

Information Spreading Model Based on Worth Quantification Index

Xiao-Jian Lu

Department of Power Engineering, North China Electric Power University, Baoding, Hebei, P.R. China

Keywords: Information spread, Analytic Hierarchy Process (AHP), Epidemic Model

Abstract. In this paper, two models, including the information value quantification model and the epidemic model, are built to explore how social information spreads and works, especially through the networks. The information value is quantitatively determined by means of Analytic Hierarchy Process (AHP), based on which we compare the information spread process to the virus infection procedure and develop an epidemic model to investigate the relationship between the flow and the value of information. The result shows that the information spread is positively related to its value. A large amount of data in the past applying to the model, together with the comparison of the reality, validates the feasibility and the correctness of the presented model.

1. Introduction

Social networks are information networks that describe how individuals or entities interact with one another. With the constantly development of science and technology, information networks also evolve, making information spread further and wider, and playing a more and more important role in our daily life. Therefore, it is significant to study on the information spread process and the factors which affect it.

Generally, people focus on why some information can be widely known, while others are just be known by only a few people. It is shown that the politeness of the language, and the attractiveness of the information [1-2] greatly affects the spreading scope and speed of the information. By far, few people have analyzed the relationship between the inherent value and the information spread, which needs a detailed modeling work for a refine analysis.

In this paper, we set up an epidemic model to assess the information spread capacity. This model uses AHP to quantify the value of different information and compares the information spread to the virus infection process.

2. Description of the model

2.1 The model of quantizing the information's value based on the AHP

Information value is an essential attribute, which can be analyzed in two first grade indexes: the value of content and the quality of writing. The content is a primary index to judge the information value and can be described by 11 second grade indexes, i.e., the truth, importance, timeliness, veracity, accuracy, completeness, pertinence, foreseeability, necessity, low-down and novelty. The quality of writing is also one of indexes to evaluate information value, which includes 5 second grade indexes: specific theme, deliberate thought, accurate language, brief expression and comprehensive characters.

We use AHP to determine the weight of these indexes.

Firstly, comparing the importance of the second grade indexes for the first grade indexes to get the Comparative Matrix Pairs [3].

The importance scale of the indexes are represented in the tab.1:

Tab.1 The importance scale of the Comparative Matrix Pairs

Scale	Meaning
1	The two elements are equally important.
3	The first element is slightly important to the next element.
5	The first element is clearly important to the next element.
7	The first element is very important to the next element.
9	The first element is strongly important to the next element.

2, 4, 6, 8 express the intermediate value of the above judgments.

If the ratio of important of elements i and j is a_{ij} , the ratio of important of elements j and i

$$\text{is } a_{ji} = \frac{1}{a_{ij}}$$

Then, we compare the importance of indexes on the same hierarchy with the above to get the Comparative Matrix Pairs [3]. The Comparative Matrix Pairs about content value and write quality are shown separately as follows [4]:

$$G_1 = \begin{bmatrix} 1 & 2 & 1 & 3 & 5 & 3 & 4 & 5 & 7 & 3 & 5 \\ 1/2 & 1 & 1 & 2 & 1 & 3 & 1 & 3 & 5 & 1 & 3 \\ 1 & 1 & 1 & 3 & 5 & 3 & 3 & 1 & 5 & 2 & 5 \\ 1/3 & 1/2 & 1/3 & 1 & 1 & 2 & 1 & 3 & 1/2 & 1 & 2 \\ 1/5 & 1 & 1/5 & 1 & 1 & 1 & 2 & 1 & 1/2 & 1/3 & 1/2 \\ 1/3 & 1/3 & 1/3 & 1/2 & 1 & 1 & 1 & 1 & 1/3 & 1/5 & 2 \\ 1/4 & 1 & 1/3 & 1 & 1/2 & 1 & 1 & 3 & 1 & 1/3 & 5 \\ 1/5 & 1/3 & 1 & 1/3 & 1 & 1 & 1/3 & 1 & 1/2 & 1/3 & 1 \\ 1/7 & 1/5 & 1/5 & 1 & 2 & 3 & 1 & 2 & 1 & 1/2 & 3 \\ 1/3 & 1 & 1/2 & 1 & 3 & 5 & 3 & 3 & 2 & 1 & 1 \\ 1/5 & 1/3 & 1/5 & 1/2 & 2 & 1/2 & 1/5 & 1 & 1/3 & 1 & 1 \end{bmatrix} \quad (1)$$

$$G_2 = \begin{bmatrix} 1 & 4 & 1/2 & 1 & 1 \\ 1/4 & 1 & 1/3 & 1/4 & 1/4 \\ 2 & 3 & 1 & 1 & 2 \\ 1 & 4 & 1 & 1 & 1 \\ 1 & 4 & 1/2 & 1 & 1 \end{bmatrix} \quad (2)$$

Then, make consistency check to these matrixes by the software of Matlab. The consistent ratio of G1 and G2 are 0.094 and 0.027 respectively. Thus, they can pass the consistency check.

Finally, the characteristic vectors of the maximum characteristic roots of above matrixes are normalized to get the weight of the indexes, and their weight vectors of content value and write quality are shown separately as follows:

$$W_1 = (0.218, 0.119, 0.169, 0.065, 0.051, 0.043, 0.070, 0.045, 0.068, 0.108, 0.041) \quad (3)$$

$$W_2 = (0.203, 0.063, 0.299, 0.232, 0.203) \quad (4)$$

By consulting literatures, the indexes and their weight are represented in the tab.2.[2]

Tab.2 the indexes and their weight

I indexes	II indexes	index weight
Content value 0.7	truth	0.218
	importance	0.119
	timeliness	0.169
	accuracy	0.065
	precision	0.051
	completeness	0.043
	pertinence	0.070
	foreseeability	0.045
	necessity	0.068
	low-down	0.108
	novelty	0.041
write quality 0.3	specific theme	0.203
	deliberate thought	0.063
	accurate language	0.299
	brief expression	0.232
	comprehensive characters	0.203

From the data in table 2, the value of different information (V) can be obtained quantitatively by assigning the indexes coefficients (e_i) for a piece of information.

$$V = 0.7 \sum w_{1,i} \cdot e_i + 0.3 \sum w_{2,i} \cdot e_i \quad (5)$$

Here, e_i is the value of every index about the value of content and the writing quality, which can be got through the scores given by experts. And w_i is the weight of every index. So we can get the value of a piece of information.

2.2. the information spreading model

By analyzing the theories of information, we can get some qualitative conclusion:

- The more value the information has, the further the information spreads.
- The more communication between people, the wider the information spreads.

We find that these characteristics are similar to viruses. On the one hand, the more infectivity of viruses, the more people are infected. On the other hand, the more contact with infected people, the more the patients are. So we found Epidemic Model (SIR) to simulate information spread [4]. According to Epidemic Model, we set three functions as:

$$\begin{cases} s(t) + i(t) + r(t) = 1 \\ N \cdot [i(t + \Delta t) - i(t)] = \lambda \cdot N \cdot s(t) \cdot i(t) \cdot \Delta t - \mu \cdot N \cdot i(t) \cdot \Delta t \\ N \cdot [s(t + \Delta t) - s(t)] = -\lambda \cdot N \cdot s(t) \cdot i(t) \cdot \Delta t \end{cases} \quad (6)$$

Where:

N is the total number of locals.

$i(t)$ is the percentage of the spreaders of information.

$s(t)$ is the percentage of the people who have not gotten information.

$r(t)$ is the percentage of the people who have gotten information, but they aren't interested in it, which means that they couldn't transmit the information to others.

λ is the percentage of the people who are told the information per day.

μ is the percentage of the people who aren't interested in information per day.

We let

$$\begin{cases} \sigma = \lambda / \mu \\ i(0) = i_0 \\ s(0) = s_0 \end{cases} \quad (7)$$

We can derive

$$i(s) = (s_0 + i_0) - s + \frac{1}{\sigma} \cdot \ln\left(\frac{s}{s_0}\right) \quad (8)$$

From these equations, we can get fig.1.

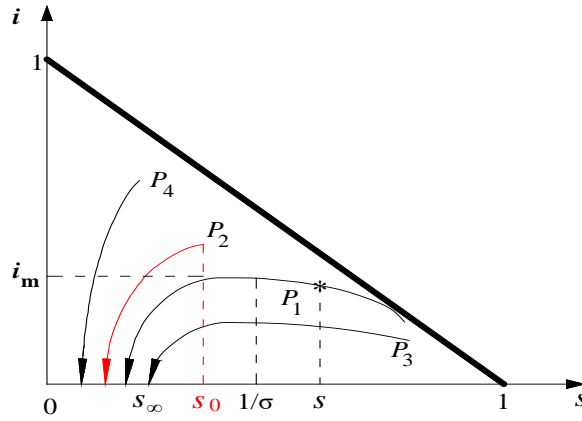


Fig.1 The curve of equation (7)

The figure shows follow results:

- When $s(t) < 1/\sigma$, the value of $i(t)$ will decrease to 0 finally, which means that information can't spread.
- When $s(t) > 1/\sigma$, the value of $i(t)$ will first increase and then decrease to 0, which means that information can spread.
- When $s(t) = 1/\sigma$, the value of $i(t)$ is maximum.

Since information spread is positively correlated to its value, we assume

$$\lambda = k \cdot V. \quad (9)$$

We define that T is the time when $i(t)$ is maximum. So $s(t) = 1/\sigma$ can be rewritten as

$$s(T) = 1 - i(T) - r(T) = \mu / (k \cdot V). \quad (10)$$

Equation (9) shows the relationship between the value and spread of information. We can see from it that if the information has more value, its maximum of spreaders is bigger, which is content with reality.

3. Reliability Analysis

We focus on the propagation process of the information, which is about Volvo Ocean Race happened in 2012, in Tumblr. The data comes from the question in the fifth China's online challenge. According to it, nearly 200,000 persons have read the information after the internet marketer who have 2,000,000 fans pushed. And nearly 25,000 persons have replied, for 60% of which will forward this information. We found the probability of spreading is 0.75% by calculating., and the probability of immunity is 99.25% [6].

3.1 Quantizing the value of the information.

According to the analysis about information's value, we assign the value of index to every second grade indexes through the way of public scoring. We get the value vector format of 11 second grade indexes of about the content value is: [0.91 0.65 0.82 0.88 0.56 0.78 0.66 0.69 0.34 0.52 0.76], and the value vector format of 5 second grade indexes of about the quality of writing is: [0.85 0.87 0.64 0.55 0.42].

Then according to the formula (5): $V = 0.7 \sum w_{1,i} \cdot e_i + 0.3 \sum w_{2,i} \cdot e_i$, we can get this information's value:

$$V = 0.6944 \quad (11)$$

3.2 The solution base on the above models for the question

According to the formula (6), (7), (8) got by the epidemic model, and the formula (11), we can get: $1/\sigma = 0.3216$, which is lower than $s(t)$ in the beginning. It expresses that the percentage of people spreading the information is increasing in the beginning and then reducing, which accords with the truth that the message propagating will be gradually damping after the spread in the highly speed.

According to the character of the information, we can get the coefficient of the spreading k : $k = 0.85$. By the formula (10): $s(T) = 1 - i(T) - r(T) = \mu / (k \cdot V)$, we can get the time T when the percentage of the spreaders of information arrived at max is 1.16 day, and then the spread is reducing. This is according with the character of information's spread. The detailed process is shown roughly as fig.2.

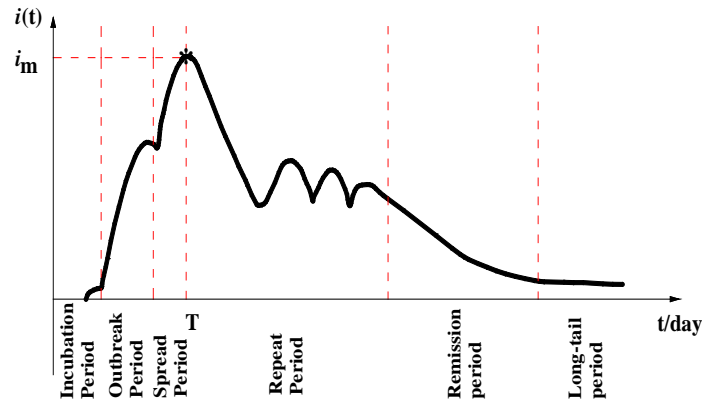


Fig.2 The percentage of spreaders on different time

3.3 Results and analysis

We determined the information value by means of Analytic Hierarchy Process (AHP). We can get the max percentage of the people who have not gotten information $s(T)$ because the maximum of the $i(t)$ is known. And then we plot the fig.3 which have a universal nature:

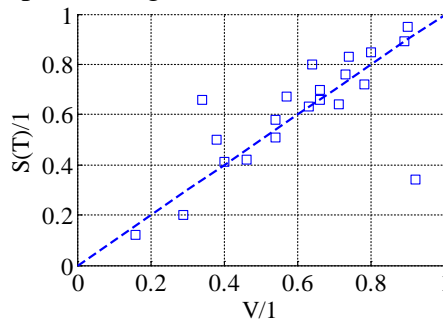


Fig.3 The scatter plot of the relationship between $s(T)$ and V

The correlation coefficient is 0.7161 by computation. So $s(T)$ is positively correlated with V . This is consistent with above conclusion. So we validate that our model is general and reliable.

4. Conclusion

This paper explains how information value influences its spread, and we conclude the following.

- (1) Information value can be determined quantitatively according to its content and writing quality.
- (2) In a certain area, the scope of information spread can be determined approximately by its value.
- (3) Evaluating the spreading process of information by the information value quantification model and the epidemic model have the universality.

References

- [1] Zuo Zhao; Shengbing Zhang, "Research on influence of weak ties to information spreading in online social networks," in *Communications and Information Technologies (ISCIT), 2013 13th International Symposium on*, vol., no., pp.210-215, 4-6 Sept. 2013
- [2] A. Clementi, A. Monti, F. Pasquale and R. Silvestri, "Information Spreading in Stationary Markovian Evolving Graphs," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 9, pp. 1425-1432, Sept. 2011. Chiang Kao, Ralph E. Steuer
- [3] Decai, Huang; Liangzhong, Shen, "New method for constructing comparison matrix based on the proportion scales in the AHP," in *Systems Engineering and Electronics*, Journal of , vol.14, no.3, pp.8-13, Sept. 2003
- [4] Jeonghwan Jin; Rothrock, L.; McDermott, P.L.; Barnes, M., "Using the Analytic Hierarchy

Process to Examine Judgment Consistency in a Complex Multiattribute Task," in *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* , vol.40, no.5, pp.1105-1115, Sept. 2010

- [5] Weijun Hong; Zhipeng Gao; Yuwen Hao; Xiaoxue Li, "A novel SCNDR rumor propagation model on online social networks," in *Consumer Electronics - Taiwan (ICCE-TW), 2015 IEEE International Conference on*, vol., no., pp.154-155, 6-8 June 2015
- [6] . G. Chamberlain, J. Czika, N. D. Miller and G. N. Williams, "The IEEE Oceanic Engineering Society at Forty: The Challenges of an Evolving Society," in *IEEE Journal of Oceanic Engineering*, vol. 33, no. 1, pp. 1-54, Jan. 2008.