Quality Estimation for Japanese Haiku Poems Using Neural Network

Shinji Kikuchi, Keizo Kato, Junya Saito, Seiji Okura,
Kentaro Murase, Takaya Yamamoto and Akira Nakagawa
FUJITSU LABORATORIES LTD.
4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki, Kanagawa 211-8588, Japan
{skikuchi, kato.keizo, saito.junya, okura.seiji, kmurase, t.yamamoto, anaka}@jp.fujitsu.com

***Abstract*—We propose a method to estimate the artistic quality of Haiku (Japanese style short poem) texts using a machine learning approach. Based on the assumption that the artistry of a text stems from its sound factors as well as its meanings, we first constructed two types of vector models, a word-based model and a syllable-based model, converted from Haiku texts. Next, we conducted machine learning for these two models using a convolutional neural network to construct a Haiku quality estimation function. We then evaluated the precision of quality estimation for 40,000 Japanese Haiku poems obtained from a Haiku community site, assuming that the number of "likes" given from viewers to a Haiku corresponds to its artistic quality. Through the experiment, we confirmed that by conducting a quality estimation based on the consensus between different models, we can improve the precision of quality estimation up to 0.64. We also found that if we evaluate Haiku poems for which we have high confidence in quality estimation certainty, the F-measure of the estimation improved from 0.57 to 0.64.**

Keywords—Haiku; machine learning; sentiment analysis; neural network; text mining

I. INTRODUCTION

Along with the progress in natural language processing techniques, various tools and technologies have been developed for helping with text editing. For example, Cheng et al. [1] proposed an example-based proofreading approach for Chinese-Japanese translation. This type of approach works well for identifying grammatical errors and can be a great help in business document editing.

On the other hand, there are also various types of texts whose quality cannot be evaluated only by their grammatical correctness. For example, poems and song lyrics are usually evaluated from their artistic quality rather than grammatical rules. If we can realize the artificial sensitivity to evaluate the artistic quality of texts by machine, it would broaden the application areas of text editing assistance and text generation techniques. However, implementing an artistic evaluation function for texts is quite a challenging problem, since artistry cannot be determined by explicit grammatical patterns.

While there are a variety of artistic style texts, Japanese Haiku [22] has one of the simplest styles, consisting of only 17 syllables. Because of its simplicity, we suppose that Haiku is one of the best starting points to develop and evaluate a text artistry evaluation function. If we can evaluate the quality of Haiku poems using computers, we can enhance our approach to more complex targets such as other poems and song lyrics. Based on these assumptions, we decided our goal in this research would be to develop a method to evaluate the quality of Haiku poems.

In order to achieve the goal, we made an assumption that the artistic characteristics of Haiku and other texts (e.g. poems or songs) stem from its sound factors as well as its meanings. Based on this assumption, we supposed that if we take into account the words and their sounds in texts, we can improve the estimation accuracy of text artistry. We also assume that it is difficult to identify some explicit strategies or heuristics that can be used to make high-quality artistic texts, because artistic content cannot be created by simply abiding by certain rules such as grammar rules. Therefore, we decided that the black-box approach was suitable for the quality estimation, whereby we construct quality estimation functions from a dataset including artistic texts, and obtain quality ratings for them using machine learning techniques.

Based on this assumption, we designed a quality evaluation method for Haiku texts using the machine learning approach. In our approach, we first collected Haiku texts from a Haiku community site along with the number of "likes" given from viewers for each Haiku. Assuming that the number of "likes" reflects the artistic quality of the Haiku, we divide the set of Haiku texts into two categories: high quality and low quality. Next, we convert the Haiku texts into vector models in the following two ways: (1) a word-based approach, converting each word into a vector and (2) a syllable-based approach, assigning a vector to each syllable. Then we conduct machine learning using a convolutional neural network (CNN) [5] for both of these vectors to construct a model to estimate the quality (high or low) of a Haiku poem from its text. After constructing the estimation model, we evaluate the estimation accuracy of each Haiku's quality by cross-validation of the set of Haiku.

The main contributions of this paper can be summarized as follows.

- Proposal of a text artistry estimation method using a machine learning approach in which we model texts in a word-based and syllable-based approach.

- Accuracy evaluation of the text artistry estimation method using 40,000 Haiku poems.

The rest of this paper is organized as follows. In Section II we first provide an overview of Haiku poems and explain the reason why they are suitable examples for text artistry evaluation. Next, in Section III we explain the details of our text quality estimation method using the machine learning approach. We then show the evaluation results for the accuracy of quality estimation for a large number of Haiku poems in Section IV. We also discuss the difficulties in quality estimation of Haiku based on the evaluation results. After providing an account of related studies in Section V, we conclude our findings in Section VI.

II. Overview of Haiku

Haiku [22] is a style of short Japanese poem developed in the 17th century. Figure 1 shows an example of Haiku. This is one of the most famous Japanese Haiku texts, written by Haiku master Basho Matsuo, depicting a scene where he saw a frog dive into an old pond. There are the following two basic rules for Haiku poems.

(1) Haiku consists of three phrases. These phrases have 5, 7 and 5 syllables respectively. Therefore, a traditional Haiku consists of 17 syllables.

(2) Haiku include one word representing a season, called "Kigo" (seasonal word). The definition of the seasonal words and their seasons are on a specific list called Saijiki [23]. For example, "frog" is the word representing spring in the Haiku in Figure 1.

Haiku is short and simple, which makes Haiku quite a popular pastime activity in Japan. For example, most of the Japanese national newspapers have Haiku sections, to which several thousand readers submit their Haiku every week [2]. The selection is so competitive that only a small fraction (less than 1%) of submitted Haiku is published in the newspapers after thorough reviews by Haiku masters. There are also a lot of Haiku community websites [3] to which people submit their Haiku and give comments and ratings based on their feelings.

While there are various types of artistic texts such as poetry and song lyrics, we believe that Haiku is quite a good starting point for the research and development of text artistry evaluation technology for the following reasons.

- Compared with other artistic texts, it has a short and simple format (5-7-5 syllable pattern). This simple format has the advantage in modeling texts that we can analyze them by computer. If we can evaluate the quality of Haiku, we can enhance our approach to more complex targets such as other poems in more flexible formats as well as song lyrics.
- Haiku is so popular that we can obtain index values representing its artistic qualities in several ways. For example, we can find Haiku poems praised by many viewers on Haiku community websites. We can also find a large number of good-quality Haiku poems in certain national newspapers that have undergone review by Haiku masters.

Text (Japanese):	古池や	蛙飛び込む	水の音
Pronunciation :	Hu ru i ke ya,	Ka wa zu to bi ko mu,	Mi zu no o to
Translation :	An old pond,	a frog dives in,	sound of splash

Figure 1: An example of Haiku by Basho Matsuo

III. Haiku Quality Estimation Method

In this section, we explain our quality estimation method for Haiku in detail.

Figure 2 shows an overview of our Haiku quality estimation method using machine learning. As explained in Section I, based on the assumption that the sound of text pronunciation as well as the text's meaning is an important factor having an impact on the artistry of texts, we conduct machine learning for both words and syllables in texts.

The quality estimation method consists of the following four steps: (1) collecting Haiku texts and their ratings, (2) decomposing texts into words and syllables, (3) converting them into vector representations, (4) conducting machine learning to construct models to determine the relation between vectors and ratings, and (5) estimating the rating of a target Haiku text based on the models. The details of these steps are as follows.

Step 1: Data collection

First we collect Haiku texts along with ratings for them from data sources such as websites or newspapers. The ratings are the parameters representing a Haiku's artistic quality. There are various types of rating index such as scores marked by experts or the number of "likes" rated by a large number of viewers. For simplicity, here we suppose that each text has either "high quality" or "low quality" as its rating.

Step 2: Text decomposition

We decompose Haiku texts in two ways: word by word and syllable by syllable. In word-by-word decomposition, we

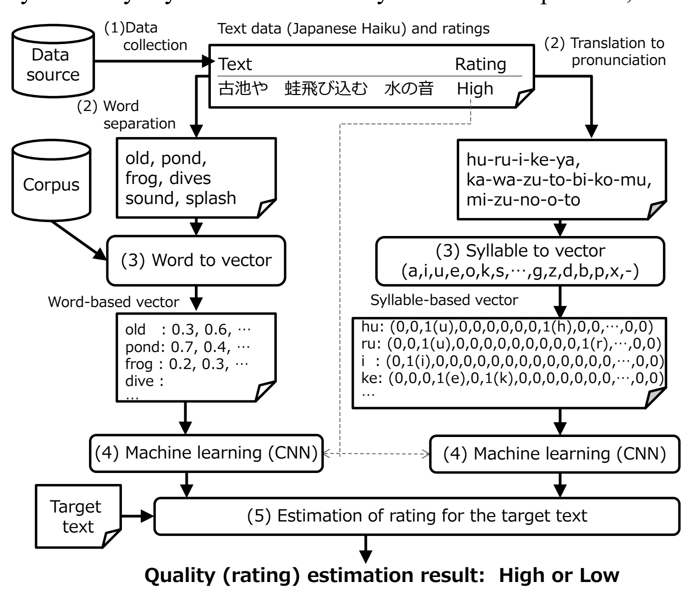

Figure 2: Overview of Haiku quality estimation method

simply extract each word from Haiku texts. For texts written in languages such as Japanese, in which words in phrases are not separated by blanks, we use the morphological analysis tool MeCab [4], which can identify each word in a sentence. For syllable-based decomposition, we extract syllables from the texts by referring to pronunciation dictionaries. In the case of Japanese language for Haiku, a syllable basically consists of a combination of a vowel (a, i, u, e, o), a consonant (k, s, t, n, h, m, y, r, w, g, z, d, b, p), and additional factors (double consonant, contracted sound and prolonged sound) represented by (nn, x, -). For example, by decomposing the Haiku in Figure 1, we obtain a word set {古(old), 池(pond), …} and a pronunciation set {hu, ru, i, ke, ya, …}.

Step 3: Conversion into vectors

Here we convert the word sets and pronunciation sets into vectors so that we can input them into our machine learning component. For the word set, we prepare a vector corpus recording a large number of words and their vector expressions beforehand. While various tools can be used to translate words into vectors, we used word2vec [14] because of its convenience and performance. By using the vector corpus as a dictionary, we convert each word in the word sets into a corresponding vector representation with 200 dimensions. For the pronunciation sets, we convert each syllable into a vector with 22 dimensions. Each dimension represents one of the pronunciation characters (a, i, u, e, o, k, s, t, n, h, m, y, r, w, g, z, d, b, p, nn, x, -). If a syllable has some corresponding vowels, consonants or additional factors, we set the corresponding dimension of the vector expression as one. For example, the vector expression for the first syllable "hu" in the Haiku in Figure 1 is (0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0) since the syllable consists of a consonant "h" and a vowel "u".

Step 4: Machine learning

After obtaining the word-based and syllable-based vectors, we conduct machine learning for them independently to identify models representing the relation between these vectors and their quality rating (high or low). Here we use the CNN proposed by Kim [5] and Britz [6], which has been used for sentiment analysis [10] for texts.

Figure 3 shows an overview of the CNN for Haiku classification for word vectors. It consists of three layers; a convolution layer, a max-pooling layer and a softmax layer. First, in the convolution layer, we calculate the convolutions for Haiku texts represented by vectors. Here we use several filter convolutions with different sizes. For example, when we conduct convolutions with a filter whose size is two, two consecutive words in Haiku are convoluted into one value in the sliding window manner. Next, in the max-pooling layer, the maximum value in the vector for each filter size is selected and sent to the softmax layer. Finally, in the softmax layer, using the fully connected neural network, we obtain the likelihood of the Haiku being classified as a certain quality category (high or low).

By comparing the output results of CNN with the actual category of Haiku, we can determine whether the estimation of a Haiku's quality is correct or not. If not, we back-propagate

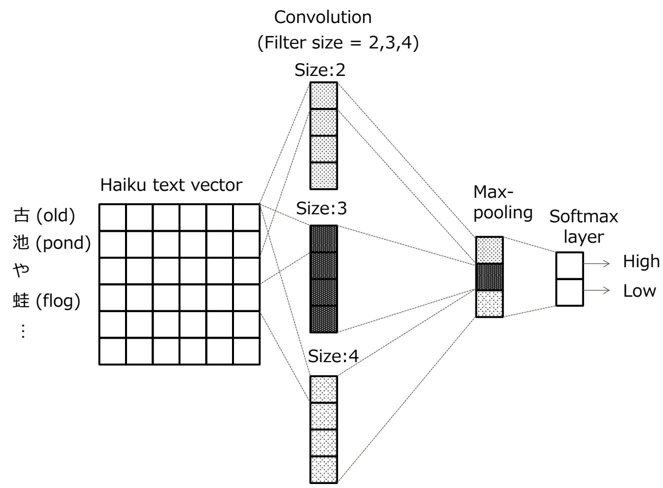

Figure 3: Convolutional neural network (CNN)

the result for CNN. For the learning algorithm, we used Adam optimizer. We can conduct the learning for syllable-based vectors in the same way just by replacing the word-based vector input with syllable-based vectors.

Step 5: Quality estimation for target text

Once we complete the machine learning for Haiku and their quality categories using word-based and syllable-based vectors, we can finally conduct the quality estimation for Haiku texts. Here, we input Haiku text into the CNNs for word-based and syllable-based vectors. Once we obtain the estimation results from the CNNs, we determine the estimated quality of the Haiku based on consensus. If the estimation results from both CNNs are "high", we conclude that the text's estimated quality label is high. Likewise, if both CNNs estimate the quality of a Haiku as "low", we judge that the quality category of the Haiku is low. If the two CNNs disagree about the estimated quality, we conclude that the CNNs do not have enough confidence to determine the quality of the Haiku. In this case, the estimated quality category of the Haiku is "uncertain".

IV. Evaluation of the Accuracy of Haiku Quality Estimation

Here we explain the implementation and the accuracy evaluation for our Haiku quality estimation method. We first explain the Haiku dataset we used for the evaluation. Next we show the implementation and the metrics for the accuracy evaluation. Then we show the evaluation results and give some considerations for them.

A. Haiku Dataset

For evaluation, we collected Haiku texts from the Haiku community site "Photo Haiku Circle by Seiichi Morimura" [3]. From the site, we obtained 40,000 Haiku poems submitted from 2011 to 2015. Each Haiku sample includes Japanese Haiku text, Japanese Kana format representing its pronunciation information, and the number of "Likes" (integer) rated by viewers. Assuming that the "Likes" ratings correspond to the Haiku's quality, we categorized a Haiku whose number of "likes" exceeds the half-year average as "high" quality, otherwise we categorized them as "low". As a result, we obtained 18,691 high-quality Haiku and 21,309 low-quality Haiku. Therefore, the baseline for the precision of estimation accuracy is 0.53 (=21309/40000), which can be achieved by just guessing that all Haiku have low quality.

B. Implementation

Here we explain our implementation regarding the translation of Haiku texts into vector representations, the construction of CNNs, and the configurations for them.

As explained in the previous section, we translate Haiku texts into word-based vectors and syllable-based vectors. However, in the translation into word-based vectors, the types of corpuses used to define the vectors for words can affect the estimation accuracy. Therefore, we prepared two different corpuses (Wikipedia corpus and Haiku corpus) for word-based vector translations. As a result, we obtained the following three types of vector representations for each Haiku text.

(a) **Word-based vector with Wikipedia**

We translated each word in the Haiku into a vector in 200 dimensions by referring to the dictionary obtained by applying word2vec to the Japanese Wikipedia corpus (January 18, 2015) [13]. If a Haiku contains a word which is not in the corpus, we translated it to a zero vector.

(b) Word-based vector with Haiku

We translated each word in the Haiku into a vector in 200 dimensions using a dictionary obtained by applying word2vec to a Haiku corpus derived from 300,000 Haiku texts on another Haiku community site, "Haishi-salon" [7]. Since the Haiku texts on this site do not have ratings information such as the number of "likes", we used them only for the corpus, and not for the dataset for evaluation.

(c) Syllable-based vector

As explained in the previous section, we translated each syllable in Haiku derived from Japanese Kana format into a 22-dimension vector.

As for CNN, we implemented the machine learning function by CNN shown in Figure 3 with one convolutional layer, one max-pooling layer and one softmax layer using Google's Tensorflow [12][21] by referring to the study of Kim [5] and Britz [6]. Figure 4 summarizes the hyperparameters for machine learning. We set the dimension of vectors for Haiku syllables as 22×35. The number of columns (22) comes from the definition of the syllable-based vector above. We set the number of rows (35) based on the number of syllables in a Haiku text. While traditional Haiku has only 17 syllables in it, as explained in Section II, sometimes some Japanese characters representing contracted sounds are not counted as one syllable in a Haiku. As a result, our syllable-based vector model sometimes needs more than one 22-demension vector for one syllable if a Haiku text has contracted sounds. Therefore, we prepared 35 rows including some margins for the vectors representing a Haiku with 17 syllables.

Dimensions	: 200x20 (word), 22x35 (syllable)
Filter size	: 3,4,5,6,7,8,9
Number of filters	: 32
Dropout rate	: 0.5
Batch size	: 100
Number of epochs	: 15000

Figure 4: Hyperparameters for CNN machine learning

C. Metrics for accuracy evaluation

Here we formalize the metrics for accuracy evaluation. First, we define a set of Haiku texts and their actual quality categories, vector models for estimation and the categories estimated by a vector model as follows.

- $T = \{t_1, t_2, ...\}$: A set of Haiku texts which are targets for quality estimation.
- $C = \{High, Low\}$: A set of categories representing the quality of Haiku.
- $f : T \rightarrow C$: The mapping function from Haiku text T to its actual quality category C. If a Haiku t has a high quality, $f(t) = High$. Otherwise, $f(t) = Low$.
- $M = \{W, H, S\}$: The set of vector modeling methods. Each alphabet (W, H, S) in the set represents a word-based vector model using the Wikipedia corpus, a word-based vector model with the Haiku corpus and the syllable-based vector model defined in Section IV.B, respectively.
- $\hat{f} : T \times M \rightarrow C$: The mapping function from the Haiku text and the vector modeling method to the category estimated by our approach. If a Haiku t's category is estimated as "High" using vector model m, then $\hat{f}(t,m) = High$. Otherwise, $\hat{f}(t,m) = Low$.

From the above definitions, if $f(t) = \hat{f}(t,m)$, we conclude that the quality of Haiku t is correctly estimated by the model m. Based on these definitions, we also define the consensus-based estimation in which we determine the estimated quality of Haiku only if all of the categories estimated by different vector models are the same. If some models disagree in their estimation, we estimate that the quality of the Haiku is difficult to estimate. In this case, we categorize the Haiku's estimated quality as "Uncertain". For $N \subseteq M$,

$$\hat{g}(t,N) = \begin{cases} c & (\forall n \in N, \hat{f}(t,n) = c) \\ \text{"}Uncertain\text{"} & (otherwise) \end{cases}. \qquad (1)$$

From this definition, if $f(t) = \hat{g}(t,N)$, we conclude that the quality of the Haiku t is correctly estimated by the consensus between the vector models in N.

Next, we define the metrics for estimation accuracy. Here we use three common accuracy indexes: weighted precision, weighted recall and weighted F-measure. Precision and recall are common metrics in evaluation for clustering accuracy, and the F-measure is their harmonic mean. These values can be calculated for each category. We can derive their weighted value (precision, recall and F-measure) by calculating the weighted average for Haiku samples with high and low quality. Before explaining the details of these indexes, we present the following notations for simplicity.

- $s(c) = \{t \in T \mid f(t) = c\}$: The set of Haiku whose actual categories (the ground truth) are c.
- $\hat{s}(c,N) = \{t \in T, N \subseteq M \mid \hat{g}(t,N) = c\}$: The set of Haiku whose categories are estimated as c by the models in N.

Using these notations, we represent precision and recall for each category c using a set of estimation models $N \subseteq M$ as follows. The precision represents how many correct results exist in the set of texts which are estimated as category c. The recall represents how many correct results cover the samples whose actual category is c.

$$Precision\,(c,N) = \frac{|\,s(c) \cap \hat{s}(c,N)\,|}{|\,\hat{s}(c)\,|} \tag{2}$$

$$Recall(c,N) = \frac{|\,s(c) \cap \hat{s}(c,N)\,|}{|\,s(c)\,|} \tag{3}$$

Using these, we calculated the weighted precision $WPre$ and weighted recall $WRec$. These indexes represent the estimation accuracy for all categories with weighting, in proportion to the number of samples for each category.

$$WPre(N) = \frac{1}{|\,T\,|} \cdot \sum_c (|\,s(c)\,| \cdot Precision\,(c,N)) \tag{4}$$

$$WRec(N) = \frac{1}{|\,T\,|} \cdot \sum_c (|\,s(c)\,| \cdot Recall(c,N)) \tag{5}$$

By using these parameters, we calculated the weighted F-measure WF , which is the harmonic mean of the weighted precision and weighted recall.

$$WF(N) = \frac{2 \cdot WPre(N) \cdot WRec(N)}{WPre(N) + WRec(N)} \tag{6}$$

In addition to the above metrics, in order to evaluate the effect of Haiku whose estimated quality is "uncertain", we define the weighted recall and F-measure, from which we eliminate the counts of Haiku with "uncertain" estimated quality as follows.

$$\hat{u}(c,N) = s(c) - (s(c) \cap \hat{s}(Uncertain, N)) \tag{7}$$

$$Recall'(c,N) = \frac{|\,s(c) \cap \hat{s}(c,N)\,|}{|\,\hat{u}(c,N)\,|} \tag{8}$$

$$WRec'(N) = \frac{\sum_c (|\,\hat{u}(c,N)\,| \cdot Recall'(c,N))}{\sum_c |\,\hat{u}(c,N)\,|} \tag{9}$$

$$WF'(N) = \frac{2 \cdot WPre(N) \cdot WRec'(N)}{WPre(N) + WRec'(N)} \tag{10}$$

D. Evaluation results

For evaluation, we conducted 10-fold cross-validation for our 40,000 Haiku samples, in which we used 36,000 Haiku as a learning set and 4,000 Haiku as a test set, repeating it for 10 different test sets. We conducted the experiment on CentOS7 (64 bit) installed on a desktop PC with an Intel Xeon CPU (3.60 GHz) and 8 GB memory. The experiment took about 12 hours.

Table I shows the estimation results $\hat{s}(c,N)$ for each quality category (High, Low and Uncertain) by the combination of three vector models (Wiki-corpus-based word model, Haiku-corpus-based word model and syllable-based model). From this table, we can see that for all vector models, the number of Haiku whose actual and estimated quality are both "High" is larger than the number of Haiku whose estimated quality is "High" but the actual quality is "Low". The same thing can be said for the "Low" category. Therefore, we can conclude that all of these models and their combinations have quality estimation capabilities that are to some extent better than random selections.

Figure 5 shows the plots for the weighted precisions and the weighed recalls calculated from the values in Table I. The black circles represent the normal weighted precisions and recalls calculated using Equations (4) and (5), while the white triangles are the plots with normal weighted precision using Equation (4) and the weighted recalls only for the Haiku whose estimated categories are not "uncertain", calculated by Equation (9). From this graph, we can find that the precisions of quality estimation by a single vector model do not have significant improvement from the baseline (0.53). However, we also find that we can improve the precisions by utilizing the consensus of different vector models. When we used all three vector models, we achieved a precision of 0.64, while the recall decreased to 0.28. The drawback comes from the Haiku samples whose qualities are estimated as "uncertain", since these Haiku are not correct estimations for "High" nor "Low" categories. If we eliminate the Haiku samples whose qualities

TABLE I. Estimation results for Haiku quality

Vector model	Actual Haiku quality	Estimated Haiku quality		
		High	Low	Uncertain
Wiki	High	9360	9331	-
	Low	7745	13564	-
Haiku	High	9489	9202	-
	Low	7767	13542	-
Syllable	High	8632	10059	-
	Low	7403	13906	-
Wiki + Haiku	High	6128	5970	6593
	Low	4211	10008	7090
Haiku + Syllable	High	5350	5920	7421
	Low	3649	9788	7872
Syllable + Wiki	High	5251	5950	7490
	Low	3559	9720	8030
Wiki + Haiku + Syllable	High	3812	4127	10752
	Low	2239	7574	11496

are estimated as "uncertain" from the calculation of recalls by using Equations (8) and (9), their weighted recalls improve as shown by the white triangles in Figure 5, up to 0.64 when we used all three vector models.

We can also see the same trends in F-measures. Figure 6 shows the values of F-measures for all Haiku and Haiku whose estimated categories are not "uncertain" derived from the values in Table I and Equations (6) and (10). In this graph, combining vector models decreases the F-measure due to the existence of "uncertain" quality Haiku, while the F-measure increases up to 0.64 if we limit the Haiku samples whose estimated quality is not "uncertain". From these results, we can conclude that by utilizing the consensus of different models, we can improve the estimation accuracy for Haiku for which we have a certain confidence in quality estimation (i.e. the estimation results are not "uncertain").

E. Consideration

The following are considerations from the evaluation results.

(1) Difficulties in Haiku evaluation

Certain sentiment analysis studies such as [10] present some outstanding outcomes for positive/negative opinion

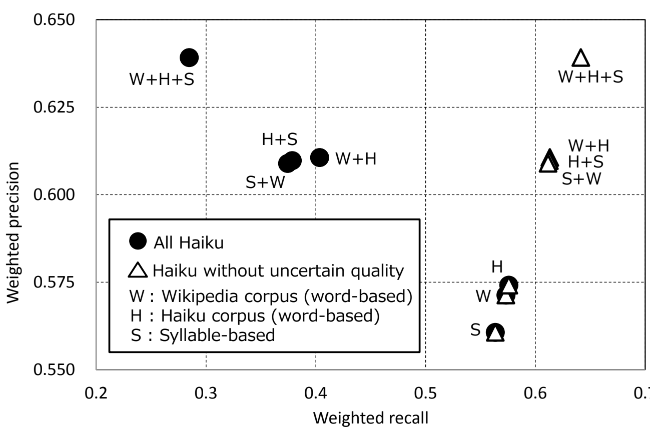

Figure 5. Weighted recall and precision in quality estimation accuracy

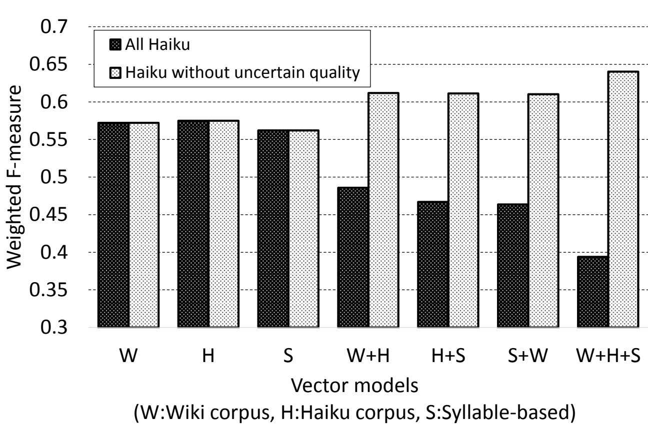

Figure 6: Weighted F-measure of quality estimation

estimation (e.g. over 0.76 by Kim [5]), while the accuracy in our Haiku quality estimation is relatively lower than this. We consider that the notable difficulty in Haiku evaluation compared with other sentiment analyses is the lack of a typical pattern. For example, there are typical words or sentences for positive and negative texts (e.g. positive: awesome, terrific and fantastic; and negative: terrible, disappointing and awful). The fact that Haiku and other artistry texts have no specific words for making them impressive provides us with great challenges when evaluation their quality.

Another possible obstacle in Haiku quality estimation is the length of Haiku. While the simple and short style of Haiku helps us make vector models concisely, if the sentences are short, it is difficult to find the factors distinguishing texts with high quality from texts with low quality. We suppose that in order to extract significant information from short texts, we need to enhance our models so that they contain richer information. One possible approach is including information regarding the structure of Haiku texts (e.g. subject, object and verb).

(2) Combination of different models

Even though there were several difficulties, as presented above, evaluation based on consensus between different models improved the precision of the quality estimation. This result suggests that by conducting the estimation from various viewpoints (e.g. word and pronunciation), it is possible to capture the underlying characteristics for impressive artistic texts. While this seems to be a promising outcome, we have to take into account its drawbacks. For example, while we achieved an F-measure for 0.64 for Haiku not having a quality of "uncertain", in this case we had 22,248 "uncertain" Haiku samples, which constituted more than half of all samples. If we can measure confidence levels for Haiku quality estimation by methods other than our current consensus-based approach, we might be able to contribute to improving the quality estimation accuracy by reducing the number of Haiku of "uncertain" quality.

V. RELATED WORK

Various types of research have been conducted in the area of automated generation and evaluation of poetry. For Haiku poems, the Haiku generation assisting tool "Hitch Haiku" has been developed by Tosa et al. [8] and Wu et al. [9]. While their approach has a scoring function for automatically generated Haiku, it is based on simple deterministic rules defined manually such as the usage of idioms and onomatopoeias. Greene et al. [16] proposed a method to translate poems into different languages taking into account their rhythmic characteristics (word-stress patterns), although it cannot evaluate good rhythmic patterns in poems as we did in our approach by evaluating poems using the syllable-based model.

Some research has been conducted to identify the artistic beauty in poems. Kao et al. [19] compared poems written by professional poets and ones by amateurs. By identifying the differences in probabilities of word counts between them, they revealed the characteristics of poems by professionals. One example is that professional poems tend to have more references to concrete objects (e.g. trees, rooms, flowers) than amateur poems. Herbelot [18] also determined word frequencies in poems, articles in Wikipedia and random texts so that they could differentiate these texts. While these approaches can be quite useful in identifying common characteristics good poems might have, it does not mean that we can determine whether or not a certain poem is of a good quality. For example, just because we conjure up a poem that includes a lot of references to concrete objects does not mean that the poem is good quality. He et al. [17] proposed a statistical approach to generate and evaluate poems in Chinese classical style using the BLEU metric [20] used for automatic evaluation of machine translation systems. They generated poems based on BLEU metrics and evaluated the quality of the poems by human judgement. While they confirmed that the quality of poems calculated by their approach conforms with the results of quality judgement by humans, the number of samples used for the evaluation was quite small (40 poems).

Methods to determine the emotions or feelings of people towards entertainment content have also been researched in the area of sentiment analysis and opinion mining. For example, Pang et al. [11] conducted a prediction of movie ratings by analyzing review comment texts. Similar to our approach, their approach used machine learning to determine the quality category (good/bad) entertainment content. But they used the text from reviews (opinions of critics) to evaluate the quality of an item, while we used the content itself (Haiku texts) to determine the quality. Mihalcea et al. [15] applied classification algorithms (e.g. Naïve Bayes and Support Vector Machine) to humor recognition tasks in which they differentiated one-liner joke texts from other types of texts from Reuters, proverbs and British National Corpus. While their formulation of the classification problem is similar to ours, they do not take into account the quality of the content itself. In other words, they could not tell which joke was funnier than others.

Compared with the above approaches, the notable characteristics in our research were that we used the machine learning approach to determine the quality (high or low) of Haiku and conducted an evaluation of quality estimation accuracy with a large number (40,000) of Haiku texts.

VI. CONCLUSION

We developed a quality estimation method for Haiku using a machine learning approach. In this approach, we first translated each Haiku into word-based and syllable-based vectors. Next, we conducted machine learning for these vectors using a CNN model. For the quality estimation models for different types of vectors, we conducted an evaluation of quality estimation accuracy for 40,000 Haiku texts obtained from a Haiku community site. As a result, we confirmed that by utilizing the consensus of estimation results obtained from different vector models, we achieved an F-measure of 0.64 in the estimation when discounting Haiku samples which did not have consensus in the qualities estimated by the applied vector models.

As future work, first we are going to conduct further experiments with a larger number of Haiku with ratings by Haiku masters. While we assumed that the number of "likes" given to Haiku on a website from viewers reflects their quality, it is still possible that the Haiku masters' opinions differ from that of the majority of amateurs. The masters' opinions will help us implement key factors to identify good Haiku in our quality estimation method. Next, we are going to apply other algorithms (e.g. recursive neural network) and text modeling methods (e.g. modeling text structures such as subject, object and verb) to improve the quality estimation accuracy. Since our research is still in its primitive stage, we only used one algorithm (CNN) and two vector models (word-based and syllable-based). If we can compare the results obtained by different algorithms with different vector models, this would be quite helpful in finding the optimal quality estimation method. Finally, by utilizing these improvements, we are going to develop an automated Haiku generation function. If we can construct a model representing the implicit knowledge and expertise of Haiku masters, it will be an indispensable asset for creating Haiku with high enough quality to be endorsed by notable Haiku masters. It will be a great challenge to implement artistic creativity on computers.

Since appreciating the beauty in poetic texts is one of the most sensitive cognitive skills humans have, we believe that our research into Haiku quality estimation will be able to contribute to broadening the research areas of computational intelligence for human-like intelligence.

REFERENCES

[1] Yuchang Cheng and Tomoki Nagase, "An Example-Based Japanese Proofreading System for Offshore Development," in Proceedings of COLING, 2012.

[2] Haiku Weekly, "Interview with Takako Usami, a Haiku section editor at Asahi Newspaper," http://weekly-haiku.blogspot.jp/2011/08/blog-post_8618.html (in Japanese), 2011.

[3] Seiichi Morimura, "Photo Haiku Circle," http://azby.fmworld.net/photohaiku/gallery2/list/index.html (in Japanese), 2015.

[4] Taku, Kudo, "MeCab: Yet another part-of-speech and morphological analyzer," http://taku910.github.io/mecab/, 2002.

[5] Yoon Kim, "Convolutional Neural Networks for Sentence Classification," in Proceedings of EMNLP, 2014.

[6] Denny Britz, "Implementing a CNN for Text Classification in TensorFlow," http://www.wildml.com/2015/12/implementing-a-cnn-for-text-classification-in-tensorflow/, 2015.

[7] Haishi-salon, http://www.haisi.com/saijiki/ (in Japanese), 2016.

[8] Naoko Tosa, Hideo Obara and Michihiko Minoh, "Hitch Haiku: An Interactive Supporting System for Composing Haiku Poem," Entertainment Computing - ICEC 2008, Proceedings LNCS Springer, pp.209-216, 2008.

[9] Xiaofeng Wu and Naoko Tosa, "New Hitch Haiku: an Interactive Renku Poem Composition Supporting Tool Applied for Sightseeing Navigation System," IFIP Entertainment Computing 2009 Proceedings, Springer, pp.191-196, 2009.

[10] Kumar Ravi and Vadlamani Ravi, "A survey on opinion mining and sentiment analysis," Know.-Based Syst. 89, C (November 2015), 14-46. DOI=http://dx.doi.org/10.1016/j.knosys.2015.06.015, 2015.

[11] Bo Pang and Lillian Lee, "Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales," in Proceedings of the 43rd Annual Meeting on

Association for Computational Linguistics (ACL '05). Association for Computational Linguistics, Stroudsburg, PA, USA, 115-124, 2015.

[12] Google, Tensorflow, https://www.tensorflow.org/, 2015.

[13] Wikipedia (Japanese), https://ja.wikipedia.org/

[14] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean, "Distributed Representations of Words and Phrases and their Compositionality," in Proceedings of NIPS, 2013.

[15] Rada Mihalcea and Carlo Strapparava, "Making Computers Laugh: Investigations in Automatic Humor Recognition," in Proceedings of the Joint Conference on Human Language Technology / Empirical Methods in Natural Language Processing (HLT/EMNLP), 2015.

[16] Erica Greene, Tugba Bodrumlu and Kevin Knight, "Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation," in Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp.524-533, 2010.

[17] Jing He, Ming Zhou and Long Jiang, "Generating Chinese Classical Poems with Statistical Machine Translation Models," in Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, pp.1650-1656, 2012.

[18] Aurelie Herbelot, "The Semantics of Poetry: A Distributional Reading," Literary and Linguistic Computing, 2014.

[19] Justine Kao and Dan Jurafsky, "A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry," NAACL Workshop on Computational Linguistics for Literature, 2012.

[20] Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu, "BLEU: a Method for automatic evaluation of machine translation," in Proc. of the 40th Meeting of the Association for Computational Linguistics, 2002.

[21] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu and Xiaoqiang Zheng, "TensorFlow: A system for large-scale machine learning," Google Brain, 2016.

[22] Haiku (in Wikipedia), https://en.wikipedia.org/wiki/Haiku

[23] Saijiki (in Wikipedia), https://en.wikipedia.org/wiki/Saijiki