別　紙　３

平成２５年度北海道大学情報基盤センター共同研究成果報告書

１．研究領域番号　　A5 デジタルコンテンツ

２．研 究 課 題 名　　ソーシャルメディアセマンティクス及び集合知の評価に関する研究

Evaluation of social media semantics and collective intelligence

３．研 究 期 間　　平成　25 年 4 月　1 日　～　平成　　26 年　3 月　31 日

４．研 究 代 表 者

氏　　名	所属機関・部局名	職　　名	備　　考
Frederic Andres	NII コンテンツ科学研究系	准教授	

５．研 究 分 担 者

氏　　名	所属機関・部局名	職　　名	備　　考
田邊　鉄	**北大情報基盤センター**	**准教授**	
平林義治	〃	〃	
田中　譲	**北大大学院　情報科学研究科**	**教授**	
Asanee Kawtrakul	**カセサート大学（タイ）工学部**	**准教授**	

６．共同研究の成果

下欄には，当該研究期間内に実施した共同研究の成果について，その具体的内容，意義，重要性等を，共同研究申請書に記載した「研究目的」と「研究計画・方法」に照らし，800 字～1,000 字で，できるだけ分かりやすく記載願います。文章の他に，研究成果を端的に表す図表を貼り付けても構いません。なお、研究成果の論文・学会発表等を行った実績（発表等の予定を含む。）があれば、あわせて記載して下さい。

The research target was to evaluate social media semantics and collective intelligence. The explosion of social media data in semantic web formats since the introduction of the Linked Data principles, and the corresponding growth of open data initiatives all over the world now in many other countries both point to an opportunity for social media systems to be built which integrate structured social media data, build on semantic and collective intelligence technologies and provide new services for e collaborative problem-solving and social project management.

The evolution of project management from static to dynamic management approach has enabled easy updates. Furthermore, the integration of social network within social project management enables new best practice modeling methods. It's obvious - being recognized and doing some actions means that the project workflow is progressing. Widespread the project scope shared and accessible as a SaaS service affects the size and depth level of its project structures. It could have negative impact on the project and weaker usability.

Usability of delivered information and knowledge depends on dynamically created project member profiles as a result of project activities mining and particularly project user navigation and patterns discovery process. Knowledge acquired from the social project management server log files in order to generate navigation and actions patterns embraces useful as well as non-relevant association rules and has to be validated. Therefore the ultimate goal of social project management mining of the method allowing for generated knowledge refinement.

These are very specific features of social project management mining procedures: temporal and massive data input, big differentiation of project members types. They have direct impact on generated knowledge about social project management and forms and in turn should be considered in knowledge validation framework. List of constrains useful in knowledge validation processes includes: action duration, document view duration, total session time, include and exclude items, support and confidence and time and date.

All the mentioned constraints should be considered as potential criteria of discovered knowledge about project member navigation and actions patterns. On the other hand the considered constraints are useful in search space reduction in a case of frequent actions..

We evaluated recommendation systems as they help to address information overload by using discovered user navigation patterns knowledge gained from web server log files. Let us see the problems to use them in social project management. A problem to apply association rule recommendation systems (RS) to project management mining is located in the input dataset. It is often sparse because for any given project member visit or object rank, it is difficult to find a sufficient number of common items in multiple user profiles. As a consequence, a RPS system has difficulty to generate recommendations, especially for social project management applications.

In order to solve above problem, some standard dimensionality reduction techniques have been applied to improved performance. Sarwar et al. have explored one technology called Singular Value Decomposition (SVD) to reduce the dimensionality of recommender system databases. The second experiment compares the effectiveness of the two recommender systems at predicting top-n lists, based on a real-life customer purchase database from an e-commerce application. Finally, results suggest that SVD can meet many of the challenges of recommender systems, under certain conditions. Ad hoc exclusion of some potential useful items can be one of known deficiencies of this and other reduction of dimensions solutions hence they will not appear in the final results. Two solutions addressing this problem were proposed by Fu et al. The first solution assumes to rank all the discovered rules based on the degree of intersection between the left-hand side (antecedent) and active session of the user. Then, the SurfLen (client-server system) generates the top k recommendations. In addition to deal with sparse datasets - if users browsing histories intersect rarely, the system is unable to produce recommendations - the algorithm for ranking association rules was presented. The second solution takes advantage of collaborative filtering. The system is able to find “close neighbours” who represent similar interest to an active user. Then, based on that, a list of recommendations is generated. Many collaborative filtering systems have few user ranks (opinions) compared to the large number of available documents. Sarwar et al. defined and implemented an integration model for content-based ranks into a collaborative filtering.

In addition, metrics for assessing the effectiveness filter bots and a system were identified and evaluated.

Lin et al. produced a collaborative recommendation system based on association rules framework. The framework provides two measures for evaluating the association expressed by a rule: confidence and support. Moreover, the system generates association rules among users as well as among items.

We evaluated the model spaces

The social project management space can be considered as universe U which consists of a sequential sets (Pi) where i =1...M. Each of sets Pi corresponds to unique user session where as each element of Pi is user’s request for single page shared by a project. We consider only such subsets A of Pi (A⊆Pi⊆U) which appeared often enough. The frequency of subset A is defined as support (or support ratio) denoted as support (A)=|{i ; A⊆Pi}|/M. A “frequent” set is a set which support satisfies the minimum support value, denoted as *Minsupport*. It is a project member dependent value, often defined as cut-off as well. We developed and applied an Apriori-like algorithm to mine frequent item sets that is based on level-wise search. It scans a database recursively – a growing collections A are generated using smaller collections, if and only if ∃a∈A A =B∪ {a }. An association rule is an implication of the form A→B , where B⊆Pi⊆U and

（研究成果のつづき）

The results are a web infrastructure necessary for large-scale evaluation of social media semantic-based retrieval and collective intelligence sharing methodologies and benchmarking. We focused in encouraging education and research in social media retrieval based on large open test collections; in setting up an open community (CI@PracticeDay) among industry, academia, and government for the exchange of research ideas on social media access, and in increasing the availability of appropriate evaluation techniques for use by industry and academia, including development of new evaluation techniques more applicable to current social media systems.