Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 5601–5608
Marseille, 20-25 June 2022

Developing a Dataset of Overridden Information in Wikipedia

Masatoshi Tsuchiya, Yasutaka Yokoi
Toyohashi University of Technology
1–1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, Japan
{tsuchiya, yokoi}@is.cs.tut.ac.jp

Abstract

This paper proposes a new task of detecting information override. Since all information on the Web is not updated in a timely manner, the necessity is created for information that is overridden by another information source to be discarded. The task is formalized as a binary classification problem to determine whether a reference sentence has overridden a target sentence. In investigating this task, this paper describes a construction procedure for the dataset of overridden information by collecting sentence pairs from the difference between two versions of Wikipedia. Our developing dataset shows that the old version of Wikipedia contains much overridden information and that the detection of information override is necessary.

Keywords: Information Override; Recognizing Textual Entailment

1. Introduction

The vast amount of information provided through the Web has become an indispensable tool in our daily lives. The appropriate selection of this information is necessary because it is unrealistic to refer to all information (Marshall and Shipman, 1997; Macskassy and Provost, 2001). In particular, since all information on the Web is not updated in a timely manner, it is essential to discard information overridden by another information source. As an example, consider the two sentences shown in Table 1. Since *Oshida Station* was discontinued in 2016, as described in the sentence s_r, it is inappropriate to refer to the sentence s_t for the transportation method to Asagishi Aza Oshida, Morioka City, in 2022.

s_t	As of the year 2010, *Oshida Station* is an unmanned station in Asagishi Aza Oshida, Morioka City, Iwate Prefecture, Japan, with three train stops per day.
s_r	*Oshida Station* was an unmanned station in Asagishi Aza Oshida, Morioka City, Iwate Prefecture, Japan, which was discontinued in the year 2016, and is not in use as of the year 2018.

Table 1: Example of overridden information. Because the reference sentence s_r has overridden the information in the target sentence s_t, the information in the sentence s_t is outdated, and the sentence s_t cannot be referred to by a user in 2022.

The Web contains much outdated information like the sentence s_t, which is an obstacle to information utilization. To support the use of such information, methods have been proposed to organize and present this information sequentially (Fung et al., 2007; Wang et al., 2008). These studies, however, do not provide a way to automatically determine whether a particular piece of information is inappropriate for current reference from a users' perspectives.

Based on the above discussion, we propose a new task detecting **overridden information** like the sentence s_t. There are two possible settings for detecting overridden information: one is to determine whether the reference sentence s_r has overridden the target sentence s_t using both the target sentence s_t and the reference sentence s_r as input, and the other is to determine whether the target sentence s_t has been overridden using only the target sentence s_t as input. Considering the original goal, which is to support information usage by automatically detecting overridden information, the latter setting, which takes only the target sentence as input, is more reasonable. However, to realize the latter setting, it is necessary to solve the subtask to discover the reference sentence that overrides the target sentence at the same time, which is expected to make the whole task extremely difficult. Therefore, this paper will focus on the former setting to determine whether the reference sentence s_r has overridden the target sentence s_t, using the target sentence s_t and the reference sentence s_r as input.

The significant contributions of this paper are the following three points:

- This paper proposes a new task of detecting overridden information and formalizes it as a binary classification problem to determine whether a reference sentence has overridden a target sentence (Section 1).
- This paper offers a formal definition of information override between two sentences while relating it to textual entailment (Section 2).
- This paper proposes a procedure to construct a dataset of overridden information by collecting sentence pairs from the difference between two versions of Wikipedia (Section 3) and reports the construction result (Section 4).

2. Definition of Information Override

This section discusses the formal definition of information override while relating it to textual entailment.
Intuitively, information override is making the information in the old sentence outdated when the new sentence has overridden the old sentence, as shown in Table 1. For further analysis, let us suppose the following sentence s_n.

$s_n =$ Morioka City constructed a 13-kilometer sidewalk from Oshida Station to Asagishi Station, which became available for use on May 15, 2015.

These three sentences, s_t, s_r, and s_n, share the same topic, *Oshida Station*; however, the relation between the sentence s_t and the sentence s_r and the relation between the sentence s_t and the sentence s_n are entirely different. Although the sentence s_r can override the station operation status described in the sentence s_t, the sentence s_n can never override it. This example suggests that the information override relation exists not for all sentence pairs but only for specific sentence pairs.
This observation leads us to two issues: first, what kind of sentence pair information override relation exists, and second, what kind of phenomenon occurs in those sentence pairs.

2.1. Condition of Information Override

This section discusses the condition of sentence pairs with information override relations through observations of pairs with and without information override relations.
Let us consider the following sentence pair, s_1 and s_2, as an example pair with information override.

$s_1 =$ The tallest building in Japan is Yokohama Landmark Tower, as of the year 2010.
$s_2 =$ The tallest building in Japan is Abeno Harukas, as of the year 2015.

Sentence s_2 has obviously overridden the information in sentence s_1 about *the tallest building in Japan*, and it is natural to think that the information in sentence s_1 is outdated from the user's point of view in 2022.
For a formal discussion of this phenomenon, let us consider the content of the sentence s as a tuple consisting of the content x that is valid regardless of time and the time t that the sentence s is focused on.

$$s = \langle x, t \rangle$$

In this view, it is possible to decompose the content of sentence s_1 into the tuple of x_1 and t_1 as follows.

$s_1 = \langle x_1, t_1 \rangle$
$x_1 =$ The tallest building in Japan is Yokohama Landmark Tower.
$t_1 =$ as of the year 2010

It is also possible to decompose the content of sentence s_2 into the tuple of x_2 and t_2 as follows.

$s_2 = \langle x_2, t_2 \rangle$
$x_2 =$ The tallest building in Japan is Abeno Harukas.
$t_2 =$ as of the year 2015

Let us suppose sentence $s_{1,2}$ which is generated from sentence s_1 by replacing its time only with the time of sentence s_2.

$s_{1,2} = \langle x_1, t_2 \rangle$
$=$ The tallest building in Japan is Yokohama Landmark Tower as of the year 2015.

Note that sentence $s_{1,2}$ contradicts sentence s_2.
Next, we consider the following sentence, s_3, which does not override sentence s_1.

$s_3 =$ The tallest building in the world is Burj Khalifa as of the year 2015.

It is easy to decompose sentence s_3 into the tuple in the same procedure as for sentence s_2.

$s_3 = \langle x_3, t_3 \rangle$
$x_3 =$ The tallest building in the world is Burj Khalifa.
$t_3 =$ as of the year 2015

Because the topic of sentence s_1, *the tallest building in Japan*, and the topic of sentence s_3, *the tallest building in the world*, are entirely different, sentence s_3 can never override the information in sentence s_1. This intuition reinforces the analysis that sentence $s_{1,3}$ is equal to sentence $s_{1,2}$ and does not contradict sentence s_3.

2.2. Relation between Information Override and Textual Entailment

This section defines information override based on the above observations and explains the relation between information override and textual entailment.
Previous studies, including (Marelli et al., 2014; Bowman et al., 2015), define the recognition of textual entailment as a 3-class classification problem that takes both a premise sentence s_p and a hypothesis sentence s_h as input and returns three values: entailment, contradiction, and independence. The following function f is its formal definition.

$$f(s_h, s_p) = \begin{cases} \text{if } s_h \rightarrow s_p \vee s_p \rightarrow s_h \\ \quad \text{entailment} \\ \text{if } s_h \wedge s_p = \phi \\ \quad \text{contradiction} \\ \text{otherwise} \\ \quad \text{independence} \end{cases} \quad (1)$$

Note that this widely adopted definition ignores the direction of the entailment relation. In order to keep the latter discussion of information override simple, this definition is also adopted in this study.

The target sentence s_t and the reference sentence s_r are decomposed into tuples of their content and their focusing times as follows:

$$s_t = \langle x_t, t_t \rangle \quad (2)$$
$$s_r = \langle x_r, t_r \rangle \quad (3)$$

Based on the observation in Section 2.1., we define the detection of overridden information as a binary classification problem that takes two sentences, a target sentence s_t and a reference sentence s_r, as input and returns binary values. The following function g is its formal definition.

$$g(s_t, s_r) = \begin{cases} \text{if } f(s_t, s_r) = \text{contradiction} \\ \quad \vee f(\langle x_t, t_r \rangle, s_r) = \text{contradiction} \\ \quad \vee f(s_t, \langle x_r, t_t \rangle) = \text{contradiction} \\ \quad \text{override} \\ \text{otherwise} \\ \quad \text{neutral} \end{cases} \quad (4)$$

The above definition of information override is consistent with user intuition in most cases, but requires consideration for information that changes over time. Suppose the following three sentences, s_4, s_5, and s_6, are examples of such information.

$s_4 =$ The number of applicants to Touto University for the 2014 academic year is 2,000.
$s_5 =$ The number of applicants to Touto University for the 2015 academic year is 3,000.
$s_6 =$ The number of applicants to Touto University for the 2015 academic year increased by 1,000 compared to the previous year.

Note that sentence s_5 and the combination of sentence s_4 and sentence s_6 are other representations of the same fact, "The number of applicants to Touto University for the 2015 academic year is 3,000."

First, we analyze the relation between sentence s_4 and sentence s_5. Following the procedure described in Section 2.1., sentence s_4 is decomposed into the tuple consisting of its content x_4 and its focusing time t_4 as follows:

$s_4 = \langle x_4, t_4 \rangle$
$x_4 =$ The number of applicants to Touto University is 2,000.
$t_4 =$ for the 2014 academic year

Sentence s_5 is also decomposed into the tuple consisting of its content x_5 and its focusing time t_5 as follows:

$s_5 = \langle x_5, t_5 \rangle$
$x_5 =$ The number of applicants to Touto University is 3,000.
$t_5 =$ for the 2015 academic year

Suppose sentence $s_{4,5}$, which is generated from sentence s_4 by replacing its time only with the time of sentence s_5.

$s_{4,5} = \langle x_4, t_5 \rangle$
$=$ The number of applicants for the 2015 academic year of Touto University is 2,000.

Since sentence $s_{4,5}$ contradicts sentence s_5, the definition in Eq. 4 considers that sentence s_5 has overridden the information in sentence s_4 and that the information in sentence s_4 is outdated, even though the sentence s_4 is still true as the information for the 2014 academic year.

Second, we move on to an analysis of the relation between sentence s_4 and sentence s_6. It is possible to decompose sentence s_6 into a tuple as follows:

$s_6 = \langle x_6, t_6 \rangle$
$x_6 =$ The number of applicants to Touto University increased by 1,000 compared to the previous year.
$t_6 =$ for the 2015 academic year

Sentence $s_{4,6}$ is equal to sentence $s_{4,5}$ since the time of sentence s_6 is equal to the time of sentence s_5. Because sentence $s_{4,6}$ does not contradict sentence s_6, the definition in Eq. 4 considers that sentence s_6 has not overridden sentence s_4 and that the information in sentence s_4 is not outdated. In other words, this analysis concludes that the relation between sentence s_4 and sentence s_5 and the relation between sentence s_4 and sentence s_6 are different, even though sentence s_5 and the combination of sentence s_4 and sentence s_6 are other representations of the same fact.

Although this strange behavior may be caused by the incompleteness of the definition of information override in Eq. 4, we employ it in the following discussion. The reason is that it was useful as the guideline for actual annotation work in many cases.

3. Construction Procedure of the Dataset

This section describes the procedure to construct a dataset of overridden information. It consists of two parts: the first part is to prepare a text source, and the second part is to collect target and reference sentence pairs from the text source.

3.1. Selection of Text Source

This section discusses a text source for collecting target and reference sentences with information overrides and explains why this paper focuses on Wikipedia as the text source.

Considering the original goal, which is to support information usage by automatically detecting overridden information, it would be appropriate to collect sentences written in an accessible style, such as blog posts by individuals, as the target sentences, and sentences written in a reliable and standardized style, such as newspaper articles, as the reference sentences. In realizing this setup, the target sentences must be collected from blog posts and the reference sentences from newspaper articles. However, it is difficult to collect target sentences including overridden information from many normal

# of articles exist on December 11, 2014	943,488
# of articles exist on December 20, 2018	1,132,813
# of created articles	195,828
# of deleted articles	6,503
# of articles that exist in both versions	936,985

Table 2: Statistics of articles on Japanese Wikipedia. Target and reference sentences were collected from 936,985 articles that existed in both versions.

sentences of blog posts. It is also difficult to collect the reference sentences corresponding to the target sentences because there is no mapping between blog posts and newspaper articles.

Crowdsourcing is widely employed to create many sentence pairs that meet the research objectives (Bowman et al., 2015). In this method, a worker is presented with a target sentence (or a reference sentence) and is asked to compose a reference sentence (or a target sentence) suitable for the given sentence. However, (Gururangan et al., 2018; Tsuchiya, 2018) point out that free sentence composition by human workers may cause hidden bias. Collecting sentences from an existing text source is necessary to avoid this kind of bias.

Based on the above discussion, this paper focuses on the difference between two versions of Wikipedia as the text source for collecting target and reference sentences. If Wikipedia articles are updated to reflect changes in their describing items, newer articles are expected to override information about older ones. In particular, modified sentences in a new version of articles are expected to cause information override, instead of unmodified sentences that exist in both versions. Therefore, we propose collecting modified sentences of the new version of articles as reference sentences and sentences in the old version of articles as target sentences by comparing two versions of Wikipedia articles that describe the same item.

3.2. Collecting Target and Reference Sentences

This section describes how to collect target and reference sentences from Wikipedia. Since Wikipedia is continuously updated for various reasons, the difference between the old and new versions contains a huge number of modified sentences, most of which are not related to information override. Therefore, we need several devices to collect sentence pairs related to information override.

Table 2 shows the article-level difference between Japanese Wikipedia on December 11, 2014, and Japanese Wikipedia on December 20, 2018. 195,828 articles were created, and 6,503 articles were deleted during the period between the two versions. Because it is difficult to discover articles related to these created and deleted articles from the old version of Wikipedia, we focused on 936,985 articles that existed in both versions when collecting target and reference sentences.

A huge number of modified sentences were discovered by comparing the two versions of Wikipedia. Because it is not feasible to annotate all of them, the following four filtering steps were employed:

1. To collect sentence-to-sentence changes,
2. To collect changes containing either references to updated articles or time and date expressions,
3. To ignore changes caused by minor editing, and
4. To collect changes from high quality articles.

The first step was to collect sentence-to-sentence changes. Suppose a Wikipedia article is updated to reflect changes in the concept being decribed. In this case, it is considered that the new version of the article has overridden the information in the old version of the article, whether in part or whole. Since such article-level information override is too difficult to analyze, we will not cover it in this study and limit our discussion to sentence-level information override. Therefore, only single-sentence changes were collected by comparing the new and old versions of the article text, and multi-sentence changes were ignored.

The second step was to collect changes containing either headwords of articles which were updated during the period between the two versions or date expressions referring the same period[1]. Since there are many reasons why Wikipedia articles are updated, as described in (Yang et al., 2017), many changes are not related to information override, and a device to pick differences related to information override is needed. We assumed that differences related to information updated during the concerned period were most likely related to information override. The second step, which was designed based on this assumption, collected changes that satisfied the above condition and gave us 193,142 sentence pairs.

The third step was to employ the agreement ratio of the target and reference sentences to ignore changes caused by minor editing, such as adding/removing commas or fixing typos. Where $L(s)$ is the length of the sentence s and $C(s, s')$ is the longest common sequence of two sentences, the agreement ratio of the target sentence s_t and the reference sentence s_r is defined as follows:

$$r(s_t, s_r) = \frac{L(C(s_t, s_r)) \times 2}{L(s_t) + L(s_r)} \quad (5)$$

Collecting only sentence pairs whose agreement ratio was less than or equal to 0.6 gave us 15,648 sentence pairs.

The final step was to prioritize differences in high quality articles to collect significant sentence pairs preferentially. Wikipedia's internal PageRanks was employed to sort articles following the same procedure of (Rajpurkar et al., 2016). Finally, the 9,600 sentence pairs were obtained as manual annotation targets.

[1] Date expressions referring the concerned period were automatically extracted by hand-crafted regular expressions.

	Override	Neutral	Sum.
Entailment	430 (4.5%)	3,765 (39.2%)	4,195 (43.7%)
Contradiction	1,359 (14.2%)	0 (0.0%)	1,359 (14.2%)
Independence	1,301 (13.6%)	2,745 (28.6%)	4,046 (42.1%)
Sum.	3,090 (32.2%)	6,510 (67.8%)	9,600 (100.0%)

Table 3: Distribution of textual entailment labels and information override labels. Because 32.2% of target pairs were judged to be overridden information, it was revealed that the old version of Wikipedia contained much overridden information.

Article title	Target sentence	Reference sentence	Textual entailment label	Information override label
Exceisior Cafe	As of March 2014, it operates three stores in Tokyo and Saitama.	As of October 2016, it operates only one store in Saitama.	independence	override
Machinori (rental bicycle)	As of November 2012, it operates 19 service stations.	As of May 2017, it operates 22 service stations, including its office.	independence	override
President of Italy	The current holder is Giorgio Napolitano.	The current holder is Sergio Mattarella.	contradiction	override
Okinawa Urban Monorail	It is scheduled to open in the spring of 2019.	It was scheduled to open in the spring of 2019, but it was announced in May 2018 that it would be in the summer of 2019 at the earliest.	contradiction	override
Kunitachi Station	The ticket gates were consolidated to one under the elevated tracks as of January 13, 2013.	The ticket gates were consolidated to one under the elevated tracks as of January 13, 2013, but Nonowa Gate was built on the west side on April 24, 2016, bringing the number of ticket gates to two.	entailment	override
Komeri Co.	As of August 2011, it operates stores in all prefectures except Okinawa.	As of July 2018, it operates stores in all prefectures except Okinawa.	independence	neutral
Nihon University Itabashi Hospital	It is operated by Nihon University Educational Corporation.	It is an affiliated hospital of Nihon University School of Medicine.	entailment	neutral

Table 4: Examples of annotated sentence pairs

4. Construction Result

This section reports the construction result of the dataset of overridden information with the described procedure and discusses the difficulty of detecting overridden information based on the preliminary experimental results using the existing NN models proposed for recognizing textual entailment.

4.1. Annotation

This section explains our annotation results and examples of overridden information.

We manually assigned three types of textual entailment labels (entailment, contradiction, and independence) and two types of information override labels (override and neutral) to 9,600 sentence pairs collected by the procedure described in Section 3.2. Table 3 shows that 32.2% of the reference sentences were judged to override the information in the target sentences. This result means that many pieces of information on Japanese Wikipedia on December 11, 2014 were overridden by Japanese Wikipedia on December 20, 2018 and an automatic detection of information override is needed.

To check the quality of the annotation work, the inter annotator agreement was measured. We collected 300 sentence pairs from articles sorted by Wikipedia's internal PageRanks and asked two individual annotators to annotate them. The agreement ratio for textual entailment labels was 83.7%, and the agreement ratio for information override labels was 88.0%. Since these agreement ratios are high enough, we conclude that the annotation work was stable and reliable.

Table 4 shows several examples of annotated sentence pairs[2]. The sentence pair collected from *Exceisior Cafe* is the simplest example of information override. Its target sentence described the store operation status in March 2014, whereas its reference sentence described it in October 2016. Since these two sentences described

[2]The sentence pairs collected from Japanese Wikipedia are written in Japanese, but Table 4 shows their English translations for an explanation.

information at completely different moments, there is no entailment relation between them. The definition of information override in Eq. 4, however, considers that the reference sentence has overridden the store operation status described in the target sentence.

The sentence pair collected from *President of Italy* in Table 4 is an example of hidden time information. Its target sentence described the president of Italy in 2014, whereas its reference described the president of Italy in 2018. A superficial comparison of these two sentences without considering time information would suggest that they are contradictory. When considering time information, it is natural to think that the reference sentence has overridden the information in the target sentence.

Table 3 shows that 4.5% of sentence pairs were judged as demonstrating the entailment and information override relation. The sentence pair collected from *Kunitachi Station* in Table 4 is an example of such a sentence pair. Both sentences explained that the number of ticket gates was decreased to one on January 13, 2013; therefore, it is possible to entail the target sentence from the reference sentence when the reference sentence is given as a premise sentence. Because the entailment direction is ignored, as shown in Eq. 1, this sentence pair was judged as being connected by an entailment relation. Moreover, the reference sentence held newer information on the number of ticket gates compared to the target sentence. It is natural to think that the reference sentence has overridden the information of the target sentence.

These examples show that textual entailment and information override are different types of relations between sentences.

4.2. Difficulty of Detecting Overridden Information

This section shows the experimental results to check the difficulty of the task setting of this study. Because the detection of information override is deeply related to the recognition of textual entailment, as already described in Section 2, two LSTM models for recognizing textual entailment were adopted.

The first model (henceforth denoted as the parallel LSTM model) was proposed by (Bowman et al., 2015) for recognizing textual entailment. The following equations define this model.

$$
\begin{aligned}
\boldsymbol{h}_{p,i} &= \mathrm{LSTM}_p(W_e x_{p,i} + W_{hp}\boldsymbol{h}_{p,i-1}) \\
\boldsymbol{h}_{h,i} &= \mathrm{LSTM}_h(W_e x_{h,i} + W_{hh}\boldsymbol{h}_{h,i-1}) \\
\boldsymbol{l}_1 &= \tanh(W_1[\boldsymbol{h}_{p,|\boldsymbol{x_p}|}, \boldsymbol{h}_{h,|\boldsymbol{x_h}|}] + B_1) \\
\boldsymbol{l}_2 &= \tanh(W_2\boldsymbol{l}_1 + B_2) \\
\boldsymbol{l}_3 &= \tanh(W_3\boldsymbol{l}_3 + B_3) \\
\boldsymbol{y} &= \mathrm{softmax}(\boldsymbol{l}_3)
\end{aligned}
$$

The first step is to convert a premise sentence $\boldsymbol{x}_p$ (or a reference sentence) and a hypothesis sentence $\boldsymbol{x}_h$ (or a target sentence) into embedding vectors using the word embedding matrix W_e, which is initialized with the 200-dimension vectors trained from Japanese Wikipedia by (Bojanowski et al., 2016). The second step is to convert the embedding vectors into two 100-dimension sentence vectors with LSTMs, and they are concatenated into a 200-dimension vector. The remaining steps are to predict a textual entailment label (or an information override label) with three tanh fully connected layers and then to apply the softmax function.

The second model (henceforth denoted as the sequential LSTM model), which was proposed by (Rocktäschel et al., 2015) for recognizing textual entailment, is defined as follows.

$$
\begin{aligned}
\boldsymbol{h}_{p,i} &= \mathrm{LSTM}_p(W_e x_{p,i} + W_{hp}\boldsymbol{h}_{p,i-1}) \\
\boldsymbol{h}_{h,0} &= \mathrm{LSTM}_h(W_{hh}\boldsymbol{h}_{p,|\boldsymbol{x_p}|}) \\
\boldsymbol{h}_{h,i} &= \mathrm{LSTM}_h(W_e x_{h,i} + W_{hh}\boldsymbol{h}_{h,i-1}) \\
\boldsymbol{l} &= \tanh(W_l\boldsymbol{h}_{h,|\boldsymbol{x_h}|} + B_l) \\
\boldsymbol{y} &= \mathrm{softmax}(\boldsymbol{l})
\end{aligned}
$$

In the second model, two LSTMs are sequentially connected. Thus, it is possible to consider that the memory cells of these LSTMs are directly modeling a recognition process unlike the parallel LSTM model. All vectors of the sequential LSTM model are 100-dimension. Although (Rocktäschel et al., 2015) proposed the variants with an attention layer between the premise and hypothesis sentences, the attention-less model was employed in this experiment because of its simplicity.

Before the experiment, it was necessary to split the dataset. Our dataset, shown in Table 3, was randomly divided into three subsets for the experiments. Table 5 shows that 8,000 sentence pairs were prepared for training, 800 sentence pairs for development, and 800 sentence pairs for testing, respectively.

Table 6 shows the classification accuracy of the above two NN models. The baseline row shows the accuracy when the majority label was selected. Because the accuracy of recognizing textual entailment was still lower than their performance against English textual entailment datasets, these results suggested that more training pairs were required. Due to the poor accuracy in detecting overridden information, it was suggested that more training instances or sophisticated models were needed.

5. Conclusion

This paper proposed a new task detecting information override and formalized it as a binary classification problem that takes a sentence pair as input. Through observations of sentence pairs with information override, we provided a formal definition of information override while relating it to textual entailment. According to this definition, the construction procedure of the dataset of overridden information, which collected target and reference sentences from the difference between two versions of Wikipedia, was proposed. Our developing dataset has revealed that the old version of

	Training set		Development set		Test set	
	Override	Neutral	Override	Neutral	Override	Neutral
Entailment	354 (4.4%)	3,168 (39.6%)	38 (4.8%)	313 (39.1%)	38 (4.8%)	284 (35.5%)
Contradiction	1,118 (14.0%)	0 (0.0%)	104 (13.0%)	0 (0.0%)	137 (17.1%)	0 (0.0%)
Independence	1,094 (13.7%)	2,266 (28.3%)	107 (13.4%)	238 (29.8%)	100 (12.5%)	241 (30.1%)
Sum.	8,000		800		800	

Table 5: Statistics of divided sub datasets.

Model	Textual entailment	Information override
Parallel LSTM Model	0.639	0.803
Sequential LSTM Model	0.630	0.809
Baseline	0.426	0.656

Table 6: Classification accuracy. Because the accuracy achieved by the existing NN models for recognizing textual entailment was lower than their accuracy against English textual entailment datasets, these results suggested that more sentence pairs are required to train these models. Because their accuracy in detecting overridden information was also poor, it was suggested that more training instances or sophisticated models were required.

Wikipedia contained much overridden information and that detection of information override is necessary.

Because our dataset is still developing, several points must be examined. Since there are many cases of information override between single sentences and multiple sentences in the actual data, it is necessary to relax the limitation of this study, which considers only sentence-level information override. Although it is challenging to collect pairs of contradictory sentences from a real world text sources, Table 3 suggests that the difference between the two versions of Wikipedia is a promising source. We will investigate ways to build various datasets by analyzing the reasons for updating Wikipedia.

6. Bibliographical References

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2016). Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Fung, G. P. C., Yu, J. X., Liu, H., and Yu, P. S. (2007). Time-dependent event hierarchy construction. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '07, page 300–309, New York, NY, USA. Association for Computing Machinery.

Gururangan, S., Swayamdipta, S., Levy, O., Schwartz, R., Bowman, S., and Smith, N. A. (2018). Annotation artifacts in natural language inference data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 107–112, June.

Macskassy, S. A. and Provost, F. (2001). Intelligent information triage. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '01, pages 318–326, New York, NY, USA. Association for Computing Machinery.

Marshall, C. C. and Shipman, F. M. (1997). Spatial hypertext and the practice of information triage. In *Proceedings of the Eighth ACM Conference on Hypertext*, HYPERTEXT '97, pages 124–133, New York, NY, USA. Association for Computing Machinery.

Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P. (2016). SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392.

Rocktäschel, T., Grefenstette, E., Hermann, K. M., Kociský, T., and Blunsom, P. (2015). Reasoning about entailment with neural attention. In *Proceedings of The International Conference on Learning Representations (ICLR2016)*.

Tsuchiya, M. (2018). Performance impact caused by hidden bias of training data for recognizing textual entailment. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Wang, C., Zhang, M., Ru, L., and Ma, S. (2008). Automatic online news topic ranking using media focus and user attention based on aging theory. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, page 1033–1042, New York, NY, USA. Association for Computing Machinery.

Yang, D., Halfaker, A., Kraut, R., and Hovy, E. (2017). Identifying semantic edit intentions from revisions in Wikipedia. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2000–2010.

7. Language Resource References

Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D. (2015). A large annotated corpus for learning natural language inference. In *Proceedings of* *the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642, Lisbon, Portugal, September. Association for Computational Linguistics.

Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R., and Zamparelli, R. (2014). A SICK cure for the evaluation of compositional distributional semantic models. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 216–223, Reykjavik, Iceland, may. European Language Resources Association (ELRA).