The Global Banking Standards QA Dataset (GBS-QA)

Kyunghwan Sohn
KAIST
Daejeon, South Korea
sony0125@kaist.ac.kr

Sunjae Kwon[*]
University of
Massachusetts
Amherst, MA, USA
sunjaekwon@cs.umass.edu

Jaesik Choi
KAIST
Daejeon, South Korea
jaesik.Choi@kaist.ac.kr

Abstract

A domain specific question answering (QA) dataset dramatically improves the machine comprehension performance. This paper presents a new Global Banking Standards QA dataset (GBS-QA) in the banking regulation domain. The GBS-QA has three values. First, it contains actual questions from market players and answers from global rule setter, the Basel Committee on Banking Supervision (BCBS) in the middle of creating and revising banking regulations. Second, financial regulation experts analyze and verify pairs of questions and answers in the annotation process. Lastly, the GBS-QA is a totally different dataset with existing datasets in finance and is applicable to stimulate transfer learning research in the banking regulation domain.

1 Introduction

Global banking standards have been strictly organized and gradually evolved over time to reflect the changes of financial environment and new risks emerged. The Basel Committee on Banking Supervision (BCBS) which is one of international rule setters governs to create and revise the international banking standards in cooperation with member countries and international financial supervisory agencies.

Creating new rules and revising current standards require clear communications with related parties such as financial supervisory authorities and market players over the globe. In the process of this work, Frequently Asked Questions (FAQs) are formally constructed and disclosed in order to help understand better how to implement provisions in practice. The FAQs contains questions from market players and answers from the BCBS. However, original pairs of questions and answers in the FAQs are not standardized so that they need to be re-organized and revised in order to become

qualified practices on the corresponding provisions. Considering that financial regulations are directly influential to the market and they require expertise to interpret and understand the context in the FAQs, financial regulation experts should be involved.

We propose a new Global Banking Standards QA dataset (GBS-QA). The GBS-QA has three values. First, given that GBS-QA is composed of actual questions and answers in practice, both market players and regulators find it quite useful at work as additional guidelines. Second, financial regulation experts participate in this study to construct revised QA pairs which can be properly applied into the NLP model. Third, the GBS-QA is a totally different dataset with existing datasets in finance and is applicable to stimulate transfer learning research in the banking regulation domain.

This paper is constructed as follows. Related work is reviewed in Section 2. The process of collecting and annotating the GBS-QA is described in Section 3. The experimental results of the pre-trained language model and its domain-adaptive model on GBS-QA are revealed in Section 4.

2 Related work

There have been many researches on constructing new datasets in various domains and utilizing them on their own purpose. In finance, financial news sentiment dataset (Ding et al., 2015) and financial opinion mining and questions answering dataset (Chen et al., 2020) are widely utilized. However, since these are basically rooted from financial news and general narratives in financial reports, they are not equipped with terminology in financial regulation and of course are not qualified for doing tasks in financial regulation domain.

Aside from construction of new datasets, how to make AI-enabled NLP models be more understandable and explainable is one of top priorities. Recently centered on medical domain, text entailment(Abacha and Demner-Fushman, 2016; Fei

---

[*] He engaged in this work when he was in KAIST.

Figure 1: An Example of provisions, questions and answers in GBS-QA

et al., 2021) and deep neural reasoning(Han et al., 2021) are applied to do question answering tasks and combination of these approaches and input medical text contributes to achieve the goal of enhancing interpretability and explainability. From a perspective of few-shot setting, few-shot textual entailment framework combined with graphical structure in data also shows its effectiveness in the downstream tasks(Yin et al., 2020).

In addition, domain-adaptive pretraining is proposed as one of the guaranteed way to lead to performance gains (Ding et al., 2015) in domain-specific tasks. FinBERT (Andrew and Gao, 2007) and BioBERT (Lee et al., 2020), for example, are types of domain-adaptive models.

3   GBS-QA

GBS-QA is collected from the BCBS website[1]. The BCBS framework comprises the 14 standards and each standard has its provisions and the associated FAQs which are matched with the corresponding provisions. As shown in Figure 1, Risk-based capital requirements (RBC) which is one of the standards contains 4 sub-categories and each sub-category has a varying number of provisions with different length. The provision of 30.12, for example, raises two questions from market players so that the BCBS answers these questions with detailed explanations.

Overview of a GBS-QA construction process follows Figure 2. Starting from BCBS framework, all standards and the associated FAQs are automatically filtered out of the BCBS website. These data are organized and preprocessed by human annotation process. Human annotation includes organizing provisions, matching the provisions with the

corresponding FAQs and reviewing questions and answers in a reconciled manner. This review is conducted by independent annotators which consists of five financial regulation experts. From the human annotation process, pairs of questions and answers are classified into four types which include 1) Binary answerable type, 2) WH type, 3) How type and 4) Conditional type. Upon question classification, questions are revised into "Binary answerable type" and answers are labelled into "Yes" or "No" to corresponding questions according to the GBS-QA classification guideline in Appendix A. After completing the whole process in Figure 2, GBS-QA is constructed as a set of questions and answers in banking regulation domain.
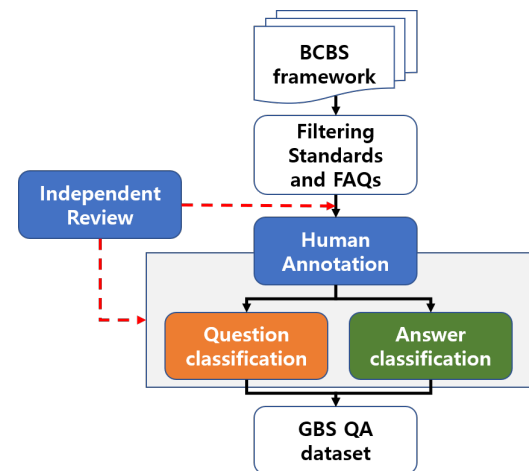

Figure 2: Overview of a GBS-QA construction process.

3.1   Automatic filtering of standards, questions and answers

The BCBS website provides a well-structured platform. From web crawling, data is collected by filtering "basel_navigations_standard_selection", "basel_paragraphs", "faqs_and_footnotes" and "basel_faq".

[1]https://www.bis.org/basel_framework/

20

3.2 Human Annotation

Human annotation mainly covers question classification and answer classification. This annotation process is conducted by five financial experts independently. Since it requires expertise in interpreting international banking regulations, annotators should be financial regulation experts who are qualified for this task. In the beginning, guideline is set to make it transparent and consistent for them to annotate questions and answers. The guideline is described in Appendix A. In compliance with the guideline, reviewing all pairs of questions and answers is implemented and then classifying questions and answers is followed. After constructing a dataset of questions and answers individually, five outcomes are compared and pairs are chosen throughout majority voting. Kappa statistics (Gururangan et al., 2020) is used as quantitative metric.

3.2.1 Question classification

Questions raised in the process of creating or revising standards are intended to make the corresponding provisions clearer with additional explanations in order to help market players to implement them compliantly. Given that answers are directly influential to market, they are supposed to be transparent and concise based on attached references.

Considering the characteristics of pairs of questions and answers in the BCBS framework, we reach to a conclusion that if all the questions are revised to be answered with "Yes" or "No", the revised questions can be added to the corresponding provisions and ultimately become additional provisions. The question of "What is the difference between (the jurisdiction of) "ultimate risk" and (the jurisdiction of) "immediate counterparty" exposures?", for example, can be transformed into "Is it correct that the difference between "ultimate risk" and "immediate counterparty" exposures means that ... ?" and the meaning phrase or sentence becomes additional information. The point is to determine if certain behavior is compliant with regulation or not and answers in the FAQs can be regarded as qualified provisions interpreted from the BCBS.

To transform the questions to be answerable with "Yes" or "No, question type should be analyzed and classified in the beginning. In this paper, we propose four question types which include 1) Binary answerable type, 2) WH type, 3) How type and 4) Conditional type. The definitions of four types are as follows. Examples are described in Appendix B.

1. **Type1 : Binary answerable type**
   A question which is answerable with "Yes" or "No".

2. **Type2 : WH type**
   A question which starts with "What", "Where", "Why" and "When" or has same nuance to all "WH".

3. **Type3 : How type**
   A question which starts with "How".

4. **Type4 : Conditional type**
   A question which is not answerable with "Yes" or "No" in general. However, if certain condition is met, this question becomes Type1.
   [Example]
   [Q] Can subordinated loans be included in regulatory capital?
   [A] As long as the subordinated loans meet all the criteria required for Additional Tier 1 or Tier capital, it is correct.

According to the annotation guideline in Appendix A, questions in Type2, Type3 and Type4 are transformed into new questions which can be answerable with "Yes" or "No" as same as Type1. Type2 and Type3 have specific phrase or sentence as the corresponding answer. In this case, the questions are simply revised into the sentence structure of "Is it correct that ..." or "Does it mean that ...". Type4 can be answerable as same as Type2 if certain conditions are met. In this regard, the condition captured from the context of the answer is put in front of or at the end of the original question with "if", "in cases where" and "as long as".

3.2.2 Answer classification

In line with question classification, answer classification is implemented by five annotators independently.

3.2.3 Classification result

Total number of pairs of questions and answers in the GBS-QA is 186. Question type distribution is shown in Table 1. Type1 accounts for 39% at the highest ratio and Type4 secondly ranks with 30% in the GBS-QA. The coverage ratio of "Yes" is 70%, whearas that of "No" is 30%.

Total	Type1	Type2	Type3	Type4
186	72(39%)	36(19%)	23(12%)	55(30%)

Table 1: Question type distribution

3.3 Comparison with other datasets in finance

The GBS-QA is quite different from other datasets in finance as shown in Figure 3. Comparing top 4 words among three datasets, GBS-QA contains the words of "capital", "risk", "bank", "Tier" which are closely related to banking regulations whereas FiQA includes "stock", "money", "company" and "credit" and PhraseBank covers "company", "profit", "net" and "sales". FiQA focuses on stock market and PhraseBank deals with corporate financial performance.
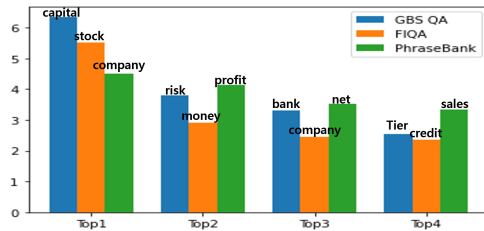

Figure 3: top 4 words distribution of GBS-QA, FiQA and PhraseBank.

In addition, in top 100 words, jaccard similarity of GBS-QA with the other two datasets remarks 0.10 and 0.05 respectively. Especially, the top 10 words in GBS-QA do not have any significant intersection with others. It implies that words distribution of GBS-QA is far away from the existing datasets in finance.

3.4 Important issues on GBS-QA

One of core parts in constructing the GBS-QA is how to revise the original questions to be answerable with "Yes" or "No" without significant loss of information in the original pairs of questions and answers. It is quite difficult to measure how much information is lost in this transforming process. However, if it is guaranteed that revised questions contain correct phrases or sentences in the original answers, it could tell that there is little loss of information in a qualitative manner.

Among four types of questions proposed in the section 3.2.1, Type3 (How type) and Type4 (conditional type) are much harder to transform the original questions because they require complete understanding on the context of answers in order to extract correct phrases or sentences concisely. In this regard, before disclosing the GBS-QA, it needs more effort to thoroughly review all QA pairs. Eventually the GBS-QA should be confirmed by the BCBS.

4 Model Architecture

The experiments aim to show that GBS-QA can be applicable to stimulate transfer learning on the classification task. Considering uniqueness and expertise of GBS-QA, domain-adaptive language models which are trained with all provisions in current standards and all answers are used to do this task.

As shown in Figure 4, it starts from pre-trained models such as RoBERTa (Liu et al., 2019) and ELECTRA(Clark et al., 2020). By training domain knowledge of banking standards, pre-trained models are evolved into post-trained model which is qualified for dealing with GBS-QA. Next, these post-trained models are fine-tuned for classification task with pairs of questions and answers("Yes" or "No") and the associated provisions. The final model is called BankReg QA model.
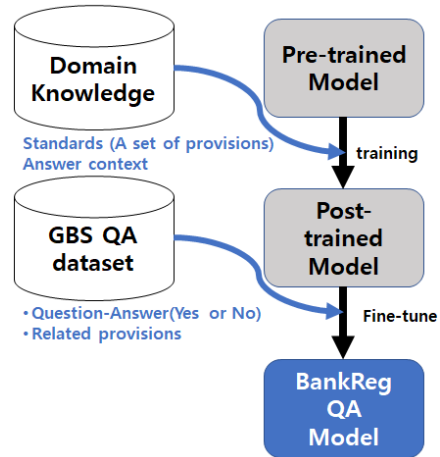

Figure 4: Overview of a BankReg QA model construction process.

5 Experiments

Due to lack of data points in the GBS-QA, there is no significant difference in experimental results between pre-trained and post-trained model in the GBS-QA. Moreover, some questions have longer length than 512 which is maximum length in the model and the other questions require interpretation of equations to calculate risk factors or risk exposure so that the models can not handle them properly. However, the post-trained model shows slightly lower performance in both FiQA and PhraseBank dataset.

6 Conclusion

We propose a new QA dataset which is called GBS-QA in the banking regulation domain. The

GBS-QA contains pairs of questions and answers re-organized and verified by financial regulation experts. Extra information extracted from the GBS-QA can become qualified practices on the associated provisions in the regulation. From the analysis and experiments, the GBS-QA shows that it has a different words distribution compared to other existing datasets in finance and it can be applicable in the NLP models.

7 Future Work

First of all, GBS-QA consists of high quality but relatively small number of QA pairs. Therefore, it is necessary to expand the size of the dataset. However, it requires tremendous cost to label large-scale data by qualified financial experts. Semi-supervised approaches can be one possible way to constuct large-scale data with minimal cost. We can consider to use a bootstrapping that iteratively expands data based on good quality seed data. In addition, we can utilize a relationship among provisions. This relationship can be drawn as a graph and the graphical structure can be used in training and fine-tuning. Lastly, studies on text entailment and non monotonic reasoning are applicable to analyzing and understanding financial regulation in the NLP tasks.

Acknowledgements

References

Asma Ben Abacha and Dina Demner-Fushman. 2016. Recognizing question entailment for medical question answering. In *AMIA Annual Symposium Proceedings*, volume 2016, page 310. American Medical Informatics Association.

Galen Andrew and Jianfeng Gao. 2007. Scalable training of L1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. Fine-grained financial opinion mining: A survey and research agenda. *arXiv preprint arXiv:2005.01897*.

Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*.

Hao Fei, Yuanpei Guo, Bobo Li, Donghong Ji, and Yafeng Ren. 2021. Adversarial shared-private model for cross-domain clinical text entailment recognition. *Knowledge-Based Systems*, 221:106962.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*.

Zhongyi Han, Benzheng Wei, Xiaoming Xi, Bo Chen, Yilong Yin, and Shuo Li. 2021. Unifying neural learning and symbolic reasoning for spinal medical report generation. *Medical Image Analysis*, 67:101872.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Wenpeng Yin, Nazneen Fatema Rajani, Dragomir Radev, Richard Socher, and Caiming Xiong. 2020. Universal natural language processing with limited annotations: Try few-shot textual entailment as a start. *arXiv preprint arXiv:2010.02584*.

A Annotation guideline

This annotation guideline contains a set of process to do question classification and answer classification. Question classification means that every question is categorized into one of pre-determined types considering context of questions and answers in a reconciled manner. The four types are as follow.

1. **Type1 : Binary answerable type**
   A question which is answerable with "Yes" or "No" without critical difficulty.

2. **Type2 : WH type**
   A question which starts with "What", "Where", "Why" and "When" or has similar nuance to all "WH". Specific phrase or

sentence is an answer to the corresponding question.

3. **Type3 : How type**
A question which starts with "How". Interpretation or explanation is an answer to the corresponding question.

4. **Type4 : Conditional type**
A question which is not answerable with "Yes" or "No" in general. However, if certain condition is met, this question becomes Type1.

Question classification is solely dependent on experts' viewpoint and interpretation from scratch. In order to verify the classification results from five independent annotators, it follows majority voting and reliability test with Kappa statistics.

One of the biggest challenges is how to revise questions into Type1. To save time and effort, one leading annotator does transform the questions in Type2, Type3 and Type4. Type2 and Type3 follows that firstly the phrase or sentence or a set of sentences in the context of the corresponding answer is identified and new question is constructed with the sentence structure of "Is it correct that ..." or "Does it mean that ...". These structures are considered quite simple. However, we all agree that it is one of the most effective and intuitively understandable approach. Type4 is revised by adding the condition captured in the context of the corresponding answer in front of or at the end of the questions with "If + condition". Upon question classification, answers are labelled into "Yes" or "No" at the end.

B Examples of question types

B.1 Type1: Binary answerable type

1. **original question**
Regarding CAP10.11(16), consider a bank that issues capital out of a foreign subsidiary, and wishes to use such capital to meet both the solo requirements of the foreign subsidiary and include the capital in the consolidated capital of the group. Is it correct that the relevant authority in jurisdiction of the consolidated supervisor must have the power to trigger write-down / conversion of the instrument in addition to the relevant authority in the jurisdiction of the foreign subsidiary?

2. **original answer**
Yes, this is correct.

B.2 Type2: WH type

original question
What is the difference between (the jurisdiction of) "ultimate risk" and (the jurisdiction of) "immediate counterparty" exposures?

original answer
The concepts of "ultimate risk" and "immediate risk" are those used by the BIS' International Banking Statistics . The jurisdiction of "immediate counterparty" refers to the jurisdiction of residence of immediate counterparties, while the jurisdiction of "ultimate risk" is where the final risk lies. For the purpose of the countercyclical capital buffer, banks should use, where possible, exposures on an "ultimate risk" basis.

revised question
Is it correct that the jurisdiction of "immediate counterparty" refers to the jurisdiction of residence of immediate counterparties, while the jurisdiction of "ultimate risk" is where the final risk lies?

revised answer
Yes

B.3 Type3: How type

original question
How is the final bank-specific buffer add-on calculated?

original answer
The final bank-specific buffer add-on amount is calculated as the weighted average of the countercyclical capital buffer add-on rates applicable in the jurisdiction(s) in which a bank has private sector credit exposures (including the bank's home jurisdiction) multiplied by total risk-weighted assets. The weight for the buffer add-on rate applicable in a given jurisdiction is the credit risk charge that relates to private sector credit exposures allocated to that jurisdiction, divided by the bank's total credit risk charge that relates to private sector credit exposures across all jurisdictions. Where the private sector credit exposures (as defined in RBC30.13(FAQ1)) to a jurisdiction, including the home jurisdiction, are zero, the weight to be allocated to the particular jurisdiction would be zero.

revised question
Is it correct that the final bank-specific buffer add-on amount is calculated as the weighted average of the countercyclical capital buffer add-on rates applicable in the jurisdiction(s) in which a bank has private sector credit exposures (including the

bank's home jurisdiction) multiplied by total risk-weighted assets?

revised answer

Yes

B.4 Type4: Conditional type

original question

Can subordinated loans be included in regulatory capital?

original answer

Yes. As long as the subordinated loans meet all the criteria required for Additional Tier 1 or Tier 2 capital, banks can include these items in their regulatory capital.

revised question

As long as the subordinated loans meet all the criteria required for Additional Tier 1 or Tier 2 capital, can subordinated loans be included in regulatory capital?

revised answer

Yes