Asymmetric information embedding

Amin Shokrollahi

Abstract In the Information Embedding Problem one is given a piece of data which can be altered only conditionally, for example only at certain places. One is then asked to embed an arbitrary message into the data by only applying admissible changes to the data. These changes lead to a distortion which is to be kept low. In this short note, we introduce an “asymmetric” version of information embedding in which the file is regarded as a string over a finite alphabet, and admissible changes on the alphabet elements are modeled by a directed graph. We introduce embedding techniques based on list-decoding algorithms for algebraic–geometric codes, and analyze their performance.

1 Introduction

Let $G = (V, E)$ be a directed graph in which for all $v \in V$ we have $(v, v) \in E$, and let $\hat{G}$ be its transitive closure. An example is the $q$-line consisting of $q$ nodes $0, 1, \ldots, q - 1$ such that there is an edge between $i$ and $i + 1$ for $i = 0, \ldots, q - 2$. Figure 1 gives an example for the case $q = 5$.

Let $P(S)$ denote the set of subsets of a set $S$. An Information Embedding Scheme (IES) of signature $(n, k; G)$, denoted $\text{IES}(n, k; G)$, consists of a pair $(E, D)$ of polynomial time computable functions $E : \mathbb{F}_q^n \times \mathbb{F}_q^k \to \mathbb{F}_q^n$ and $D : \mathbb{F}_q^n \to P(\mathbb{F}_q^k)$ called encoding and decoding functions, such that:

1. For all $v \in \mathbb{F}_q^k$ and $x \in \mathbb{F}_q^n$ and all $i = 1, \ldots, n$ we have that $(y_i, x_i) \in E$, where $y = E(x, v)$.
2. There exist $\delta, \gamma > 0$ (depending possibly on $q$ but not on $n$ or $k$) such that if $v \in \mathbb{F}_q^k$ and $x \in \mathbb{F}_q^n$ are chosen uniformly at random, then $\Pr[v \notin D(\mathbb{E}(v, x))] \leq \gamma \exp(-n\delta)$.

We call the quantity $k/n$ the “rate” of IES($n, k; G$). We are interested in IESs of high rate. More precisely, we would like to find the value of $A(G)$, where

$$A(G) := \sup \{ R \mid \exists \text{ Information Embedding scheme of rate } R \text{ for } G \}.$$ 

It is possible that for some graphs there is no IES. For example, suppose that the graph has only self-loops and no edges between the nodes. Then, the only encoding function is the one mapping $(v, x)$ to $x$, regardless of $v$. As a result, a given $x$ does not convey any information about $v$; the decoder $D$ can only guess the value of $v$, with error $1 - 1/2^k$. This error does not decay polynomially in $n$, hence there is no matching decoding function for $\mathbb{E}$.

In this paper, we will always assume that the vertex set $V$ of the graph $G$ is a set with $q$ elements, and denote this set either by $\{0, \ldots, q-1\}$ (as in the case of the $q$-line) or by the field $\mathbb{F}_q$. Moreover, our encoding function $\mathbb{E}$ is composed of two functions: an encoding function $\varphi$ which maps $v$ to an element $c = \varphi(v)$ of a suitably chosen linear $[n, k]_q$-code $C$, and a map $\pi$ which maps the pair $(x, c)$ to the vector $(y_1, \ldots, y_n)$, where

$$y_i = \begin{cases} x_i & \text{if } (c_i, x_i) \notin E \\ c_i & \text{else.} \end{cases}$$

In other words, whenever we are allowed to replace $x_i$ with $c_i$, we will do so opportunistically. Whether this scheme is successful or not depends on the statistics of $x$, and that of $c$. For example, suppose that the underlying graph is the $q$-line, and that $c = (q-1, q-1, \ldots, q-1)$. Then $\mathbb{E}(x, c) = x$ for all $x$, and hence $\mathbb{E}(x, c)$ conveys essentially no information about $c$.

We will therefore assume that we know something about the statistics of $c$. More precisely, we assume that we know the signature of $c$, defined as the real vector $(s_\beta \mid \beta \in \mathbb{F}_q)$ such that

$$s_\beta = \frac{\#\{i \mid c_i = \beta\}}{n}.$$