Abstract

The advantages of the multimedia make the video news presented believable and impressed to the viewers when the personal opinions and ideological perspectives hidden in the contents still cause the effect. To reduce the risk of the misleading, based on a Material-Opinion model, we propose a method of detecting the inconsistent news items reporting the same event when the viewer is watching one of them. In the Material-Opinion Model, main participants filmed as the materials are presented to the viewer through the video stream, which is used to support the arguments put forward. Based on this model, given a series of multimedia news items reporting a same event, we explore inconsistency between any two of them by computing their dissimilarities of materials and of opinions. Material-dissimilarity is based on the appearance of the main participants in the video. Opinion-dissimilarity is calculated as the vector difference of two vectors consisting of the argument points extracted from the closed captions. If one of the dissimilarities is high and the other is low, we consider that there exists the inconsistency as a result. We also show some experimental results to validate the proposed methods.

1 Introduction

People are exposed to a large number of multimedia news every day. Most of them rely on the news reports to know what happened in the world, especially the important global events. The simple way of obtaining information by watching video news from TV programs or on Internet websites turns to be popular and welcome because of its clear, deeply-considered and convinced exposition. Moreover, the multimedia news, as an impressive combination of vision, audio and text, also provides the attractive entertainment obtaining the high acceptance from the viewers. Those advantages help the multimedia news popular and welcome as the widespread believable press to the viewers.

However, as one of the most valuable media transferring the messages, multimedia news is also facing the challenge of the bias issue as the traditional newspaper is. The entertainment wears down the curiosity of chasing the facts behind the story, and the presentation of vision, sound and text also makes it more difficult to perceive the concealment or distortion. For the news agency’s intentions that more or less exist in the news items [5, 15], unfortunately, it is not feasible or recommended of viewers watching massive news items to find the most objective one. In order to help viewers be aware of the misleading, we are studying technologies on inconsistency exploration between two news
items by using a Material-Opinion Model. As a usage scenario, viewers watching a multimedia news will be alerted if there are some other news items reporting the same event but unusually different from the current one. These differences are the inconsistency that we aim to discover in this paper, of helping the viewers maintain objective viewpoints by presenting the inconsistent items. It also helps reduce the redundant by neglecting the multimedia news items reported similar with the current one.

In our Material-Opinion Model, contents of the multimedia news items are classified into two parts as the material part as the segments filmed by the camera, and the opinion part as the arguments put forward by the announcers. For any two of the multimedia news items reporting a same event, we compute the dissimilarity of materials and the dissimilarity of opinions by using the two video clips and the corresponding closed captions. Supposing the given multimedia news item $c_1$ as the analyzing target, according to the dissimilarities, inconsistency between $c_1$ and the compared item $c_2$ is detected if: a) The material-dissimilarity is large when the opinion-dissimilarity is small, which indicates that $c_2$ provides additional information to help the viewer understand the opinion reported in $c_1$. b) The material-dissimilarity is small when the opinion-dissimilarity is large, which suggests that $c_2$ provides different viewpoint on the event and the double check of $c_2$ can help the anti-bias.

Because most of the descriptions in the news item is related to some particular entities, in the computation of the material-dissimilarity, we focus on the visual descriptions of the main participants of the event as the more valuable part of the video. We also illustrate our method of extracting the main participants who are the frequently exposed persons in the videos and the closed captions, which is the pretreatment of our inconsistency exploration. Currently we analyze the human participants only because of the numerical superiority. Other reasons are that most of the non-human participants (countries, organization, etc.) are represented by one or more persons (spokesman, leader, manager, etc.) in the video and we are not caring about the inclusion relations of the participants till now. To compute the material-dissimilarity of two multimedia news, we first extract the main participants (persons) of the event and then calculate the ratio of the LCS (Longest Common Subsequence) [7] of two character strings transferred from the main-participant-labeled segments of the two video clips.

In the opinion-dissimilarity computation, we extract the arguments in the closed captions and calculate the scores per each argument. Polarity scores between two multimedia news items is calculated as the opinion-dissimilarity by using the accordance sequences consisting of the argument points of both news items. As a result, argument vector of a news item will be formed and the opinion-dissimilarity will be computed by using such argument vectors based on vector space model.

In the experiments, we collected 87 multimedia news items consisting of video clips and their closed captions reporting 10 events. In order to explore inconsistency between two multimedia news items reporting a same event, we compare them by computing their material-dissimilarity and opinion-dissimilarity. These experimental results demonstrate our methods. There are many researches to assist reading news articles. NewsInEssence[1] and Columbia’s Newsblaster[2] summarized news articles from multiple sources into one virtual news article. Ma et al. proposed a comparative query generation method for cross-media search between text news and video news[3]. They considered news of different media to be complementary and their method is to search for complementary news.

These researches aim at complementing differences between news articles. On the other hand, some researches aim at making users conscious of differences between news articles. Nadamoto et al. focused on differences between news articles about the same topic in different countries, and developed a bilingual comparative web browser (B-CWB)[4]. Users can compare news articles in two countries and find their differences of the content. Park et al. thought that there were some bias in news articles, and developed the system NewsCube to classify the news articles by their aspects and present them to users [5]. Aoki et al. proposed author intention model and developed the system to compare the intention of different authors [6]. They focused on polarity appearing in news articles, and used it to express author intention. In other words, they focused on not only what was reported in news articles but also how it was reported. Therefore, we adopt their model to compute opinion-dissimilarity.

There are proposed sequence-based matching methods for similarity measure of videos [8, 9, 10]. Kim et al. proposed a paper suggesting the method of LCS (longest common subsequence) performed better mean average precision then the method of ED (edit distance) in the news videos because video has stronger temporal coherence and continuity [11]. So we use the LCS method to compute the material-dissimilarity.
2 Analyzing Multimedia News in a Material-Opinion Model

Studying technologies on inconsistency exploration of two news items consisting of video clips and their closed captions, we propose a Material-Opinion Model for the contents analysis. Here, the “Material” part is the visual descriptions filmed by the camera as a reflection of the real world which is believable despite the aggrandizement or forgery in films. The “Opinion” part consists of the arguments put forward by the announcer (news agency), which can be extracted from the text in the corresponding closed captions. Supposing \( c_k \) as a multimedia news item,

\[
\begin{align*}
    c_k &= (M_k, O_k) \\
    M_k &= \{ \text{seg}_p | \text{seg}_p \in \text{VideoClips} \} \\
    O_k &= \{ \text{arg}_q | \text{arg}_q \in \text{Arguments} \}
\end{align*}
\]

where \( k \) is the item number of the multimedia news reporting a same event. \( p \) is the number of the segment in item \( c_k \). \( M_k \) is the sorted set saving the segments of each video. \( O_k \) is the arguments set extracted from the text in closed captions.

To explore the inconsistency between two news items \( c_1 \) and \( c_2 \), we calculate the material-dissimilarity \( \text{Dis}_M(c_1, c_2) \) and the opinion-dissimilarity \( \text{Diss}_O(c_1, c_2) \). As shown in Figure 1, according to their locations, there are four typical patterns as follows.

- **Case 1**: The two news items are similar with each other in both text and video. The achieved unity indicates that the “same” multimedia item is not necessary to be double-checked.

- **Case 2**: The opinions reach an agreement when different materials are reported in the news items. This kind of the multimedia news can be provided to a viewer to observe the event in a different aspect by different materials. On the other hand, such kind of video can be used to provide additional information (materials) in order to help understand the opinion.

- **Case 3**: Different videos are supporting different viewpoints, in which it is hard to say whether there appears inconsistency or not. However, when the two dissimilarly scores of texts and videos are very large, the two multimedia news items are possibly reporting different events and are incomparable.

- **Case 4**: Similar materials are supporting different opinions. Viewer should double-check the contents without swallowing everything in the contents. This kind of different multimedia news can provide us additional information for understanding news from different viewpoints and anti-news bias.

![Figure 1. Comparison in material and opinion](image)

In our system, inconsistency will be detected if the result of comparisons is in **Case 2** or in **Case 4**. The news items detected with **Case 2** and **Case 4** are helpful for understanding news reports, especially, **Case 4** is from the anti-bias aspect and **Case 2** is from the aspect of providing detailed evidence (materials) supporting the opinions when the latter is more serious because it points out the credibility gap in news items. The specific operations are introduced in Section 3.

3 Detecting Inconsistency

3.1 Computing Opinion-dissimilarity

To compute opinion-dissimilarity between news items, we modify and extend the author intention model [6]. We make feature vectors by using it and define opinion-dissimilarity between news items. In this subsection, we first introduce the original author intention model and then describe the method of computing opinion dissimilarity based on it.

3.1.1 Author Intention Model

In author intention model, author intention is expressed by a set of 3-tuples of argument point, polarity, and strength. An argument point is expressed by
an observation. Author intention is defined as follows. A pair of an argument point and polarity, is called as positively (negatively) mentions the argument point. A set of keywords. Polarity means polarity about an argument point to be added to $P_t_i$. We iteratively add $cp$ whose evaluation value becomes maximum to $P_t_i$ until all of the evaluation values are less than or equal 0. Then, $P_t_i$ is a set of argument points.

### 3.1.2 Opinion-dissimilarity

We compute opinion-dissimilarity by extending Aoki et al.’s author intention model. In their approach, they count the number of observation sentences corresponding to each combination of author, argument point and polarity, and they obtain author intention. Instead of this, we count the number of observation sentences corresponding to each combination of news item, argument point and polarity, and we obtain opinion in a new item. If there are $n$ argument points, there are $2n$ 3-tuples for each news item. Suppose that the following opinion in news item is given:

$$
\text{op} = \{(pt_1, P, \text{str}(pt_1, P)), \ldots, (pt_n, P, \text{str}(pt_n, P)), (pt_1, N, \text{str}(pt_1, N)), \\
\ldots, (pt_n, N, \text{str}(pt_n, N))\}
$$

(9)

We make the following 2n-dimensional feature vector to express opinion in a news item $c_k$:

$$
O_{k} = (\text{str}(pt_1, P), \ldots, \text{str}(pt_n, P)), \\
\text{str}(pt_1, N), \ldots, \text{str}(pt_n, N))
$$

(10)

The opinion-dissimilarity between news items is defined as sine of opinion feature vectors as follows.

$$
\text{Dis}_{o}(c_1, c_2) = \sqrt{1 - \cos^2(O_1, O_2)}
$$

(11)

$$
= \sqrt{1 - \frac{(O_1 \cdot O_2)^2}{|O_1|^2|O_2|^2}}
$$

(12)

Where, $O_1$ and $O_2$ are opinion feature vectors of news item $c_1$ and $c_2$ respectively.

### 3.2 Computing Material-dissimilarity

The first question of analyzing the material part of a multimedia news item is how to select and construct its features. Video clips in the news items offer not only the streams of the fix-sized pixel information but also the appearance of the concepts (objects shown in the frames). Furthermore, there are also abundant connotations included in the appearing patterns and particulars that make the differences between two items.
Considering that most of the news are reported around at least one entity as a person, a location, an organization, etc., which could be regarded as the main participant of the event, we think that the appearance of these main participant(s) in the video is the most valuable parts. Video clips can be thought of the sequence of the participants’ segments where they appear. For the feasibility and the efficiency of the video comparing, we compute the material-dissimilarity by comparing the sequences of the main participants’ segments rather than the pixel-based methods. Here, the main participants should be: a) important enough to be frequently mentioned in the contents; b) applicable to all the news reporting the same event. Currently we focus on the participants who are human being only because of the face matching method help make correlation between textual description (closed captions) and visual description (video clips). We use and improve the method of stakeholder extraction [16] for the material analysis. In our method of computing material-dissimilarity, we first extract the main participants from the items reporting the same event, and then compare the two videos by using their visual descriptions represented related to the participants to compute the video dissimilarity for the inconsistency detection.

In this subsection, we first describe how to extract participants from video contents with closed captions. Then, we show the method of computing the material-dissimilarity between two multimedia news items.

### 3.2.1 Extracting Main Participants

In the news videos reporting a same event, the main participants (persons) are supposed to appear in the videos a lot. For the entities being filmed for a long period of time but not being the important ones itself in the event (the guard of the president, etc.), names of the main participants should also be mentioned in the responding closed captions frequently. Because of the freedom of the different aspects, the major concerns and the reported persons are not identical in all the news items. Our method of extracting the main participants (persons) in the event is as follows.

1. Extract the candidates. After we extract the candidate names from the text (closed captions) and detect the candidate faces from the video, we identify the faces with a face matching method by using the database saving persons names and its faces. Extracted candidate names (person names) are possibly the matching names of the faces appearing in the same segment of the video. We also calculate the possibility score per each candidate name and the importance score per each candidate face for the further analysis.

First, we calculate the similarity of frames extracted by time interval and segment the video clips by using the twins-comparison method [14] detecting the boundaries for both camera motion and rapid change.

Second, focusing the textual descriptions in the given closed captions, we use ChaSen\(^1\) to check if there is person names in the sentences (in Japanese). According to the time information in the closed captions, we can separate the texts by the segmented boundaries and label each segment with the candidate names extracted in the text.

Third, we calculate the probability scores by using the frequency that the candidate name occurs in the segments. The text range for discovering candidate name in \(p^{th}\) segment is the textual descriptions corresponding to the \(p^{th}\) and \((p - 1)^{th}\) segments. The probability score of candidate \(s_i\) in the segment \(seg_p\) is computed as follows.

\[
ps_{s_i,seg_p} = fre_{s_i,seg_{p-1}} \times w_p + fre_{s_i,seg_p} \times w_c + fre_{s_i,seg_{p+1}} \times w_n
\]

(13)

where, \(w_c > w_p > w_n\) are pre-specified weight parameters. High possibility is that the face detected in the frame belongs to the name mentioned in the current segment. \(fre_{s_i,seg_p}\) denotes the frequency of \(s_i\) appearing in segment \(seg_p\). \(ps_{s_i,seg_p}\) help us rank the candidate names by their possibilities of identifying the person face.

Due to the short duration of segments, the ambiguous description, and the non-strictly corresponding between video and its closed caption, possibility is that there is no name extracted in the three segments. In such case, we extend the text range to the beginning of the video. It is to say that, we compute \(ps_{s_i}\) as follows.

\[
ps_{s_i} = fre_{s_i,above} \times w_{above}
\]

(14)

where, \(fre_{s_i,above}\) is the frequency of \(s_i\) appearing in the text ranging from current segment to the first one.

Fourth, we detect candidate faces appearing in \(2^{nd}\) frame, middle frame and \((n - 1)^{th}\) frame for each segment if there are fade-in and fade-out method in the boundaries. \(n\) is the number of extracted frames of the segment. We use joint Haar-like features [13] with multiple cascade files for the better

---

\(^1\)http://chasen.aist-nara.ac.jp/chasen/distribution.html
precision and recall. If there are two faces detected highly overlapped with each other, the smaller one will be neglected. Then we calculate the importance score $IS$ of each face. Two factors are the position and the area. The faces in the middle field and the faces with large area seek more attention than the others. Therefore, they are more likely to be the major participants in the video. To reduce the time cost in the face matching method, we compare the faces with each other extracted from the same news items and merge the similar ones. The comparisons take the use of the Affine-SIFT method [12] that if the matching number of two faces comparison is larger than the threshold, we think the two faces are from the same person.

Finally, we use the face matching method to identify the faces extracted in the segments. We build a database saving person names and its faces for the comparison by using the Affine-SIFT method [12]. For each face detected, we make the list of the possible names (=candidate names extracted from the textual descriptions in the corresponding closed captions) and compare the face with the faces extracted from the database that belong to the possible names one by one. If all the face matching failed, we search the faces from the whole database for the identification. Names of the identified faces are labeled to the segments as the results. If the face belongs to the news announcer, the name is also labeled for the further video comparison but the person himself is out of the main participant candidates in the event.

2. Calculate the exposure degrees. We calculate the textual exposure degree and the visual exposure degree of each candidate (identified face) extracted from each news items. Textual exposure degrees are computed by using the sentence information in the closed captions where the candidate name is mentioned in; visual exposure degrees are computed by the candidates’ presenting information in the video clips.

**Textual Exposure Degree** We calculate the textual exposure degree by using the character length of the related closed captions. As shown in Figure 2, for each sentence in the text of the closed captions, if the candidate is mentioned, we consider that the sentence is related and its character length is added up for the computation. The exposure length of the textual description of candidate $s_i$ in multimedia item $c_k$ is computed as follows.

$$ T(s_i, c_k) = l(s_i, c_k) $$ (15)

where $l(s_i, c_k)$ is the length of characters related to candidate $s_i$ in news item $c_k$. If $s_i$ does not exist in the news item, $T_{s_i, c_k} = 0$. Let $l(c_k)$ be the length of characters in the closed caption of news item $c_k$, the textural exposure degree of candidate $s_i$ in item $c_k$ is computed as follows.

$$ T(s_i, c_k) = \frac{T_{s_i, c_k}}{l(c_k)} $$ (16)

**Visual Exposure Degree** We calculate visual exposure degrees by using visual information about segments in which the candidate appears. As shown in Figure 3, the time duration of related segments and the importance score of each face are used to computed the degrees. Visual exposure degree $V(s_i)$ of candidate $s_i$ in item $c_k$ is given by function below.

$$ V(s_i, c_k) = \frac{1}{d(c_k)} \cdot \sum_{p=1}^{q} \left( d(s_i, seg_p, c_k) \cdot IS(s_i, seg_p, c_k) \right) $$ (17)

where $d(s_i, seg_p, c_k)$ is the time duration of segment $seg_p$ if it is related to candidate $s_i$; $IS(s_i, seg_p, c_k)$ is the face importance score corresponding to candidate $s_i$ in segment $seg_p$, and $d(c_k)$ is the time duration of the news item $c_k$. If candidate $s_i$ does not exist in segment $seg_p$, then $d(s_i, c_k) = 0$.

3. Filter the candidates. After merging the candidates from each multimedia items reporting the same event, we calculate the average textual exposure degree and average visual exposure degree of each candidate as the two factors to filter the candidates with low probability of being a main participant in the event. To exclude the persons that have descriptions but low exposure degrees
in either textual or visual descriptions, we specify two thresholds, $T_{\text{textual}}$ and $T_{\text{visual}}$, for textual and visual descriptions, respectively. For each candidate, we calculate the average of the textual exposure degrees and the average of the visual exposure degrees as follows.

$$T(s_i) = \frac{\sum_{c=1}^{l} T(s_i, c_k)}{l}$$

$$V(s_i) = \frac{\sum_{c=1}^{l} V(s_i, c_k)}{l}$$

where $T(s_i, c_k)$ and $V(s_i, c_k)$ are the textual and visual exposure degrees of candidate $s_i$ in item $c_k$ and $l$ is the number of items reporting the same event. If $T(s_i) > T_{\text{textual}}$ and $V(s_i) > T_{\text{visual}}$, then candidate $s_i$ is a main participant in the event.

### 3.2.2 Material-dissimilarity

When comparing two video clips in the event, we think the appearing sequence in the video is also an important feature that the different appearing patterns indicate different enumerate and cause different effects. For example, there is a hint if the sequence is that “Participant A accepted the proposal” was shown at first and “Participant B rejected the proposal” in the next. Foreexample, thereisahintifthesequenceisthat“Par-
dicate different enumerate and cause different effects.

The appearing sequence in the video is also an important feature that the different appearing patterns in the event so deeply, but we can tell if there is high possibility that they are reporting different events.

For computing the material-dissimilarity between two news items, first of all, for each segment in the video, we make a participant set $S_p$ consisting of the main participant(s) extracted in the segment. $p$ is the number of the target segment. If there is no main participant extracted, $S_p = \emptyset$. After that, we transfer each video to a character string as a chain of the repeated $Sid_p$. $Sid$ is an integer identifying the different participant set. Same participant set shares the same $Sid$ in all the news items reporting a same event. If there is no main participant extracted in the segment $p$, $Sid_p = “N”$. Because of the announcer faces extracted in the segments are different from each other because of the different post-casters, we define $Sid_p = “-”$ if in the segment $p$ there is no main participant but an announcer making the introduction. The repeating counts rely on the frame amount of each segment with settled time interval. The character string of each video is the ordered chain of the repeated $Sid_p$ of the segments. Our comparison of two videos is actually the comparison of the two transferred character strings. Finally, the material dissimilarity $Dis_m(c_1, c_2)$ of two news items $c_1$ and $c_2$ is calculated as follows.

$$Dis_m(c_1, c_2) = \frac{|\text{LCS}(c_1, c_2)|}{ll(c_1, c_2)}$$

where $\text{LCS}(c_1, c_2)$ is the LCS [7] of multimedia news $c_1$ and $c_2$. $ll$ is the longest length of the character string of the two videos. Respectively, $0 \leq Dis_m(c_1, c_2) \leq 1$.

### 3.3 Detecting Inconsistency between two Multimedia News Items

After calculating the opinion dissimilarity and the material dissimilarity of the given two multimedia news, as shown in Figure 1, we say there appears inconsistency between them if the result of the comparison locates in area $Case2$ or area $Case4$. Area $Case2$ indicates the two news items are similar in material and dissimilar in opinion. Area $Case4$ indicates that they are similar in opinion and dissimilar in material. In $Case1$ the two multimedia news have the similar contents and the same opinion. In $Case3$ there is high dissimilarity between the two news items that there is high possibility that they are reporting different events.

Because the comparison whose result located in the middle part is hard to be classified into the four cases, we set four thresholds for the inconsistency exploration in the experiments. If $Dis_m(c_1, c_2) \geq \vartheta^u_M$ and $Dis_o(c_1, c_2) \geq \vartheta^u_O$, or $Dis_m(c_1, c_2) \leq \vartheta^l_M$ and $Dis_o(c_1, c_2) \geq \vartheta^l_O$, then the system will consider the result that there is the inconsistency between news item $c_1$ and $c_2$. Here, $\vartheta^u_M$ and $\vartheta^u_O$ are the upper and lower thresholds of material-dissimilarities. $\vartheta^l_M$ and $\vartheta^l_O$ are the upper and lower thresholds of opinion-dissimilarities.

### 4 Experiments

To validate our methods, we carried out three experiments: 1) Experiment of computing opinion-dissimilarity; 2) Experiment of computing material-dissimilarity; and 3) Experiment of inconsistency detection between two multimedia news items. As shown in...
Table 1. Experiments data set

<table>
<thead>
<tr>
<th>Event</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hatoyama’s pledge of 25% emissions cut</td>
</tr>
<tr>
<td>2</td>
<td>Summit between Japan &amp; USA</td>
</tr>
<tr>
<td>3</td>
<td>Obama’s United Nations speech</td>
</tr>
<tr>
<td>4</td>
<td>Itinerary of IOC Meeting determined</td>
</tr>
<tr>
<td>5</td>
<td>Obama’s visit to Japan</td>
</tr>
<tr>
<td>6</td>
<td>Working Group meeting</td>
</tr>
<tr>
<td>7</td>
<td>Summit between China and USA</td>
</tr>
<tr>
<td>8</td>
<td>Futenma Problem (domestic agreement)</td>
</tr>
<tr>
<td>9</td>
<td>Futenma Problem (overseas agreement)</td>
</tr>
<tr>
<td>10</td>
<td>Conference in Copenhagen (COP15)</td>
</tr>
</tbody>
</table>

Table 1, after we segmented the recorded Japanese TV programs (in Japanese) into news items, we collected the news items reporting same events. Events were determined by 6W1H features. After that, 10 events consisting of 87 multimedia news items are selected as the test data set of the experiments. The selected video news in the data set last more than 250 minutes and the average duration is about 173 seconds. These multimedia news items were from ANN News, NNN Straight News, NNN News REALTIME, NHK Ohayo Nippon, NHK NEWS (SYOHGO), News 7, and News Watch 9.

For the face matching method in the video comparison, we built a database consisting of 406 face pictures (in JPEG) of 23 persons. The size of each picture is $100 \times 100$ pixels.

4.1 Computing Opinion-dissimilarity

To evaluate the method of computing opinion-dissimilarity, we compared its results with user rating of dissimilarity between news items. The users read the closed captions of news items, and they rated dissimilarity between any pairs of closed captions of news items on the following criteria:

0: There are no differences.
1: There are small differences.
2: There are differences.

We computed the Spearman’s ranking correlation between computed opinion-dissimilarity and user rating for each new event.

First, we calculate Spearman’s ranking correlation between the scores by two users for each news events. We show the result in Table 2. Here, event means id of news event, and correlation means Spearman’s ranking correlation between scores by two users. From Table 2, there are positive-correlations in most news events. There are, however, negative-correlations or no correlations in some news events. It means that there are some news items which give different impression to the users. Therefore, we did not use the news events whose correlation between users’ rating is less than 0.2 for the evaluation. We also removed the data whose rating is inconsistent between users.

Spearman’s ranking correlations between opinion-dissimilarity and users’ rating are shown in Table 3. In Table 3, “-” means that we cannot calculate Spearman’s ranking correlation. Average score of the correlation is 0.31. It means that there are positive-correlation between the proposed opinion-dissimilarity and user rating. In some news events, there are negative-correlations. One of the considerable reasons is the texts are too short to extract opinions from news items. We’ll study this issue in the near future.

4.2 Computing Material-dissimilarity

To evaluate video dissimilarity between news videos, we also compared it with the user rating. The criterion is the same with that of the opinion-dissimilarity evaluation. The precision, recall and the F-measure are shown in Table 4 in which $\vartheta_M$ is the threshold of the
from 7 events. Here, \( \vartheta \) exploring inconsistency between 81 pairs of news items in the system. If the two news items are considered to be different with each other in main concern. If the graduated score is larger than 5, it is determined that there is inconsistency between the two news items as the material-dissimilarity. If \( Dis(c_1, c_2) \geq \vartheta^u_M \) and \( 0 < Dis(c_1, c_2) < \vartheta^l_M \) or \( Dis(c_1, c_2) < \vartheta^u_M \) and \( Dis(c_1, c_2) \geq \vartheta^l_O \), the system considers the result that there is inconsistency between news item \( c_1 \) and \( c_2 \).

In Section 3.1.2, an issue of failing to extract opinions in the short closed captions has become apparent. Despite the little cases of the two vectors of the argument-points match with each other perfectly, we added the \( Dis(c_1, c_2) > 0 \) as an additional condition.

From the results we can see that when \( \vartheta^l_O = \vartheta^u_O = 0.10 \) and \( \vartheta^l_M = \vartheta^u_M = 0.50 \), precision and recall are 0.68 and 0.65 which are the best in the results. In the line of \( \vartheta^l_O = 0.7, \vartheta^u_O = 0.13, \vartheta^l_M = 0.45, \vartheta^u_M = 0.55 \), the recall decreased to be 0.57 because of the narrowed sufficient conditions of the inconsistency exploration in the system. The results suggest that we can merge the upper and lower thresholds into one for the dissimilarity of either material or opinion.

In our method of discovering the inconsistency, comparing results with both high material-dissimilarity and high opinion-dissimilarity are not decided to be the inconsistent ones. These multimedia news items possibly reported events which are unconcerned with the current one. However, the evaluators may consider these items are reporting different events of a topic and will judge them as the relevant results. This manner will affect the experimental results. For example, two news items may respectively report the Japan-China and Japan-US top-level meetings during G20 summit. Although a user may consider these items provide different information on a certain topic (G20 summit), our system will not handle this case.

### 5 Conclusions and Future Work

In this paper we propose a new method discovering the inconsistency between two multimedia news items consisting of video clips and its closed captions reporting a same event. We propose a Material-Opinion Model for the inconsistency analysis of the filmed main-participant-related visual description and the opinion put forward by the announcers in the contents. The experiment results show the performance.

Our future work is: 1) Normalizing the evaluation criterion of the opinions and materials in the contents for analyzing the two dissimilarities on the same plane; 2) Proposing a method calculating the inconsistency scores by using both opinion-dissimilarity and material-dissimilarity; and 3) Extracting special ap-

| Table 4. Computing material-dissimilarity |
|---|---|---|---|
| \( \vartheta_M \) | Precision | Recall | F-measure |
| 0.25 | 0.75 | 0.99 | 0.85 |
| 0.30 | 0.75 | 0.98 | 0.85 |
| 0.35 | 0.75 | 0.97 | 0.85 |
| 0.40 | 0.75 | 0.95 | 0.84 |
| 0.45 | 0.76 | 0.93 | 0.84 |
| 0.50 | 0.78 | 0.90 | 0.84 |
| 0.55 | 0.79 | 0.85 | 0.82 |
| 0.60 | 0.80 | 0.80 | 0.80 |

| Table 5. Exploring inconsistency |
|---|---|---|---|---|---|---|---|
| \( \vartheta^l_O \) | \( \vartheta^u_O \) | \( \vartheta^l_M \) | \( \vartheta^u_M \) | Precision | Recall |
| 0.07 | 0.10 | 0.45 | 0.50 | 0.66 | 0.58 |
| 0.07 | 0.10 | 0.45 | 0.55 | 0.66 | 0.58 |
| 0.10 | 0.10 | 0.50 | 0.55 | 0.68 | 0.65 |
| 0.07 | 0.13 | 0.45 | 0.50 | 0.65 | 0.57 |
| 0.07 | 0.13 | 0.45 | 0.55 | 0.65 | 0.57 |
| 0.10 | 0.13 | 0.45 | 0.50 | 0.68 | 0.63 |
| 0.10 | 0.13 | 0.45 | 0.55 | 0.68 | 0.63 |
| 0.10 | 0.13 | 0.50 | 0.55 | 0.68 | 0.63 |
paring patterns in textual and visual descriptions of the multimedia news items.

Acknowledgment

This work is supported in part by the National Institute of Information and Communications Technology. This research is partly supported by the research for the grant of Scientific Research (No.20700084, 20300042) made available by MEXT, Japan.

References


