Clustering Editors of Wikipedia by Editor's Biases

Akira Nakamura, Yu Suzuki, and Yoshiharu Ishikawa
Graduate School of Information Science, Nagoya University
Furo, Chikusa, Nagoya, Aichi 464–8603, Japan
Email: {nakamura, suzuki, ishikawa}@db.ss.is.nagoya-u.ac.jp

***Abstract*—Wikipedia is an Internet encyclopedia where any user can edit articles. Because editors act on their own judgments, editors' biases are reflected in edit actions. When editors' biases are reflected in articles, the articles should have low credibility. However, it is difficult for users to judge which parts in articles have biases. In this paper, we propose a method of clustering editors by editors' biases for the purpose that we distinguish texts' biases by using editors' biases and aid users to judge the credibility of each description. If each text is distinguished such as by colors, users can utilize it for the judgments of the text credibility. Our system makes use of the relationships between editors: agreement and disagreement. We assume that editors leave texts written by editors that they agree with, and delete texts written by editors that they disagree with. In addition, we can consider that editors who agree with each other have similar biases, and editors who disagree with each other have different biases. Hence, the relationships between editors enable to classify editors by biases. In experimental evaluation, we verify that our proposed method is useful in clustering editors by biases. Additionally, we validate that considering the dependency between editors improves the clustering performance.**

***Keywords*—Bias, Wikipedia, Edit Histories, Peer Reviews**

I. INTRODUCTION

Wikipedia[1] is an Internet encyclopedia and one of the most successful user-generated content (UGC). Wikipedia has high comprehension and the latest information because anyone can edit articles. However, there are not only high quality editors but also low quality editors for this reason. Low quality editors would write improper statements. Therefore, many articles that have low quality texts, such as false descriptions. Additionally, editors' biases are reflected in edit actions because editors act on their own judgments. In such a case, the articles have low credibility. For these reasons, the credibility of Wikipedia has been seen as a problem.

In order to solve this problem, we proposed a technique for automatically measuring the credibility of texts [1] based on the assumption that the texts which remain beyond many edits are credible. Our system enables users to refer to the credibility of each text.

However, measuring the credibility is difficult if many people have opinions to the articles, such as articles about politics, famous person, and common sense of specific countries. This is because, if there are many opinions and is no consensus about a topic, we cannot decide whether each opinion is credible or not. Wikipedia has the principle "Neutral point of view[2]." All editors must edit based on this principle. However, the neutral point of view would vary depending on the country, the culture and the person. *Edit warring* between several groups is due to difference in neutral viewpoints. Each group believes that their own opinion must be correct. From the viewpoint of the one, the texts written by the other seems wrong. When two groups of editors in an article have different opinions and conflict with each other, one group would delete texts written by the other group, and then the edit warring occurs between two groups of editors. In this case, the majority group may be judged to be more credible than the minority group. If people have multiple viewpoints in an article, majority rule is inappropriate for judging the correctness of each viewpoint. Therefore, we need to develop how to aid users to judge the credibility of texts without calculating the credibility of each text.

To accomplish this goal, we propose a method

[1]http://www.wikipedia.org/

[2]http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

978-1-4799-2902-3/13 $31.00 © 2013 IEEE
DOI 10.1109/WI-IAT.2013.50

for clustering editors in an article by biases, and visualize the relationship of editors by network graph. We distinguish texts' biases by using editors' biases. If each text is distinguished such as by colors, users can utilize it for the judgments of the text credibility.

Our system uses relationships between editors: agreement and disagreement. We identify editors' relationships from *edits of leaving* and *edits of deleting*. Editors leave texts based on fact or bias which is similar to their own. On the other hand, editors delete texts including some errors or based on bias which are opposed to their own. Therefore, we can consider that editors leave texts written by editors that they agree with, and delete texts written by editors that they disagree with. Furthermore, we assume that the stronger editors agree or disagree, the more editors leave or delete in quantity. Based on this assumption, we can determine the type of relationships between editors and the strength of the relationships from their edit actions. In addition, we can consider that editors who agree with each other have similar biases, and editors who disagree with each other have different biases. Hence, the relationships between editors enables to classify editors by biases.

II. Related Work

First, we introduce related work about assessing the quality or the credibility of Wikipedia articles. Edit history based approach [2], [3] is a major approach for assuming the quality of articles. The key idea of this approach is that the longer texts remain beyond multiple edits, the more credible the texts and editors. Their system measures the editor credibility, and the system measures the article credibility by using the editors' credibility. We [1] also measured the credibility of articles in Japanese Wikipedia using this approach. The important point of this approach is, when texts are deleted, our system considers the credibility of the editor who deleted the texts to measure the credibility of the editor who inserted the texts. These studies are similar to our study proposed in this paper in that the systems utilize the edit histories to aid users to judge the credibility of each text in articles. However, in these methods, a minority opinion might be excluded as an incredible opinion because their systems do not distinguish agreement relationships between editors but measure the credibility of editors based on the degree of remaining texts. Therefore, we do not measure the credibility of editors, but group editors who agree with each other. Our study differs from their studies in this respect.

Next, we introduce the related work about biases and editor network in Wikipedia. Paolo et al. [4], [5] assumed that editors's biases arise from difference of languages, and they researched this task. Their studies are similar to our study in the assumption that editors have various viewpoints. However, in our study, we focus on the bias of editors' opinions and how the bias affects the contents of Wikipedia. Jesus et al. [6] investigated the bipartite network of articles linked by common editors. This approach is similar to our approach in visualizing networks in Wikipedia, but we do not consider relationships between articles and editors because we focus on editor network in an article. Brandes et al. [7] classified editors and visualized editors' network in the edit warring by defining editors as nodes and disagreement relationships between editors as edges. Their techniques are similar to our proposed method in that an edge represents an editor and a node represents the relationship between editors and the system weights edges based on data extracted from the edit history, but is different in not using agreement relationships for clustering. In addition, their method of weighting and clustering are also different from our method.

III. Relationship between Editors

In this section, we describe how to classify editors in one article by editors' biases. Fig. 1 shows the overview of our proposed system. First, the system finds how many characters editors leave and delete from the edit history. Next, the system calculates the relationships for each pair of editors using the extracted features. Finally, the system classifies editors into several clusters based on the relationships between editors. Consequently, each of clusters consists of editors with similar biases.

A. Data Extraction from the Edit History

We extract statistical data about the editors' actions of leaving and deleting from the edit history.

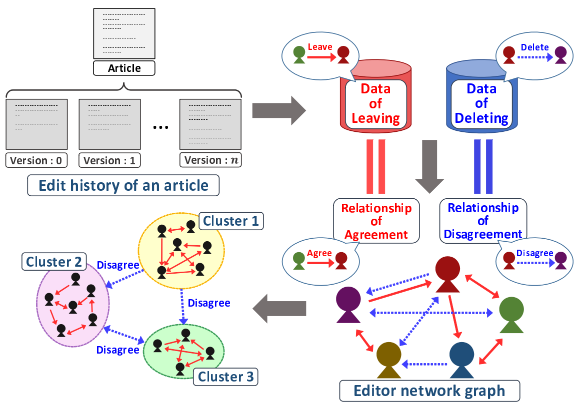

Fig. 1. Overview of our proposed system.

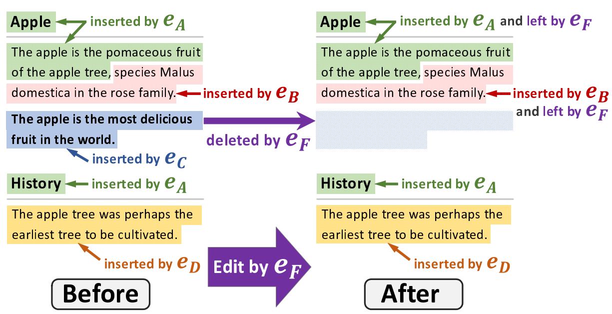

Fig. 2. An example of the edit by e_F.

The edit history data includes the complete text and editor's name in every version. Therefore, by analyzing the edit history, we can find which texts are inserted and who inserts the texts. In addition, we can find who leaves or deletes the texts.

Example 1: Fig. 2 gives an example of the edit. Editor e_A, e_B, e_C, and e_D wrote several texts before the edit by e_F. Then, e_F deletes the texts written by e_C and leaves the rest of the texts written by e_A, e_B, and e_D.

In this edit, e_F deletes 41 characters written by e_C except special characters such as space, period, comma, etc., and leaves 82 characters. The system regards only characters in the sections edited by e_F as left characters. This is because editors may not browse the whole of an article when the article has a lot of texts.

B. Calculation of the Edit Weights between Editors

In this section, we calculate the edit weights $W(e_s, e_t)$ between two editors e_s and e_t by using the statistics which are extracted from the edit history. $W(e_s, e_t)$ represents the strength of relationship (agreement or disagreement) from one editor $e_s \in E$ to another editor $e_t \in E$, where E is a set of all editors in the article.

As mentioned earlier, we identify the strength of relationships from edits of leaving and deleting. First, we will introduce $W_l(e_s, e_t)$ and $W_d(e_s, e_t)$. $W_l(e_s, e_t)$ represents the strength of the agreement relationship and is calculated from edits of leaving. $W_d(e_s, e_t)$ represents the strength of the disagreement relationship and is calculated from edits of deleting. In order to calculate of these weights, we consider the edit dependency between editors. When e_s leaves texts written by e_t mainly, and when e_t's texts are left mainly by e_s, the weight of leaving has a high value. On the other hand, when e_s leaves texts written not only by e_t but also by various editors, and when e_t's texts are left by various editors including e_s, the weight of leaving has a low value. The weight of deleting is also similar. We define these two weights as follows:

$$W_l(e_s, e_t) = \frac{\log(c_l(e_s,e_t)+1)}{\log(C_l^{sou}(e_s)+1)} \left(\log_{10} \frac{C_l^{art}}{C_l^{tar}(e_t)} + 1\right), \quad (1)$$

$$W_d(e_s, e_t) = \frac{\log(c_d(e_s,e_t)+1)}{\log(C_d^{sou}(e_s)+1)} \left(\log_{10} \frac{C_d^{art}}{C_d^{tar}(e_t)} + 1\right). \quad (2)$$

Now, we define $c_l(e_s, e_t)$ as the total number of characters written by e_t and left by e_s in an article. Additionally, We define the total number of characters which are left by e_s as $C_l^{sou}(e_s) = \sum_{e_j \in E} c_l(e_s, e_j)$ Then, we define the total number of left characters written by e_t as $C_l^{tar}(e_t) = \sum_{e_i \in E} c_l(e_i, e_t)$. $C_l^{art} = \sum_{e_i \in E} \sum_{e_j \in E} c_l(e_i, e_j)$ is the total number of left characters in the article. We calculate $W_d(e_s, e_t)$ in the same way.

We calculate $W_l(e_s, e_t)$ based only on edits of leaving and $W_d(e_s, e_t)$ based only on edits of deleting, that is, we do not consider the correlation between leaving and deleting. Without considering the correlation, when the number of characters which is written by e_t and left by e_s is particularly more than the number of characters which is written by e_t

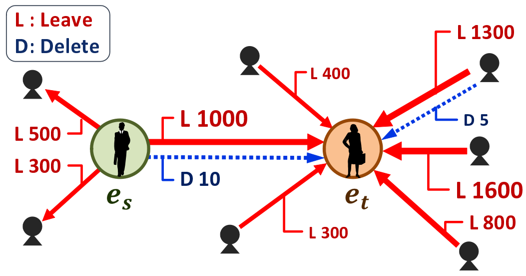

Fig. 3. An example of the edit relationship between e_s and e_t for Example 2.

TABLE I. THE VALUES OF $W_l(e_s, e_t)$ AND $W_d(e_s, e_t)$ IN FIG.3

$c_l(e_s,e_t)$	$C_l^{sou}(e_s)$	$C_l^{tar}(e_t)$	C_l^{art}	$W_l(e_s,e_t)$
1000	1800	5400	50000	1.81
$c_d(e_s,e_t)$	$C_d^{sou}(e_s)$	$C_d^{tar}(e_t)$	C_d^{art}	$W_d(e_s,e_t)$
10	10	15	10000	3.82

and deleted by e_s, the weight of leaving from e_s to e_t may be higher than the weight of deleting from e_s to e_t. Therefore, it is inappropriate to calculate weights of the edit action by a simple method such as a subtraction of the weight of deleting from the weight of leaving.

Example 2: In Fig. 3, e_s agrees with e_t, leaves 1000 characters and deletes 10 misused characters. When a lot of editors agree with e_t, $C_l^{tar}(e_t)$ becomes high and the weight of leaving becomes low. In addition, when e_t's texts are rarely deleted, $C_d^{tar}(e_t)$ becomes low and the weight of deleting becomes high. Moreover, when e_s rarely deletes, $C_d^{tar}(e_t)$ becomes even lower and the weight of deleting becomes even higher. As shown in Table I, $W_d(e_s, e_t)$ becomes higher than $W_l(e_s, e_t)$ even though e_s agrees with e_t. If we decide the edit weights by only using $W_l(e_s, e_t)$ and $W_d(e_s, e_t)$, the system judges e_s disagrees with e_t.

It is necessary to consider the correlation between edits of leaving and deleting so as to decide the edit weights appropriately. We normalize the weight of leaving and the weight of deleting so that the weight of leaving/deleting becomes low when the number of left/deleted characters is extremely lower than the number of deleted/left characters. We define the ratio of the number of left characters to the sum of the number of left and deleted characters, and the ratio of the number of deleted characters to the sum as follows:

$$r_l(e_s, e_t) = \frac{c_l(e_s, e_t)C_d^{art}}{c_l(e_s, e_t)C_d^{art} + c_d(e_s, e_t)C_l^{art}} \quad (3)$$

$$r_d(e_s, e_t) = \frac{c_d(e_s, e_t)C_l^{art}}{c_l(e_s, e_t)C_d^{art} + c_d(e_s, e_t)C_l^{art}} \quad (4)$$

It is inappropriate to deal equally with an edit of leaving n characters and an edit of deleting n characters. This is because multiple editors can leave the same text, and on the other hand, only one editor can delete the same text. Therefore we normalize by multiplication of $c_l(e_s, e_t)$ by C_d^{art} and multiplication of $c_d(e_s, e_t)$ by C_l^{art}. We use these ratios for normalization of the weight of leaving and the weight of deleting. Specifically, we apply $R_l(e_s, e_t)$ and $R_d(e_s, e_t)$ to reflect both $W_l(e_s, e_t)$ and $W_d(e_s, e_t)$ in the edit weights.

$$R_l(e_s, e_t) = \frac{1}{1 - \log_2 r_l(e_s, e_t)} \quad (5)$$

$$R_d(e_s, e_t) = \frac{1}{1 - \log_2 r_d(e_s, e_t)} \quad (6)$$

Finally, we define the definitive weight of edit from e_s to e_t as follows:

$$W(e_s, e_t) = R_l(e_s, e_t) \cdot W_l(e_s, e_t) - R_d(e_s, e_t) \cdot W_d(e_s, e_t). \quad (7)$$

The edit weight becomes a positive value if e_s agrees with e_t, and becomes a negative value if e_s disagrees with e_t. Furthermore, when this relationship is strong, the absolute value of the edit weight gets a big value.

IV. VISUALIZATION OF EDITOR NETWORK GRAPH

In this section, we describe a method for clustering of editors and visualizing the relationships among editors.

A. Clustering

We make graphs of editors' relationships, where nodes are editors, edges are the relationships of agreement or disagreement and weights of the edges are strengths of the relationships. Now, we distinguish each group of editors who have similar biases by clustering of the graph. In this approach, we assume that one editor has only one bias.

Newman [8] proposed a clustering technique that is a bottom-up approach using modularity Q that represents the quality of clustering. We use a technique adapted to this study based on the Newman's technique. Let e_{ij} be a ratio of positive weights from one cluster i to another cluster j to all positive weights, where i and j are the cluster ID. Therefore, e_{ii} represents a ratio of positive weights in only cluster i to all positive weights. $\overline{e}_{ij}$ is a ratio of negative weights from cluster i to cluster j to all negative weights. Now, we define e^*_{ij} as follows:

$$e^*_{ij} = e_{ij} - \overline{e}_{ij}. \tag{8}$$

We define an indicator of clustering quality as follows:

$$Q^* = \sum_i (e^*_{ii} - a_i b_i), \tag{9}$$

where $a_i = \sum_j e_{ij}$ and $b_i = \sum_j e_{ji}$. Therefore, $a_i b_i$ is the expected value of e_{ii} when edges are connected randomly. We define an increment of Q^* by the combination of cluster i and cluster j as follows:

$$\Delta Q^*_{ij} = e^*_{ij} + e^*_{ji} - a_i b_j - a_j b_i. \tag{10}$$

First, we distribute all editors to separate clusters. Next, we integrate cluster i and cluster j that have maximum ΔQ^*_{ij}, and we repeat the same process until ΔQ^*_{ij} is less than 0.

B. Visualization

Fig. 4 shows the editor network graph in the article "Nuclear power plant[3]". Red edges and blue edges represent the agreement relationships and the disagreement relationships, respectively. Colors of nodes differ per cluster. Editors in the same cluster are generally connected by red edges. In addition, the clusters that consist of only one node and are located in a corner are connected by a lot of blue edges. This is because these editors are vicious editors such as vandals who delete all texts or are deleted by all other editors. This means that an editor who is a vandal or writes low quality texts is isolated from other clusters.

In Fig. 4, the editors who oppose nuclear power plant mainly belong to the yellow-colored cluster located in the bottom right, and the editors who approve nuclear power plant mainly belong to the purple-colored cluster located in the bottom. In this way, editors are classified into clusters by editors' biases. We state detail of clustering results in the next section.

[3]http://ja.wikipedia.org/wiki/原子力発電所

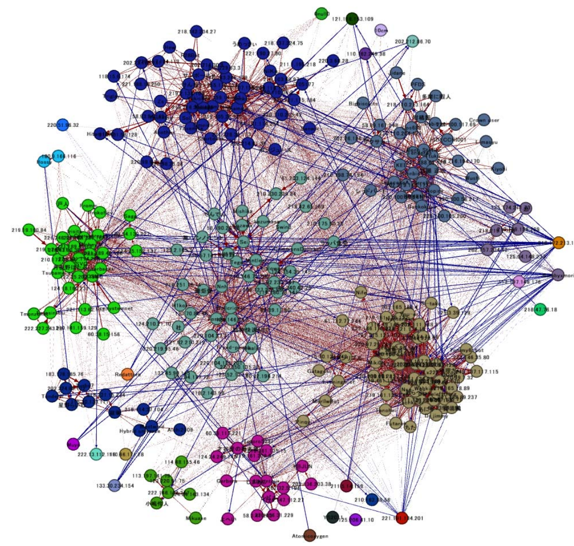

Fig. 4. Editor network graph of the article "Nuclear power plant"

V. Experiments

In order to confirm the effectiveness of our editor classification method, we conducted an experimental evaluation. The experimental procedure is as follows:

1) Our proposed system extracts statistics data about leaving, deleting and inserting texts from the edit history of an article.
2) The observer manually labels active editors. These labels represent editor's biases.
3) The system computes the edit weights between editors by using the statistical data.
4) We make a graph in which nodes are editors, edges are the relationships of edit and weights of the edges are the edit weights.
5) By clustering of the graph, the system classifies editors by bias.
6) We compare the clustering result and editors' labels in order to evaluate the experiment.

Step 1) and 3) are mentioned in section III. Step 4) and 5) are mentioned in section IV. In step 2), the observer manually labeled editors based on the edit history. We choose 7 articles where editors can be classified into three groups, such as positive, negative, and even. In addition, we restrict targets to the editors writing a lot of texts. This is because it is comparatively easy to label only editors who write a lot of texts. According to our former research [1], about 80% of texts are written by the top 20% of active editors in Japanese Wikipedia. Hence, we regard the top 20% of the most active editors as targets for evaluation.

In our experiment, we used a baseline method of clustering by the simple weighting scheme as follows:

$$W(e_{\rm s}, e_{\rm t}) = \frac{c_l(e_{\rm s}, e_{\rm t})C_d^{art} - c_d(e_{\rm s}, e_{\rm t})C_l^{art}}{c_l(e_{\rm s}, e_{\rm t})C_d^{art} + c_d(e_{\rm s}, e_{\rm t})C_l^{art}} \quad . \quad (11)$$

We classify editors by using three methods: this baseline method, random clustering and our proposed method.

A. Data Sets

In our experiments, we used the edit history of Japanese Wikipedia dumped on October 27, 2012[4]. As mentioned above, the targets for the experiment are articles where editors are able to be classified roughly into three groups: positive, negative and even. In the edit history, we can often observe vandals who behave vicious actions such as deletion of all texts. We label these editors "vandal," however, because vandals disagree with each other, we decide that each vandal has his or her own label. Table II shows the names of articles used in the experiment and the numbers of editors in each label. "all" is the number of editors in each article, and "sum" is the number of the top 20% of the most writing editors. The representations of bias labels in Table II are negative, positive and even, but, the labels are practically various based on each article. For example, the article "Nuclear power plant" has three groups: anti-nuclear power, pro-nuclear power and even. Anti-nuclear power means negative, and pro-nuclear power means positive. In the article "AKB48," there is no positive-editors because no positive-editor was observed in the top 20% of the most writing editors.

B. Evaluation measure

The most popular measures for cluster evaluation are $Purity$, $InversePurity$ and a harmonic mean (F*-measure*) of $Purity$ and $InversePurity$ [9]. $Purity$ evaluates precision of the most frequent label in each cluster produced by clustering. $InversePurity$ evaluates maximum recall of each true cluster produced by labeling. We introduce these measures for evaluating our clustering results.

C. Results and Discussions

As shown in Fig. 5, we calculated evaluation values of each clustering result. The vertical axis shows $Purity$, $InversePurity$, and F*-measure*. The horizontal axis shows article IDs. The green bars represent the random clustering, the blue bars represent the baseline method, and the red bars represent the proposed clustering method. Fig. 5 shows the evaluation results based on $Purity$, $InversePurity$ and F*-measure*. A point in common among these graphs is that the evaluation values of the baseline method and our proposed method are higher than the evaluation values of the random clustering in most cases. Therefore, using relationships between editors is valid for clustering of editors by biases. Additionally, the evaluation values of our proposed method are higher than the evaluation values of the baseline method in most cases. In the comparison using F*-measure*, 5 of 7 articles show that the proposed method is better than the baseline method, and two evaluation values in the rest of articles are approximately the same degree. From these results, the proposed method in which the edit dependency

TABLE II. TARGET ARTICLES AND THE NUMBER OF TARGET EDITORS IN EACH LABEL

ID	article name	neg.	pos.	even	vandal	sum	all
1	Nuclear power plant	16	4	37	3	60	302
2	Nanking Massacre	20	20	19	3	62	310
3	AKB48 (Girl group)	10	0	75	2	87	436
4	Kazuko Hosoki	32	1	32	2	67	337
5	Senkaku Islands	5	12	46	11	74	376
6	Emperor in Japan	5	2	17	3	27	137
7	Comfort women	14	14	28	1	57	291

[4]http://dumps.wikimedia.org/jawiki/20121027/jawiki-20121027-pages-meta-history.xml.bz2

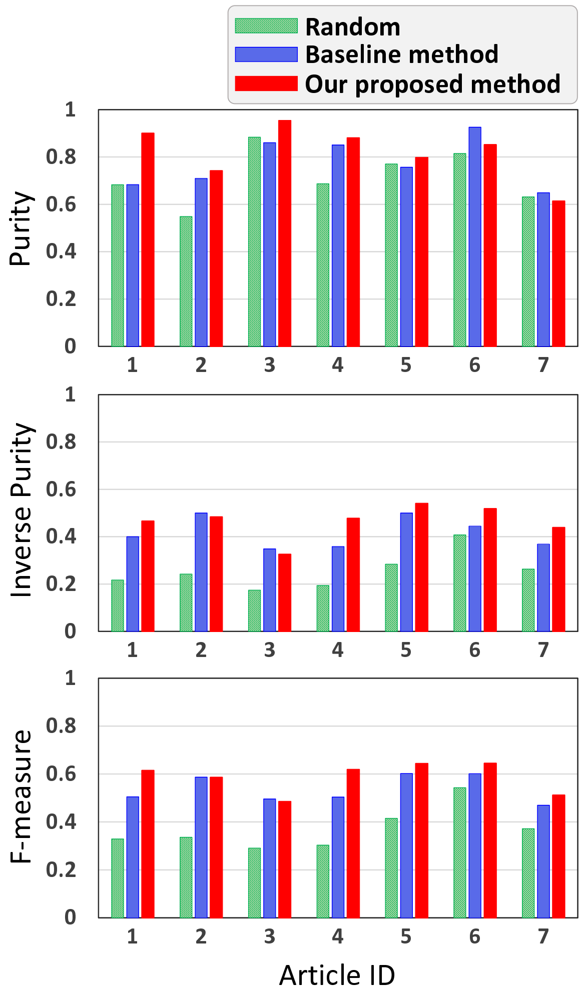

Fig. 5. Evaluation results based on *Purity*, *InversePurity* and *F-measure*

between editors is reflected improves the clustering performance. Superior editors write facts and even contents, and consequently superior editors' texts are left by biased editors easily. For this reason, even editors and biased editors are likely to be classified into the same cluster if the edit weights are simply decided as with Eq. (11). In our proposed method, when an editor leaves a text written by another editor who is left by many editors, the edit weight comparatively indicates a low value. As a result, the possibility of error clustering decreases, and the clustering performance is improved.

Table III shows the clustering result of the article "Nuclear power plant" by our proposed method. The values of every bias in each cluster represent the number of editors. From this table, we can see that editors in the negative group are mainly classified into cluster 3, and editors in the positive group are classified into cluster 8. As you can see in Table III, the number of clusters is more than the number of labels. The number of labels is 6: negative, positive, even, and three vandals, whereas the number of clusters is 11. In fact, if we count clusters composed of only editors who are not included the top 20%, the number of clusters are 30. Therefore the recall decreases, and this reduction makes $InversePurity$ low. In the article "Nuclear power plant," $Purity$ is 0.9, but $InversePurity$ is 0.467. These results suggest that purity of each cluster is high, but editors having the same label are divided into multiple clusters. This feature is common to results of other articles.

Two reasons for subdivision of clusters are considered. The first reason is the severity of ΔQ^*_{ij} at Eq. (10) used in the clustering. In our proposed method, we deal with negative weights, which represent the strength of disagreement relationships between editors. In order to reflect the negative weights in clustering, we define ΔQ^*_{ij}, however, the combination condition of clusters became severe. We can consider that this is one of the reasons for the subdivision.

The second reason is the abstractness of labeling by hand. In the experiment, we roughly classify editors into three types of labels. However, editors have more subdivided biases practically. For example, the clustering result in Table III shows that editors who oppose nuclear power are mainly classified into cluster 1 and cluster 3, but cluster 1 differs from cluster 3 in the features of editors.

TABLE III. CLUSTERING RESULT OF THE ARTICLE "NUCLEAR POWER PLANT"

Cluster ID	neg.	pos.	even	vandal
1	1	0	0	0
2	0	0	6	0
3	14	0	4	0
4	0	0	5	0
5	0	0	7	0
6	0	0	0	1
7	1	0	7	0
8	0	4	1	0
9	0	0	0	1
10	0	0	7	0
11	0	0	0	1

Cluster 3 consists of editors who oppose government cost estimates and claim danger of nuclear power. On the other hand, cluster 1 consists of editors who describe Fukushima Daiichi nuclear disaster. Furthermore, difference is seen between clusters of even editors. The number of even clusters is 5, but editors who eagerly form contents are concentrated in cluster 10. The texts written by the editors in cluster 10 occupy 63% of the whole texts in the latest version. In this way, the number of clusters in the proposed method is more subdivided than the number of labels.

VI. Conclusion

In this paper, we proposed a method of clustering editors by bias, and visualized the relationship of editors using network graph. We calculated the strength of relationships between editors from the data of leaving and deleting, and used the relationships to classify editors by bias. In the evaluation experiment, we introduced $Purity$, $InversePurity$ and F*-measure* to measure the effectiveness of our proposed method. Consequently, we verified the relationships of agreement or disagreement between editors that identified from leaving/deleting relationships are useful in clustering editors by bias. Additionally, we validated that weighting based on the dependency between editors improves the clustering performance.

We will show several future directions for our study. Our method utilizes only characters of texts for clustering, but we can also use the features of contents. Sachan et al. [10] reported that using not only link structure but also contents improves the clustering performance. The features of contents enable us to decide the edit weights more properly.

We extracted the relationships between editors from the edit history, but there are few pairs of editors who have the relationships. Prediction of edge between editors who have no relationship is an effective solution for this problem. In a social graph that contains both positive edges and negative edges, a prediction technique of edges' natures is proposed by Leskovec et al. [11]. An increase of edges in a graph by prediction of edges would improve the clustering performance.

In our study, we classified editors by editors' biases, but we do not find each cluster's kind of bias. We plan to detect characteristics of the clusters from editors' texts and label the clusters automatically.

Acknowledgment

This research is supported by KAKENHI (25280039, 23700113) and the Telecommunications Advancement Foundation (TAF).

References

[1] Y. Suzuki, "Effects of Implicit Positive Ratings for Quality Assessment of Wikipedia Articles," *Journal of Information Processing*, vol. 21, no. 2, pp. 342–348, 2013.

[2] B. T. Adler and L. de Alfaro, "A Content-Driven Reputation System for the Wikipedia," *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pp. 261–270, 2007.

[3] M. Hu, E.-P. Lim, A. Sun, H. W. Lauw, and B.-Q. Vuong, "Measuring Article Quality in Wikipedia: Models and Evaluation," *CIKM '07: Proceedings of ACM International Conference on Information and Knowledge Management*, pp. 243–252, 2007.

[4] P. Massa and B. F. Scrinzi, "Manypedia: Comparing Language Points of View of Wikipedia Communities," *WikiSym '12: Proceedings of the 2012 international symposium on Wikis*, 2012.

[5] E. S. Callahan and S. C. Herring, "Cultural bias in Wikipedia content on famous persons," *Journal of the American Society for Information Science and Technology*, vol. 62, pp. 1899–1915, 2011.

[6] R. Jesus, M. Schwartz, and S. Lehmann, "Bipartite Networks of Wikipedias Articles and Authors: a Meso-level Approach," *WikiSym '09: Proceedings of the 2009 international symposium on Wikis*, 2009.

[7] U. Brandes, P. Kenis, J. Lerner, and D. van Raaij, "Network Analysis of Collaboration Structure in Wikipedia," *WWW '09: Proceedings of the 18th international conference on World Wide Web*, pp. 731–740, 2009.

[8] M. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, vol. 69, p. 066133, 2004.

[9] Y. Zhao and G. Karypis, "Criterion Functions for Document Clustering: Experiments and Analysis," University of Minnesota, Department of Computer Science / Army HPC Research Center, Tech. Rep., 2001.

[10] M. Sachan, D. Contractor, T. A. Faruquie, and L. V. Subramaniam, "Using Content and Interactions for Discovering Communities in Social Networks," *WWW '12: Proceedings of the 21st international conference on World Wide Web*, pp. 331–340, 2012.

[11] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting Positive and Negative Links in Online Social Networks," *WWW '10: Proceedings of the 19th international conference on World Wide Web*, pp. 641–650, 2010.