

Conversational Knowledge Process for Social Intelligence Design

Toyoaki Nishida

Dept. of Intelligence Science and Technology,
Graduate School of Informatics, Kyoto University,
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
nishida@i.kyoto-u.ac.jp
<http://www.ii.ist.i.kyoto-u.ac.jp/~nishida/>

Abstract. The Internet and ubiquitous network technologies have succeeded in connecting people and knowledge over space and time. The next step is to realize knowledgeable communities on the ubiquitous network. Social Intelligence Design is a field of research on harmonizing people and artifacts by focusing on social intelligence, defined as the ability of actors and agents to learn and to solve problems as a function of social structure and to manage their relationships with each other. In this paper, I present a computational approach to understanding and augmenting the conversational knowledge process that is a collective activity for knowledge creation, management, and application where conversational communications are used as a primary means of interaction among participating agents. The key idea is conversation quantization, a technique of approximating a continuous flow of conversation by a series of conversation quanta that represent points of the discourse. Conversation quantization enables to implement a rather robust conversational system by basing it on a large amount of conversational quanta collected from the real world. I survey major results concerning acquisition, annotation, adaptation, and understanding of conversation quanta.

1 Introduction

The Internet and ubiquitous network technologies have succeeded in connecting people and knowledge over space and time. In order for networked people and knowledge to be creative, proper communication functions for supporting the community knowledge process need to be implemented, for communities play a central role in knowledge creation.

Social Intelligence Design is a field of research on harmonizing people and artifacts by focusing on social intelligence, defined as the ability of actors and agents to learn and to solve problems as a function of social structure and to manage their relationships with each other [1].

Social Intelligence Design aims to integrate understanding and designing social intelligence. Engineering aspects of Social Intelligence Design involve design and implementation of systems and environments, ranging from group collaboration support systems that facilitate common ground building, goal-oriented interaction

among participants, to community support systems that support a large-scale online discussions. Scientific aspects involve cognitive and social psychological understanding of social intelligence, attempting to provide a means for predicting and evaluating the effect of a given communication medium on the nature of discussions, interaction dynamics, and conclusions.

Our approach places a particular emphasis on the use of conversational communications in the knowledge process, for conversation is the most natural and effective means for people to manifest their social intelligence such as heuristically producing stories from different points of view, making tacit-explicit knowledge conversion, and entraining participants to the subject.

In this paper, I present a computational approach to understanding and augmenting the conversational knowledge process that is a collective activity for knowledge creation, management, and application where conversational communications are used as a primary means of interaction among participating agents.

The key idea is conversation quantization, a technique of approximating a continuous flow of conversation by a series of conversation quanta that represent points of the discourse. Conversation quantization enables to implement a rather robust conversational system by basing it on a large amount of conversation quanta collected from the real world. I survey major results concerning acquisition, annotation, adaptation, and understanding of conversation quanta.

2 Conversation Quantization

Conversation Quantization is a technique of articulating a continuous flow of conversation by a series of objects called conversation quanta each of which represents a point of the discourse. Conceptually, it consists of extraction, accumulation, processing, and application of conversation quanta (Fig. 1). The extraction of conversation quantum results from identification and encoding of coherent segments of interactions in a conversational situation. The extracted conversation quanta are accumulated in a server, processed whenever necessary, and applied to other conversational situations. The application of a conversation quantum in a target situation involves production of conversational sequence or other form of presenting the content of information stored in the conversation quantum.

Conversation quantization allows for implementing a conversation system by reusing a collection of conversation quanta gathered from real/hypothetical conversation situations. Given a conversational situation, a conversation quantum that best matches it will be sought from the collection of conversation quanta, and one role of the participants of the retrieved conversation quantum can be replayed by an embodied conversational agent, and other roles will be mapped to the participants in the given conversational situation. Such an algorithm is relatively easy to implement and rather robust in nature.

The granularity and size of conversation quanta essentially depend on the context and background knowledge of the observer. Although the detailed investigation of the nature of conversation quantization is left for future, we conjecture, based on experiments made so far, that each conversation quantum roughly corresponds to a small talk often identified in the discourse of daily conversations.

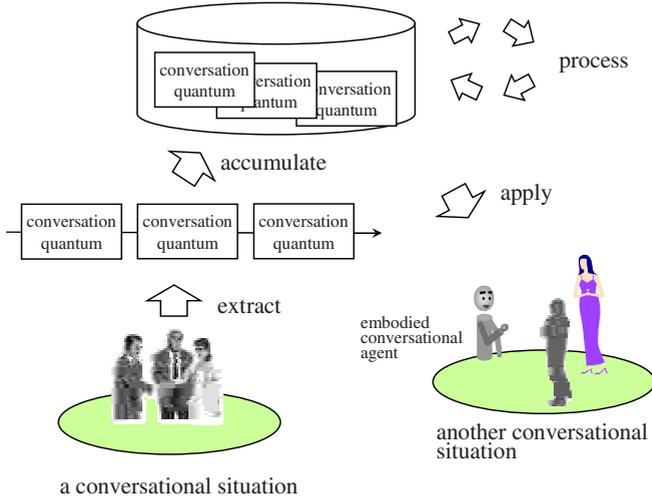


Fig. 1. Conversation quantization -- the concept

The implementation of conversation quantization depends on the data structure for representing conversation quanta. One could use plain video clips as representation but its utility in retrieving and processing would be quite limited and a large cost would be required in retrieving, editing, and applying conversation quanta. Alternatively, a deep semantic representation using logical formulas or case frames would not be ideal due to the expense and their limited capability of representing nonverbal information. A reasonable implementation appears to use annotated videos and images to represent a conversation quantum.

The current focus of our research is on acquisition, annotation, adaptation, and understanding of conversation quanta (Fig. 2). The simplest method of acquiring conversation quanta is manual input by an external observer. Alternatively, a more sophisticated method of extraction may involve automatic identification of a coherent segments of interactions from a conversational situation, and automatic annotation to the objects and events. Conversation quanta may well be generated from other media such as written documents using natural language processing techniques.

There are generally two methods of presenting the content of a conversation quantum. One is temporal expansion that generates a temporal sequence of information presentation actions. By dynamically switching among multiple conversation quanta upon the other participants' utterances, we can give the system an interactive flavor. An alternative one is spatial expansion that generates information landscape which is a spatial representation of the content of conversation quantum. An information landscape enables the user to visually grasp the global nature of knowledge, explore the information space, and accommodate new information at an appropriate place.

It is critical for us to establish a means for evaluating the effect of information tools in the user community. Social Intelligence Quantity (SIQ) is a measure of intelligence collectively possessed by a group of people. With a proper characterization of SIQ,

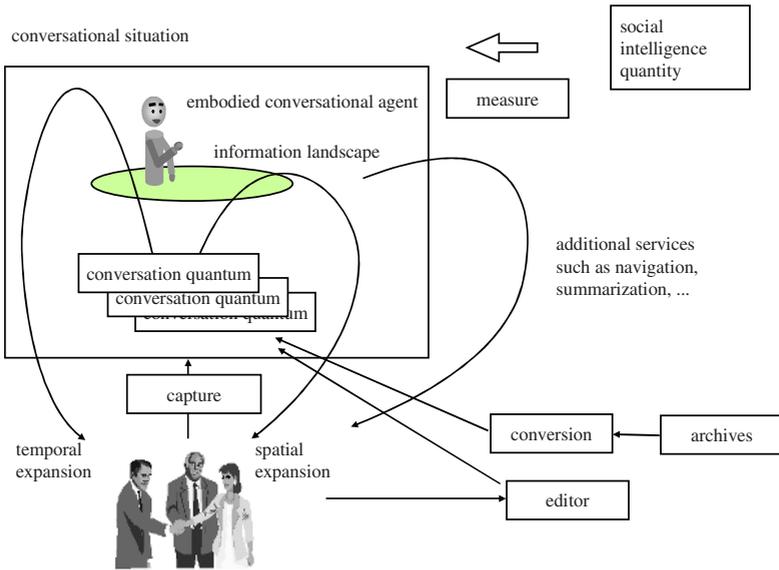


Fig. 2. A general architecture of a conversational system based on conversation quantization

the effect of a given information tool can be measured as a difference of the SIQ of the users for before and after the information tool is applied.

In the following sections, I overview the systems implemented the idea of based on conversation quantization.

3 EgoChat

EgoChat is a system for enabling an asynchronous conversational communication among community members [2,3]. EgoChat implements both temporal and spatial expansion of conversation quanta (Fig.3).

In EgoChat, a simple implementation of conversation quantization using data structures called knowledge cards, stories and knowledge channel was employed. A knowledge card is relatively self-contained package of tacit and explicit knowledge, enabling one to implement a rather robust conversational system without introducing a complex discourse generation algorithm (Fig.4). A story is a sequence of knowledge cards, representing a complete story based on a plot. In a story, preceding and succeeding knowledge cards give a discourse to each knowledge card.

EgoChat is based on the talking-virtualized-egos metaphor. A virtualized ego is a conversational agent that talks on behalf of the user. Each virtualized ego stores and maintains the user's personal memory as a collection of knowledge cards and presents them on behalf of the user. It helps the user not only develop her/his personal memory but also better understand other members' interest, belief, opinion, knowledge, and way of thinking, which is valuable for building mutual understanding. We use a powerful dialogue engine that permits a virtualized ego to answer questions by search-

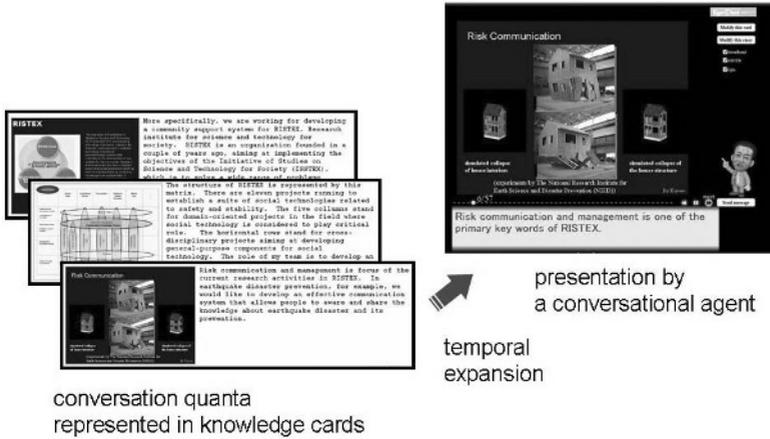


Fig. 3. Implementation of conversation quantization theory in EgoChat

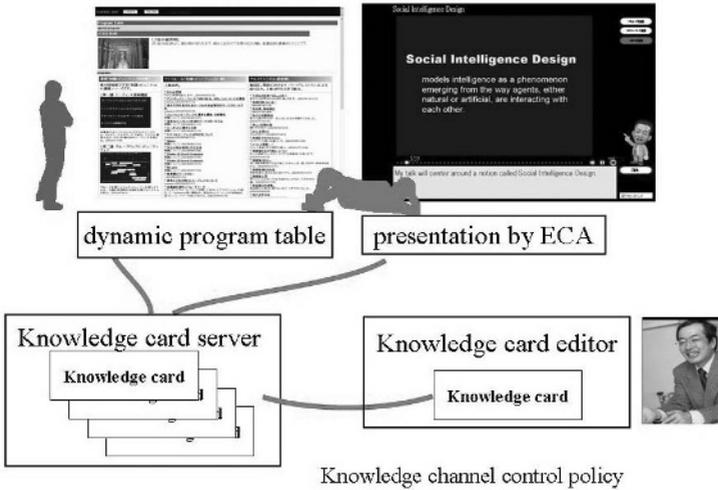


Fig. 4. The architecture of the EgoChat system

ing for the best match from a potentially large list of question-answering pairs prepared in advance.

EgoChat provides a couple of unique features. First, it integrates personal and interpersonal knowledge life cycles. At earlier stages of the lifecycle when knowledge is not well captured, the user might want to describe her/his idea as one or more knowledge card and have the virtualized ego present them for personal review. After a while when the knowledge becomes clear and evident to the user, s/he might want to delegate her/his virtualized ego to her/his colleagues to present the idea and ask for critiques. The automatic question answering facility of EgoChat will encourage the

audience to ask questions or give comments. The virtualized ego can reproduce the question-answering session with the audience so the owner can review the interactions. It will highly motivate the owner to supply more knowledge or improve existing knowledge for better competence.

Second, EgoChat allows for specifying a channel policy that is used to define the control strategies of the sender and the receiver. Four types of strategies are identified depending on whether the strategy is about the order of programs in a single program stream or about the way multiple program streams are mixed, and whether the program scheduling is static or dynamic. The skeleton of the actual flow structure of knowledge cards for a given pair of the sender and receiver is determined by resolving constraints of their channel policies. It can be visually presented to the user by a dynamic program table.

4 The Sustainable Knowledge Globe

A persistent memory system is an approach to spatio-temporal expansion of conversation quanta. By establishing a long-term relationship with a persistent memory that can coevolves with the user's biological memory, the user, as we believe, will be able to find easily an appropriate place to accommodate new information and make it ready for later use.

We are developing a system called the Sustainable Knowledge Globe (SKG) that permits the user to build her/his own customizable intellectual world by spatially arranging information (Fig. 5) [4].

The user can move around on the surface of the globe in search for interesting items and create/reconfigure landmarks consisting of a visual image and a text for later reference.

Presentation by an embodied conversational agent is a temporal representation of knowledge in the sense that the major axis behind the representation is a temporal evolution of a story. In contrast, the configuration of knowledge items on the surface

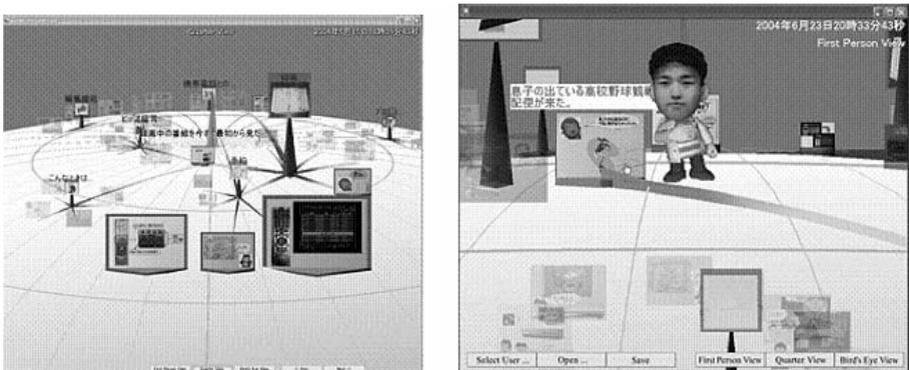


Fig. 5. Screenshots of Sustainable Knowledge Globe (SKG). A quarter view of the globe on the left, and the first person view on the right

of the globe is a spatial representation. SKG allows the user to switch between temporal representation that allows for in-depth causal understanding of the issue and spatial representation that facilitates global and geometric understanding of the subjects.

5 Capturing Conversation Scenes by Multiple Sensors

A sophisticated method of capturing conversation quanta in the real world environment is proposed by Sumi et al [5]. They implemented a smart room environment where conversational activities can be captured by environment sensors (such as video cameras, trackers and microphones ubiquitously set up around the room) and wearable sensors (such as video cameras, trackers, microphones, and physiological sensors). In order to supplement the limited capability of sensors, LED tags (ID tags with an infrared LED) and IR tracker (infrared signal tracking devices) are used to annotate the audio/video data with the positional information. Significant intervals or moments of activities are defined as interaction primitives called "events". Currently, five event types are recognized, including "stay", "coexist", "gaze", "attention", and "facing" (Fig.6). Events are captured by the behavior of IR trackers and LED tags in the room. For example, a temporal interval will be identified as a (joint) attention event when an LED tag attached to an object is simultaneously captured by IR trackers worn by two users, and the object in focus will be marked as a socially important object during the interval.

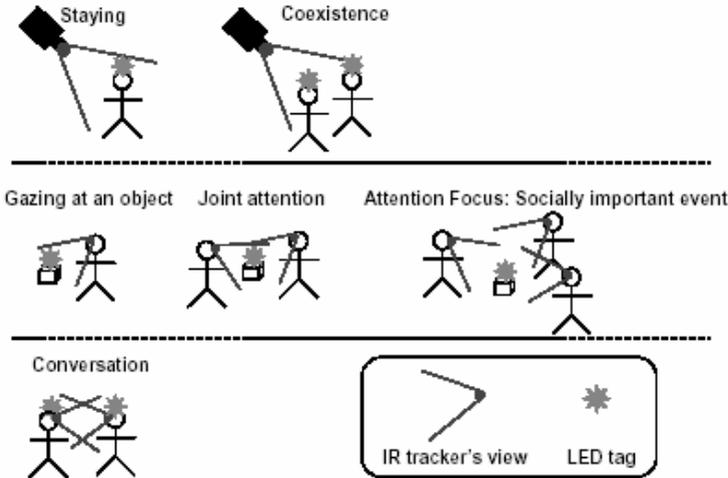


Fig. 6. Interactive Primitives [5]

They also invented a novel method of spatially displaying captured video clips called Spatio-Temporal Video Collage. A virtual 3D space is used as a medium for re-experiencing the conversations in the environment captured from multiple view-

formation obtained by accessing the TTS. The output from the ABS is a set of animation instructions that can be interpreted and executed by the RISA animation system.

As an optional function in CAST, we have also built a component called the Enhancement Generator that automatically adds highlighting animations to agent gesture animations after the gestures are determined by the ABS. Two types of highlighting methods are incorporated to emphasize synchronization between verbal (speech) and nonverbal (gesture) behaviors.

One is superimposition with beat gesture. Beat gestures simply emphasize one part of an utterance without representing the meaning of a word. To visualize synchronization between the emphasized words and a beat gesture, the Enhancement Generator adds a superimposition of the emphasized words to the agent's beat gesture animation.

The other is illustrative animation with metaphoric gesture. When a specific shape of gesture is assigned to a metaphoric gesture, it will be emphasized by illustrative animations, such as an arrow and a line. If the emphasized concept implies motion or movement, such as "increase" or "decrease," the direction of the movement will be illustrated by an arrow animation. If the emphasized concept expresses a static state, a motionless picture will be used to emphasize the gesture. For example, when the agent is performing a "long" gesture, a rectangle shape is shown near the agent's hands to emphasize the length.

7 IPOC

IPOC (Immersive Public Opinion Channel) [8,9] is a successor of SPOC. IPOC allows for expanding conversation quanta in a virtual immersive environment. Users can interact with conversational agents in a story-space, which is a panoramic picture background and stories are embedded in the background (Fig.8). The embedded stories are presented on demand by the user or spontaneously according to the discourse. Stories are used to represent a discourse structure consisting of more than one conversation quantum.

In order to generate a more complex set of nonverbal behaviors in an immersive environment, theories of nonverbal communication are extensively studied and incorporated in the agent behavior generation system. Sidner proposed conversation management, collaboration behavior, and engagement behaviors as communicative capabilities required for collaborative robots [10]. Conversation management includes abilities of turn taking, interpreting the intentions of participants in the conversation, and updating the state of the conversation. Collaboration behavior determines agent's next action in order to accomplish the goal for the conversation and the collaboration with the user. Engagement behaviors consist of initiating a collaborative interaction, maintaining the interaction, and disengaging from the interaction.

The Interaction Control Component (ICC) of the IPOC system interprets inputs from a speech recognizer and a sensor system, and generates verbal and nonverbal behaviors performed by conversational agents. The major components of ICC are the Conversation Manager (CM) that maintains the history and the current state of the conversation, the Collaboration Behavior Generation Module (CBG) that selects the next Card to be read and determines agents' behaviors in telling a story, and Engage-

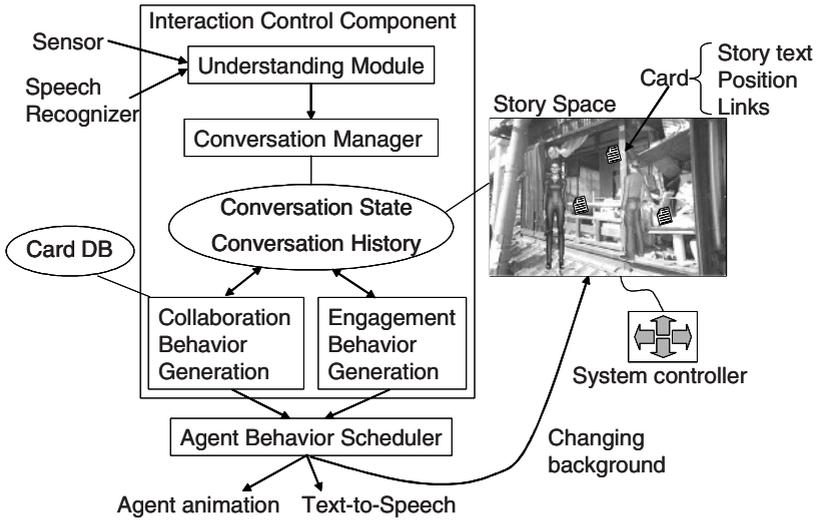


Fig. 8. The architecture of IPOC [9]

ment behavior generation (EBG) that determines appropriate engagement behaviors according to the state of the conversation.

8 Weblog Analyzer

Fukuhara develops a Weblog analyzer for understanding the social concern using Weblog articles [11]. The technique might be applied to extracting trends from a large collection of dynamically changing conversation quanta.

The Weblog analyzer automatically collects Weblog articles and displays the recent trends of social concern as a temporal transition pattern of the number of Weblog articles that co-occur with representative keywords. The Weblog analyzer consists of a database, an httpd server, and several Perl scripts for collecting and retrieving Weblog articles. The Weblog analyzer collects RSS (RDF Site Summary¹) files of Weblog sites. An RSS file contains the title, the summary, the date of publish, the author, and the category of articles. The current system is programmed to collect RSS files every 10 minutes, from personal Weblog sites, news sites, and governmental Web sites. The current Weblog analyzer started to collect articles from 18 March 2004, and acquired more than a million articles by July 12, 2004.

The temporal behaviors of the observed social concern can be classified into five categories: periodic (several peaks appear periodically), gradual increase (a peak appears gradually), sensitive (has a keen peak), trailing (the social concern persists after one or several matters occur), and others (Fig.9).

¹ <http://web.resource.org/rss/1.0/spec>

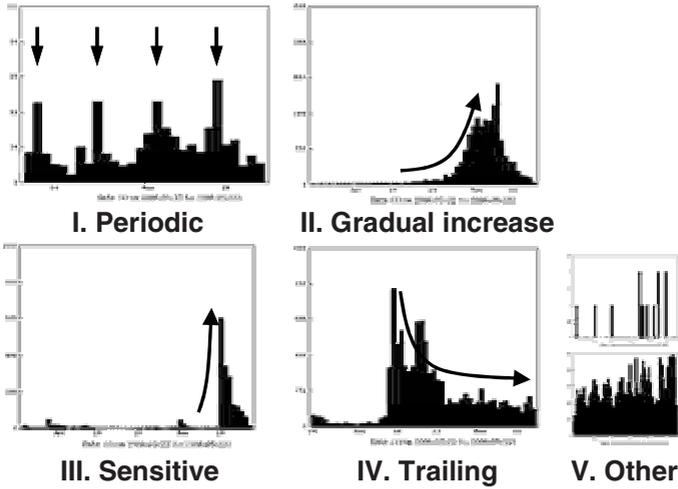


Fig. 9. Patterns of social concern transition over time [11]

9 Social Intelligence Quantity

Social Intelligence Quantity (SIQ) is a framework of standardized quantitative measurement of social intelligence [12]. Our approach combines the qualitative evaluation consisting of questionnaire and protocol analysis, and the quantitative evaluation consisting of the network log analysis, the factorial experiment and the standardized psychological scale. SIQ was first introduced as a means for evaluating communication tools. SIQ consists of SIQ-Personal and SIQ-Collective.

SIQ-Personal specifies the individual’s personal attitudes to the society. SIQ-Personal is measured with the individual’s information desire and intention to participate in the community. Matsumura investigated in detail the information desire and identified that the information acquisition desire consists of the interpersonal relation desire, the trend information acquisition desire, the information publication desire, the information monopoly desire, and the information acquisition desire [13].

Matsumura also investigated the structure of the individual’s intention to participate in the community, in terms of the estimate of the community, the evaluation of tools, and the intention. He introduced seven factors (i.e., the intention of active participation, the intention of continuous participation, the benefits of using tools, the interest in the tool, the clarity of others’ idea, the contribution to their community, and the understanding of the community’s status) and proposed a causal model based on the statistic analysis of the test users’ subjective evaluation (Fig. 10).

These results indicated that the intention to participate in a community was influenced by the benefit received by using a communication tool. The benefit was affected by factors associated with understanding the state of a community. The individual’s subjective contribution had a strong effect on the benefits of using a tool. The

individual's subjective contribution depends on the understanding of the state of a community and the clarity of others' opinion. Thus, in order to evaluate a communication tool for supporting a community, we should examine not only users' subjective evaluation of a communication tool but also their subjective evaluation of a community.

SIQ-Collective represents the community's status of information and knowledge sharing or knowledge creation. SIQ-Collective is measured with the amount of information in the community, and the objective indices such as the diversity and the convergence of information.

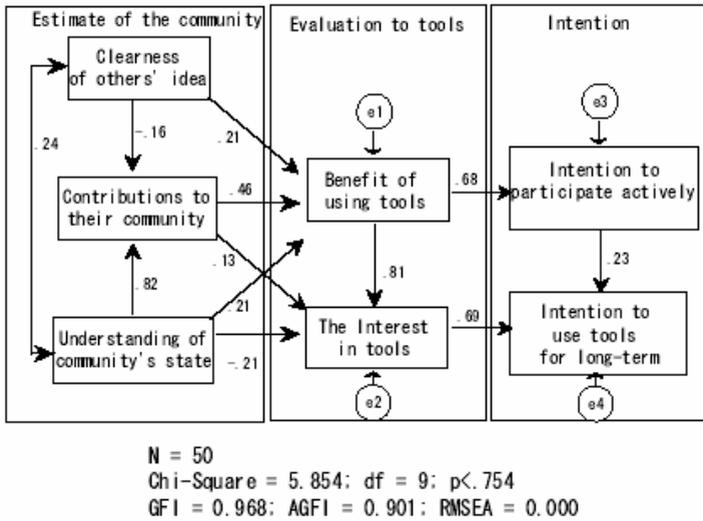


Fig. 10. A causal model concerning the intention of participating in a community and using a tool [13]

10 Discussion and Future Perspectives

I propose to take a data-driven approach by introducing a conceptual framework based on conversation quantization, rather than resorting to an algorithmic-centric framework. I believe a conceptual framework is suitable for capturing different approaches employing different computational frameworks with the same orientation. By basing the entire framework on a vague notion of quantization, it is possible to organize pieces of research into a rather comprehensive framework of conversation systems even though the author of each piece does not have an explicit image of the whole. In the previous sections, I have surveyed representative work on acquisition, annotation, adaptation, and understanding of conversation quanta, in attempt to integrate a vast varieties of attempts regarding conversational systems in a coherent framework.

There are many interesting work left for the future research. Among others, we need to build a more detailed and elegant theory of conversation quantization. We have already obtained an empirical characterization of conversation quanta, but we do not have a systematic and theoretical account of conversation quanta. A more sophisticated theory of conversation quanta will permit us to better design the representation and basic operation for conversation quanta. It may well enable us to predict the cost and effect of building a conversation system based on conversation quantization.

Another big concern is automatic capture of conversation quanta. In addition to the work described so far in this paper, there are a number of interesting work. Minoh and Kakusho developed a robust method for automatically recognizing communicative events in a real class room by integrating audio and visual information processing [14]. Rutkowski studied a method for monitoring and estimating efficiency of the human-human communication based on recorded audio and video by computing correlation between sender activities and receiver responses [15]. Taniguchi and Arita used computer vision techniques to realize the notion of real time human proxy to virtually create a classroom for distributed learning environments [16]. Kurohashi and Shibata integrate robust natural language processing techniques and computer vision to automatically annotate videos with closed caption [17]. Nakamura developed an intelligent video production system that can compute the best matching between the given index-scenario and video stream captured by camera to produce an annotated video [18]. Those work should be incorporated into the theory of conversation quantization.

11 Conclusion

In this paper, I have presented a computational approach to understanding and augmenting the conversational knowledge process that is a collective activity for knowledge creation, management, and application where conversational communications are used as a primary means of interaction among participating agents. I have introduced the notion of conversation quantization, a technique of approximating a continuous flow of conversation by a series of conversation quanta that represent points of the discourse. Conversation quantization enables to implement a rather robust conversational system by basing the conversational system the large amount of conversational quanta collected from the real world. I have surveyed major results concerning acquisition, annotation, adaptation, and understanding of conversational quanta.

References

1. B T. Nishida. Social Intelligence Design for Web Intelligence, Special Issue on Web Intelligence, IEEE Computer, Vol. 35, No. 11, pp. 37-41, November, 2002.
2. H. Kubota, T. Nishida: Channel Design for Strategic Knowledge Interaction, in Proceedings KES 2003, pp. 1037-1043, 2003.
3. Hidekazu Kubota, Jaewon Hur, and Toyoaki Nishida: Agent-based Content Management System, in Proceedings of the 3rd Workshop on Social Intelligence Deisgn (SID 2004), CTIT Workshop Proceedings, pp. 77-84, 2004.

4. Toyoaki Nishida, Yasuyuki Sumi, Hidekazu Kubota, Hung-Hsuan Huang: A Computational Model Of Conversational Knowledge Process, to be presented at 1st International Workshop on "Intelligent Media Technology for Communicative Intelligence", affiliated with 4th National Conference on Multimedia and Network Information Systems, Szklarska Poreba, Poland, September 16-17, 2004.
5. Yasuyuki Sumi, Kenji Mase, Christof Mueller, Shoichiro Iwasawa, Sadanori Ito, Masashi Takahashi, Ken Kumagai, Yusuke Otaka: Collage of Video and Sound for Raising the Awareness of Situated Conversations, to be presented at International Workshop on Intelligent Media Technology for Communicative Intelligence in Warsaw, Poland September 13-14, 2004
6. Nakano, Y. I., Murayama, T., and Nishida, T.: Multimodal Story-based Communication: Integrating a Movie and a Conversational Agent. Vol.E87-D No.6 pp.1338-1346 2004/6.
7. Q. Li, Y. Nakano, M. Okamoto, and T. Nishida: Highlighting Multimodal Synchronization for Embodied Conversational Agent, the 2nd International Conference on Information Technology for Application (ICITA 2004), 2004.
8. Yukiko I. Nakano, Masashi Okamoto, and Toyoaki Nishida: Enriching agent animation with Gestures and Highlighting Effects, to be presented at International Workshop on Intelligent Media Technology for Communicative Intelligence in Warsaw, Poland September 13-14, 2004
9. Yukiko I. Nakano, Toshiyasu Murayama, and Toyoaki Nishida: Engagement in Situated Communication By Conversational Agents, to be presented at 1st International Workshop on "Intelligent Media Technology for Communicative Intelligence", affiliated with 4th National Conference on Multimedia and Network Information Systems, Szklarska Poreba, Poland, September 16-17, 2004.
10. Sidner, C. L., C. Lee, and N. Lesh: Engagement Rules for Human-Robot Collaborative Interactions, in Proc. IEEE International Conference on Systems, Man & Cybernetics (CSMC), Vol. 4, pp. 3957-3962, 2003
11. Tomohiro Fukuhara and Toshihiro Murayama: An Analysis Tool for Understanding Social Concerns using Weblog articles, to be presented at 1st International Workshop on "Intelligent Media Technology for Communicative Intelligence", affiliated with 4th National Conference on Multimedia and Network Information Systems, Szklarska Poreba, Poland, September 16-17, 2004.
12. Koji Yamashita and Toyoaki Nishida: SIQ (Social Intelligence Quantity): Evaluation Package for Network Communication Tools, APCHI 2002 -- 5th Asia Pacific Conference on Computer Human Interaction - Beijing, China, 1-4 November 2002.
13. Ken'ichi Matsumura: The Measures for the Evaluation of Communication Tools: the Causality between the Intention and Users' Subjective Estimation of Community, in Proceedings of the 3rd Workshop on Social Intelligence Design (SID 2004), CTIT Workshop Proceedings, pp. 85-90, 2004.
14. Minoh, S.Nishiguchi:"Environmental Media - In the Case of Lecture Archiving System," Proc. Int. Conf. Knowledge-Based Intelligent Information & Engineering Systems (KES2003), Vol.II, pp.1070-1076 (2003).
15. Rutkowski, M., Seki, S., Yamakata, Y., Kakusho, K., and Minoh, M.: "Toward the Human Communication Efficiently Monitoring from Captured Audio and Video Media in Real Environment," Proc. Int. Conf. Knowledge-Based Intelligent Information & Engineering Systems (KES2003), Vol.II, pp.1093-1100 (2003).

16. Arita, D. and Taniguchi, R.: Non-verbal Human Communication Using Avatars in a Virtual Space, in Proc. Int. Conf. Knowledge-Based Intelligent Information & Engineering Systems (KES2003), pp. 1077-1084, Sep. 2003.
17. Shibata, T., Kawahara, D., Okamoto, M., Kurohashi, S., and Nishida, T.: Structural Analysis of Instruction Utterances, in Proceedings KES 2003, pp. 1054-1061, 2003.
18. Ozeki, M., Izuno, H., Itoh, M., Nakamura, Y., and Ohta, Y.: Object Tracking and Task Recognition for Producing Interactive Video Content --- Semi-automatic indexing for QUEVICO, in Proc. Int. Conf. Knowledge-Based Intelligent Information & Engineering Systems (KES2003), pp.1044-1053, 2003