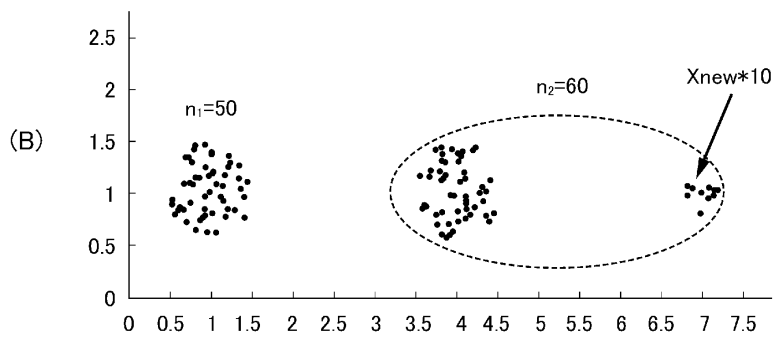
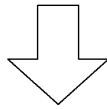
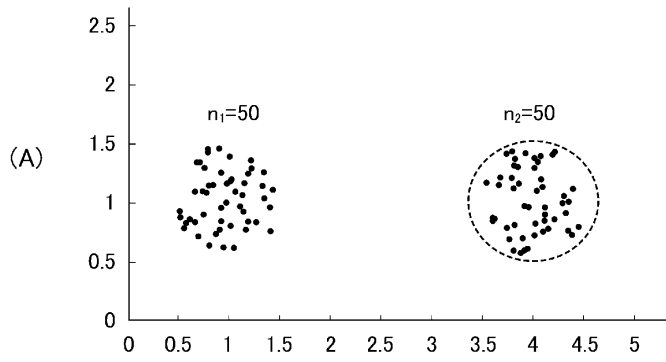
도면15



도면16



도면17



도면18

