Collective Intelligence and Uncertain Knowledge Representation in Cloud Computing

Liu Yuchao\textsuperscript{1,2}, Zhang Haiyu\textsuperscript{3}, Ma Yutao\textsuperscript{2,4}, Li Deyi\textsuperscript{1}, Chen Guisheng\textsuperscript{2}
\textsuperscript{1}Department of Computer Science and Technology, Tsinghua University, Beijing 100084, P. R. China
\textsuperscript{2}Institute of Electronic System Engineering, Beijing 100141, P. R. China
\textsuperscript{3}College of Command Automation, PLA University of Science and Technology, Nanjing 210007, P. R. China
\textsuperscript{4}State Key Lab of Software Engineering, Wuhan University, Wuhan 430072, P. R. China

Abstract: The lasting evolution of computing environment, software engineering and interaction methods leads to cloud computing. Cloud computing changes the configuration mode of resources on the Internet and all kinds of resources are virtualized and provided as services. Mass participation and online interaction with social annotations become usual in human daily life. People who own similar interests on the Internet may cluster naturally into scalable and boundless communities and collective intelligence will emerge. Human is taken as an intelligent computing factor, and uncertainty becomes a basic property in cloud computing. Virtualization, soft computing and granular computing will become essential features of cloud computing. Compared with the engineering technological problems of IaaS (Infrastructure as a service), PaaS (Platform as a Service) and SaaS (Software as a Service), collective intelligence and uncertain knowledge representation will be more important frontiers in cloud computing for researchers within the community of intelligence science.

Key words: collective intelligence; social annotation; common sense; uncertain knowledge representation;

I. INTRODUCTION

In recent years, Internet has made great development, and soaked into other fields deeply. George Gilder’s Law predicts that the growth speed of bandwidth is even 3 times more than Gordon Moore’s Law of the performance of computer chip [1]. Information technology development inclines to depend on the Internet, social development and people’s life also incline to rely on it. Since 2000, the development of Web Services, semantic web and Web 2.0 has accelerated the formation of Internet computing environment. The computing environment shifts from Turing Machine to the Internet, domain oriented and service oriented net-worked software replaces the traditional PC software, the human machine interaction mode transforms from “computer centric interaction” to “human centric interaction”. In short, the lasting evolution of computing environment, software engineering and interaction mode, has gradually fostered cloud computing.

Cloud computing is a paradigm of Internet com-
puting in which dynamical, scalable and virtualized resources are provided as services. In 2007, IBM firstly brought forward the concept of “cloud computing”, then, many companies pushed out their plans for cloud computing. Just like in industrialization age, manufacturing is turned into social common infrastructure with socialization, intensive management and specialization, nowadays, cloud computing is making information technology and information service be social common infrastructure. With virtualization technology, cloud computing can offers all kinds of information services for the public by means of multi tenancy mode. As we known, SaaS, PaaS and IaaS are three main service types in cloud computing.

When all companies are paying more attention to the services and products in cloud computing, as a researcher in the field of intelligence science, we must have an idea of participation, interaction and collective intelligence in cloud computing. If handled appropriately, data about Internet based communication and interactivity could revolutionize our understanding of collective human behaviors [2].

The rest of this paper is organized as follows: Section II introduces the change of Internet resources configuration and human machines interaction mode. In Section III, we point out social annotation is an important way of human interaction on the Internet. Uncertainty is an important feature in cloud computing, and Section IV introduces an uncertainty representation model of common sense. Section V is the conclusion.

II. CHANGES OF INTERNET RESOURCES CONFIGURATION AND HUMAN MACHINE INTERACTION MODE

2.1 Internet resources configuration evolution

Over the past forty years, Internet has shown a remarkable ability to adapt to changes in technologies and applications. From Client/Server mode to virtual service, we can review the process of Internet resources configuration evolution (see Figure 1). In early Client/Server mode, client and server together share in disposal missions and application software includes two parts. In Browser/Server mode, client is reduced into a unified browser and more work is assigned to server. With the increasing of applications on the Internet, lots of companies and departments set up their own servers, then this causes the blowout of the type and number of servers, such as mail server, data server, security server and media server, etc. Therefore, many Internet application companies are bothered by the management and maintenance of thousands of servers, and server hosting market arises at the historic moment. However, the utilization rate of servers is still less than 15%, and how to reduce cost and to provide high utilization rate of resources becomes the main problem for server hosting companies. In order to solve this problem, servers are integrated and virtualized, virtual computing and storage units instead of servers and hard disks are provided. For example, the utilization rate of Google clusters can reach 85%. It is convinced that cloud computing provides a mode to optimize Internet resources configuration and improve their utilization rates.

![Fig.1 Internet resources configuration evolution](image)
Cloud computing provides virtualized and abundant on-demand data centers, software download and maintenance centers, computing centers and security services centers. Under the rules of the Internet and network science [3], after a period of technology accumulation, cloud computing may become a new revolution after the PC revolution in 1980s and Internet revolution in 1990s. In cloud computing, users buy services and pay for use, while needn’t buy software and hardware resources, implying that they do not care about service providers, but only pay close attention to whether their demands can be satisfied. As to a large number of small- and medium-sized enterprises, they no longer need to invest a large amount of money on software and hardware apparatus. Computing, storage and bandwidth become common infrastructure. Virtual organization, assignment and usage mode, will be helpful to Internet resources optimization, energy saving and emission reduction, and to realize green computing.

2.2 Human machine interaction mode evolution

Internet and WWW [4] have become a platform which gathers human thinking and communication. The number of services, users and information resources is increasing rapidly, and all of the people around world can communicate with each other in an ultra large size. These extend the apperception and cognition of human beings, and brake away from the bound of time and space. With the popularization of mobile Internet, diversified and individualized terminal devices such as Apple iPad, iTouch, and iPhone emerge gradually over time. Human machine interaction mode has taken great changes, e.g., the peripheral changed from keyboard to mouse, and then to touch sensitive screen, voice input equipment, etc, which transformed “computer centric interaction” to “human centric interaction”. Ten years ago, we need a week to make a personal website by coding, but nowadays a child can make a beautiful blog page with rich features by click and drag operation only in an afternoon. Properties of real time, interactivity, low cost, individualization and position sensing capability of mobile Internet foster a rapid growing service requirement for clients. People can think, have mind, and as if they participate in network interaction conveniently, web pages can be linked via not only hype links, but also lots of annotation information, such as tags, comments, collection, and even personal profiles. Internet has become a media of human communication. If handled properly with the Internet based interaction behaviors and community clustering that emerge from crowds, we may find a new way to the uncertainty processing and recognition simulation, and take a new revolution of our understanding of collective human behaviors too. Turing Model [5] regarded as the basis of modern computer did not take account of the uncertainty of interaction. Internet full of uncertainty, without centralized control, cannot be seen as a super Turing machine or the sum of several Turing machines. Participation, online interaction and collective intelligence are three important differences in cloud computing from Turing model. Wikipedia\(^1\) is just a platform for people to share in their knowledge by editing own words or modifying others words, and a concept will be formed in this interaction process.

III. Social annotation is an important way of human interaction on the Internet

Some new kinds of network culture and behaviors were born in WWW. Podcast is a personal broadcast station; Witkey is a public place where people open his ability to help others or ask for help; Blog is a personal information releasing system which drives the birth of moving blog, picture blog and group blog. Twitter\(^2\) is a kind of micro blog, and its message size is a limited number. However, in Iron election, it becomes a platform where lots of people participated in, and foreign interveners can affect the Iron politics by using the opposition’s sensibility and behaviors [6]. If you have any interesting
or difficult things, share it on the Internet. After data sources are published into network, they will be attached by lots of annotation information tagged by net users via kinds of annotation systems. Thus, social annotation is a kind of cognition from public collective behaviors. The data resources include picture, video, audio, text and webpage, future events, products and websites, and everything in realistic world that can be mapped into Internet data resources.

The four layers of social annotation are tag, comment, scholium and wiki, as shown in Table I, which affect data sources in different degree. Each layer can express different individual common sense, experience, hobby, and cognition ability. A same object with different users may be enhanced the content correlations too. The public interaction forms collective behaviors, which may generate communities and collective intelligence.

Social annotation will change passive searching to active pushing. Although search methods have advanced greatly in recent years, pictures and texts recognition is still very hard for computers. Users have been tired of thousands of web pages returned by search engine only for a simple query. Finding manually with the clicks on “next page” is not the best services of a search engine. How to organize the results and answer queries on demands, this is a new challenge for search engines. As we know, similarity ranking and static ranking, e.g., HITS [7], Page Rank [8], are two main aspects for web searching research. With the development of Web 2.0 technologies, users’ annotations on the Internet have a great increasing. Social annotations are emergent useful information that can be used in various ways. Some work has been done on exploring social annotations for visualization[9], semantic web[10], enterprise search[11] etc. World Wide Web Conference Committee has been provided to utilize social annotations effectively to optimize web search[12].

We have already accumulated massive Log files and important records about users and their behaviors, which is not a bad thing if we use it with the right control of privacy protection. The evolving architectures of searching centers, which are viewed as an important cloud computing technology, help us to employ pages storing, processing, marching, and ranking functions as well. Based on previous data, search engines should be able to be cleverer: taking a user’s previous search words as tags to identify him/her, and differentiating users' interaction behaviors. Actively pushing related and useful information to users based on their personality will take the place of passively waiting for user queries.

### Table I Four layers of social annotation

<table>
<thead>
<tr>
<th>Layers</th>
<th>Data resources</th>
<th>Aim</th>
<th>Characteristic</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiki</td>
<td>Directly modification or delete, etc</td>
<td>To express understanding of content or consummate</td>
<td>Change data sources directly</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Scholium</td>
<td>Objectively commons or modification</td>
<td>To express understanding of content, or prompt some kinds of modification or evaluation</td>
<td>Change annotation data, and change data sources</td>
<td>Google Docs</td>
</tr>
<tr>
<td>Comment</td>
<td>Subjective commons or association of ideas</td>
<td>To express understanding or attitude of some content</td>
<td>Only change annotation data</td>
<td>Lots of websites have added the function</td>
</tr>
<tr>
<td>Tag</td>
<td>Mostly objective tagging</td>
<td>For easy search, or prompt others browse</td>
<td>Only change annotation data</td>
<td>Flickr, Delicious</td>
</tr>
</tbody>
</table>
IV. COLLECTIVE INTELLIGENCE EMERGING FROM ONLINE INTERACTION

In the age of cloud computing, participation and online interaction with social annotations become the important activities of people in a virtual society on the Internet. People's collective behaviors will lead to communities and form commonsense knowledge, which is just the formation process of collective intelligence.

4.1 Uncertainty of commonsense knowledge and Community

Commonsense knowledge hints the information of intelligent behavior, such as decision, consequence and abstraction etc. When speaking of human intelligence, we often refer to the expression of commonsense knowledge. However, the number of common sense is boundless, and there isn't an absolute ambit between commonsense knowledge and professional knowledge. Furthermore, which one of commonsense knowledge is chosen in the cognition process is full of uncertainty, and hard to formalize. Thus, the representation of commonsense knowledge is always one of foundational difficulties in the research of artificial intelligence (AI). There are two main research methods in traditional AI. One is proposed by McCarthy to establish a logical system of common sense, and to introduce a complete set of formalized systems. The other is proposed by Feigenbaum to realize an AI plan by constructing a huge scale of commonsense knowledge. However, because of the huge vast range of commonsense knowledge and uncertainty, the classification is intercross, thus it is hard to construct a commonsense warehouse automatically, and lots of commonsense knowledge is hard to express, suggesting that all of these become realistic difficult of the two knowledge approaches[13].

On the Internet, online search and on time interaction of users are happening all the time in different virtual communities. Network community is a clustering of people who have the same interests or professions.

For the people involved, nodes in the network possess their own activities, and the communications between nodes have local effects, thus leading to the asymmetry of times and mass. At the end, the clustering tends to a few communities. Hence, the local effect means that a node affects its neighbors mostly, and the longer a path is, the less the effect has. With the evolvement of network, the effect of topology and nodes may change, thus the partition is uncertainty, reflecting the uncertainty of communities.

In the field of complex network, there are lots of other community partition methods, such as GN algorithm[14] and spectral dimidiate[15] etc. Figure 2 is an example of communities of Web Services annotation (description field), where we extract words from their annotations (test collection is OWL S TC³). What we show in Figure 2 is the result of words clustering, which is corresponding to different fields, such as travelling, economics, medicine etc. Communities emerge from the interaction of crowds, without programming and harmonization, needn't the command of any third part to express common sense. People have certain understanding for concerned affair, which is the common sense knowledge of certain professions. The process that common sense forms communities is just the result of collective intelligence.
4.2 Typical instance of collective intelligence

Internet builds a common computing environment of human communication and more and more Web application are taking collective intelligence into network. On the Internet, people are both the users and providers of information.

Taken image recognition as an example, it is natural and easy for human, but for an image search engine based on content identification, it is rather difficult. At present, sometimes some search engine enterprises need to hire people to classify images and tag them. In fact, what we are facing is the Internet with a great deal of users, if we can ask Internet users to participate in the process, not only computers are linked by the Internet, but also people do. For example, Luis von Ahn developed ESP games in Refs. [16 17], where he presents the concept of human computing. An ESP gamer will generate image tags in gaming, and this tag can be used to clean up the images. The basic principle of human computation is just gathering human intelligence which is idle in the each angle of the Internet to resolve some difficult problems that Turing machine intelligence is hard to solve.

We also make an experiment. As shown in Figure 3 (a), participators are asked to classify twelve photos of human and pets shown in random order, and the funny classification result returned by participators is shown in Figure 3 (b), where human and pet with similar features are matched. The decisions of most participators are accordant, but it is too difficult for us to find a good image recognition algorithm to do it. Indeed it is a good algorithm if it only can distinguish human and pets images. So we may find that collective intelligence is more natural and with some uncertainty, for it just comes from human intelligence.

Wikipedia is an online encyclopedia created by Internet users, which is another typical instance of collective intelligence. In the wiki model, any registered legal users can create or edit items to propose individual idea about the item. In this mode, item with a rather high level [18] will be formed by collective intelligence. Just like the cognition process of human’s exploration on a certain concept, after lots of seeking, interacting, modifying and evolution, the item under collective editing can get into a relatively more correct answer. The answer can be accepted by most of participators, thus can come into a steady and correct edition.

In collective intelligence, different community common sense decides different annotation; furthermore, we can not only annotate images, test or multimedia, but also search in cross media by annotation[19]. Besides, we can direct crowds’ annotation too, such as RGPS (Role, Goal, Process and Service) meta model [20], which abstracts the knowledge into the corresponding four layers, and it is easier for us to analyze and apply the annotations.

V. UNCERTAINTY REPRESENTATION OF COMMON SENSE IN COLLECTIVE INTELLIGENCE

How to represent and analyze the uncertainty of
common sense formed in collective intelligence is the critical problem. Randomness, fuzziness, vagueness and incompleteness are main uncertainty. The greatest difference between human intelligence and the intelligence of other creatures lies in the fact that the former can, with the help of language, carry on knowledge accumulated over thousands of years. Fuzzy sets [21] proposed by Zadeh is a tool to measure the fuzziness of knowledge, which defines membership degree, membership function, and fuzzy logic to realize computing with words. Rough set [22] introduced by Pawlak focuses on the vagueness and incompleteness of knowledge. It is another important method for uncertainty representation based on knowledge classification, which defines a concept by two crisp sets called upper and lower approximation sets.

Most uncertainty representation methods study how to abstract concepts from data, but in human cognition process, the process from concepts to data is also an important part. Furthermore, human knowledge comes from the nature and society, so the research of uncertainty knowledge cannot break away the randomness of the objective world. Cloud model [23] is a cognitive model of uncertainty based on probability theory. The cloud model can realize the transformation between a qualitative concept and quantitative data by cloud generator algorithms. Compared with fuzzy sets, this model can construct an integrated cognition process, and doesn’t concern a crisp membership degree, and produces random numbers automatically through computer algorithm. Different from type 2 fuzzy sets [24] focusing on the fuzziness of membership degrees, cloud model discussed the randomness of that. Besides, the cloud model utilized 2 order normal distribution to produce membership degrees, which can depict the collective effect in human cognitive process. The formalized definition is:

Cloud Model: Let $U$ be a universal set described by precise numbers, and $C$ be the qualitative concept related to $U$. If there is a number $x \in U$, which randomly realizes the concept $C$, and the certainty degree of $x$ for $C$, i.e. $\mu(x) \in [0, 1]$, is a random value:

$$\mu : U \rightarrow [0, 1] \quad \forall x \in U \quad x \rightarrow \mu(x)$$

Then the distribution of $x$ on $U$ is defined as a cloud, and every $x$ is defined as a cloud drop.

As shown in Figure 4, different people have different understanding about “Young”, and a precise normal distribution cannot reflect the essential relevancy. Cloud model can describe this uncertainty, meanwhile, demonstrates the basic certainty of uncertainty.

Cloud model includes three main algorithms. Forward Cloud Generator (FCG) algorithm transforms a qualitative concept with $E_x$, $E_n$ and HE into a number of cloud drops representing the quantitative description of the concept. Reverse Cloud
Generator (RCG) algorithm can be used to abstract three numerical characters ($E_x$, $E_n$, HE) representing the concept from a number of cloud drops. Cloud Transform algorithm can be used to implement the concepts transformation in different granularity.

Obviously, there are numbers of quantitative data produced in the process of information sharing and online interaction, the common sense formed in collective intelligence can be represented by qualitative concept, the qualitative and quantitative transform thought of cloud model just according with collective behavior and common sense. This will help us to study on collective intelligence in future work.

VI. CONCLUSIONS

Information sharing, online interaction with social annotations and collective intelligence in cloud computing are three main differences from Turing model. Internet becomes the platform of information sharing and interaction. Collective intelligence emerges from people’s online interaction, which can apperceive and utilize human common sense. With the development of cloud computing, a common computing environment is built up rapidly, which is the basis of collective intelligence. Representation and inference of uncertain knowledge will become a critical topic of cloud computing and semantic computing. The uncertainty research in semantic computing, and the mathematical expression of qualitative concepts is one of the basic scientific problems. Cloud model provides a qualitative and quantitative transform model to represent the common sense in collective intelligence. In future, we hope to find a better application to establish a more natural and proper way to resolve the difficulties of artificial intelligence in cloud computing.

Acknowledgements

This work was supported by National Key Basic Research Program of China (973 Program) under Grant No.2007CB310804 and China Post doctoral Science Foundation under Grants No. 20090460107, 201003794.

Notes


References

[13] LI D, DU Y. Artificial Intelligence with Uncertainty [M],


Biographies

Liu Yuchao, an engineer of the Institute of Electronic System Engineering, China. He received his B.S. degree, M.S. degree in computer science from the PLA University of Science and Technology, Nanjing, China in 2002 and 2005 respectively. He is currently pursuing his Ph.D. degree at Computer Science and Technology Department, Tsinghua University, China. His research interests include cloud computing, granular computing and Artificial Intelligence. E mail: liuyc08@mails.tsinghua.edu.cn

Zhang Haisu, received his B.Sc. and M.Sc. degrees in computer application technology from PLA University of Science and Technology, China, in 2004 and 2007, respectively. Now, he is the doctoral candidate at PLA University of science and technology, China. His research interest mainly covers network data mining. Email:zhanghaisu@139.com

Ma Yutao, is now a lecturer at State Key Lab of Software Engineering (SKLSE) in Wuhan University, China as well as a post doctor researcher of the Institute of Electronic System Engineering. He received the Ph.D. degree from the Wuhan University in 2007 and his M.S. and B.S. degrees from the Wuhan University of Science and Technology, China in 2004 and 2001, all in computer science. His current research interests include software metrics, software evolution and the interdisciplinary research between software engineering and complex networks. Email: mmyytt_2000@163.com

Li Deyi, professor and PhD supervisor at Tsinghua University, China and president of Chinese Association of Artificial Intelligence. His research interest covers artificial intelligence, complex network and cloud computing.

Chen Guisheng, a senior engineer of the Institute of Electronic System Engineering, China. He received his Ph.D. from Tsinghua University, China in 2005. His research interest covers artificial intelligence and complex network.