One Model for 'Gossip' Distribution

Kristijan TABAK

Rochester Institute of Technology, RIT Croatia, Damira Tomljanovića Gavrana 15, 10000 Zagreb

***Corresponding author e-mail** kxtcad@rit.edu

In this paper I present one model for information distribution (gossip maybe), where original information could be (deliberately or not) changed. Idea is to define information and to describe its transport. Formulas are provided to describe probability of correct information after certain number of steps. This paper ends by posting some interesting research problems.

Introduction & Definitions

Definition: Information is an n-array

$$\omega = (\omega_1, \ldots, \omega_n) \in \{0,1\}^n.$$

We also need next:

Definition: Transport vector is

$$\Pi = (p_1, \ldots, p_n) \in \langle 0, 1]^n,$$

While

$$\Omega_\Pi = (\Omega_{\Pi_1}, \ldots, \Omega_{\Pi_n})$$

is transport function, defined as

$$\Omega_\Pi = (\Omega_{\Pi_1}(\omega_1), \ldots, \Omega_{\Pi_n}(\omega_n)),$$

where $\Omega_{\Pi_i}(\omega_i)$ is a random variable defined as

$$\Omega_{\Pi_i}(\omega_i) = \begin{pmatrix} \omega_i & 1-\omega_i \\ p_i & 1-p_i \end{pmatrix}$$

Example: If $\Pi = (1,1,\ldots,1)$ then $\Omega_\Pi(\omega) = \omega$, but if $\Pi = (0,0,\ldots,0)$ then $\Omega_\Pi(\omega) = I - \omega$, where $I = (1,1,\ldots,1)$.

Results

Let us define matrix P_i, by which we describe probabilities of statuses on i-th bit. We have

$$P_i = \begin{pmatrix} p_i & 1-p_i \\ 1-p_i & p_1 \end{pmatrix},$$

where $(P_i)_{11}$ means probability that ω_i was read as ω_i. Similarly goes for other three probabilities. To be more precise, if we assume that $X_{i,m}$ is a status of i-th bit after m transfers, we have

$$P(X_{i,n+1} = \omega_i \mid X_{i,n} = \omega_i) = p_i,$$

$$P(X_{i,n+1} = 1-\omega_i \mid X_{i,n} = 1-\omega_i) = p_i,$$

$$P(X_{i,n+1} = \omega_i \mid X_{i,n} = 1-\omega_i) = 1-p_i,$$

$$P(X_{i,n+1} = 1-\omega_i \mid X_{i,n} = \omega_i) = 1-p_i.$$

Lemma: Let $X_{i,m}$ be a status of i-th bit after m steps. Then

$$P(X_{i,m+1} = a \mid X_{i,m} = b) = p_i||a-b|-1| + (1-p_i)|a-b|.$$

Using linear algebra we can show that P_i^m is

$$P_i^m = \begin{pmatrix} \frac{1}{2}+\frac{1}{2}(2p_i-1)^m & \frac{1}{2}-\frac{1}{2}(2p_i-1)^m \\ \frac{1}{2}-\frac{1}{2}(2p_i-1)^m & \frac{1}{2}+\frac{1}{2}(2p_i-1)^m \end{pmatrix}$$

On the other hand, we have another result.

Lemma: Let $X_{i,m}$ be a status of i-th bit after m steps. Then probability of sending 'a' if 'a' was received is

$$P(X_{i,m} = a \mid X_{i,m+1} = a) = \frac{p_i P(X_{i,m} = a)}{P(X_{i,m+1} = a)}.$$

Proof: We use formula for conditional probability.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A)} \frac{P(A)}{P(B)} = P(B \mid A) \frac{P(A)}{P(B)}.$$

Using this, we get

$$P(X_{i,m} = a \mid X_{i,m+1} = a) = P(X_{i,m+1} = a \mid X_{i,m} = a) \frac{P(X_{i,m} = a)}{P(X_{i,m=1} = a)}.$$

But, on the other hand, we have

$$P(X_{i,m=1} = a \mid X_{i,m} = a) = p_{aa} = p_i,$$

hence our assertion is proved. □

Using power of P_i we also proved that

$$P(X_{i,m} = \omega_i \mid X_{i,0} = \omega_i) = \frac{1}{2} + \frac{1}{2}(2p_i - 1)^m,$$

$$P(X_{i,m} = 1-\omega_i \mid X_{i,0} = \omega_i) = \frac{1}{2} - \frac{1}{2}(2p_i - 1)^m.$$

Next result tells us how to calculate probability that information has been transferred as original had been sent.

Theorem: Let X_m be a state of information after m steps, where $\Pi = (p_1, \ldots, p_n)$ is a transport vector. Then, after introducing the notation

$$P(X_m = \omega) := P(X_m = \omega \mid X_0 = \omega),$$

we have

$$P(X_m = \omega) = \frac{1}{2^n} \prod_{i=1}^{n} (1 + (2p_i - 1)^m).$$

Proof: It is clear that we have

$$\begin{aligned} P(X_m = \omega) &= P(X_m = \omega \mid X_0 = \omega) = \\ &= P(X_{1,m} = \omega_1 \mid X_{1,0} = \omega_1) P(X_{2,m} = \omega_2 \mid X_{2,0} = \omega_2) \times \cdots \times \\ &\times P(X_{n,m} = \omega_n \mid X_{n,0} = \omega_n) = \\ &= \left(\frac{1}{2} + \frac{1}{2}(2p_1 - 1)^m\right)\left(\frac{1}{2} + \frac{1}{2}(2p_2 - 1)^m\right) \cdots \left(\frac{1}{2} + \frac{1}{2}(2p_n - 1)^m\right) = \\ &= \frac{1}{2^n} \prod_{i=1}^{n} (1 + (2p_i - 1)^m). \end{aligned}$$

This gives us the desired result.

There is interesting consequence of previous result.

Theorem: Let X_m be a state of information after m steps, where $\varPi = (p_1, \ldots, p_n)$ is a transport vector. Then, following statements hold:

1. $P(X_{2m} = \omega) > 0,$
2. $p_i = 0 \Rightarrow P(X_{2m+1} = \omega) = 0.$

Proof: One can notice that

$$(2p_i - 1)^{2m} \geq 0 \Rightarrow 1 + (2p_i - 1)^{2m} > 0 \Rightarrow P(X_{2m} = \omega) = \frac{1}{2^n} \prod_{i=1}^{n} (1 + (2p_i - 1)^m) > 0.$$

On the other hand, if $p_i = 0$ for some $i \in \{1, 2, \ldots, n\}$, then

$$(2p_i - 1)^{2m+1} = (-1)^{2m+1} = -1.$$

Therefore $1 + (2p_i - 1)^{2m+1} = 0$, which gives us $P(X_{2m+1} = \omega) = 0$.

This proves our assertion.

Previous result claims, that if one segment of information or a story is always deliberately changed, we should accept information after event number of transports, otherwise we have smaller probability of detecting true information.

Now let us introduce function that describes probability of correct information transport.

$$g(p_1, \ldots, p_n) = \log_2 P(X_m = \omega).$$

For the sake of simplicity let us introduce another notation:

$$\widetilde{p_i} = 2p_i - 1, \; i = 1, 2, \ldots, n.$$

Then, we may write

$$\begin{aligned} g(p_1, \ldots, p_n) &= \log_2 P(X_m = \omega) = \\ &= \sum_{i=1}^{n} \log_2 (1 + \widetilde{p_i}^m) - n = \\ &= \widetilde{g}(\widetilde{p_1}, \ldots, \widetilde{p_n}). \end{aligned}$$

Now, our task is to check extreme values for probability described by previously introduced function.
We find derivatives by each variable. Then we have

$$\frac{\partial \widetilde{g}}{\partial \widetilde{p_i}} = \frac{1}{(1 + \widetilde{p_i}^m)\ln 2} \cdot m \cdot \widetilde{p_i}^{m-1} = 0 \Rightarrow \widetilde{p_i}^{m-1} = 0 \Rightarrow \widetilde{p_i} = 0 \Rightarrow p_i = \frac{1}{2}.$$

But, one can notice following sequence of conclusions:

$$p_i \in \langle 0, 1] \Rightarrow \widetilde{p_i} \in \langle -1, 1] \Rightarrow \widetilde{p_i}^m + 1 \in \langle 0, 2] \Rightarrow \log_2(\widetilde{p_i}^m + 1) \in \langle -\infty, 1].$$

Then, after substitution

$$\log_2(\widetilde{p_i}^m + 1) = x_i \in \langle -\infty, 1],$$

we get

$$g(p_1, \ldots, p_n) = \sum_{i=1}^{n} x_i - n \leq 0,$$

On the other hand it is clear that

$$g(p_1, \ldots, p_n) = \sum_{i=1}^{n} x_i - n > -\infty.$$

Natural question what is probability of correct information transport after many transports.

Theorem: Let X_m be a state of information after m steps, where $\varPi = (p_1, \ldots, p_n)$ is a transport vector. Let

$$k = |\{i \mid p_i = 1\}|,$$

then, after infinitely many steps probability of correct information transport is given by

$$\lim_{m \to \infty} P(X_m = \omega) = \frac{1}{2^{n-k}}.$$

Proof: We had

$$P(X_m = \omega) = \frac{1}{2^n} \prod_{i=1}^{n} (1 + (2p_i - 1)^m).$$

Therefore, we may write

$$P(X_m = \omega) = \frac{1}{2^n} \prod_{i=1,\; p_i < 1}^{n} (1 + (2p_i - 1)^m) \times \prod_{i=1,\; p_i = 1}^{n} (1 + (2p_i - 1)^m).$$

One can see that

$$p_i \in \langle 0, 1\rangle \Rightarrow (2p_i - 1)^m \in \langle -1, 1\rangle \Rightarrow \lim_{m \to \infty} (2p_i - 1)^m = 0,$$

therefore

$$\lim_{m \to \infty} P(X_m = \omega) = \lim_{m \to \infty} \left(\frac{1}{2^n} \prod_{i=1,\; p_i < 1}^{n} (1 + (2p_i - 1)^m) \times \prod_{i=1,\; p_i = 1}^{n} (1 + (2p_i - 1)^m) \right) =$$

$$= \left(\frac{1}{2^n} \prod_{i=1,\; p_i < 1}^{n} (1) \times \prod_{i=1,\; p_i = 1}^{n} (1 + 1) \right) = \frac{1}{2^n} \times 2^k = \frac{1}{2^{n-k}}.$$

This gives us our proof. □

Next result claims that is better to accept information after even number than after odd number of transports.

Theorem: Let X_m *be a state of information after* m *steps,* $\varPi = (p_1, \ldots, p_n)$ *is a transport vector. Then*

$$P(X_{2m = \omega}) > P(X_{2m+1} = \omega).$$

Proof: First, notice that

$$\begin{aligned} P(X_m = \omega) &= \frac{1}{2^n} \prod_{i=1}^{n} (1 + (2p_i - 1)^m) = \frac{1}{2^n} \prod_{i=1}^{n} (1 + \widetilde{p_i}^m) = \\ &= \frac{1}{2^n} \prod_{i=1,\; \widetilde{p_i} \neq 0}^{n} (1 + \widetilde{p_i}^m) = \frac{1}{2^{n-k}} \prod_{i=1,\; \widetilde{p_i} \neq 0,1}^{n} (1 + \widetilde{p_i}^m), \end{aligned}$$

where $k = |\{p_i \mid p_i = 1\}|$. So, from now on through the proof, we may assume that $\widetilde{p_i} \neq 0, 1$. Thus we have

$$\widetilde{p_i}^{2m} - \widetilde{p_i}^{2m+1} = \widetilde{p_i}^{2m}(1 - \widetilde{p_i}) > 0,$$

because

$$-1 < \widetilde{p_i} < 1.$$

Therefore, we have

$$(1 + \widetilde{p_i}^{2m}) > (1 + \widetilde{p_i}^{2m+1}) > 0,$$

hence, after multiplication we have

$$\prod_{i=1}^{n} (1 + \widetilde{p_i}^{2m}) > \prod_{i=1}^{n} (1 + \widetilde{p_i}^{2m+1}),$$

thus, finally

$$P(X_{2m = \omega}) > P(X_{2m+1} = \omega).$$

This proves our theorem. □

Finally, we present one lower bound for probability of correct information.

Theorem: Let X_m be a state of information after m steps, where $\varPi = (p_1, \ldots, p_n)$ is a transport vector and $p_i \geq p$. Then

$$P(X_m = \omega) \geq \left(\frac{1 + \widetilde{p}^m}{2} \right)^n.$$

Proof: One can see that following sequence of conclusions work

$$p_i \geq p > 0 \Rightarrow 2p_i - 1 \geq 2p - 1 \Rightarrow \tilde{p}_i \geq \tilde{p} > -1 \Rightarrow \tilde{p}_i^m \geq \tilde{p}^m > -1.$$

Then we have

$$P(X_m = \omega) = \frac{1}{2^n} \prod_{i=1}^{n} (1 + \tilde{p}_i^{\,m}) \geq \frac{1}{2^n} \prod_{i=1}^{n} (1 + \tilde{p}^m) = \left(\frac{1 + \tilde{p}^m}{2}\right)^n.$$

Further Research

Now we will say something about distribution of information over complete graph

$$K_n = (V, E),\ V = \{1, 2, \ldots, n\},\ E = \{(i, j) \mid i, j = 1, 2, \ldots, n\},$$

with vertices and edges, at which every two vertices are connected.

Definition: Distribution of information in graph K_n is sub graph, which is a tree with n vertices. If distribution of information has no sequence $i \to j \to k,$, where $i > j,$ then we will say that it is directed distribution of information.
Hence, it is natural to introduce notation

$$DI(n) = \{D \leq K_n \mid D \text{ is a distribution of information}\},$$
$$DDI(n) = \{D \leq K_n \mid D \text{ is a directed distribution of information}\}.$$

It is quite clear that

$$DDI(n) \neq DI(n).$$

Using Cayley's formula we can prove that

$$|DI(n)| = n^{n-2}$$

.

There are some topics that could be interesting to cover, like, to describe probability for returning information to someone that have already heard it. Also, it would be of great interest to offer some algebra to classify and count direct distributions of information. All of that probably could be used in determining, so called, returning probability of an information.

References

Asmussen, S. (2000). *Ruin Probabilities.* Singapore, London. *World Scientific.*
Bhat, U.N. and Miller, G.K. (2002). *Elements of Applied Stochastic Processes*, 3rd edition. Wiley, New York.
Doob, J.L. (1953). *Stochastic Processes*, Wiley & Sons, New York.
Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd edition. Wiley & Sons, New York.