

Formalisation of Contested Collective Intelligence

Gangmin Li¹

Abstract—Contested Collective Intelligence (CCI) is a knowledge acquisition model used in capturing implicit knowledge in the web environment. This paper reports a work on formalization of the CCI model. In CCI model, contested collective intelligence comes from interactions in forms of debate and argument among web users. Their ideas are contested, conflicts are resolved, ambiguity is clarified and finally a common sense has been made by means of claims and counter claims. A knowledge network is constructed reflect this process. We formalize this knowledge networks with an argumentation structure in order to study the related calculi to support knowledge fusion, synchronization and reuse.

Keywords—Contested collective intelligence, formalisation of knowledge model, knowledge capturing, web intelligence.

I. INTRODUCTION

CONTESTED Collective Intelligence (CCI) Model in Knowledge Acquisition (KA) is to simulate a community of knowledge users, monitoring their reactions to the environment and interaction with each other with a hope that the valuable knowledge can be captured by recording the whole process [1]–[3]. This interaction is also observed in the web environment [4], [5]. It is a common view that Internet and Social Networks like Twitter and Facebook contain huge amount of knowledge. Many advanced technologies developed to scan text in post in public forums. Other complicated algorithms are used to filter valuable signals and insights from noise to work out what users are referring to and even with what emotional tone. Methods in data and text mining have been success in capturing simple and explicit expressed knowledge such as sentiment analysis on the web [6], [7]. However they are not effective in capturing organisational knowledge which important insights can only be revealed and emerged after a long interaction process among knowledge users [8], [9]. This is what CCI model seeking to capture, store and reuse. There are many theoretical issues needs to be resolved so that CCI model can be widely used to leverage other existing models.

The paper is organized in the following manner. Section two represents the CCI model for construction of a knowledge networks. Section three is the formalizations of the CCI Model. It enables advanced knowledge computing methods to be deployed for knowledge services. It is a foundation for knowledge fusion, sharing and inference. Therefore some

advanced knowledge services such as search can be provided in section four. Finally section five presents our lessons learnt and future works.

II. CONTESTED COLLECTIVE INTELLIGENCE MODEL

Research into Contested Collective Intelligence seeks to develop a conceptual foundation, which will increase our capability to make sense, and to construct a sociotechnical infrastructure, which can capture collective intelligence combining contributions from many sources in the web environment [11]–[14].

Based on the previous study, valuable knowledge is often revealed in a form of continues interactions between knowledge users to clarify facts, argue about meaning and debate about truth through claims and counterclaims in a social environment such as the Web. In CCI model, each claim is an individual user's specific view, experience or insight of an issue. The process, which the claims are made to interact with other users to ensure a mining, is called sensemaking [15], [16]. This process is a continuing and sometimes a spiral process. It will continue until a stable state is reached where a neutral understanding of a meaning in the community has agreed.

The core of the CCI model is its three key elements: claims, sensemaking process, and commonly agreed knowledge called “contested collective intelligence”. A claim, in form of a text statement in which a true value is clearly stated by the claim maker, is the basic entity in CCI, which is also called a concept. Claim has properties such as author, date and the source that specify the where the supporting document is. The sensemaking process is the interactions between users. It is represented as a chain of claims linked with certain relations, such as debate logics in the “argument structure” [3], [5], [13], [17]. This rhetoric relation has fixed label, type, polarity, weight and directions. Contested collective intelligence is a label of commonly agree and accepted knowledge. It is represented as a subset of larger knowledge networks.

Fig. 1 shows an example of knowledge networks using CCI model. Where, notes are concepts from multiple sources and different users. They are connected with the fixed text labels as cues to the nature of argumentation relationship. The relations also have complicated properties to reflect the connections. In many cases users emphasis on the same concepts or relations among concepts are regarded as enhancements of the part of the networks. Other cases, the extension of the networks has been created. A lot of cases connections between previously patched networks have been newly established. In the same time, some

¹ Gangmin Li is with the Department of Computer Science and Software Engineering in Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215123 China (phone: +86 512 8816 1510; e-mail: gangmin.li@xjtlu.edu.cn).

negative connections can be suggested. These all reflect the arguments and the debate among knowledge users in the sensemaking process. Contested collective intelligence is the part or whole knowledge networks that satisfy a search criterion.

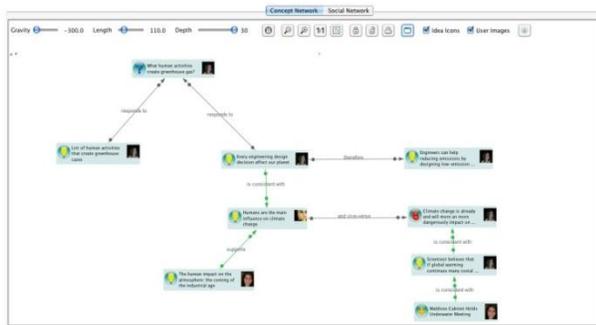


Fig. 1 Network of claims and the relationships between them in CCI model¹

Fig. 1 only shows a small part of a large network. It is a focused view on connections where debates have been displayed in details. Other views including a bird's eye view of the whole networks where only network structure is showed but the details are omitted.

III. FORMALIZATION OF THE CONTESTED COLLECTIVE INTELLIGENCE MODEL

Formalizing CCI model is necessary. Only this is done, some basic calculus can be defined and complex knowledge computation and search services can be provided.

A. Node

A node in knowledge networks define by CCI is a claim on a concept. It is a 5-tuple.

$$C = (V, A, T, S, P) \quad (1)$$

Where, V is a concept set, A is an author set and its element is the originator of the concept, T is the time when the concept is defined. S is the supporting document identifier showing where the concept originally comes from, P is a set of concept types. It is easy to find out that a claim is a basic unit in the knowledge representation in CCI. It has property of time and space. The sets associated with a claim can be regarded as mapping functions from a concept to its corresponding feature's domain. They together provide computational foundations for knowledge computations such as redundant removal, valuation, integration and inferring. Types in CCI model can be view, fact, hypothesis, question, phenomena and data etc.

¹ The figure is quoted from [1] De Liddo, A., Sandor, A. & Buckingham Shum, S. (2012). Contested Collective Intelligence: rationale, technologies, and a human-machine annotation study. Computer Supported Cooperative Work (CSCW), 21(4-5), pp. 417-448.

B. Edge

Edges in knowledge networks in CCI represent relationships between connected nodes. The relationship defined in CCI is the relationship following "Argumentation Structure". It can be formalised as a 6-tuple.

$$R = (E, A, T, S, P, \omega) \square \square \square \square \quad (2)$$

Where, E is a set of directional edges. Its element is a triple (u, v, r) , where $u, v \in V$ are claims, $r \in R$ are relations. It means that every pair of nodes has been assigned to one or more relations. That is to say any pair of claims is connected by one or more relations. As seen in the node definition, here A is a set of authors who is the originator of the connections; T is set of time stamps specifying when the connections are created; S is the supporting document identifier, which is used to show where the connections are derived from; P is a set of types showing the semantic type of the connection. Based on the argument structure, there are mainly two types: supportive and opposite. Under these two major types, there are more fine-grain types to support detailed categorizations on types such as causal, similar, evidence and sub-class, etc. ω is a mapping function representing weight of the relations.

C. Knowledge Networks

Once we have node and edge definitions, it is simple to define knowledge networks. Clearly, knowledge networks are heterogeneous networks where their nodes and edges have time and space dimensions. It means any part of knowledge networks can have a life-span. Their "True" value is always relative to its context. Furthermore, the knowledge networks always have a list of associated mapping functions and calculus. Given time sets T and space set S , node types A and edge types R , Knowledge Networks GT, S are defined in the following 8-tuple,

$$G_{T,S} = (V, E, \gamma, \rho, \theta, \tau, \phi, \delta) \quad (3)$$

Where, V is a node set; E is an edge set; there is a list of argumentation relations triple (u, v, r) , $u, v \in V$ and $r \in R$. Also, $\gamma: V \rightarrow P_V$ is a mapping function on node set. It returns node types. In general, a node v represents a concept, therefore $\gamma(v)$ represents the types of the concept v and $\gamma(v) \in P_V$.

$\rho: E \rightarrow P_E$ is a mapping function on edge set. It returns edge types. In general, an edge e represents an argumentation relation, therefore $\rho(e)$ represents the types of the relations e and $\rho(e) \in P_E$.

$\theta: V \rightarrow 2^T$ is a time mapping function on node set. It returns specific time stamp of a node. It can be used to describe node's life span. A node generally represents a concept therefore it is used to describe a concept's life span.

$\tau: E \rightarrow 2^T$ is a time mapping function on edge set. It returns specific time stamp of an edge. It can also be used to describe an edge's life span. An edge is generally representing an argumentation relation therefore it is used to describe the life span of a specific relation in argumentation structure.

$\varphi: V \rightarrow S$ is a space mapping function on node set. It returns space information of a node. It is actually a document identifier showing the concept is derived from a specified document.

$\delta: E \rightarrow S$ is a space mapping function on edge set. It returns space information of an edge. It is actually a document identifier showing the argument relations are derived from a specified document.

IV. KNOWLEDGE COMPUTATION

With the formal definitions on knowledge networks, its node and edges, we are able to provide knowledge networks' computation to support its construction, integration and services. Based on Marcus Kracht's argument structure [17], an argument is a triple:

$$\langle X: \nabla \otimes: \begin{bmatrix} CASE: & nom \\ NUM: & pl \end{bmatrix} \rangle \quad (4)$$

Where, X is a concept variable; ∇ and \otimes define the argument logic and computation of the concept. ∇ is the condition that X holds; \otimes is the argument operator that indicates the direction of the argument. The matrix contains the properties and the specific cases of the concept. It is a complex expression. However it follows simple first-order logic in representing the causes and the results of a given concept. In the first-order logic a basic argument structure is a pair $\langle D, I \rangle$, D is a concept set and I is a function assigning to a relation R with $\Omega(R) = n$, which is a subset of D^n and to a function f with $\Omega(f) = n$, which is a function from D^n to D . Here, Function $\Omega: Rel \cup Fun \rightarrow \omega$, is the first-order logic (FOL). Rel is a set of relations; Fun is a set of functions and ω denotes the set of natural numbers.

Notice that we have two special cases, namely relations of arity zero and functions of arity zero. By definition, a relation of arity zero is a subset of D^0 , which we take to be $\{\emptyset\}$. Hence there exist two such relations, \emptyset and $\{\emptyset\}$. A function of arity zero is by construction interpreted by a function from D^0 to D . Since $D^0 = \{\emptyset\}$, we get that the function is uniquely identified by $I(f)(\emptyset)$. This is why these functions are also called constants. We interpret that relation of arity zero is no argument relations and the functions of arity zero is a constants that can be interpreted as no ambiguity and a common consent holds. Marcus Kracht represents complicated argument relations with "λ-calculus". Although it provides a foundation for argument verification and integration it comes with a cost of the computation complexity.

We have proposed three basic calculi for knowledge computations. They are knowledge verification, knowledge integration and knowledge inferring.

A. Knowledge verification

Knowledge verification happens after some concepts are captured from a document with text mining techniques. It is generally consisting of operations of check its contents, verify its text and unify its format. The purpose of the knowledge verification is to reduce noise, remove redundancy and resolve conflict. The key is to define equality "=" in a given domain.

Generally, assume we have two claims C1 and C2 describing concepts v1 and v2, equality C1=C2 is defined as:

$$\begin{aligned} \text{Definition of equality (=): } C1 = C2 \text{ iff } v1 = v2, \text{ and} \\ \gamma(v1) = \gamma(v2), \\ \tau(v1) = \tau(v2), \\ \varphi(v1) = \varphi(v2) \end{aligned} \quad (5)$$

This definition guarantees the same claims on the same concepts. Despite that they may have different authors (no restriction on authors), the two concepts have the same time and space restrictions and even they have a same type. This is too restrict to be useful in finding the same concepts that they may have different time stamps, come from different authors and different sources even have different types. These conditions reflect the same concepts but from different authors and have different views on it. A relaxed equality is defined as follows:

Definition of loosely equality (\triangleq):

$$C1 \triangleq C2 \text{ iff } v1 \approx v2 \quad (6)$$

where, \approx represents similarity. If two concepts are same, then the two claims are same too despite they may have different time stamps, come from different authors and different sources even have different types. This definition is loose but it is useful in knowledge integration. The essence of the two claims defined above is the reflection of the fact that the concepts are fundamentally talking the same thing. They are the ones need to be integrated to avoid the redundancy.

B. Knowledge Integration.

Knowledge integration is vital in resolve problem of patched networks. Integration includes node integration and edge integration. Node integration is to find a same node and using it as an anchor to extent existing knowledge network to include newly created networks. Edge integration has two cases in two different scenarios. One is that a new relation is created on existing two nodes. In this case, the integration of the new relation is simply adding it into the existing edge set. The other scenario is that the newly created relation already exist, in this case the integration is to increase the weight of existing edge in one unit. They are denoted respectively as follows.

For knowledge networks $G_{T,S}$, a new node v , and a new edge e then the Integration $INT(G_{T,S}, v)$, $INT(G_{T,S}, e)$ and $INT(G_{T,S}, v) \stackrel{\text{def}}{=} G_{T,S} \rightarrow G'_{T,S}: V' = V \cup v$. i.e.

$$G'_{T,S} = (V', E, \gamma, \rho, \theta, \tau, \varphi, \delta) \quad (7)$$

and,

$$INT(G_{T,S}, e) \stackrel{\text{def}}{=} G_{T,S} \rightarrow G'_{T,S}: E' = E \cup v, \text{ when } e \notin E.$$

$$INT(G_{T,S}, e) \stackrel{\text{def}}{=} G_{T,S} \rightarrow G'_{T,S}: \omega(e) = \omega(e) + 1, \text{ when } e \in E.$$

$$G'_{T,S} = (V, E', \gamma, \rho, \theta, \tau, \varphi, \delta) \quad (8)$$

The general integration can be done through these basic integrations.

C. Knowledge Inferring.

Knowledge inferring is advanced operations on a given graph following argument structure under certain time and space constraints. Apart from the conventional propositional connectives: \neg , \wedge , \vee , and \rightarrow , Lambda Calculus (λ -Calculus) is useful but out of scope of this report.

V. CONCLUSION AND FUTURE WORKS

We have made the case that Contested Collective Intelligence (CCI) can be considered as a significant and distinctive method for knowledge acquisition and knowledge services in the Web environment. Research in sensemaking and the modelling of dialogue and debate, motivates a conceptual model of CCI and its formalization which together begins to address the hard problem of the knowledge acquisition, integration and services.

This paper reports some preliminary works on this direction. It lays a foundation for complex knowledge computation and integration. With these utilities we can start to addressing difficult problems: 1) distinction between repetition and redundancy on both concepts and connections; 2) weight on connections and their calculations; 3) integration on patching knowledge networks; 4) and lastly the search on circular networks.

Several strands of ongoing work seek to advance the research. We are studying the sentence structure and syntactic markers to improve cue spotter. More efforts will be focused on analysing the property of the same or nearest concepts not only on words and phrases, but also on semantics. It is a hope that adopting existing similarity measurements on meaning including general term and local dialects can help to identify concepts redundancy and repetition. Similar approach but focused on the structure of the networks they are in can be taken to evaluate the weight of a connection. Advance machine learning methods and algorithms are investigated to provide a higher level knowledge services such as search on circular networks and integrate with external networks.

ACKNOWLEDGMENT

Research and development reported in this paper are continuation of Cohere project in UK. Cohere was being developed to enable Collective Intelligence for the Open Educational Resource movement, as part of the OLnet Project (<http://olnet.org>).

REFERENCES

- [1] De Liddo, A., Sandor, A. & Buckingham Shum, S. (2012), 'Contested Collective Intelligence: Rationale, Technologies, and a Human-Machine Annotation Study', *Computer Supported Cooperative Work: the journal of collaborative computing*, vol. 21, no. 4-5, pp. 417-448.
- [2] Buckingham Shum and De Liddo, Anna. (2010). Collective intelligence for OER sustainability. *Collective Intelligence in Organizations*. February 6-10 Savannah.
- [3] Buckingham Shum, S., Selvin, A., Sierhuis, M., Conklin, J., Haley, C. and Nuseibeh, B. (2006a). Hypermedia Support for Argumentation-Based Rationale: 15 Years on from gIBIS and QOC. In: *Rationale Management in Software Engineering* (Eds.) A.H. Dutoit, R. McCall, I. Mistrik, and B. Paech. Springer-Verlag: Berlin
- [4] De Liddo, Anna, G.Li and Buckingham Shum, Simon (2010). Cohere: A prototype for contested collective intelligence. In: *ACM Computer Supported Cooperative Work (CSCW 2010) - Workshop: Collective Intelligence In Organizations - Toward a Research Agenda*, February 6-10, 2010, Savannah, Georgia, USA.
- [5] Buckingham Shum, S., Uren, V., Li, G., Domingue, J., Motta, E. . (2003) Visualising internetworked argumentation. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. P. A. Kirschner, Buckingham Shum, S. and Carr, C. London, Springer-Verlag, 185-204.
http://dx.doi.org/10.1007/978-1-4471-0037-9_9
- [6] Xindong Wu. (2013) Data Mining with Big Data. *International Conference on Big Data and Cloud Computing* December 28-30, 2013 Xiamen, China.
- [7] V. Terziya,O. Shevchenko, M. Golovianko, (2014) "An Introduction to Knowledge Computing".
- [8] Ikujiro Nonaka, Georg von Krogh, (2009) Perspective—Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory. *Organization Science* 20(3):635-652. <http://dx.doi.org/10.1287/orsc.1080.0412>
<http://dx.doi.org/10.1287/orsc.1080.0412>
- [9] Kiku Jones and Lori N. K. (2009) Leonard. From Tacit Knowledge to Organizational Knowledge for Successful. KM. W.R. King (ed.), *Knowledge Management and Organizational Learning*, *Annals of Information Systems* 4, DOI 10.1007/978-1-4419-0011-1_3, © Springer Science+Business Media, LLC
- [10] Knight, S., Buckingham Shum, S. & Littleton, K. (2013), 'Collaborative sensemaking in learning analytics', *CSCW and Education Workshop (2013): Viewing education as a site of work practice, co-located with the 16th ACM Conference on Computer Support Cooperative Work and Social Computing (CSCW 2013)*,..
- [11] De Liddo, A. & Buckingham Shum, S. (2013), 'Improving online deliberation with argument network visualization', *Digital Cities* 8.
- [12] De Liddo, A. & Buckingham Shum, S. (2013), 'The Evidence Hub: Harnessing the Collective Intelligence of Communities to Build Evidence-Based Knowledge', *Large Scale Ideation and Deliberation Workshop, 6th International Conference on Communities & Technologies*.
- [13] Buckingham Shum, S. (2008). *Cohere: Towards Web 2.0 Argumentation*. 2nd International Conference on Computational Models of Argument, 28-30 May 2008, Toulouse. IOS Press: Amsterdam.
- [14] Iandoli, L., Quinto, I., De Liddo, A. & Buckingham Shum, S. (2012), 'A Debate Dashboard to Enhance Online Knowledge Sharing', *VINE The Journal of Information and Knowledge Management Systems*, vol. 42, no. 1, pp. 67-93.
- [15] Sensemaking Workshops, *ACM Computer-Human Interaction (CHI) Conferences* (2012)
- [16] Klein, G., Moon, B. and Hoffman, R.F. (2006). Making sense of sensemaking 1: Alternative Perspectives. *IEEE Intelligent Systems*, 21(4), 70-73.
<http://dx.doi.org/10.1109/MIS.2006.75>
- [17] Marcus Kracht, "Agreement Morphology, Argument Structure and Syntax" 3rd Ed, Los Angeles, September 2005.