김일성종합대학학보 주체109(2020)년
수 학 제66권 제4호

γ－피복을 리용한 모든 최대α－간격반복들의 한계에 관한 한가지 계산방법

강진웅, 조영선

위대한 령도자 **김정일**동지께서는 다음과 같이 교시하시였다.

《새로운 과학기술분야를 개척하기 위한 사업도 전망성있게 밀고나가야 합니다.》
(《**김정일**선집》 증보판 제11권 138페지)

모든 최대α－간격반복들의 한계를 구하는것은 최근 생명정보학을 비롯한 여러 분야에서 중요한 문제의 하나로 제기되고있다.

선행연구[1, 2]에서는 최대α－간격반복들의 모임을 조건에 따라 몇가지 부분모임으로 가르고 매 부분모임의 한계를 구하는 방법으로 길이가 n인 문자렬에 대한 최대α－간격반복들의 개수가 추상적으로 $O(\alpha^2 n)$, $O(\alpha n)$이라는 사실을 밝혔다. 선행연구[3]에서는 최대α－간격반복들의 한가지 부분모임인 γ－피복을 리용한 최대α－간격반복들의 한계가 구체적으로 $3n/\gamma$이라는것을 증명하였다.

론문에서는 γ－피복값을 새롭게 정의하고 그에 기초하여 γ－피복을 리용한 모든 최대α－간격반복들의 한계를 구하였다.

정의 1[1] 임의의 실수 $\alpha \geq 1$과 반복문자렬 $(u_\lambda,\ u_\rho)$에 대하여 그 주기가 반복되는 부분문자렬길이의 α배를 넘지 못하며 반복되는 부분문자렬들의 왼쪽, 오른쪽문자들이 각각 서로 다를 때 $(u_\lambda,\ u_\rho)$를 최대α－간격반복이라고 부른다.

정의로부터 $b(s)$, $e(s)$가 문자렬 s의 시작위치, 마지막위치를 나타내고 $w[i]$가 w의 i번째 문자를 나타낸다면 $q = b(u_\rho) - b(u_\lambda)$라고 할 때 $q \leq \alpha |u_\lambda| = \alpha |u_\rho|$이고

$$w[b(u_\lambda)-1] \neq w[b(u_\rho)-1],\ w[e(u_\lambda)+1] \neq w[e(u_\rho)+1]$$

이다.(그림 1)

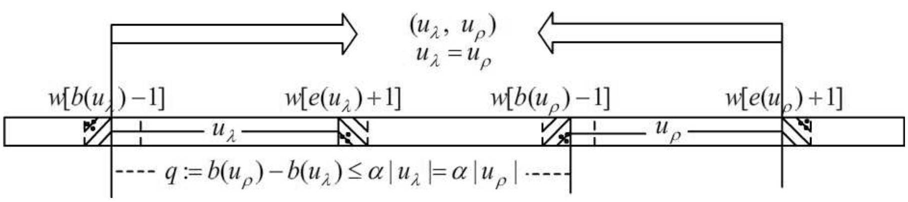

그림 1. 최대α－간격반복렬 $(u_\lambda,\ u_\rho)$

최대α－간격반복 $(u_\lambda,\ u_\rho)$를 $\mathbf{Z}^2$의 점 $(e(u_\lambda),\ q)$로 보내는 넘기기를 φ로 표시하자.[1] $C_n := \{(x,\ y) \in Z^2 \mid 1 \leq y \leq n-1,\ 1 \leq x \leq n-y\}$이라고 하자. 그러면 $e(u_\lambda) + q = e(u_\rho) < n$이므로 임의의 $(e(u_\lambda),\ q) \in C_n$이다. 즉 G_α를 최대α－간격반복전체의 모임이라고 할 때 $\varphi(G_\alpha) \subseteq C_n$이다.

정의 2[3] 2차원자리표계의 점 $(x,\ y),\ (x',\ y')$와 임의의 $\gamma \in (0,\ 1]$에 대하여

$$\begin{cases} x-\gamma y \le x' \le x \\ y-\gamma y \le y' \le y \end{cases}$$

가 성립할 때 $(x,\ y)$는 $(x',\ y')$를 γ−피복한다고 말한다.(그림 2) 그리고 γy를 분포길이라고 부른다.

정의 2에서 알수 있는바와 같이 점 $(x,\ y)$가 γ−피복하는 점들의 y자리표 $y' \ge 0$을 만족시키며 고정된 γ에 대하여 점들의 y자리표값에 따라 분포길이가 달라지면서 γ−피복하는 점의 개수가 차이난다.

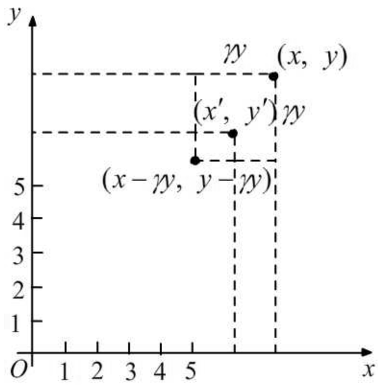

그림 2. $(x',\ y')$를 γ−피복하는 $(x,\ y)$

먼저 C_n에서 γ−피복과 관련한 점모임농도의 한계를 구하기 위하여 필요한 몇가지 보조정리들을 정식화한다.

보조정리 1 임의의 $I=[\psi-1/\gamma,\ \psi)$, $\gamma,\ \psi \in \mathbf{R}$, $\gamma \in (0,\ 1]$에 대하여

$$|I \cap \mathrm{Z}| = \begin{cases} \lfloor 1/\gamma \rfloor + 1, & 0 < \psi - \lfloor \psi \rfloor \le \delta \\ \lfloor 1/\gamma \rfloor & ,\ 기타 \end{cases} \tag{1}$$

이다. 여기서 $I \cap \mathbf{Z} = \{i \in \mathbf{Z} \mid i \in I\}$, $\delta := 1/\gamma - \lfloor 1/\gamma \rfloor$이다.

보조정리 2 $g_\gamma(i) := |\{y \in \mathbf{N} \mid (i-1)/\gamma \le y < i/\gamma\}|$, $1 \le i \le \lceil n\gamma \rceil$인 $g_\gamma : \mathbf{N} \to \mathbf{N}$과 임의의 비증가정값함수 $f : \mathbf{N} \to \mathbf{R}^+$에 대하여

$$\sum_{i=1}^{\lceil n\gamma \rceil} (f(i) g_\gamma(i)) \le \sum_{i=1}^{\lceil n\gamma \rceil} f(i)/\gamma$$

이다. 여기서 $\gamma \in (0,\ 1]$이다.

정의 2에 기초하여 모든 최대 α−간격반복들의 한계계산에서 중요하게 쓰이게 될 γ−피복값을 정의한다.

정의 3 점모임 C와 $\mathbf{Z}^2$의 점 p에 대하여 p를 γ−피복하는 C의 점 $(x',\ y')$가 $(i-1)/\gamma \le y' < i/\gamma\ (i \in \mathbf{N})$를 만족시킬 때 p는 γ−피복값 $1/i^2$을 가진다고 말한다. 만일 p가 C의 어느 한 점에 의해서도 γ−피복되지 않는다면 γ−피복값을 0으로 정의한다.

그리고 점 $p(x,\ y)$의 γ−피복값을 $w(p)$ 또는 $w(x,\ y)$로 표시한다.

y자리표값이 부수인 점은 그 어떤 점에 의해서도 γ−피복되지 않으므로 γ−피복값이 항상 0이다.

이때 점 p가 C의 여러 점들에 의하여 γ−피복된다면 p는 γ−피복값을 여러개 가지게 되는데 하나만 가지게 하자면 C에 일정한 조건을 주어야 한다.

지금 γ−피복과 관련한 다음의 정리를 증명한다.

정리 1 임의의 실수 $\gamma \in (0,\ 1]$과 $C \subseteq C_n$에 대하여 C의 임의의 두 점이 같은 점을 γ−피복하지 않는다면 $|C| < n\pi^2/(6\gamma)$이다. 특히 $\gamma = 1$이면 $|C| < n(\pi^2/6 - 3/4)$이다.

증명 정의 1로부터 γ−피복되는 점의 y자리표가 0이 될수는 있지만 부의 값은 가질수 없으며 정리의 조건으로부터 γ−피복하는 점의 y자리표는 1부터 가능하다. 또한 C의 임의의 두 점이 같은 점을 γ−피복하지 않으므로 $\mathbf{Z}^2$의 모든 점의 γ−피복값은 유일

하게 된다.

정의 2로부터 $(i-1)/\gamma \le y' < i/\gamma$ 이고 $i-1 \le \gamma y' < i$ 이며

$$\begin{cases} x'-i < x'-\gamma y' \le x'-(i-1) \\ y'-i < y'-\gamma y' \le y'-(i-1) \end{cases}$$

이다. 그러면

$$\begin{cases} x'-i < x'-\gamma y' \le x'-(i-1) \le x \le x' \\ y'-i < y'-\gamma y' \le y'-(i-1) \le y \le y' \end{cases}$$

를 얻으며 결국 C 에 속하는 점 $(x',\ y')$ 에 의해 $\gamma-$피복되는 점은 모두 i^2 개이고 그것들의 $\gamma-$피복값들의 합은 1로 된다.(그림 3) 즉 C 의 점 하나에 그에 의해 $\gamma-$피복되는 점들의 $\gamma-$피복값합 1이 대응된셈이다. 결국 $|C|$ 의 한계를 $\mathbf{Z}^2$ 에서 모든 점들의 $\gamma-$피복값들의 총합으로 보아도 무방하므로 이제부터 $\mathbf{Z}^2$ 의 모든 점의 $\gamma-$피복값합의 한계를 구하겠다.

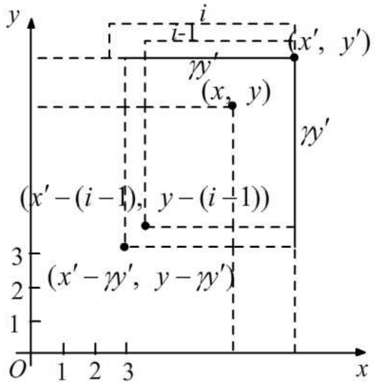

그림 3. $(x',\ y')$ 에 의해 $\gamma-$피복되는 i^2 개의 점

우선 $(i-1)/\gamma \le y < i/\gamma$ 인 y 를 고정하면 $(\cdot,\ y)$ 인 점들의 $\gamma-$피복값합은 기껏 n/i^2 이다.

$(\cdot,\ y)$ 인 점들을 $\gamma-$ 피복하는 점의 y 자리표는 $[(i-1)/\gamma,\ i/\gamma)$ 에 속할수도 있고 그보다 더 큰 구간들에 들수도 있으므로 $(\cdot,\ y)$ 인 점들의 $\gamma-$피복값은 기껏 $1/i^2$ 로 된다.

C_n 에서 x 의 정의역이 $1 \le x \le n-y$ 이므로 다음과 같이 나눌수 있다.(그림 4)

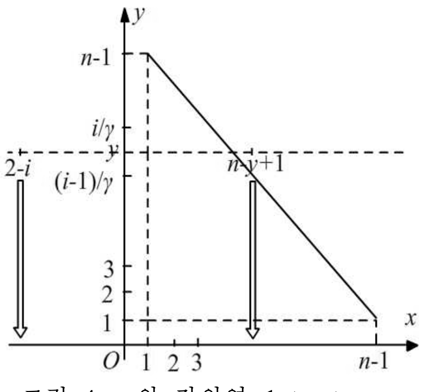

그림 4. x 의 정의역 $1 \le x \le n-y$

$x \ge n-y+1$ 인 점들에 대해서는 C 의 점이 피복할수 없으므로 $\gamma-$피복값이 0이고 $1 \le x \le n-y$ 인 $n-y$ 개의 점들의 $\gamma-$피복값은 기껏 $1/i^2$ 이다.

$x \le 0$ 인 점들에 대하여 보면 $i-1 \le \gamma y < i$ 이므로 $\gamma-$피복될수 있는 가장 작은 x 자리표값은 $(1,\ y)$ 에 의해 피복되는 $2-i$ 이다. 즉 $2-i \le x \le 0$ 인 $i-1$ 개의 점들의 $\gamma-$피복값은 기껏 $1/i^2$ 이고 $x < 2-i$ 인 점들에 대해서는 $\gamma-$피복값이 0으로 된다.

따라서 $(\cdot,\ y)$ 인 점들의 $\gamma-$피복값합은

$$\sum_{x \in Z} w(x,\ y) \le (n-y+i-1)/i^2 \le (n-y+\gamma y)/i^2 \le n/i^2$$

으로 된다.

다음 고정했던 y 를 변화시키면서 i 를 보다 구체적으로 따져 $\mathbf{Z}^2$ 에서의 $\gamma-$피복값합의 한계를 계산해보자.

먼저 $\gamma = 1$ 인 경우를 보자.

$(\cdot,\ 0)$ 인 점들을 1−피복하는 점 $(x',\ y')$ 는 무조건 $y' = 1$ 이므로 $i-1 \le y' < i$ 라는데로부

터 $i=2$로 되며 $(\cdot,\ 0)$인 점들의 $\gamma-$피복값합은 기껏 $n/2^2$이다.

$y\ge 1$인 $(x,\ y)$를 1－피복하는 $(x',\ y')$는 $y'=y$도 가능하므로 $i=y+1$ 즉 $y\ge 1$인 $(\cdot,\ y)$인 점들의 $\gamma-$피복값합은 기껏 $n/(y+1)^2$이다.

그러므로

$$\sum_{(x,\ y)\in\mathbf{Z}^2} w(x,\ y)\le n/2^2+n\sum_{y=1}^{n}(1/(y+1)^2)<n/4+n\sum_{i=2}^{\infty}(1/i^2)=n/4+ \tag{2}$$
$$+n\pi^2/6-n=n\pi^2/6-3n/4=n(\pi^2/6-3/4)$$

이다. 마지막으로 $\gamma<1$인 경우를 보자.

우에서와 마찬가지로 따지면 $1\le i\le\lceil n\gamma\rceil$에 대하여 $(i-1)/\gamma\le y<i/\gamma$인 점 $(x,\ y)$의 $\gamma-$피복값은 기껏 $1/i^2$이므로 $Y_i:=\{y\in N\,|\,(i-1)/\gamma\le y<i/\gamma\}$라고 할 때 전체 $\gamma-$피복값합

$$\sum_{(x,\ y)\in\mathbf{Z}^2} w(x,\ y)\le\sum_{i=1}^{\lceil n\gamma\rceil}n\,|Y_i|/i^2$$

이다.

이제 Y_i가 들어간 오른변을 정돈하자.

$f(i):=n/i^2$이라고 하면 $f(i)$는 비증가정값함수로 되며 보조정리 2로부터

$$\sum_{(x,\ y)\in\mathbf{Z}^2} w(x,\ y)\le\sum_{i=1}^{\lceil n\gamma\rceil}\frac{n\,|Y_i|}{i^2}\le\sum_{i=1}^{\lceil n\gamma\rceil}(n/i^2)/\gamma<\sum_{i=1}^{\infty}\frac{n}{i^2\gamma}=\frac{n}{\gamma}\sum_{i=1}^{\infty}\frac{1}{i^2}$$

을 얻는다. $\sum_{i=1}^{\infty}(1/i^2)=n\pi^2/6$이므로

$$\sum_{(x,\ y)\in\mathbf{Z}^2} w(x,\ y)<n\pi^2/(6\gamma) \tag{3}$$

이 성립한다. 식 (2)와 (3)으로부터 정리의 결과를 얻는다.(증명끝)

보조정리 3 $(x,\ y)\in\varphi(G_\alpha)$이면 $(x+1,\ y)\notin\varphi(G_\alpha)$이다.

보조정리 3을 리용하면 정리 1로부터 $\gamma-$피복을 리용한 모든 최대$\alpha-$간격반복들의 한계를 계산할수 있다.

정리 2 임의의 $\gamma\in(0,\ 1]$과 $C\subseteq\varphi(G_\alpha)$에 대하여 C의 임의의 두 점이 같은 점을 $\gamma-$피복하지 않는다면 $|C|<n(\pi^2/6-1/2)/\gamma$이다.

증명 $\gamma=1$이라면 정리 1로부터 $|C|<n\pi^2/6-3n/4<n\pi^2/6-n/2$이다.

$\gamma<1$인 경우를 보자.

전체 평면에서의 점들의 $\gamma-$피복값합을 계산하기 위해 우선 y축에서의 첫 $1/\gamma$구간 즉 $E:=\{(x,\ y)\in Z^2\,|\,1\le x\le n,\ 0\le y<1/\gamma\}$의 점들의 $\gamma-$피복값합을 따로 따져보자.

$y<1/\gamma$인 점들은 $\gamma y<1$이므로 자기자체에 의해서만 $\gamma-$피복되는 점들 즉 $\gamma-$피복값이 1인 점들이 있을수 있기때문이다.

E에 속하는 점들의 $\gamma-$피복값합을 정확히 구할수 없으므로 그것이 평균적으로 기껏 얼마인가를 따지자.

E에 속하는 점들의 개수는 기껏 n/γ이다.

$(x, y) \in E \cap C$ 들의 $\gamma-$피복값은 1이며 $(x, y) \in E \setminus C$ 들에 대해서는 최대로 가질수 있는 $\gamma-$피복값은 $1/2^2$ 이다.

만일 $E \cap C = \varnothing$ 이면 E 에 속하는 모든 점들의 $\gamma-$피복값합은 기껏 $(1/4)|E|$이다.

만일 $(x, y) \in E \cap C$ 인 점이 있다면 $w(x, y) = 1$ 이고 보조정리 3으로부터 $(x+1, y) \notin C$ 이다.

이때 $w(x+1, y) = 0$ 이면 두 점 (x, y) 와 $(x+1, y)$ 는 $\gamma-$피복값합이 1이다.

이때 $w(x+1, y) > 0$ 이면 $(x+1, y)$ 를 $\gamma-$피복하는 점 $(x', y') \in C \setminus E$ 는 적어도 4개의 점을 $\gamma-$피복하게 되며 그 4개의 점들속에는 $(x+1, y) \notin C$ 도 있다. 그러면 E 에 속하는 세 점 (x, y), $(x+1, y)$, $(x+2, y)$ 의 $\gamma-$피복값합은 기껏 $1+1/4+1/4=3/2$ 이다.

이로부터 E 에 속하는 점들의 평균 $\gamma-$피복값은 기껏 1/2임을 알수 있다.
즉 E 에 속하는 점들의 $\gamma-$피복값합은 기껏 $(1/2)|E| = n/(2\gamma)$ 이다.

$(i-1)/\gamma \le y < i/\gamma,\ i \ge 2$ 부터는 정리 1의 증명방법을 그대로 적용한다. 즉 비증가함수 $f(1) := n/2,\ f(i) := n/i^2 (i \ge 2)$ 라고 하자. 그러면 결국

$$\sum_{i=1}^{\lceil n\gamma \rceil} (f(i) g_\gamma(i)) \le n(1/2 + \sum_{i=2}^{\lceil n\gamma \rceil} (1/i^2))/\gamma < n(\pi^2/6 - 1/2)/\gamma$$

를 얻는다.(증명끝)

참 고 문 헌

[1] Roman Kolpakov et al.; Proc. CPM., LNCS, 8486, 212, 2014.

[2] Maxime Crochemore et al.; arXiv:1509.01221v3[cs.FL], 2, 2015.

[3] Pawel Gawrychowski et al.; Theoret. Comput. Sci., 62, 1, 162, 2018.

주체109(2020)년 6월 5일 원고접수

A Computing Way about Bound for All Maximal $\alpha-$Gapped Repeats by $\gamma-$Cover

Kang Jin Ung, Jo Yong Son

In this paper, we improved a computing way of partitioning the set of all maximal $\alpha-$gapped repeats, defined $\gamma-$cover weight and computed a bound for all maximal $\alpha-$gapped repeats by $\gamma-$cover at $n(\pi^2/6-1/2)/\gamma$.

Keywords: $\gamma-$cover, maximal $\alpha-$gapped repeat