Mining Useful News Information Based on Reader Feedback for Building News Communities

Chih-Ming Chen¹, Wei-Chia Chiu²

¹Graduate Institute of Library, Information and Archival Studies, National Chengchi University ²Department of Computer Science, National Chengchi University NO.64, Sec.2, ZhiNan Rd., Wenshan District, Taipei City 116, Taiwan (R.O.C)

¹ chencm@nccu.edu.tw; ² chiu.wei.jia@gmail.com

Abstract- Due to the rapid development of information and communication technologies (ICTs), the Internet has become one of the most important communication media for journalism. Models of reader reception of information have changed with websites providing convenient and interactive user interfaces for communicating information instantly. However, this evolution has generated an unprecedented challenge for traditional communication media such as print newspapers. Online news sites need to create unique content to attract readers, and need to develop engaging community management services with Web 2.0 interactive mechanisms to compete with other websites and retain user attention. This work presents a novel online news platform that facilitates the construction of a University Press reader community. This platform can automatically analyze the reader community dataset of University newspapers, including an opinion deviation indicator, popularity indicator, and topicality indicator for each news story. Based on fuzzy inference, the proposed online news platform can select targeted news stories using these three indicators to identify top news stories that promote debate and interactivity within a reader community and promote communication efficiency and reader engagement. Experimental results reveal that the proposed interactive mechanisms satisfy the needs of most readers and correctly display top news stories that readers find interesting. Additionally, the proposed online news platform can assist journalists in understanding reader needs while promoting online social interaction.

Keywords- News Community; Data Mining; Fuzzy Inference; Users' Interaction and Feedback; Web 2.0 Technologies

I. INTRODUCTION

In applying collaborative recommender systems to the journalism field, many studies have focused on improving schemes that promote successful communication on online journalism platforms ^[1] ^[2] ^[3]. The rapid emergence of information and communication technologies (ICTs) has made the World Wide Web (WWW) to become an extremely important communication medium for news publication. The WWW has grown at a faster rate than any previous communication technology. Most newspapers now publish daily news on their websites, build digital news archives, and develop advanced portals for WWW news publication. The features of online journalism, such as

instant interactivity and the use of hypertext, have markedly impacted traditional press management and news content production. Newsrooms have had to extend working hours to report on breaking news and update stories. Thus, news editors have repackaged content to meet the requirements of online newsreaders. Currently, the competitors for online journalism are major commercial Web 2.0 portals that efficiently integrate diverse information, emphasize community interaction, and provide unique content for general users. If online news content remains unchanged, then competing with these new Web 2.0 portals is difficult. In other words, online journalism needs to create unique content to attract readers, and develop engaging community management services using same mechanisms as Web 2.0 sites to interact with readers and retain user attention. Pavlik and Sagan [4] pointed out that most WWW-based publications are evolving rapidly along with new technologies. Particularly, the idea of new media is transforming from traditional journalism business into digital age. One important function of online journalism is to provide an interactive and virtual space for public opinion, debate, and discussion. Some scholars have proposed that journalism in the digital age provides new opportunities for news media to re-define their roles in forming public opinion in democratic societies ^[5]. That is, news media should change their role from content providers to facilitators of public debate.

Innovative applications, such as implicit/explicit feedback and collaborative filtering recommendation systems, have been considered as improved mechanisms for information communication in public forums on online journalism platforms ^{[1] [2] [3]}. Additionally, Debnath, Ganguly, and Mitra ^[6] proposed a hybridization of collaborative filtering and content based recommendation system in which attributes used for content based recommendations are estimated from a set of linear regression equations obtained from a social network graph. Compared to the above-mentioned research, this work proposed a data mining scheme, which can mine useful implicit and explicit user data based on readers' online surfing profiles to construct a novel independent online news platform with community interactive mechanisms. This work compares the proposed online news platform with the original online University Press publication platform at National Chengchi University (http://www.uonline.nccu.edu.tw/) to confirm whether this

platform can facilitate construction of a news community. In contrast to the traditional news production process based on newsrooms, the proposed online news platform is based on user preferences by mining user profile data to automatically choose news stories for site homepage. That is, this work uses the novel online news publication system with a news story recommendation scheme to create a reader-centered design. First, this work develops a user-friendly platform supporting a public discussion forum for readers in the University Press community. Second, based on records of online browsing and social networking profiles of users, this work provides a novel tool that allows journalists and editors to survey online reader patterns when setting community agendas to assist in future newsroom operation. This work also develops efficient interactive mechanisms for the reference of future web-based journalism platforms. This work has the following goals.

- (1) Construct a journalism platform with appropriate interactive mechanisms that facilitate public discussion.
- (2) Mine reader profile data from the online journalism platform for managing engaged news communities.
- (3) Mine news information that facilitates reader interaction and the development of public agendas that affect reader perceptions.

II. LITERATURE REVIEW

A. News on the Internet

The World Wide Web (WWW) is now a ubiquitous communication medium^[7]. Pavlik and Sagan^[4] argued that the evolution of WWW-based news has three developmental stages. In the first stage, news content is simply transferred from print to the Internet. In the second stage, journalists start creating content with attributes such as hyperlinks, such that readers can instantly access another story by simply clicking a mouse. Moreover, interactive features, such as search mechanisms and customization, also enable readers to choose a news category according to their preferences. In the third stage, WWW-based news publishers outline online content based on unique website features, and use new forms of storytelling to fulfill reader needs for information in various forms. Pavlik and Sagan^[4] claimed that the third stage is the main developing trend of online news. That is, online news sites cannot simply publish print stories on homepages. To be competitive with major portal websites, the content of online news websites must be unique to attract reader attention. Notably, the advantages of journalism among mass communication media often involve dealing with public affairs closely related to daily life from social, political, and economical perspectives; this will prove to be the most unique feature of online journalism among information providers of Web 2.0 portals on the WWW. Therefore, news media should change their role from content providers to facilitators of public debate. Generally, online news has the following five main features: immediacy, accessibility, customization, low cost, and interactivity. Immediacy is immediate access and delivery of breaking

news. Accessibility means a reader can access news without the limitations of space and time. Customization is the degree to which news display or delivery fits personal preferences. Additionally, the cost of producing or accessing online news is considered more economical than traditional broadcasting of news. Additionally, online news is more interactive than traditional broadcasted news. Compared to the interactive mechanisms of television, radio, and newspapers, such as call-in shows and mail from audiences, the interactivity of online news has advantages in terms of reduced cost, and no time or space limitations. Furthermore, readers can customize their interactive spaces specifically on the employed collaborative based filtering recommendation mechanism. Readers can also provide feedback, share opinions, and discuss issues with others on public forums via ratings or comments. Moreover, readers can easily find preferred information via the innovative interactivity mechanism.

Goldhaber [8] proposed the concept of "attention economy," indicating that the idea of information economy cannot be applied to today's "information flux age" within the WWW networked society. Goldhaber [8] claimed that "scarceness" is the most valuable commodity and "attention" refers to the need for scarce and valuable commodities in selecting useful information in overloaded information and the digital flux environment. Understanding Internet user attitudes and their cognitive behavior is the key to understanding information communication on the Internet in the scope of the "attention economy." Davenport and Beck [9] noted that successful websites must embed rich information and interactive mechanisms with "stickiness" to attract reader attention. Thus, content must be relevant, convenient, and engage users; hence, users would naturally network with a virtual community, send feedback, such as useful social networking data, to a platform for further reference. As relevance and convenience are common characteristics of website portals, community engagement is the most advantageous feature of Web 2.0 portals. Therefore, the roles and influences of newsreaders or audiences are transforming in this information communication age. The traditional understanding of information usage and user satisfaction is no longer sufficient for interpreting information communication. Bruns ^[10] argued that the production chain is constructed by a linear process of "product-distribution-usage" in the industry age, and coined the term "produsage" to represent the combination of production and usage in the Internet age. The idea of production is no longer associated with used and familiar products in the industrial age, and has been replaced by dynamic produsage processes.

B. Group Discussion, Decision-Making Behavior and Crowd Wisdom Enabled by Virtual Communities

A large number of studies have focused on virtual communities (VCs) as the Internet has become the major social communication channel. Rheingold ^[11] described online communities as virtual groups of people exchanging thoughts, ideas and agendas via the Internet. Hagel and Armstrong ^[12] proposed the idea of a VC from a thematic

perspective; that is, a VC is a group of people on the Internet who manage a particular topic. Additionally, the number of participants in a VC typically increases over time. Adler and Christopher ^[13] and Igbaria, Shayo and Olfman ^[14] defined a VC from an instrumental perspective, and defined a VC as a group of people interacting via a certain medium/tool to construct a VC. Additionally, Lee, Vogel and Limayem^[15] addressed VCs from the interactive perspective, defining a VC as a network space supported by technology. The community must be based on communication and interaction among group members to build relationships. To summarize, a VC comprises elements of theme/agenda, media, people and interaction. Applying these features to examine community networking via web journalism has already been featured with the theme/agenda attracting interested people participating in VCs. However, to promote reader discussion, factors, such as instrument/medium, need to be constructed to facilitate opinion sharing and debate.

The perspectives for public opinion and agenda setting on the Internet can be categorized as optimistic and pessimistic studies ^[16]. Optimistic studies argued that the Internet provides democracy with a promising future, while pessimistic studies argued that the Internet will generate new conflicts and misunderstanding. Rheingold ^[11] argued that the Internet can transmit thoughts, facilitate debates about agendas, and mobilize mass opinion. Johnson ^[17] indicated that the Internet is the best tool for citizens participating in government policy decision-making. However, some scholars have argued with this point. For example, Sunstein ^[18], who examined 20 left- and right-wing blogs, discovered that the two readers groups were polarized after online debates. From Sunstein's ^[19] perspective, people with similar thoughts tend to stay together and share opinions they already hold; thus deepening biases and/or stereotypes. Shenk^[20] found fault with web users and the overflow of information on the Internet; that is, when users cannot find useful information, no effective discussion will take place, deepening public bias on the Internet as a new biased public sphere.

While many scholars have argued that crowd psychology is easily influenced or fooled, Surowiecki and Silverman^[21] argued that crowd decision-making based on an aggregation of information is better than that by a single group member. However, according to Surowiecki and Silverman^[21], not all crowds are wise. The following four criteria can be applied to identify crowd intelligence: opinion diversity, independence, decentralization, and aggregation. The first three criteria focus on keeping one's opinion independent within a group, while the fourth focuses on the necessary mechanism for transforming different personal decisions into a collective decision. When these criteria are not met, crowd intelligence is poor because members within a group easily take on the opinions of others, resulting in homogeneity of thought, not diversity of thought.

This study summarizes several principles for creating a novel online news platform that supports the expression of public opinion and debate and is based on the mechanisms of VCs. To form an ideal VC in which all readers can freely express their opinions about web news in public forums, the proposed online news system should, as Surowiecki and Silverman^[21] suggested, support a diversity of opinion, and the independence, decentralization, and aggregation of readers who visit an online news site. To encourage readers to express their opinions, one important feature of the proposed online news system is to mine the most opinion deviating news story from reader ratings and displays them as top news stories on a news site. To avoid reader opinions influencing those of other readers, the proposed online news system does not display previous ratings from others unless readers are sharing their ratings; hence, the independence of each reader is maintained. Additionally, the proposed online news system can take the frequency with which news story is browsed as a popularity indicator in the inference process; this indicator functions as an implicit feedback that does not depend on the explicit ratings of readers. That is, this study uses the proposed features to maintain opinion diversity in online communities. Last, the proposed online news system integrates all implicit and explicit feedback from readers, such as rating, browsing, mouse clicking and commenting, to explore useful added-value news information that can facilitate the construction of a news community.

C. User Feedback and the Collaborative Filtering Recommendation System for Web-Based Information Systems

Web-based information systems can have two types of feedback—explicit feedback and implicit feedback ^[22]. Explicit feedback is from users who provide subjective ratings or written descriptions after browsing and assessing an item. In contrast, for implicit feedback, the system estimates the degree of user interest in a particular item by monitoring user online behavior, such as the duration of time spent viewing a document, or mouse clicking on a digital object. The study by White, Rutheven, and Jose^[23] compared explicit relevance feedback and implicit relevance feedback in the application of query formulation in web item retrieval. Implicit feedback is based on user search behavior, while explicit feedback is from users marking documents based on their interest. In their study, the experimental results show that implicit feedback performs as well as explicit feedback, but strong evidence is needed to determine whether implicit feedback and explicit feedback are interchangeable. A study by Morita and Shinoda^[24] of an information filtering system for Internet news demonstrated a correlation between reading duration and rated interest. Their study utilized user behavior monitoring and measurement for user profile acquisition and feedback to filter out news stories that were not interested to users. The experimental results demonstrated that the score calculated by their proposed measurement scheme predicted user interest with high precision. White, Rutheven, and Jose^[23] compared the relevance of feedback from two systems experimentally to determine system effectiveness based on explicit and implicit feedback. In contrast to explicit feedback, which is generated by clicking a box to indicate that a document is relevant, the implicit feedback system re-ranked the list of relevant documents when a user moved a mouse over a document title. The system then generated a

list of possible expansion terms, which was added to a user's original query. The query expansion terms generated a new set of relevant documents. Joachims *et al.* ^[25] implemented eye-tracking techniques to assess the relevance of mouse clicking to query results. The mouse clicking records can view as a reliable reference in Web search results, but the view of point was only partially supported. Nichols ^[26] discussed the advantages of implicit rating by applying user browsing behavior for information filtering. The experimental evidence underscores the potential of implicit feedback as an alternative to explicit evaluation. The explicit model and implicit model have both data-mining advantages and drawbacks. Generally, most researchers agree that the explicit model gives more precise outcomes than the implicit model. However, the explicit model also generates a user cognitive burden or changes user reading habits.

To avoid problems generated by limited and improper use of explicit and implicit feedback, some studies [27] suggested using an information filter in system modeling, i.e., a collaborative filtering recommendation system, with reference to a user's social environment. The collaborative recommendation system is based on group members sharing evaluations. By referencing the opinions of others, the utility of filtered information will increase for each end user. Generally, the recommendation system is based on the concept that each evaluation is consistent with each user's future preference. The recommendation system is an information filtering technique for displaying information items, such as news stories, web pages, films, music, and books, that are likely consistent with a user's preferences. This system utilizes user community opinions, comments or feedback and helps individuals effectively identify content of interest from a potentially overwhelmingly amount of choices [28].

Tapestry, the first system to apply collaborative filtering technology to a recommendation system, was developed at the Xerox Palo Alto Research Center ^[29]. Tapestry was originally devised to handle the increased amount in incoming e-mails at workplaces. Additionally, Konstan et al. ^[30] implemented the GroupLens project for collaborative filtering of Usenet news. Their study showed that readers who spend considerable time reading a news article are likely to rate it highly, providing evidence of the correlation between time spent reading and explicit reader ratings. Avery and Zeckhause^[27] argued, in considering the negative factors associated with free-riding, that a centralized collaborative filtering mechanism that has features, such as implicit feedback, is necessary to avoid the cost of reading low-value messages, ensuring that sufficient and informative evaluations are generated. A social information filtering networked system developed by Shardanand and Maes^[31], the Ringo, makes personalized recommendations of music albums to users according to similarities between the interest profile of a user and those of other users. The Stanford digital library project, Fab, ^[32] was designed as a hybrid system incorporating content-based recommendation and collaborative recommendation methods. Users of this system can set their own profiles documenting their preferences, and rate content after reading. The system can recommend articles based on the profiles of other users with similar preferences. If the number of similar users for making a recommendation is lacking, the system retrieves data from previous ratings made by users, and measures content similarity to make recommendations.

Although the collaborative recommendation system has been studied and implemented on various academic or commercial websites, such as Amazon and eBay, some disadvantages emerged. For example, problems such as how to determine an appropriate rating scale, encourage user contributions ^[27], deal with problematic biased evaluators ^[33], limit the number of free-riding users, and solve problems, such as data scarcity and cold start, are difficult to solve. The "cold-start problem" for recommendation system developers occurs when a system cannot gather sufficient information about user profiles. In this case, the system cannot recommend items to users; hence, a vicious cycle from system to end users is created. To solve such problems, many studies have turned their attention to implicit feedback implementation.

III. SYSTEM DESIGN

A. The Design Issues and Communication Model of the Proposed Online News System for Building Virtual News Communities

This work developed a novel news publishing platform with news community networking and interactive features, allowing readers to post comments, feedback, ratings, and annotations. Based on the disadvantages associated with polarized opinions ^[34] which are commonly triggered in news and agenda debates, the proposed online news system was designed to mine the degree of popularity based on reader browsing behavior, the degree of opinion deviation, and the degree of topicality from ratings and comments posted by readers concerning news content to support community networking functionalities. First, according to the study by Surowiecki and Silverman^[21], a key to maintaining crowd wisdom is to retain the independence of each individual within a group. Thus, this work designed this opinion deviation degree as one feature in the proposed online news system. Additionally, a reader cannot view the ratings of others until this reader rates a selected news item. Moreover, this work utilized implicit feedback based on records of mouse clicks and explicit user comments to infer news item popularity. Implicit feedback here considers the situation in that some readers may not be used to commenting on a news story; thus, mouse clicking records (i.e., implicit feedback) are an important reference for determining whether a news story is popular on the online news platform. That is, according to the study by Surowiecki and Silverman^[21], retaining the diversity of group members is a key to maintaining crowd wisdom. Furthermore, the proposed system provides features for readers who are participating actively in public debates via comments, annotation, and feedback as the basis of computing degree of topicality. The system retrieves data from the discussion forum to infer the degree of topicality. Finally, the proposed system integrates the degree of opinion deviation, the

topicality and popularity scores for an overall sequence for news display and uses the fuzzy inference scheme to draw out different perspective news as top news stories, encourages readers to share their thoughts, and avoids opinion polarization^[34].

In short, this work constructs a user-friendly interactive news platform for building virtual news communities. Based on the discussion of methodologies that promote community networking, a novel communication model is proposed (Fig. 1). This model has three key elements for building a special news operating mode that facilitates news community interaction: news content, system, and community. Among these three key elements, a web server and a server program were constructed combined with database technologies as the basic infrastructure for community interaction. The news community is a group comprised of a system manger, journalists, editors, and online readers. News content is created by journalists, and published on the web server according to news subcategories. Additionally, the knowledge organization scheme is a basis for well-organized news content. Readers can select a category of interest, read news stories, and provide feedback via the interactive mechanisms. Platform technology is the medium between the system and community, facilitating the presentation of news content and data collection of community member interaction. Community members can express their opinions and give explicit or implicit feedback that is published through the interactive mechanisms provided by the system.

Through the operation processes of the system, community, and content, the unique "Produsage" is created using dynamic interactive processes.



B. System Architecture of the Proposed Online News System

The proposed online news system has 5 agents and 3 databases (Fig. 2). The news database stores the original news stories; the user feedback database stores user explicit and implicit feedback data; and the latent information database stores inference data from user feedback. The functionalities of 5 agents are detailed as follows:



Fig. 2 System architecture of the proposed on-line news system

1) The Information-Extraction Agent:

Based on the technology for information extraction, the proposed online news system extracts news information from the University Press website. Two procedures are used. First, a software crawler is used to extract news from HTML web pages. Second, according to the extracted news stories, the system extracts and saves useful data, such as news content, publication date, and the names of journalists, in the news database.

2) The News Display Agent:

The proposed online news system infers the top news

story indicator based on the popularity indicator, opinion deviation indicator, and topicality indicator. Based on the top news story indicator, the system lists top news stories and displays them on the homepage. The system also generates a news list based on popularity and opinion deviation degree and presents this list on the proposed online news site for reader reference. Readers can browse online news stories based on system recommendations, which are based on the indicators or personal interest. In other words, the news display agent of the proposed online news system displays useful news lists based on the popularity indicator, opinion deviation indicator, and topicality indicator, and

displays implicit rating data, such as browsing frequency of each news story, rating statistics, feedback, and annotation outcomes.

3) The Explicit Feedback Recording Agent:

This system has a variety of explicit feedback mechanisms. These mechanisms also encourage interaction, including rating news stories, responding to news events, and annotating news content. Rating is on a scale of 1–5 points. Each reader can only rate a news article once and cannot see the ratings of others until he/she rates the news article. In this way, the influence from other readers is reduced. Readers evaluate news from their own perspective to ensure personal independence within a group. The explicit feedback recording agent gathers explicit feedback from readers and records it in the user feedback database.

4) The Implicit Feedback Detection Agent:

The implicit feedback detection agent in the proposed online news system monitors user reading behavior. One detection task is to record user browsing and mouse clicking. The main purpose of implicit feedback is to reduce the impact of a cold start—users do not need to make a direct response because the system automatically retrieves each user's browsing path. This approach is consistent with the need for diversity in that every user can contribute feedback to the system.

5) The Inference Agent:

The inference agent, the core of the proposed online news system, is in charge of identifying the top news stories based on implicit and explicit feedback from readers.

The operation procedures of the proposed online news system are as follows.

- **Step 1.** The information extraction agent extracts original news data from the University Press website
- **Step 2.** The information extraction agent stores extracted news content in the news database
- **Step 3.** The news display agents reads news data stored in the news data database and identifies stories in different categories for display
- **Step 4.** The categorized news stories are displayed on a graphical user interface with a designed layout for readers
- **Step 5.** The explicit feedback recording agent and implicit feedback detection agent retrieve explicit and implicit information generated by readers from the graphical user interface
- **Step 6.** The explicit feedback recording agent and implicit feedback detection agent store explicit and implicit feedback in the user feedback database
- **Step 7.** The inference agent retrieves user feedback information from the user feedback database and determines the potential utility of news information to facilitate the construction of a

news community, and stores this information in the latent information database

Step 8. The news display agent reorganizes news content stored in the latent information database. At the same time, the news display agent updates the news and sends it back to the graphical user interface for display in a new arrangement for readers

Steps 3–8 are repeated by the proposed online news system. The displayed news popularity rank, opinion deviation rank, and the rank of top news stories are the statistical inference results computed by the proposed online news system based on explicit and implicit feedback from readers and user behavior while reading online news. Importantly, the news rank changes dynamically and stabilizes slowly based on reader using behavior. By combining the proposed online news system, news content, and community, the prod usage news process ^[10] is formed gradually.

C. Inference Mechanism for Exploring Top News Events

The list of top news stories plays a significant role for readers, regardless of whether it is of print newspapers or online news. To promote interaction in the news community, the proposed system automatically generates a list of top news stories based on explicit and implicit feedback from readers. As stated by Surowiecki and Silverman ^[21], "the best collective decisions are the product of disagreement and contest, not consensus or compromise." Thus, the proposed online news system encourages debate because debate is beneficial and facilitates interaction in the news community. Restated, online debate, ratings and annotations are the result of interactive behaviors, which can increase reader motivation to contribute to the collective intelligence.

The inference agent uses the following three criteria to select top news stories.

- (1) The opinion deviation indicators: Top news stories should be those that readers disagree upon most.
- (2) The popularity indicator: Top news stories should be those that readers will debate most within the online community.
- (3) The topicality indicator: Top news stories should be those that inspire readers to actively participate in public debate.

1) Opinion Deviation Indicator:

Opinion deviation in information theory is the degree of uniformity and diversity in public opinions. Entropy is typically as a measure of information chaos. Thus, this work utilizes entropy ^[35] to quantify the degree to which readers disagree about news stories. This study adopted reader ratings of news stories to calculate the entropy value as the opinion deviation indicator, as in Eq. (1),

$$E = -\sum_{i \in T} P_i \times \log(P_i)$$
(1)

where *E* is for opinion deviation for a each news event, *T* is the set of all reader evaluations, and P_i is the probability that a reader uses the number *i* rating while reading a news event. In this work, the rating is 1–5 points; thus, $T \in \{1,2,3,4,5\}$.

When $P_1 = P_2 = P_3 = P_4 = P_5 = 1/5$, the entropy value reaches the maximum which agrees with the fact that this news has the maximum opinion deviation indicator.

2) Popularity Indicator and Topicality Indicator:

Reader feedback is in terms of popularity and topicality. First, the proposed online news system retrieves data for reader mouse clicks on news stories as reader preferences to infer the popularity of a news story. Second, the proposed online news system provides features for readers to actively participate in public debate (e.g., comments, annotation, and feedback for other readers) and refers to the topicality score. In this work, a news publication is issued by the University Press once weekly. Equations (2) and (3) are for the popularity indicator, and Eqs. (4) and (5) are for the topicality indicator.

$$P = \text{Count Views(news)}$$
(2)

$$P_n = \frac{P - P_{\min}}{P_{\max} - P_{\min}} \tag{3}$$

where P is the original data of the popularity indicator, representing the number of times a news story has been browsed, P_n is for the scaling data after normalizing the popularity indicator, and P_{max} and P_{min} represent the maximum and minimum values of the popularity indicator within one week, respectively.

$$T = \text{Count Comments(news)} \tag{4}$$

$$T_n = \frac{T - T_{\min}}{T_{\max} - T_{\min}} \tag{5}$$

where T is the original data of the topicality indicator, representing the number of times feedback and annotation are generated, T_n is the scaling data after normalizing the topicality indicator; and T_{max} and T_{min} are the maximum and minimum values of the topicality indicator within one week, respectively.

3) Indicator of Top News Stories:

The website homepage is the most important page for first-time readers. Homepage content usually influences readers most and inspires them to browse other pages. Generally, most users browse homepages for only a few seconds before continuing to the next page. For community members in an online news network, the homepage containing top news stories should be considered as the most important page. If the top news stories which can attract readers to browse are displayed on the homepage of a news site, readers will likely take part in the debate of online news community. To logically identify top news stories, this work adopts fuzzy logic inference ^[36] to infer each top news story

MIOT

indicator as basis for news display based on the popularity indicator, topicality indicator, and opinion deviation indicator from readers' explicit and implicit feedback.

4) The Determined Fuzzy Membership Functions:

Figure 3 shows the determined membership functions for the opinion deviation indicator, popularity indicator, and topicality indicator. Center points c_1 , c_2 , and c_3 are individually determined as the first quartile, average, and third quartile value based on the data distribution of each corresponding indicator. Figure 4 shows the output membership function. This work heuristically sets the center points as 0.75, 1, and 1.25 for the linguistic variables Low, Middle, and High for inferring the indicator of top news story.







Fig. 4 Membership functions of output linguistic variable for the indicator of top news story

TABLE I THE DEFINED FUZZY RULE BASE FOR MINING TOP NEWS STORIES

	Fuzzy Rule		
No.	IF		THEN
1.	$P.Low \cap T.Low \cap O.Low$	\rightarrow	S.VL
2.	$P.Low \cap T.Low \cap O.Mid$	\rightarrow	S.L
3.	$P.Low \cap T.Low \cap O.High$	\rightarrow	S.L
4.	$P.Low \cap T.Mid \cap O.Low$	\rightarrow	S.L
5.	$P.Low \cap T.Mid \cap O.Mid$	\rightarrow	S.L
6.	$P.Low \cap T.Mid \cap O.High$	\rightarrow	S.M
7.	$P.Low \cap T.High \cap O.Low$	\rightarrow	S.L
8.	P.Low \cap T.High \cap O.Mid	\rightarrow	S.M
9.	$P.Low \cap T.High \cap O.High$	\rightarrow	S.H
10.	P.Mid \cap T.Low \cap O.Low	\rightarrow	S.L
11.	P.Mid \cap T.Low \cap O.Mid	\rightarrow	S.L
12.	P.Mid \cap T.Low \cap O.High	\rightarrow	S.M
13.	P.Mid \cap T.Mid \cap O.Low	\rightarrow	S.L
14.	P.Mid \cap T.Mid \cap O.Mid	\rightarrow	S.M
15.	$P.Mid \cap T.Mid \cap O.High$	\rightarrow	S.H
16.	P.Mid \cap T.High \cap O.Low	\rightarrow	S.M
17.	$P.Mid \cap T.High \cap O.Mid$	\rightarrow	S.H
18.	P.Mid \cap T.High \cap O.High	\rightarrow	S.H
19.	P.High \cap T.Low \cap O.Low	\rightarrow	S.L
20.	P.High \cap T.Low \cap O.Mid	\rightarrow	S.M
21.	P.High \cap T.Low \cap O.High	\rightarrow	S.M
22.	P.High \cap T.Mid \cap O.Low	\rightarrow	S.M
23.	$P \operatorname{High} \cap T.Mid \cap O.Mid$	\rightarrow	S.H
24.	P.High \cap T.Mid \cap O.High	\rightarrow	S.H
25.	P.High \cap T.High \cap O.Low	\rightarrow	S.H
26.	P.High \cap T.High \cap O.Mid	\rightarrow	S.H
27.	P.High $∩$ T.High $∩$ O.High	\rightarrow	S.VH

5) Building Fuzzy Logic Rules:

Fuzzy logic rules are based on three dimensions-opinion deviation, popularity and topicality-and the membership functions are defined accordingly. Twenty-seven fuzzy rules are drawn from the principles based on the concepts proposed by Surowiecki and Silverman^[21] for encouraging different opinion strategies. These fuzzy rules are used to identify the news stories that have a high opinion deviation indicator value, are debated most by readers in the online community, and have high browsing frequencies as the top news stories. Take fuzzy rule No. 1 as an example. This rule indicates that as the popularity indicator value decreases (P.Low), the topicality indicator value decreases (T.Low) and the opinion deviation indicator value decreases (O.Low), resulting in a very low output score of top news story (S.VL).

6) Defuzzification:

The purpose of defuzzification is to transform inference results into practical values. The most common defuzzification methods are the maximum criterion method (MC), weighted average method, center-of-gravity method, height method, and quadrature method ^[36]. The center-of-gravity method is the most popular and practical; thus, this work adopts this method for defuzzification. The center-of-gravity method calculates the center of a shadow area; the formula is as follows:

$$F = \frac{\sum_{i=1}^{L} \mu(y_i) \times y_i}{\sum_{i=1}^{L} \mu(y_i)}$$
(6)

where F is the score of top news story, L is the number of divided quantity points, $\mu(y_i)$ is the membership degree of y_i to membership functions, and y_i are the value of the i^{th} quantified point. Last, the system decides whether to display a news story as one of top news stories on the homepage based on the defuzzification result.

Compared to obtaining the score of top news story by combining the three indicators (i.e. opinion deviation, popularity and topicality indicators) linearly via corresponding weights, using fuzzy logic inference can determine top news story indicator based on human expert knowledge as well as can avoid the difficulty of determining appropriately the corresponding weights in the linear combination scheme.

D. System Implementation

Figure 5 shows the homepage of the proposed online news system, NewSee, which displays a list of selected top news stories (i.e. headlines) in left center column, news categories in the menu on the top, opinion deviation is displayed on the top of the right column, and view frequency is displayed on the bottom of the right column. Readers can click on any recommended news story using the top news story list, opinion deviation list, or browsing frequency list. Additionally, readers can also go to each news category through menu on the top and search for news stories of interest. Figure 6 shows the user interface of the individual news page and its features, which include community information in the top-right box, an annotation tool, and feedback window. Additionally, the user interface also has rating bars with user rating statistics, and displays comments from other readers (Fig. 7).



Fig. 6 The user interface and its features of individual news page



Fig. 7 Feedbacks interface for reader's commenting and rating

IV. EXPERIMENTAL DESIGN

This section describes the experimental design, research limitations, and analytical tools employed in this work.

A. Experimental Participants and Environment

Most online readers are students from different departments at National Chengchi University, Taiwan; the other students are from local universities. Even though visitors may come from different locations, readers, journalists, and editors in this experiment, they are interested in community-based news information. Table 2 compares the experimental environment for both the experimental group and control group. This experiment took 2 weeks to complete. The news sources are the first and second news issues published by the University Press during the 2009 fall semester. The readers who browsed the first issue in first week of the experiment comprise the experimental group. Readers who browsed the second issue during the second week are the control group, regardless of whether they were also readers in the first week of the experiment. This experiment was conducted under natural conditions. All participation was voluntary. The students freely registered and logged on to the proposed online news platform for browsing or reading news within the two-week period. Thus, a research limitation is that the experiment cannot control every participant to simultaneously take part in the experiment of two groups within two weeks. Moreover, readers could not be asked to read the same stories, or participate in the same debates due to using different news issues during conducting this experiment. Moreover, the news content for two readers groups differed, and generated different responses. Samples are limited to readers of the University Press, which may be problematic when generalizing research findings to other online news platforms.

TABLE II THE COMPARISON OF THE EXPERIMENTAL ENVIRONMENT FOR THE EXPERIMENT GROUP AND CONTROL GROUP

Compared Item	Experimental Group	Control Group
The difference on the method of top news stories' selection	Recommended by the proposed online news system with automatically inferring top news stories based on fuzzy inference	Chosen by intern editors in newsroom of University Press
The difference on the method of rating news event	Users can only see rating outcome of others and statistics information after he/her gives rating outcome on a news story	Readers can see other people's rating and statistics information while reading a news story
The difference on the presentation of social community information	While readers reading the news, the proposed online news system will show the statistics of opinion deviation rating, popularity rating, general readers rating distribution and numbers of feedbacks. Also, the user interface displays the news list based on the ranking of the opinion deviation indicator, popularity indicator, and topicality indicator.	The user interface only shows the statistics of rating distribution and numbers of feedback for each news story
Sameness	The proposed on-line news provides the same interactive fun feedback.	actionalities in terms of annotation and

B. Analytical Tools

As this work constructed an online news system that

supports social networking for newsreaders, observing and analyzing variances in networking density between the two groups is important. Moreover, the methodology for mining

web usage is important to understand web user paths and behavior. The purpose of mining web usage is to trace browsing logs on severs, identify reader browsing patterns to further document reader behavior, and improve web sever management quality. Many tools have been developed for such mining. In this work, the web utilization miner (WUM) ^[37] was employed as the primary analytical tool. This work uses the WUM mining tool to mine web usage to compare differences between the experimental group and control group. The selection of top news stories for the homepage is typically a reader's first stop on the online news platform. Readers usually look for what they prefer to browse or look through the news categories, which results in an aggregation tree with a deep structure. Normally, a website design is considered poor when readers cannot find top news stories or news stories they prefer in the first few pages on a website; conversely, if an online news system presents top news stories newsreaders prefer, it is considered a user-friendly design. To achieve this goal, this work utilizes WUM to mine the number of times a news story has been viewed in end nodes.

V. ANALYSIS OF EXPERIMENTAL RESULTS

This section analyzes experimental results. Section A introduces the basic information of the experimental and control groups. Section B introduces the analysis of recommended news stories by the proposed online news system based on reader feedback. Section C analyzes reader log data. Finally, Section D shows participant social networking analysis in the news forum.

A. Basic Statistical Information for the Experimental and Control Groups

Participants in the experimental group and control group are readers who read the first and second issues published in 2009 by the University Press, respectively. Table 3 compares basic statistical information of the published online news stories read by readers in the experimental and control groups. In the experimental group, 9 new stories were selected as the top news stories by intern editors. The proposed online news system re-selected 9 top news stories utilizing implicit and explicit reader feedback based on the proposed fuzzy inference scheme. In total, 35 news stories were published on the proposed online news platform for the experimental group; total browsing and comment frequencies were 1569 and 148, respectively. During the second week, readers who logged on to the proposed online news system were viewed as the control group. In the control group, there are 14 news stories selected as the top news stories by intern editors from the total 50 news stories. Total browsing and comment frequencies were 1604 and 185, respectively. Table 4 shows statistical information of readers in the two groups. At the end of the experiment, 45 readers participated in the two-week experiment, among them, 28 (69%) readers participated during both weeks

TABLE III COMPARISON OF BASIC STATISTIC INFORMATION OF THE PUBLISHED ONLINE NEWS USED BY THE READERS OF THE EXPERIMENTAL AND CONTROL GROUPS

	Number of Top News Stories	Overall Number of Published News Stories	Browsing Frequency of the Published News Stories	Comment Frequency of the Published News Stories
The Experimental Group	9	35	1569	148
The Control Group	14	50	1604	185

TABLE IV THE STATISTICAL INFORMATION OF READERS IN TWO GROUPS

	Number of Participants
The Experimental Group	35
The Control Group	38
Intersection of the Experimental	28(60%)
and Control Groups	28(09%)
Union of the Experimental and	15
Control Groups	43

B. Analysis of Top News Stories Recommended by the Proposed Online News System Based on Reader Feedback

Table 5 compares the top news stories recommended by the proposed news system and those chosen by intern editors at University Press. The left column in Table 5 lists the top news stories chosen by fuzzy inference based on reader feedback, and the right column displays the top news stories selected by intern editors at University Press. The four news headlines in grey are the news stories selected by both the proposed system and intern editors. The top news stories selected by the intern editors are evenly distributed in each category, whereas the top news stories recommended by the proposed system are that four are in the "Life" section, three are in the "School" section, one is in the "Arts" section, and one is in the "New Knowledge" section. From a reader stratification perspective, the top news stories selected by the proposed system are based on reader interests and preferences, while the intern editors, as information gatekeepers, chose news using journalism perspectives.

TABLE V COMPARISON OF TOP NEWS STORIES RECOMMENDED BY THE PROPOSED ONLINE NEWS SYSTEM AND DETERMINED BY THE INTERN EDITORS OF UNIVERSITY PRESS

Coded Number for the Top News Story Recommended by the Proposed Online News System	News Category	System Recommended Top News Story based on Fuzzy Inference	News Category	Top News Story Selected by Newsroom of University Press
1	生活	交大響應週一無肉日 吃素減碳護地球	校園	公投未過 教部仍補助博弈課程
1	Life	NCTU's Monday Vegetarian Day	school	Referendum Failed, Government Funds for

				Gambling Class
2	生活 Life	慶中秋 鄉民揪團河濱大會烤 Barbecue in Moon Festival	國際 International	澳政府辦圓桌會議 改善國際教育 The Australian Government's Roundtable Conferences for International Education Improvement
3	校園 School	公投未過 教部仍補助博弈課程 Referendum Failed, Government Funds for Gambling Class	校園 School	設備不足 北大承諾改善 Beijing University(China) Promise to Improve School's Facilities
4	新知 New Knowledge	古早人乀智慧:天然無患子 輕鬆去汙除 垢 The Wisdom of the Old Day: Natural Sapindus Easy Decontamination Detergent	生活 Life	慶中秋 鄉民揪團河濱大會烤 Barbecue in Moon festival
5	校園 School	教育部提高外籍生比例 學者樂觀其成 The Ministry of Education Agree Increasing the Proportion of Foreign Students, Scholars Optimistic About Its Prospects	藝文 Arts	十年反高學費歷程 《粉墨登場》 Independent Documentary Film on High Education
6	藝文 Arts	十年反高學費歷程 《粉墨登場》 Independent documentary film on high education	體育 Sports	永信杯 中原男排衛冕 北教女排三連霸 Winson Volleyball Cup, CYCU and NTUE Win the Third Consecutive Men and Women's Titles
7	生活 Life	烤肉扒柚 外籍生嚐鮮過節 Foreign Students' First Taste on Moon Festival	體育 Sports	體壇印度風 卡巴迪前進台灣 Kabbad come to Taiwan
8	校園 School	設備不足 北大承諾改善 Beijing University(China) promise to improve school's facilities	副刊 Supplement	大水退去,學生返校重建心生活 Flood Receded, the Students Back to School Lives
9	生活 Life	中興感恩敬師 黑森林變身許願園 Thanksgiving Day in NCHU	副刊 Supplement	大水退去,學生返校重建心生活 Flood Receded, the Students Back to School Lives

Table 6 shows the statistics of feedback from the experiemental group readers. The results reveal that browing frequency is 43% for the nine stories recommended by the proposed online news system. Moreover, 66% of readers provided feedback. The opinion deviation indicator score is 15.5. The browsing frequency is 31% for stories selected by intern editors, and 36% of readers provided feedback for opinion deviation indicator score of 9.8. Figure 8 shows the reader viewing frequency (popularity indicator), comment frequency (topicality indicator), and degree of disagreement (opinion deviation indicator) for the nine stories recommended based on fuzzy logic inference. The horizontal axis in Fig. 8 represents the corresponding top

news stories with coded numbers listed in Table 5. The top news stories chosen by the proposed online news system have better results for social interaction than those chosen by the intern editors; the news stories not recommended by the proposed online news system have the poorest results of interaction. The proposed top news story inference mechanism does not require that all three indicators have high scores. Taking the top news stories coded with Numbers 4, 8, and 9 as examples, the topicality indicators are relatively low (comment frequency); however, viewing frequency (popularity indicators) are as high as those for other top news stories.

TABLE VI STATISTICS OF FEEDBACK FROM THE EXPERIEMENTAL GROUP READERS

	Overall Number of News Items	Browsing Frequncy	Average Frequency of Browsing (MEAN)	Overall Number of Feedback	Average Number of Feedback	Opinion Deviation Indicator	Average Indicator of Opinion Deviation
Top News Story Recommended by the Proposed Online News System	9	681(43%)	75.67	97(66%)	10.78	15.5	1.72
Top News Story Selected by the Intern Newsroom Editors of University Press	9	490(31%)	54.44	53(36%)	5.89	9.8	1.09
Other News Not Recommended by the Poposed Oline Nws System	26	888(57%)	34.15	51(34%)	1.96	18.1	0.70
Total	35	1569	44.83	148	4.23	33.6	0.96

TABLE VII STATISTICS OF FEEDBACKS FROM THE CONTROL GROUP READERS

	Overall Number of News	Browsing Frequncy	Average Frequency of Browsing (MEAN)	Number of Reader Feedback	Avgerage Number of Feedback	Opinion Deviatrion Indicator	Average Indicator of Opinion Deviation
Top News Story Selected by the Intern Newsroom Editors of University Press	14	568(35%)	40.57	74(40%)	5.29	12.9	0.92
Other News Not Recommended by Intern Newsroom Editors of University Press	36	1036(65%)	28.78	111(60%)	3.08	25.7	0.71
Overall news	50	1604	32.08	185	3.7	38.6	0.772



Fig. 8 System recommended top news story and its feedbacks from readers

Table 7 shows statistical results for control group reader feedback for top news stories and other news stories. The popularity indicators for top news stories selected by intern editors is roughly 35% for frequency and 40% for comment frequency; the opinion deviation indicator score is 12.9. The popularity of other news stories for the control group is about 65% (viewing frequency) and 60% for topicality (comment frequency); the opinion deviation indicator score is 25.7. Figures 9, 10, and 11 compare average popularity (viewing frequency), topicality (comment frequency) and opinion deviation degree (disagreement degree) between the experimental group and control group. The length of bar chart shows the proportional values of three indicators for the top news stories (i.e. headline news), non-headline news and all news stories, respectively. Obviously, the interaction level of the experimental group is superior to the control



Fig. 11 Bar chart on the comparison of average indicator of opinion deviation for two groups

Table 8 compares opinion deviation in the experimental and control groups. Comparison results show that the average opinion deviation in the experimental group is higher than that of the control group. In the experimental group no matter which indicator. In conclusion, the proposed online news system promotes reader interaction.







Fig. 10 Bar chart on the comparison of average commenting frequency for two groups

design, readers in the experimental group cannot see the ratings of others, thereby avoiding influence from the opinions of others, while those in the control group could see the ratings of others while rating news stories. We assume that displaying the ratings and feedback from others affects rating results, particularly for controversial issues. That is, the rating mechanism in the online news system for readers in the experimental group encourages opinion diversity and the retention of one's own opinion.

TABLE VIII COMPARASION OF THE OPINION DEVIATION OF THE EXPERIEMENTAL AND CONTROL GROUPS

	Average Indicator of the Opinion Deviation
The Experimental Group	0.96
Then Control Group	0.77

C. Analysis of Reader Log Data

In this work, the WUM data-mining tool was employed to mine reader log data and request data to elucidate reader behavior. By analyzing reader browsing paths in the proposed online news system, one can evaluate the dynamic news display process of this system. Generally, if readers can efficiently browse the news platform and stop at a news story of interest, then the proposed online news system successfully fulfills the goal as platform that attracts reader attention. From web usage analysis or a system manager's perspective, when readers can get what they want to read from upper nodes in the net rather than lower nodes, then the information structure design is good. The key design of the proposed online news system with the fuzzy inference mechanism as a reference for top news stories is founded on such a consideration. Analysis by WUM shows reader browsing behavior, corresponding to top news stories and news stories in each category, in terms of statistics for the experimental and control groups.

This study site for accessing online news publication was designed based on URL specifications, such that reader requests can be sent to the system sever. Thus, the browsing paths of all participants can be recorded and analyzed. In Table 9, (a) refers to access to homepage content from the server to issue requests, (b) refers to requests users make to access to news story No. 1, (c) is a request to access news content in the "Arts" category, and (d) refers to access to news story No. 2. Even though different readers may have different browsing patterns, this measure is a very useful reference to determine whether this browsing path is effective.

TABLE IN EXAMPLE OF WEBSITE LOG FILE	TABLE	IX EXA	MPLE OF	WEBSITE	LOG FILE
--------------------------------------	-------	--------	---------	---------	----------

168.0.0.1 [06/Jan/2009:00:10:00 +0800] "GET / HTTP/1.1" 200 1000	(a)
168.0.0.1 [06/Jan/2009:00:10:30 +0800] "GET /news/1 HTTP/1.1" 200 1000	(b)
168.0.0.1 [06/Jan/2009:00:11:00 +0800] "GET /category/art/ HTTP/1.1" 200 1000	(c)
168.0.0.1 [06/Jan/2009:00:11:30 +0800] "GET /news/2 HTTP/1.1" 200 1000	(d)

Table 10 shows the overall web usage analysis of reader browsing paths from the homepage top news story list and news category to individual news content pages. In total, 174 effective browsing records exists for the experimental group, and 242 effective browsing records exists for the control group. Only nine top news stories exist for the experimental group, compared with 14 for the control group. Experimental results show that readers in the experimental group have a higher percentage (46%) of browsing top news stories than readers in the control group (41%). Conversely, readers in the control group tend to select news stories from news categories rather from the list of top news stories on the homepage. Experimental results indicate that publishing news based on system inferences will help readers efficiently find the news in which they are interested. The proposed online news system with news story recommendation moves news stories on deep nodes to front nodes, based on the reader choice, which is the ideal of crowd wisdom, and reader feedback to the system to meet the needs of readers, thereby enhancing platform "stickiness" and its ability to attract reader attention.

	Top News Story	Arts	School	Lives	Sports	Supplemental	New knowledge	International	Total
The Experimental	80	8	16	14	25	7	13	11	174
Group	46%	5%	9%	8%	14%	4%	7%	6%	1/4
The Control	99	10	31	24	24	19	16	19	242
Group	41%	4%	13%	10%	10%	8%	7%	8%	242

TABLE X ANALYSIS ON BROWSING PATH

D. Social Network Analysis

The proposed online news system has a nested feedback forum (Fig. 7) that allows readers to give feedback directly on the forum. However, instead of using this feedback design to interact with others, some readers preferred giving feedback directly in the next column while referring to or commenting on other opinions of users. To identify the interaction pattern of social networking in this experiment, the social network analysis program UCINET (http://www.analytictech.com/ucinet/) is used.

Formula (7) is the definition matrix on interaction relation of community networking in this work. The size of this matrix is $m \times m$, where m represents the number of people participating in a community, the factors on the

diagonal line are 0, $r_{i,j}$ refers to the interactive

relationship value of reader u_i to reader u_j . The formula is as follows (Eqs. (8) and (9)):

$$relation \ martix = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,m} \\ r_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ r_{m,1} & \cdots & \cdots & r_{m,m} \end{bmatrix} (7)$$
$$r_{i,j} = \begin{cases} 0 & \text{if } i = j \\ \sum_{n \in N_{i,j}} S(n, u_i, u_j) & \text{otherwise} \end{cases} (8)$$

(9)

 $S(n, u_i, u_j) = \begin{cases} 1 & \text{if } u_i \text{ to } u_j \text{ is direct interaction in the n}^{\text{th}} \text{ news} \\ \frac{1}{R(n, u_i) - 1} & \text{if } u_i \text{ to } u_j \text{ is indirect interaction in the n}^{\text{th}} \text{ news} \end{cases}$

where $N_{i,j}$ is the set of news stories about which readers u_i and u_j interact, $S(n, u_i, u_j)$ is the formula for computing the interactive relationship between readers u_i and u_j in the n^{th} news. If reader u_i directly interacts with reader u_j in the n^{th} news, i.e., reader u_i is using the designed nested feedback forum to interact with reader u_j , then the interactive relationship between reader u_i to reader u_j is scored 1 point. If reader u_i interact indirectly with reader u_j in the n^{th} news, *i.e.*, reader u_i interact indirectly with reader u_j in the n^{th} news, *i.e.*, reader u_i interact u_i did not use the nested feedback forum, but reader u_i expressed his/her opinion after multiple readers u_j expressed their opinions in the forum, the resulting score is $\frac{1}{R(n, u_i) - 1}$, where $R(n, u_i)$ represents the ranking number of reader u_i put his/her opinion on the forum for the n^{th} news.

In short, this work assigned 1 point to each reader feedback item. However, the score calculated is slightly different based on whether a reader uses the nested feedback features. If a reader uses the nested feedback, then his/her feedback scores is 1 point. If readers give feedback on independent spot, then 1 point is proportionally assigned to the previous readers who gave feedback on a news story. In this work, networking density is further computed based on the interactive relationship between each reader's node and the nodes of other readers to represent the density and frequency of social networking within the community. Formula (10) is as follows:

$$T = \frac{\sum_{i,j \in \{1,2,\dots,m\}} r_{i,j}}{|N|(|N|-1)}$$
(10)

where T is community networking density, the numerator is the sum of the interaction score for two participants, and |N| is the total number of the participants in a community.

Table 11 shows the density of the online news community in the two groups. The average community networking density computed based on reader interaction in the news agenda debate in the experimental group and control group is 0.0983 and 0.1076, respectively; the standard deviation is 0.2453 and 0.3315, respectively. This indicates that readers in the control group interacted more than readers in the experimental group, but the interactive relationships of each reader in each node in the experimental group are more even than those in the control group. In the experimental group, each reader participated in debates, while interaction among readers in the control group was more polarized, indicating that not all readers in the control group were involved in community networking activities. The community density of readers in the experimental group and control group on the interactivity of top news stories is respectively 0.0878 and 0.0674, indicating that reader interactivity about top news stories in the experimental group is better than that in the control group. This supports demonstrates that the proposed features, such as the collaborative news recommendation and fuzzy inference mechanisms, promote reader debate about news. Not only does the proposed online news system increase the value of online news for virtual community networking, it also enhancing interactivity of the formed community through debate.

	Experimental Group	Control Group	Experimental Group's Interactivities to Top News Story	Control Group's Interactivities to Top News Story
Average Density of Community	0.0983	0.1076	0.0878	0.0674
Standard Deviation of Community Density	0.2453	0.3315	0.2129	0.2123

TABLE XI THE COMPARISON ON NETWORKING DENSITY OF TWO GROUPS

VI. CONCLUSIONS AND FUTURE WORK

This work presents a novel online news system that automatically analyzes reader feedback using the opinion deviation indicator, popularity indicator, and topicality indicator to automatically list top news stories. The proposed online news system also provides social networking functionalities for readers to promote debate and interactivity to increase the efficiency of communication and engagement in a reader community. Experimental results reveal that the proposed mechanisms satisfy the needs of most readers searching for top news stories and assist journalists in understanding reader needs while promoting online social networking. Moreover, this work compares top

news stories chosen by the fuzzy inference scheme based on reader browsing, feedback, comments, and opinion deviation with the top news stories selected by an intern newsroom editor. Generally, the criteria for selecting top news stories by editors at University Press are "hot news," "objective decision" and "closely related to University life," whereas the criteria for selecting top news stories by the proposed online news system are reader browsing frequency and, more importantly, reader comment frequency and opinion deviation degree. Additionally, community networking density of readers in the experimental group on the agenda interactivity of top news stories is higher than that of readers in the control group.

Although the proposed online news system is indeed helpful in building an online news community that promotes reader interaction, several issues warrant further investigation. First, the proposed online news system does not have a perfect mechanism that can completely replace the decisions of editors and journalists when choosing top news stories; thus, whether one should combine the editorial decision process with reader needs to generate a list of top news stories should be considered. Moreover, adopting a tool that can monitor human interaction with information, such as Morae (http://www.techsmith.com/morae.asp), to understand and analyze whether a list of top news stories benefits readers is another valid research direction. Moreover, detailed surveys of reader opinions about the user interfaces are needed to improve system design and increase the user-friendliness of the online news platform. Also, although the public forum on the proposed online news system is helpful for issue debate and community building, developing a mechanism for journalists to lead public discussions in online news virtual spaces is also important. Finally, based on social network structure among news users, using the collaborated decision making to determine top news story in news communities can be explored in the future [38].

REFERENCES

- Qiu, J., Liao, L., & Li, P. (2009). News Recommender System Based on Topic Detection and Tracking. *Lecture Notes in Computer Science*, 5589, 690–697.
- [2] Das, A. S., Datar, M., Garg, A., & Rajaram, S. (2007). Google News Personalization: Scalable Online Collaborative Filtering. *Proceedings of the 16th international conference on World Wide Web*, 271-280.
- [3] Lima, W. (2011). Recommendation Systems that Establish new Forms of Representational Reality: Eliminating the Dividing Line between Information Emitter and Receptor of Journalistic Information. *International Journal of Interdisciplinary Social Sciences*, 6(3), 289-296.
- [4] Pavlik, J., & Sagan, P. (1997). The Future of Online Journalism: Bonanza or Black Hole?. *Columbia Journalism Review*, 36(2), 30-38.
- [5] Mutz, D.C. & Soss, J. (1997). Reading Public Opinion: The Influence of News Coverage on Perceptions of Public Sentiment. *The Public Opinion Quarterly*, 61(3), 431-451.
- [6] Debnath, S., Ganguly, N., & Mitra, P. (2008). Feature Weighting in Content Based Recommendation System Using

Social Network Analysis. *The 17th international conference on World Wide Web*, 1041-1042.

- [7] Morris, M., & Ogan, C. (1996). The Internet as Mass Medium. Journal of Computer-Mediated Communication, 46(1), 39-50.
- [8] Goldhaber, M.H. (1997). The Attention Economy and the Net. First Monday, 2(4), from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/arti cle/viewArticle/519/440
- [9] Davenport, T., & Beck, J. (2001). The Attention Economy: Understanding the New Currency of Business. Harvard Business School Press.
- [10] Bruns, A. (2008). Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage, Peter Lang, New York.
- [11] Rheingold, H. (2000). *The Virtual Community: Homesteading* on the Electronic Frontier. MIT press.
- [12] Hagel, J., & Armstrong, A.G. (1997). Net Gain: Expanding Markets through Virtual Communities. Harvard Business School Press, Boston, MA.
- [13] Adler, R.P., & Christopher, A. J. (2009). Internet Community Primer Overview and Business Opportunities. Retrieved November 24, 2009, from http://www.digitalplaces.biz/pages/printable_html.html.
- [14] Igbaria, M., Shayo, C., & Olfman, L. (1999). On Becoming Virtual: The Driving Forces and Arrangements. Proceedings of the 1999 ACM SIGCPR Conference on Computer Personnel Research, 27-41.
- [15] Lee, F.S.L., Vogel, D., & Limayem, M. (2002). Virtual Community Informatics: What We Know and What We Need to Know. *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2863-2872.
- [16] Hill, K., & Hughes, J. (1998). Cyberpolitics: Citizen Activism in the Age of the Internet. Rowman & Littlefield Publishers.
- [17] Johnson, S. (1998). Internet Changes Everything: Revolutionizing Public Participation and Access to Government Information through the Internet. *Administrative Law Review*, 50, 277-337.
- [18] Sunstein, C. (2004). Democracy and Filtering. Communications of the ACM, 47(12), 57-59.
- [19] Sunstein, C. (2007). *Republic. com 2.0.* Princeton University Press.
- [20] Shenk, D. (1998). Data Smog: Surviving the Information Glut. HarperOne, San Francisco.
- [21] Surowiecki, J., & Silverman, M. (2007). The Wisdom of Crowds. American Joural of Physics, 75(2), 190-192.
- [22] Sharma, S.K. & Suman, U. (2013). A Framework of Hybrid Recommender System for Web Personalisation. *International Journal of Business Information Systems*, 13(3), 284-316.
- [23] White, R., Ruthven, I., & Jose, J. M. (2002). The Use of Implicit Evidence for Relevance Feedback in Web Retrieval. *Proceedings of the 24th BCS-IRSG European Colloquium on IR Research: Advances in Information Retrieval*, 93-109.
- [24] Morita, M., & Shinoda, Y. (1994). Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 272-281.
- [25] Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F., & Gay, G. (2007). Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search. ACM Transactions on Information Systems, 25(2), Article 7.

- [26] Nichols, D.M. (1997). Implicit Rating and Filtering. Proceedings of the Fifth DELOS Workshop on Filteringan d Collaborative Filtering, 31-36.
- [27] Avery, C., & Zeckhauser, R. (1997). Recommender Systems for Evaluating Computer Messages. *Communications of the ACM*, 40(3), 88-89.
- [28] Resnick P., & Varian H. (1997). 'Recommender Systems', Communications of the ACM, 40(3), 56-58.
- [29] Goldberg, D., Nichols, D., Oki, B., & Terry, D. (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35(12), 61-70.
- [30] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., & Riedl, J. (1997). GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, 40(3), 77-87.
- [31] Shardanand, U., & Maes, P. (1995). Social Information Filtering: Algorithms for Automating "Word of Mouth". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 210-217.
- [32] Balabanović, M., & Shoham, Y. (1997). Fab: Content-Based,

Collaborative Recommendation. *Communications of the ACM*, 40(3), 66-72.

- [33] Palme, J. (1997). *Choices in the Implementation of Rating*. Oldenbourg, Vienna, Austria.
- [34] Janis, I. (1982). *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Houghton Mifflin, Boston.
- [35] Shannon, C. (2001). A Mathematical Theory of Communication. ACM SIGMOBILE Mobile Computing and Communications Review, 5(1), 3-55.
- [36] Lin, C.-T., & Lee, C.S.G. (1996). Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems, Prentice-Hall.
- [37] Spiliopoulou, M., & Faulstich L. (1999). Wum: A Web Utilization Miner. *Proceedings of the International Workshop on the Web and Databases*, 109-115.
- [38] Jiang, Y.C., Hu, J., & Lin, D. (2011). Decision Making of Networked Multiagent Systems for Interaction Structures. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(6), 1107-1121.