

보조정보에 오류가 있는 상황에서 인덱스 부호 인코딩 문제의 필요충분조건

김재원, 노종선

서울대학교 전기정보공학부, 뉴미디어통신공동연구소

kjw702@ccl.snu.ac.kr, jsno@snu.ac.kr

A Necessary and Sufficient Condition for Encoding Index Codes When Side-Information is Erroneous

Jae-Won Kim and Jong-Seon No *

Dept. of ECE, Seoul National Univ., INMC

요 약

본 논문은 보조정보를 이용하여 부호의 길이를 줄이는 인덱스 부호화 문제에서 각 수신기의 보조정보에 오류가 있는 상황일 때 인코딩에 관한 필요충분조건을 제시한다. 먼저, 일반적인 비선형 부호에 관해 인코딩에 관한 필요충분조건을 찾고 선형 부호일 경우에 대해서는 생성행렬에 관한 필요충분조건을 보인다.

I. 서 론

인덱스 부호화는 1998 년에 처음 위성통신을 위하여 도입된 개념으로 위성통신적 목적 외에 다양한 정보이론적 문제들 및 네트워크 부호 문제, 간섭 관리 문제, 분산 저장 시스템 등과의 연관성으로 인해 최근 학계에서 꾸준히 주목 받고 있다. 인덱스 부호화의 주목적은 각각의 수신기가 가지고 있는 보조정보를 이용하여 송신기가 최적의 부호 길이를 갖는 부호를 설계하는 것이다. 일반적으로 인덱스 부호를 설계하고 최적의 부호 길이를 찾는 것은 NP-hard 문제로 알려져 있다 [1].

일반적인 인덱스 부호화 문제 상황은 다음과 같이 정리된다. 인덱스 부호화 문제에서는 하나의 송신기와 다수의 수신기가 에러가 없는 브로드캐스트 채널(broadcast channel)을 통하여 정보를 전달한다. 이때, 각 수신기는 송신기가 보낼 정보들 중 일부를 보조정보로 미리 알고 있으며 송신기의 정보들 중 일부를 수신 받기를 원한다. 인덱스 부호화가 다른 부호화 기법들과 다른 점은 각 수신기들이 보조정보가 있으며 모든 정보를 수신 받기를 원하는 것이 아니라는 점이다. 또한, 송신기는 각 수신기가 어떤 정보를 수신 받기를 원하는지와 어떤 정보를 미리 갖고 있는 지 알고 있다고 가정하며 이 정보들을 통해 가장 짧은 길이의 인덱스 부호를 설계하여 송신하려 한다. 위와 같은 각 수신기들의 상태를 나타내주는 그래프를 보조정보 그래프라고 한다.

구체적으로 본 논문에서 다루는 인덱스 부호화 상황을 다음과 같이 정리하며 [2]의 표기법을 따른다. 송신기가 가지고 있는 메시지를 n 개, 수신기의 개수를 m 개로 생각한다. 즉, 송신기는 벡터 $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{F}_q^n$ 를 가지고 있으며 보조정보 그래프 G 를 알고 있다. 각 수신기는 이 중 하나의 정보를 수신 받기를 원하며 i 번째 수신기 R_i 가 원하는 메시지의 인덱스를 $f(i)$ 로 정의하고 해당 원소를 $x_{f(i)}$ 로 정의한다. 또한, R_i 가 보조정보로

미리 알고 있는 메시지들의 인덱스집합을 \mathcal{X}_i 로 정의한다. 일반적인 인덱스 부호화 상황에서는 각 메시지가 벡터 형태라든지 각 수신기가 여러 메시지를 수신 받기를 원하는 상황이 발생하지만 유한한 길이의 인덱스 부호에서는 위의 상황을 본 논문에서 정의한 상황으로 자명하게 바꿀 수 있다. 따라서 본 논문에서는 위와 같은 상황에서 인덱스 부호화 문제를 다루며 어느 수신기에서도 수신 받기를 원치 않는 메시지가 있다면 그 메시지는 제외시킬 수 있으므로 모든 메시지는 어느 수신기에서는 원해지며 $m \geq n$ 으로 생각할 수 있다. 일반적인 인덱스 부호화 상황에서는 보조정보에 오류가 없으므로 $f(i) \notin \mathcal{X}_i$ 로 가정한다. 본 논문에서는 보조정보에 오류가 있는 상황을 다루기 때문에 $f(i) \in \mathcal{X}_i$ 인 경우도 생각해 볼 수 있지만 일반적인 인덱스 부호화 문제에서 $f(i) \notin \mathcal{X}_i$ 로 가정한다는 점과 인덱스 부호화 상황에서 보조정보가 에러가 없는 브로드캐스트 채널을 통해 얻어지는 경우가 있다는 점으로 생각해 봤을 때 본 논문에서도 일반적인 상황과 마찬가지로 $f(i) \notin \mathcal{X}_i$ 로 가정한다. 또한, 각 수신기는 최대 δ_i 개의 에러가 있는 정보를 갖고 있을 수 있으며 어떤 정보가 에러인지는 알지 못한다고 가정한다. 수신기의 보조정보가 에러가 있는 상태인 경우는 흔히 있을 수 있다고 생각되는 가정이며 가장 간단한 예시로는 보조정보가 메모리에 저장되어 있을 때 메모리 에러로 인한 것을 생각할 수 있다.

보조정보 그래프는 각 수신기가 미리 알고 있는 보조정보와 수신 받기를 원하는 메시지를 나타내 주는 그래프이며 메시지 개수와 수신기 개수가 같을 때는 1 개의 노드(node)로 이루어진 단일 노드 그래프(unipartite graph)로 표현되며 다른 경우에는 메시지와 수신기 노드 2 개로 이루어진 양자 노드 그래프(bipartite graph)로 표현된다.

본 논문의 구성은 다음과 같다. II 장에서는 보조정보에 오류가 있는 환경에서 인덱스 부호의 인코딩 조건을

제시하고 이를 증명한다. 또한, 이와 관련된 예시를 보인다. III장에서는 결론을 맺는다.

II. 본론

먼저 다음과 같은 벡터의 집합을 정의하며 \mathbf{x}_{x_i} 는 R_i 의 보조정보로 이루어진 벡터를 의미한다.

$$\mathcal{L}(q, G, \delta_s) = \{ \mathbf{z} \in \mathbb{F}_q^n : \exists i \in [m] \text{ s.t. } \text{wt}(\mathbf{z}_{x_i}) \leq 2\delta_s, z_{f(i)} \neq 0 \} \quad (1)$$

송신기의 인코딩 함수를 $E(\cdot)$ 으로 표현할 때, 보조정보에 오류가 있는 상황에서 인덱스 부호 인코딩 문제의 필요충분조건은 다음과 같다.

Theorem 1: 각 수신기가 최대 δ_s 개의 에러가 있는 보조정보를 갖는 인덱스 부호화 상황에서의 인코딩 함수는 다음의 수식 (2)를 필요충분조건으로 만족한다.

$$E(\mathbf{x}) \neq E(\mathbf{x}') \quad \text{for all } \mathbf{x} - \mathbf{x}' \in \mathcal{L}(q, G, \delta_s) \quad (2)$$

증명)

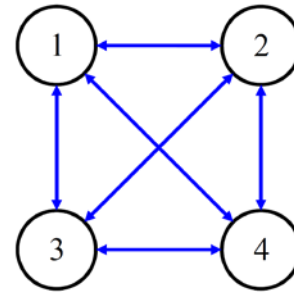
각 수신기 R_i 는 $x_{f(i)}$ 를 수신 받은 코드워드(codeword)와 보조정보 \mathbf{x}_{x_i} 를 이용하여 복호해야 한다. 그러면, 송신기의 경우는 각 수신기가 헷갈릴만한 (confusable) 메시지들을 다른 코드워드로 인코딩하면 된다. R_i 는 다른 메시지들 외에 오직 $x_{f(i)}$ 에만 관심이 있으므로 같은 $x_{f(i)}$ 값을 갖는 메시지들에 대해서는 서로 코드워드로 구분될 필요가 없다. 또한, $\text{wt}(\mathbf{x}_{x_i} - \mathbf{x}'_{x_i}) \geq 2\delta_s$ 인 두 메시지 벡터 \mathbf{x} 와 \mathbf{x}' 에 대해서는 R_i 가 보조정보로 두 메시지 벡터를 구분해 낼 수 있으므로 같은 코드워드로 인코딩되어도 상관이 없다. 따라서, 반드시 구분되어야 하는 메시지 벡터의 쌍 \mathbf{x} 와 \mathbf{x}' 은 $x_{f(i)} \neq x'_{f(i)}$ 이며 $\text{wt}(\mathbf{x}_{x_i} - \mathbf{x}'_{x_i}) \leq 2\delta_s$ 이다. 이러한 메시지 벡터 \mathbf{x} 와 \mathbf{x}' 의 차이에 대해서 표현하면 수식 (1)과 같은 집합에 포함된다.

위의 조건을 선형 부호에 대해서 논하면 다음과 같은 따름 정리를 얻는다.

Corollary 1: 각 수신기가 최대 δ_s 개의 에러가 있는 보조정보를 갖는 인덱스 부호화 상황에서의 생성행렬 A 는 다음의 수식 (3)을 필요충분조건으로 만족한다.

$$\mathbf{x}A \neq \mathbf{0} \quad \text{for all } \mathbf{x} \in \mathcal{L}(q, G, \delta_s) \quad (3)$$

아래의 예시에서 선형 부호인 경우 생성행렬 중 하나를 확인할 수 있다. $q = 2, \delta_s = 1, n = m = 4, f(i) = i$ for all $i \in [m]$ 이고 보조정보 그래프 G 는 <그림 1>과 같다.



<그림 1> $n = m = 4$ 일 때 예시의 보조정보 그래프

<그림 1>에서 각 인덱스는 수신기 및 수신기가 원하는 메시지의 인덱스를 의미하며 화살표는 보조정보의 유무를 나타낸다. 가령, 노드 1 에서 노드 2 로의 화살표는 1 번 수신기가 메시지 2 를 보조정보로 갖는다는 것을 의미한다. 이 예시의 경우는 모든 수신기가 자신이 수신 받기를 원하는 메시지 외에 모든 메시지를 보조정보로 갖고 있는 상황이다.

이 경우 $\mathcal{L}(q, G, \delta_s)$ 는 $(0,0,0,0)$ 과 $(1,1,1,1)$ 을 제외한 \mathbb{F}_2^4 상의 모든 벡터임을 쉽게 확인할 수 있다. 이 때, 아래의 생성행렬 A 는 수식 (3)을 만족하는 것을 확인할 수 있다.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad (4)$$

이 경우 인덱스 부호화를 하지 않는 경우는 메시지를 하나씩 보내어 부호 길이 4 가 필요하지만 주어진 생성행렬을 사용할 시 부호 길이가 3 으로 감소하였다. 복호화의 경우는 이 예제의 경우 간단하게 할 수 있으며 가령 1 번째 수신기의 경우 수신 받은 3 개의 값 $x_1 + x_2, x_1 + x_3, x_1 + x_4$ 각각에서 보조정보 x_2, x_3, x_4 를 빼면 x_1 의 추정치를 얻게 되는데 각 수신기는 최대 1 개의 에러가 있는 보조정보를 갖게 되므로 x_1 의 추정치는 2 개 이상이 같게 되며 이 값을 고르면 된다.

III. 결론

본 논문에서는 인덱스 부호화 상황에서 기존의 가정과 달리 보조정보에 에러가 있을 경우를 대처하는 새로운 인덱스 부호화 상황에 대하여 논하였고, 이 경우 인코딩 함수의 필요충분조건에 대해 논하였다.

ACKNOWLEDGMENT

이 논문은 2016 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2016R1A2B2012960).

참고 문헌

[1] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1479-1494, 2011.
 [2] S. H. Dau, V. Skachek, and Y. M. Chee, "Error correction for index coding with side information," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1517-1531, 2013.