Augmented Collaborative Spaces for Collective Sense Making: The Dicode Approach

Ahmad Ammari[1], Vania Dimitrova[1], Lydia Lau[1], Manolis Tzagarakis[2], and Nikos Karacapilidis[2]

[1] School of Computing University of Leeds, Leeds, UK.
[2] Research Academic Computer Technology Institute, Patras, Greece.
{ A.Ammari, V.G.Dimitrova, L.M.S.Lau, }@leeds.ac.uk, tzagara@upatras.gr, nikos@mech.upatras.gr.

Abstract. Sense making is at the heart of cognitively complex and data intensive decision making processes. It is often conducted in collective spaces through exchange of ideas, discussions, analysing situations, and exploring alternatives. This position paper proposes a novel approach to facilitate collective sense making via a collaboration platform which (a) offers multiple views to collaboration (including forums, mind maps, and argumentation structure), and (b) provides intelligent support to understand sense making behaviour by employing user and community modelling techniques. The work is conducted in the framework of the EU funded Dicode project, developing intelligent services for data-intensive collaboration and decision making.

Keywords: Collective sense making, Collaborative workspaces, Intelligent support, User and community modelling

1 Introduction

This paper proposes a novel platform to augment the synergy between human and machine intelligence in complex decision making situations. Many collaborative decision making problems have to be solved through dialoguing and argumentation among a group of people [1, 2]. In such contexts, discussions for making sense of the issues, constraints, and options are usually conducted in an unstructured manner. Sense making is a "motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively" [3]. Therefore, sense making is an inevitable path in cognitively complex and data intensive decision making processes.

Dicode[1] (**D**ata-**i**ntensive **co**llaboration and **de**cision making), an EU Framework 7 project, sets out to tackle the above challenges for three use cases. The first use case concerns a team of scientists in clinico-genomic research. The second use case involves a group of radiographers, radiologists and clinicians in a trial of rheumatoid arthritis treatment. The third use case involves public opinion monitoring on the internet for a team of brand consultants to design a campaign.

[1] Dicode website is http://dicode-project.eu/

Argumentation, as seen in Dicode, is a common activity in collective sense making process. It is valuable in shaping a common understanding of the problem and can provide the means to decide which parts of the information brought up by the decision makers will finally be the input to the solution used. Argumentation may also stimulate the participation of decision makers and encourage constructive criticism. However, discovering the connections is mainly by using tacit knowledge and the value of this activity has been largely unacknowledged. Dicode aims to address the above by user-friendly multi-view collaboration workspaces, which facilitate the exchange and sharing of ideas, opinions, comments and resources between participants. While each collaborative workspace enables an individual or a team to visualise the connections between concepts and artefacts, keeping track of the rationale behind the decision points and redeploying the accumulated knowledge in new situations is itself potentially a cognitively complex process. Hence, intelligent support will be provided by exploiting the behaviour data captured in the usage logs and by adding semantics to the content shared.

This position paper outlines a multi-faceted approach to combine human and machine intelligence for collective sense making. Specifically, we will present a novel approach to design collaborative workspaces that facilitate sense making by combining multiple views – ranging from informal (unstructured) to formal (structured). Each view facilitates different sense making aspects. Furthermore, we present a proposal how collaborative workspaces can be augmented with intelligent support utilising adaptation techniques, namely user and community modelling.

2 The Dicode Project

The goal of the Dicode project is to facilitate and augment collaboration and decision making in data-intensive and cognitively-complex settings. It will exploit and build on the most prominent high-performance computing paradigms and large data processing technologies - such as cloud computing, MapReduce [4], Hadoop[2], Mahout[3], and column databases – to meaningfully search, analyze and aggregate data existing in diverse, extremely large, and rapidly evolving sources. Building on current advancements, the solution foreseen in the Dicode project will bring together the reasoning capabilities of both the machine and the humans. It can be viewed as an innovative workbench incorporating and orchestrating a set of interoperable services that reduce the data-intensiveness and complexity overload at critical decision points to a manageable level, thus permitting stakeholders to be more productive and concentrate on creative activities. Services to be developed are: (i) scalable data mining services (including services for text mining and opinion mining), (ii) collaboration support services, and (iii) decision making support services.

In this paper, the focus is on the collaboration support services which are realised via multi-view collaborative workspaces augmented with intelligent support for collective sense making.

[2] Apache Hadoop Project http://hadoop.apache.org/
[3] Apache Mahout Project http://mahout.apache.org/

3 Multi-View Collaborative Workspace

In Dicode, three different views of collaboration workspaces (CW) are supported. These are summarised below:

- **Discussion-forum view:** In this view, the CW is displayed as a traditional *web-based forum*, where posts are displayed in an ascending chronological order. Users are able to post new messages to the collaboration workspace, which appear at the end of the list of messages. Posts may also have attachments to enable the uploading of files. Discussion-forum exhibits a very *low level of formality* and are mainly suitable to support *ideas sharing, exchange and collection*.
- **Mind-map view:** In this view, the CW is displayed as a *mind map* where users can interact with the items on the collaboration workspace. This view deploys a spatial metaphor permitting the easy movement and arrangement of items on the collaboration workspace (Fig. 1). Messages posted on the collaboration workspace in mind-map view can be one of the following types: *idea*, *comment*, *note* and *generic*. Files of any content type (e.g. pdf, jpg) can be uploaded to the CW. The set of available types can be configured and participating users will be able to define new ones. The mind-map view also provides a set of mechanisms through which: (a) items on the collaboration workspace can be related, and (b) new abstractions can be created. In particular, creation of relationships between items is facilitated by drawing directed arrows between items on the collaboration workspace. Visual cues can be used to convey semantics (e.g. red colour can indicate opposition, while green can indicate "in favour"; labels can be associated to arrows elucidating semantic relationships). Items on the CW can be aggregated, to allow a group of items to be treated as a single entity, and transformed into a single item creating new, composite items. The mind-map view aims at supporting *sense-making during data intensive and cognitive complex tasks*.

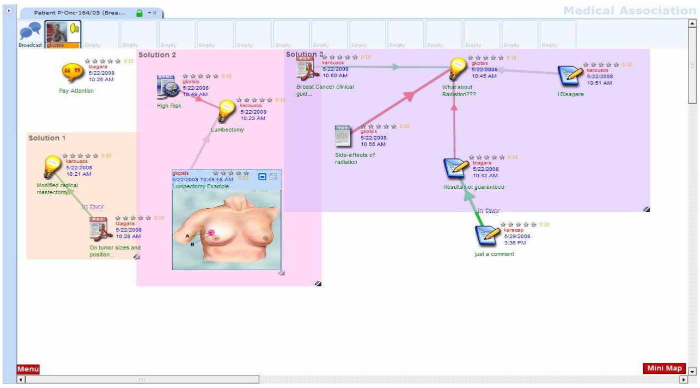

Fig. 1: Mind-map view of a collaboration workspace. Explicit relations can be created between collaboration items (arrows) or juxtaposed to express implicit/transient relationship.

- **Formal/Argumentation view:** The formal/argumentation view of the CW permits only a limited set of discourse moves for a limited set of message types whose semantics is predefined and fixed. Formal views of the collaboration workspaces exhibit a *high level of formality*. In particular, the formal view (Fig. 2) enables the posting of messages which can be of type *issue* (to indicate the decisions to be made) *alternative* (to represent potential solutions to the issues discussed) or *position* (to comment on alternatives or on other positions). Positions either support or are against alternatives and positions and their relationship are explicitly specified when users post them to the collaboration workspace. Files can be attached to positions to further support their validity. The formal view supports also the notion of preferences, used to weigh the importance of two positions and reflect the importance of one position over another. Decision making support algorithms (e.g. a voting or a multiple criteria decision making), which are associated with the CW, can take into consideration the relationships of positions as well as existing preferences and calculate which alternative is currently prevailing or which position has been defeated. The aim of the formal view is to *make the CW machine understandable* and to further support *decision making*.

Fig. 2: A formal view of the collaboration workspace shown in Fig. 1.

Every CW can be **transformed from one view into another** at any point in time by anyone participating in the collaboration. Such transformations are rule-based; a set of rules specifies how items in the source view are transformed into items of the destination view. All **discourse moves and contributions** that users create during their interaction in the CW are logged within Dicode in order to enable their further analysis by a variety of services. For each view, **log data contains** information related to the event that happened on the workspace and which includes:

- the collaboration workspace's ID and view where the event took place;
- the user's operation and the associated content (e.g. adding/updating/deleting an item, moving an item, creating relationships between items etc);
- the user who executed the operation;
- the date and time when the event occurred.

The log data in the CW will be used as an input for intelligent support algorithms.

4 Intelligent Support

Intelligent support will **augment the multi-view CWs** with machine intelligence to understand and facilitate collective sense making. Intelligent support will be provided at two levels:

- *Understanding collective sense making.* This will include user/community profiling, e.g. identifying user characteristics, discovering links between individuals, identifying common topics; discovering patterns of behaviour such as silos or dominance, extracting situations parameters.
- *Facilitating collective sense making*. This will include interface augmentation (e.g. adding visual signals to help establish situational awareness) or suggestions in the form of messages (e.g. to facilitate the exchange of ideas, point at useful patterns, highlight important situation aspects).

The following subsections propose our approach to implementing the first level of intelligent support, i.e. understanding collective sense making behaviour. This will be achieved by three functions (section 4.3) which employ descriptive machine learning and data mining algorithms and meet the key objectives as stated in section 4.2. The following section outlines how the CW log data will be enriched with semantics for user and community modelling.

4.1 Input: Augmented CW Log Data

Intelligent support will be based on the log data from the CWs which include mind mapping graphs, discussions, arguments and comments. In addition, the users' meta-data, including the users' navigational behaviour as recorded in the usage logs, as well as the searching behaviour of the users in the collaborative workspace, will be used to characterise the users and derive a user profile for each user in the community. Semantic enrichment of the user profiles is achieved by considering semantic data sources, such as domain ontologies (to identify the domain topics discussed), as well a collaboration and decision making ontology developed in Dicode (to take into account the user roles and to link sense making to decision making steps).

4.2 User and Community Modelling

Intelligent support in Dicode is underpinned by a mechanism for user and community modelling which will be outlined here. It is envisaged to be used by intelligent services which augment the CW in Dicode. For instance, a recommendation mechanism in Dicode will be able to use the output of the community modelling functions to direct to 'items' in the CW, e.g. a data set, a set of relevant discussions, a topic of interest to search for. Furthermore, the users of the CW can be pointed to a set of discussions that occurred in different times but belong to a certain topic of interest.

Objectives. The following four main objectives can be perceived for the community modelling and user profiling functions:

- **O1:** Detect topics of community discussions in the collaborative workspace.
- **O2:** Identify key characteristics of the users in the community from available data about the users, i.e. unstructured data, semantic annotations, meta-data, and use these characteristics to shape the user profile for each user within the community.
- **O3:** Quantify the strength of each characteristic for discovery of connections.
- **O4:** Discover clusters of users and interesting patterns in user behaviour by applying descriptive data mining functions, i.e. cluster analysis and association mining on the derived user profiles.

4.3 Outline of the Main Algorithms

This section will outline how **descriptive machine learning and data mining**, such as cluster analysis and association rule mining, can be applied for user and community modelling. We will group them into three main functions.

Function 1: Clustering Unstructured Data for Topic Detection

Purpose (O1). The main purpose of this function is to discover the main topics of the unstructured data, i.e. community discussions, arguments, using descriptive data mining methods, i.e. cluster analysis.

Input. Unstructured data that community users create within the collaborative workbench, as part of their collaboration activities. These include the discussion and arguments that occurred between the community users in the workbench. All the available parts of the discussions can be utilized by the function, i.e. the title of the discussion thread, main discussion body, replies by other users, tags that collaborating users attach to the discussion.

Processing. The input data will be processed as follows:

- Pre-process the input unstructured data and transform it into a term weight document matrix to be used as input for cluster analysis.
- Using the pre-processed matrix, build and train a clustering model that segments the discussions into distinct groups (clusters) based on the similarities and distances between the discussions.
- Using the profiles of the discovered clusters, detect the topic of each cluster of discussions based on the frequency of occurrence by considering the most occurring terms that occur in each cluster.

Output. There are two types of output produced by this function:

- *Clusters of discussions*, where each discussion instance will be assigned a cluster id to identify to which discovered cluster of discussions it belongs to.
- *Cluster profiles*, including the number of discussions that belong to each cluster and the most significant terms that belong to each cluster based on the frequency of occurrence.

Function 2: Deriving Key User Characteristics and Generating User Profiles

Purpose (O2 & O3). The purpose of this function is to derive the key characteristics that describe each user within the community, and weight these characteristics for every user to reflect the significance of each characteristic. These weighted user profiles will be accumulated in a community model.

Input. Data input to this function include: (a) Discussion topics that are detected using the first function described above; (b) User meta-data available from the logs and meta-data derived from the other components of the collaborative workbench, including the discussions, arguments, i.e. the author of the main body of the discussion and the authors of the replies to the main body, the mind mapping graphs, and the meta-data available from the searching behaviour in the workspace. (3) The characteristics derived from the unstructured data, i.e. topics, and the meta-data can be semantically enriched by the collaboration and decision support ontology, relevant domain ontologies, and open lexical resources, i.e. Wordnet.

Processing. This function will process the input data as follows:

- *Identify user characteristics* within the community from the available input data.
- *Compute weighted interests in the identified topics -* for each identified characteristic, the function will compute a numerical weight for each user profile that represents the significance (importance) of this characteristic to that user within the community.
- *Build a user–characteristic matrix* that could be input to further descriptive data mining functions (cluster analysis and association mining).

Output. The output of this function is a community model that includes a user profile for each user. Each user profile represents the weights of the identified characteristics for each user within the community.

Function 3: Discovering Patterns in the User Profiles

Purpose (O4). The purpose of this function is to discover hidden patterns in the user profiles for further support to collaboration and decision making, using descriptive data mining techniques.

Inputs. The input to this function is mainly the community model (user profiles) derived by the second function

Processing. This function will process the input data as follows:

- Apply *cluster analysis* methods on the derived user profiles within the community model to discover the user clusters and the user cluster profiles.
- Apply *association mining* methods on the derived user profiles within the community model to discover association hidden patterns within the user characteristics.

Output. This function mainly produces three outputs: (a) Clusters of user profiles, where each user profile instance will be assigned a cluster id to identify to which discovered cluster of user profiles each user belongs to. (b) Cluster profiles, including the number of user profiles that belong to each cluster and the characteristics' values for the average user profile, i.e. cluster centroid, for each discovered cluster. (c) Discovered hidden association patterns, including frequent characteristic-sets that list those significant characteristics that are obtained frequently by the same users, and the hidden association rules underlying these sets.

5 Related Work

The approach proposed in this paper has two main innovative aspects: (a) a new way to facilitate sense making using multiple linked views of collaborative workspaces; and (b) a novel application of user and community modelling to get an understanding of collective sense making behaviour.

Over the years, a number of systems have been developed aiming to support the process of sense making which include Debatepedia [5], Parmenides [6], ClaiMaker [7], TruthMapper [8] and Cohere [9]. Despite their powerful features, each of these systems provides only a fixed level of formality lacking the ability to adapt their environment to the needs of the collaboration. In Dicode, collaborative workspaces build on and extend the notion of spatial hypertext, which has been proposed as an alternative to navigational and semantic organisation of resources [10]. Spatial hypertext employs a spatial metaphor to organize information aiming at taking advantage of the user's visual memory and pattern recognition. Due to its ability to express ambiguity as well as transient and implicit relationships between information, it is an effective way to support information triage, i.e. the process of sorting through relevant materials and organizing them to meet the needs at hand[11]. While most existing hypertext systems permit only a single user to organize the information (e.g. VIKI [12], WARP [13]), approaches to bring spatial hypertext into the collaborative realm have only recently started to emerge [14]. Dicode will make a contribution to this stream by exploiting spatial hypertext for collective sensemaking in cases when humans need to process large volumes of heterogeneous data.

Recent research trends look at intelligent ways to support the effective functioning of close-knit communities through personalization and adaptation techniques. Modelling users within a community provides the grounds for generating group recommendations [15]. One method to support that is through detecting the topics that the collaborating users show interests in. In [16] Cheng and Vassileva derived topics of users' interests based on the resources shared by them within the community, where a reward factor is calculated to measure the relevance of each contributed resource to the topics derived. In [17], Bretzke and Vassileva modelled users' interests based on how frequently and recently users have searched for a specific area from a particular taxonomy. User relationships are then determined based on the resource downloading behaviour. A more recent approach by Kleanthous and Dimitrova [18][19] employs the metadata of the shared resources along with an ontology representing the community context and derives a semantically relevant list of interests for every user.

In Dicode, we aim to further enhance the existing topic detection approaches by exploiting a hybrid machine learning, text data mining, and semantic enrichment approach. Using as input community discussions, mind-mapping activities, and relevant ontologies, we aim to discover topics of interests that are buried within the diversity of unstructured and semi-structured contents produced by the collaborating members in the multi-view collaborative workspaces. Detected topics will then be exploited to facilitate collective sense making within the community members.

A community model can be analysed to automatically detect patterns which can be used to decide when and how interventions to the community can be done [20]. It has been shown that community patterns based on these processes can be derived from the community graph. For example, [19] have identified community patterns related to processes linked to effective knowledge sharing, such as transactive memory (how members' knowledge is related), shared mental models (shared understanding of the common goal), and cognitive centrality (influential members).

Similarly to Kleanthous and Dimitrova's work on semantically-enriched relationship detection, we will exploit semantics and ontologies to enhance the log data from CWs and get richer input about what is happening in the community. However, the community modelling approach in Dicode will take the modelling further by exploiting descriptive data mining approaches, including output from (i) statistical member segmentation, i.e. group profiles, where members assigned to the same group share a similar behavioural profile, as well as output from (ii) association rule mining, i.e. lists of the frequently co-occurring behavioural activities of the community members, in order to further improve the community pattern discovery tasks. Discovered patterns will also be used to further augment the multi-view CW for enhanced collective sense-making, knowledge sharing, and group recommendations.

6 Conclusions

We have set out an ambitious goal to exploit the synergy of machines and humans in complex cognitive situations that require making decisions involving large volumes of data. We are starting to unravel the aspects of this synergy. While data mining techniques (i.e. machine intelligence) can be exploited to process data and discover trends and patterns, human intelligence is needed to make sense of the data and take decisions. The process of sense making involves discovering connections, deriving patterns, generating alternatives, weighting possibilities. People perform these tasks in an intuitive manner using tacit knowledge. Our ultimate goal is to capture, preserve, and reuse this tacit knowledge by providing collaborative workspaces for collective sense making. In turn, we will exploit machine intelligence to analyse the human behaviour in the collaborative spaces in order to get a better understanding of the collective sensemaking process, facilitate important aspects, and support future human sense making (e.g. exploiting patterns applied earlier).

Currently, we are developing the CWs following a generic approach, which will enable the same approach to be applied to diverse use cases. The illustrations in this paper were from the exemplification of the multi-view space for a Breast Cancer research group embarking on an analysis to discover any common characteristics or trends that could be deducted from recent studies which used high-throughput technologies such as microarrays and next-generation sequencing. We plan to apply the approach presented here to support sense making in a clinical trial of Rheumatoid Arthritis treatment where a team of medical practitioners examines large data sets and analyses the effectiveness of the treatment on patients. In addition, the log data from the CWs is being analysed in line with the functions presented in here to augment CWs with intelligent support.

Acknowledgements
This publication has been produced in the context of the EU Collaborative Project "DICODE - Mastering Data-Intensive Collaboration and Decision" which is co-funded by the European Commission under the contract FP7-ICT-257184. This publication reflects only the author's views and the Community is not liable for any use that may be made of the information contained therein.

References

1. van Eemeren, F. H., Grootendorst, R., Snoeck Henkemans.F.: Fundamentals of Argumentation Theory. Erblaum, Mahwah, NJ (1996)
2. Provis, C.: Negotiation, Persuasion and Argument. Journal of Argumentation, 18(1), pp. 95-112 (2004)
3. Klein, G., Moon, B., Hoffman, R.: Making Sense of Sensemaking 1: Alternative Perspectives. IEEE Intelligent Systems, 21(4), July/August (2006)
4. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. In OSDI'04: Proceedings of the 6th conference on Symposium on Opearting Systems Design & Implementation, USENIX Association, pp. 10–10 (2004)
5. Debatepedia, http://wiki.idebate.org/
6. Atkinson, K., Bench-Capon, T.,McBurney, P.: PARMENIDES: Facilitating deliberation in democracies. Artificial Intelligence and Law, 14(4), pp. 261-275 (2006)
7. Buckingham Shum, S.J., Uren, V., Li, G., Sereno, B., Mancini, C.: Modelling Naturalistic Argumentation in Research Literatures: Representation and Interaction Design Issues. International Journal of Intelligent Systems, 22 (1), pp.17-47 (2007)
8. TruthMapping, http://truthmapping.com
9. Buckingham Shum, S.,: Cohere: Towards Web 2.0 Argumentation, Proceeding of the 2008 conference on Computational Models of Argument, p.97-108, June 21 (2008)
10. Shipman, F. M., Marshall, C. C.: Spatial Hypertext: An Alternative to Navigational and Semantic Links, ACM Computing Surveys 31(4), December (1999)
11. Marshall, C.C., Shipman, F. M. III: Spatial hypertext and the practice of information triage, Proceedings of the eighth ACM conference on Hypertext, p.124-133, April 06-11, Southampton, United Kingdom (1997)
12. Marshall, C.C., Shipman, F. M., Coombs, J. H.: VIKI: spatial hypertext supporting emergent structure, Proceedings of the 1994 ACM European conference on Hypermedia technology, p.13-23, September 19-23, Edinburgh, Scotland (1994)
13. Francisco-Revilla, L., Shipman, F. M.: WARP: a web-based dynamic spatial hypertext, Proceedings of the fifteenth ACM conference on Hypertext and hypermedia, August 09-13, Santa Cruz, CA, USA (2004).
14. Solís, C., Ali, N.: ShyWiki - A spatial hypertext wiki, Proceedings of the 4th International Symposium on Wikis, September 08-10, Porto, Portugal (2008)
15. Masthoff, J.: Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers, User Modeling and User-Adapted Interaction, 14(1), pp.37-85 (2004)
16. Cheng, R., Vassileva, J.: Design and evaluation of an adaptive incentive mechanism for sustained educational online communities, Journal of User Modeling and User Adaptive Interaction, vol. V16, no. 3, pp.321 – 348 (2006)
17. Bretzke, H., Vassileva, J.: Motivating Cooperation on Peer to Peer Networks, 9th Int. Conf. on User Modelling, Springer (2003)

18. Kleanthous, S., Dimitrova, V.: Modelling Semantic Relationships and Centrality to Facilitate Community Knowledge Sharing, Proc. of the 5th Int. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'08) Springer (2008).
19. Kleanthous, S., Dimitrova, V.: Analyzing Community Knowledge Sharing Behavior, UMAP 2010, Springer, pp.231-242 (2010)
20. Kleanthous, S., Dimitrova, V.: Detecting Changes over Time in a Knowledge Sharing Community, Proc. of the 2009 IEEE/WIC/ACM Int. Joint Conf. on WI and IAT, IEEE Computer Society Washington, DC, USA Milan, Italy (2009).