5P-09

A Consideration on Complementing Property Wikidata with ALBERT

Yuxi JIN[*1] Shun SHIRAMATSU[*1]

[*]Graduate School of Engineering, Nagoya Institute of Technology

1 Introduction

Wikidata, growing linked data which is full potential to provide reliable and powerful service, bases on the knowledge of Wikipedia. Although multiple projects required information on Wikidata, the property are not as sufficient as Wikipedia, even some property values on Wikidata are not right. Therefore, how to extract knowledge from Wikipedia automatically and add it to Wikidata is a worthy of attention.

In this paper, we study an effective method for extracting Wikidata property values. Specifically, we take the Wikipedia text as input 1, take the property question or property definition as input 2, and use the property value as the output answer. Using ALBERT for machine understanding, if there are answers in article, predict the starting and ending positions of the output, else predict [CLS] as output, and compare the two groups of predicted models.

2 Experimental Evaluation

256 articles about sports were extracted from Wikipedia. According to the content of the articles, the ALBERT question and answer model was used to automatically extract 7 sets of property values. The seven groups of properties are "country of orgin", "founded by", "inception", "number of participants", "sport", "start time", and "uses". Thence, there are 1792 pieces of data, of which 1433 pieces of data are used as the training data and 359 pieces of data are used as the test data.

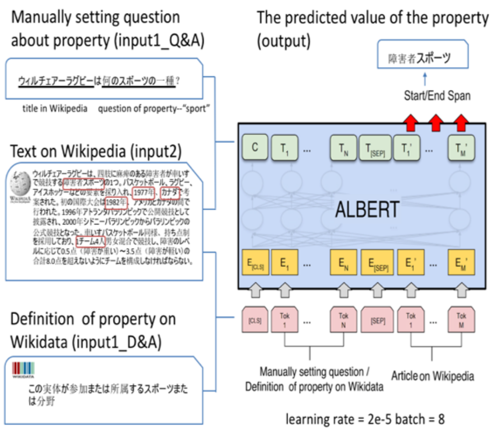

Figure 1: System of Q&A, D&A

As shown in the figure1, a total of two experiments were carried out. The first experiment was to input artificially set questions as input1 and articles on Wikipedia as input2 to obtain the value of property (Q&A) as answer. The second experiment was to use the definition of property on Wikidata as input1. Questions in this experiment, input the definition and article, and get the value of the property(D&A). The specific content of input1 is shown in Table 1.

Table1: input1 of two experiments

Input1 (Q&A)	property	Input1(D&A)
__(title of entity)はどこから発祥された？	country of orgin	製品や創作品、主題の対象物の原産国、発祥国
__は誰に発案された？	founded by	組織・宗教・場所の創設者や共同創設者
__はいつから承認された？	inception	組織・事物が設立・作成された年月日
__は何人のチーム？	number of participants	イベントへの参加者数。例：イベントに参加する人またはグループの数
__は何のスポーツの一種？	sport	この実体が参加または所属するスポーツまたは分野
__はいつから始まった？	start time	存在を開始したり、有効であると宣言された日付。おもに修飾子として使用される。
__は何を使う？	uses	主題が使う/で使われるもの

Use the Transformers package to implement the AlbertForQuestionAnswering model, and when the learning rate is 2e-5 and the batch value is 8, the loss value is the smallest. Through two indicators EM, F1 to judge the Q&A and D&A as figure2. EM is a binary measure (true/false) of whether the system output exactly matches the ground truth answer exactly. This is a fairly strict metric. F1 is a less strict metric, and it is the harmonic mean of precision and recall.

Table 2: EM, F1 of Q&A and D&A

	Loss value	EM all	F1 all	EM no_ans	Precision no_ans	EM ans	Precision ans
Q&A	3.996	35.93%	44.57%	66.88%	71.33%	12.68%	25.39%
Chance level of Q&A	11.823	0%	13.93%	0%	0%	0%	24.39%
D&A	3.510	44.85%	54.60%	68.18%	75.33%	27.32%	39.71%
Chance level of D&A	11.449	0%	20.05%	0%	0%	0%	35.12%

Evaluation value is not ideal, but compared with the chance level obtained by completely random prediction, the two sets of systems learned by ALBERT have increased significantly. "no_ans" indicates that there is no predicted answer for the property in the text, and "ans" indicates that there is an answer for the property.

Table3: F1 of Q&A and D&A in each property

	all	country of orgin	founded by	inception	number of participants	sport	start time	uses
Number of test data	359	53	42	48	53	56	45	61
F1 of Q&A	44.57%	18.87%	42.86%	42.08%	50.94%	71.93%	17.78%	50.81%
F1 of D&A	54.60%	41.51%	69.45%	45.83%	56.60%	82.46%	17.78%	52.46%

After that, a specific analysis was carried out on the prediction of each property. The table3 above shows the F1 obtained from the inspection data of each property.

It is particularly worth noting that in Table 2, whether there is a predicted answer in Wikipedia article has a huge impact on the predicted result. The predicted precision of “ans” is much smaller than the precision of “no_ans”, so the two sets of data are specifically analyzed. When a question has no answer, both the F1 and EM score are 1 if the model predicts no-answer, and 0 otherwise. Thence, use precision to analyze the data with or without answers for each property.

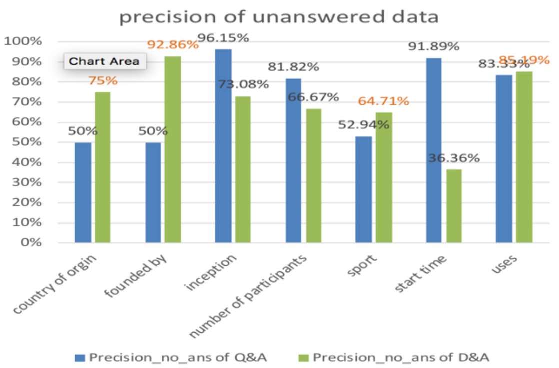

Figure 2: precision with unanswered

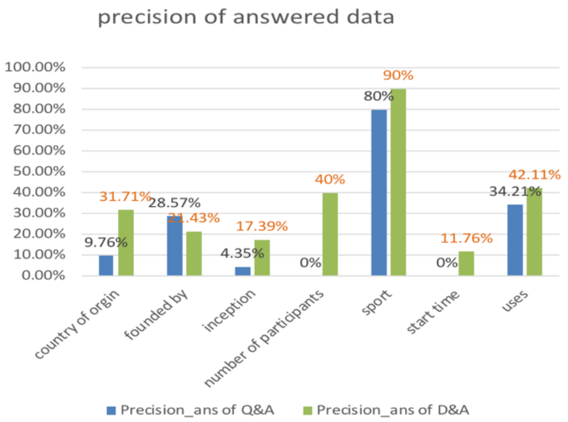

Figure 3: F1 with answered test data

Figure 2 shows the prediction accuracy of each property without answers, Figure 3 shows precision of each property with answers.

3 Discussions

Through two groups of comparisons, some observations are shown as follows.

(1) Compared with Q&A, using the definition on Wikidata as the input value of D&A indicators is better. And compared to artificially set problems, D&A can automatically extract property definitions from Wikidata, which is more applicable. D&A is not only used to infer properties related to sports, it can also be used directly to predict the value of other properties automatically.

(2) For the same property, when there are multiple answers (for example, for the ”use” of a certain sports entity, the answer will be ”マレッ ト (mallet)”, ”ball”), we set multiple same input with different answers. Therefore, during verification, although the system predicts one of multiple answers, it will still be judged as an incorrect answer. Therefore, we will conduct a more in-depth study on the automatic prediction method of multiple answers in future.

(3) Precision with no answer in the Wikipedia article is much higher than precision with answer. Moreover, in Q&A, the error analysis of the data with answers found that more than 51% of the data that should have an answer were incorrectly inferred that there was no answer (In D&A, there are more than 29% of the data that have an answer but are incorrectly inferred to have no answer). This shows that the amount of data allocated for whether there is an answer in the training data is not balanced enough.

4 Conclusion

In this paper, we put forward a consideration to use the definition of property (D&A) on Wikidata to infer the value of property from the article. In the next research work, we will adjust the proportion of whether there are answers in the training data, expand the training data, and increase the types of properties. In addition, we plan to use the semi-automatic additional data on Wikidata to extract relevant knowledge of social problem events.

Acknowledgements

This work was supported by NEDO (JPNP20006), JST CREST (JPMJCR15E1, JPMJCR20D1), and JSPS KAKENHI (17K00461).

References

[1] Lan, Zhenzhong, et al. ”Albert: A lite bert for self-supervised learning of language representations.” arXiv preprint arXiv:1909.11942 (2019).

[2] van Aken, Betty, et al. ”How does bert answer questions? a layer-wise analysis of transformer representations.” Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019.

[3] Wang, Zhiguo, et al. ”Multi-passage bert: A globally normalized bert model for open-domain question answering.” arXiv preprint arXiv:1908.08167 (2019).

[4] 李宏宇, et al. ”質問なし読解事例に対する BERT の回 答可能性に着目した機械読解難易度分析.” 人工知能学 会全国大会論文集 第 34 回全国大会 (2020). 一般社団 法人 人工知能学会, 2020.