

# STUDY ON INDEX MODEL OF COMMUNICATION BASED ON SELF-INFORMATION

○Jun KUMAGAI<sup>1</sup>, Hiroshi KANASUGI<sup>2</sup>,  
Ryosuke SHIBASAKI<sup>3</sup>

Center for Spatial Information Science, The University of Tokyo  
5-1-5, Kashiwanoha, Kashiwa City, Chiba 277-8568, Japan

Tel: +81-4-7136-4290 Fax: +81-4-7136-4292

Email: kumajun@iis.u-tokyo.ac.jp<sup>1</sup>, kumajun@iis.u-tokyo.ac.jp<sup>2</sup>  
shiba@csis.u-tokyo.ac.jp<sup>3</sup>

**KEY WORDS:** Visualization of Human Relationships, Index Model of Communication

**ABSTRACT:** We communicate with each other through various kinds of media every day. In recent years, the communication environment surrounding us has been changing with new type of communication media such as blog, SNS, online games and web sites of sharing contents.

To support for human communications, the study has been focusing on the visualization and analysis of communication or human relationships. However, as for existing studies, so far the study on strength and kinds of connections of human relationships has been superficial.

This study aims to create indexes about strength, kinds and similarity of connection of human relationships.

## 1. 1. INTRODUCTION

We communicate with each other through various kinds of media such as conversation, telephone, mobile mails, and PC mails and so on. The communication environment surrounding us has been changing with new type of communication media such as blog, SNS, online games and website for sharing contents.

While such new services appear, we have some following problems. Although we transmit information through Blog or SNS by ourselves, our data is managed by providers. We cannot manage our data by ourselves. It is difficult to find a person who has same hobbies and diversions in the website. There is no scheme to judge credibility of results of recommendation for myself objectively. We have to set the control of disclosure of information. We can't control detailed disclosure segmentation. In the present circumstances, it can be disclosed by the level (group) of "Friends ONLY" or "ALL".

To support for human communications, the study has been focusing on the visualization and analysis of communication or human relationships

As for studies on visualization of human network, Tyler studied on extraction of human relationships and communities in the company from large amount of mails. Matsuo studied on extraction of human relationships from information on the web site by web mining. Hitachi, Ltd studied on monitoring system named "business microscope" which shows a map about activities and status of communication of employees in the company with sensor technologies. As for study on presumption of strength of connection of human relationships, Nishihara studied on presumption of strength of connection of human relationships from user's words on the website or messenger.

In these studies on communication and human relationships, although some studies have been made on network of human relationship, So far the study on strength and kinds of connections of human relationships has been superficial. Although some studies have been made on analysis of mails or SNS individually, no studies have ever tried to analyze human relationships including various new types of communication media totally. Although most studies have been

made on network in organization such as company or websites, no studies have ever tried to analyze and utilize human relationships focusing on myself.

## **2. OBJECTIVE**

This study aims to lay the foundation for support services for communication by uniform management of your logs of communication, activity, hobby and diversions by yourself. As for support services for communication, we'd like to deal with two types of services. First is support service for communication based on each partners. For example, control your open information automatically, provision of communication log such as date, place, and name and so on, provision of common topics between you and partners, and change of text format or user's dictionary automatically. Second is recommendation services of persons who have valiant information for you and provision of their recommendation. For example, people search with similarity of hobbies and diversions from communication logs and activity logs.

To create these support services, there are some problems if we use your raw data such as your activities logs. It is difficult to manage logs, to compare or match between you and other users and to protect your privacy.

To protect your privacy and measure the actual status and compare between you and other users, we need to create indexes of communication logs. This study aims to create indexes about strength, kinds and similarity of connection of human relationships.

## **3. SELF INFORMATION PLATFORM**

First of all, we created the platform named "Self information platform" to accumulate and manage several communication logs. We can manage activity logs, mail logs, SNS logs, expense logs to each person.

We'll create the index model of communication with the data in the "Self information platform" and map of human relationship with the indexed data.

## **4. DATA ITEM**

As for data items, we used the following items for each communication media.

### **4.1 Face to face communication**

As activity logs, we use people name, date, time, place and category of activity. As information of expense, we use people name, date, time for meetings, place and costs.

### **4.2 Voice communication**

As voice communication logs, we use people name, date, time for speaking and type of mails (receive or sending).

### **4.3 Text communication**

As mobile phone logs, we use people name, date, subject, content number of send mails and number of receive mails. As PC mail logs, we use people name, date, subject, content, number of send mails and number of receive mails. As messenger logs, we use people name, date, content number of send comments and number of receive comments. As SNS logs, we use people name, date, subject, content and number of comments.

### **4.4 Self information**

As self information, we use basic information such as birthday, gender, married, origin, group, blood type and item lists which users have and shop lists which users visited and so on.



Fig.1. Outline of study on index models for communication based on self-information

## 5. INDEX MODEL OF COMMUNICATION

Index model of communication measures strength and kinds of connection of human relationships from various kinds of communication logs such as face-to-face communication, voice communication and text communication.

This study deals with status of connection of human relationship in two aspects. First is aspect of physical quantity of communication. Second is aspect of contents of communication.

- Connection with physical quantity of communication (number of mails, meetings etc.)
- Connection with contents of communication (contents of mails and events etc.)

### 5.1 Detail

#### 5.1.1 Mapping of human relationship with physical quantity of communication

This model considered that status of communication depends on quantity of communication. This model evaluates the status of communication by the following unit of each communication media.

Table1. Unit of each communication media

	Media	Example	Unit
High	Face-to-face Communication	Meeting	Time for meeting
↕	Voice Communication	Telephone, Skype(VoIP)	Time for speaking
	Internet Messenger	MSN Messenger, Chat	One comment
	E-Mail	PC mail, mobile mail	One mail
Low	Web	Blog, HP	One comment

A map of human relationship with indexed data by index model of communication shows status of communication according to the following steps.

- ① Comparison the quantity of indexed data per a week between the all data in the platform and the data in the time span that you want to check.
- ② Position: Mapping with the quantity of the indexed data in the time span
- ③ Direction: Increase or decrease of quantity of indexed data per a week between two time span.
- ④ Size: Rate of quantity of indexed data per a week between two time span

### 5.2.2 Labeling of connection of human relationship with contents of communication

To analyze kinds of connection of human relationship, we use the keywords which are extracted from contents and subjects of mails and events by Japanese language morphological analysis.

- ① We extract only nouns from contents of events, mails and comments by Japanese language morphological analysis.
- ② We count up the number of the same nouns extracted.

## 5.3 Index model of communication for mails

This paper especially deals with PC mails as index of communication.

### 5.3.1 Physical measurement

#### 5.3.1.1 Data items

- Date of receiving and sending mails
- Mail address
- To or Cc or Bcc
- the number of mails

#### 5.3.1.2 Measurement

- [1]. Decision of time span that you want to check  
User decides the time span to check.
- [2]. Communication point: CP  
Index model of communication (IMC) gives one point per one mail.
- [3]. Weights depend on types of mails “To or Cc or Bcc”  
IMC set the following values as the weight coefficient: W.  
To  $\Rightarrow$  1    Cc, Bcc  $\Rightarrow$  0.8
- [4]. Sum of CP in the time span to each person  
 $SCP = \text{SUM}(CP \times W)$
- [5]. Mapping of friends’ icons based on distance “D”  
IMC calculates distances to you “D” in the time span, puts icons on the map based on distance “D” far from you.  
Distance D = 1/SCP
- [6]. Average of SCP  
ACP = SCP per one week
- [7]. Comparison of ACP between two time span  
To calculate the rate of change of quantity of communication, IMC compares ACPs which calculated based on the data in time span with ACPI calculated based on all data.  
 $CCP = \text{ACPs} / \text{ACPI}$   
The status (increase or decrease) of change of quantity are expressed with direction of icons and the quantity of change are expressed with size of icons.
  1. CCP > 1: Towards myself
  2. CCP < 1: Far from myself

### 5.3.2 Measurements of content

#### 5.3.2.1 Data items

- Subject

We tried to extract keywords from contents of mail too. However as for PC mails, because users often write the subject of mails, it is enough to understand the content of mail to check only the subject. We deal with only the subject of PC mails.

#### 5.3.2.2 Measurement method

- [1]. Analysis of subjects by Japanese language morphological analysis

We extract only nouns from subjects of mails with Yahoo API of Japanese language morphological analysis.

- [2]. Count up the same nouns

We count up the same keywords in the extracted nouns from the subjects to each person. IMC labels connections of human relationship with the keywords extracted.

## 6. RESULTS

### 6.1 Sample data

We use the PC mails (receiving and sending) from 1<sup>st</sup>, October, 2006 to 31<sup>st</sup>, December, 2006 as sample data.

### 6.2 Results

As for the time span to check the status of communication, we use 16<sup>th</sup>, November, 2006 to 31<sup>st</sup>, December, 2006.

Table2 shows the results of strength of connection and rate of change of strength calculated from the sample data

Table.2 Results with PC mails

	Receive(To)	Receive(Cc)	Send(To)	Send(Cc)	Strength	Change
User1	14	4	5	1	3.57778	0.883809
User2	5	2	5	0	1.80444	0.946752
User3	3	2	2	1	1.15111	0.603962
User4	0	1	0	0	0.124444	0.086819
User5	4	1	1	2	1.15111	0.869994
. . .						
User66	0	0	0	0	0	0

Figure.2 shows results of mapping the status of communication with the indexed data. In this figure, IMC put the icon stands for myself in the center of the map and put the humanoid icon as friends with the indexed data.

Position, direction and size of friends' icons show the following.

- Distance from myself and friends' icons : Distance D

- Direction of friends' icon :

Increase or decrease of change of quantity of communication.

CCP > 1 : Inside

CCP < 1 : Outside

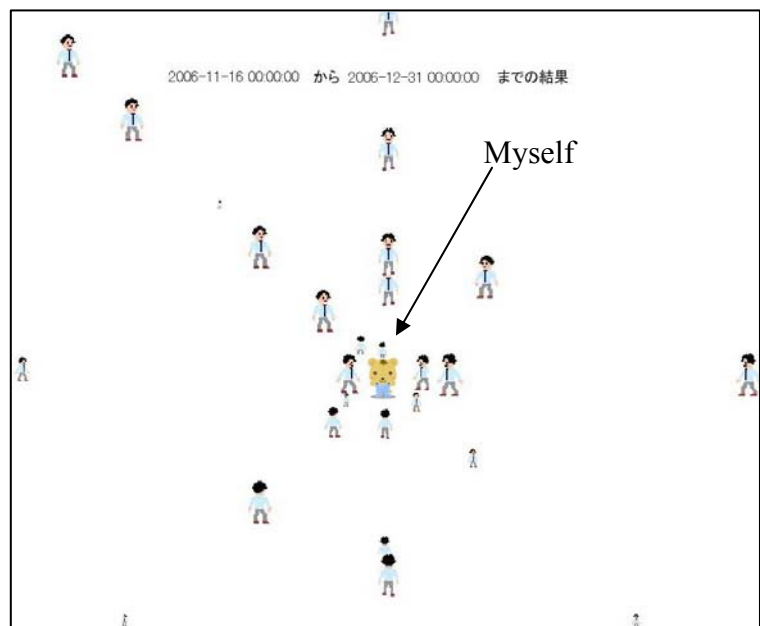


Fig.2. Result of status of communication with PC mails

- Size of friends' icons : quantity of change of indexed data.

If we click the friends' icons, we can know the keywords of the connection like figure.3.

As for keywords, important keywords which user often talks about are expressed by big fonts.

In this way, we can understand the status of communication and kinds and contents and strength of connection of human relationships with the indexed data by index model of communication. It is considered to be possible to support the communication based on this.

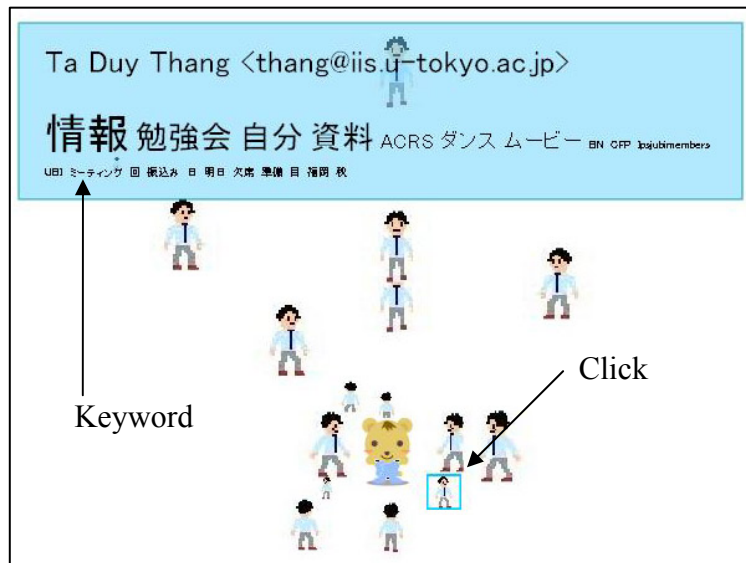


Fig.3. Status and keywords of communication with PC mails

## 7. FUTURE WORK

As for future works, we try to study on the following problems.

- Validation for index model of communication
- Adaptation of index model of communication to other communication media (Face-to-face communication, voice communication and text communication such as SNS, messengers and so on.) and validation.
- Grouping friends by keywords and kinds of connection.

## 8. REFERENCES

- [1] Joshua R.Tyler, Email as Spectroscopy: Automated Discovery of Community Structure within Organizations, 2003
- [2] Yutaka Matsuo, Junichiro Mori, Masahiro Hamasaki, Keisuke Ishida, Takuichi Nishimura, Hideaki Takeda, Koiti Hasida, and Mitsuru Ishizuka: POLYPHONET: An Advanced Social Network Extraction System, *Proc. 15th International World Wide Web Conference (WWW2006)*, (2006.5)
- [3] Hitachi, Ltd., <http://www.hitachi.co.jp/New/cnews/month/2007/06/0622.html>, Jun.22, 2007
- [4] Yoko Nishihara, Wataru Sunayama, Masahiko Yachida : Relationship Activity Estimation by Using Utterance Intention in Conversation on the Web, The 20th Annual Conference of the Japanese Society for Artificial Intelligence, 2006