Article

Comparative Study of Multiclass Text Classification in Research Proposals Using Pretrained Language Models

Eunchan Lee 1, Changhyeon Lee 2 and Sangtae Ahn 1,2,*

1 School of Electronic and Electrical Engineering, Kyungpook National University, Daegu 41566, Korea; purang2@knu.ac.kr
2 School of Electronics Engineering, Kyungpook National University, Daegu 41566, Korea; lch6694@knu.ac.kr
* Correspondence: stahn@knu.ac.kr

Abstract: Recently, transformer-based pretrained language models have demonstrated stellar performance in natural language understanding (NLU) tasks. For example, bidirectional encoder representations from transformers (BERT) have achieved outstanding performance through masked self-supervised pretraining and transformer-based modeling. However, the original BERT may only be effective for English-based NLU tasks, whereas its effectiveness for other languages such as Korean is limited. Thus, the applicability of BERT-based language models pretrained in languages other than English to NLU tasks based on those languages must be investigated. In this study, we comparatively evaluated seven BERT-based pretrained language models and their expected applicability to Korean NLU tasks. We used the climate technology dataset, which is a Korean-based large text classification dataset, in research proposals involving 45 classes. We found that the BERT-based model pretrained on the most recent Korean corpus performed the best in terms of Korean-based multiclass text classification. This suggests the necessity of optimal pretraining for specific NLU tasks, particularly those in languages other than English.

Keywords: natural language understanding; multiclass text classification; bidirectional encoder representations from transformers; transfer learning; multilingual representation learning; cross-lingual representation learning

1. Introduction

Transformer language modeling with pretraining and word representations combined with transfer learning has significantly improved many natural language understanding (NLU) tasks, such as text classification and natural language inference [1–6]. Deep-contextualized language models pretrained through masked language modeling, such as bidirectional encoder representations from transformers (BERT) [3], have demonstrated state-of-the-art performances in NLU tasks. However, these BERT-based models are primarily tested on text data in English; hence, they may not perform well on text written in other languages. In particular, the Korean language, which is an agglutinative language that requires sophisticated processing, may not be managed effectively by pretrained language models (PLMs) developed for the English language. Korean-specific PLMs such as KoBERT [7], KLUE-BERT [8], and KR-BERT [9], pretrained on numerous novel Korean-centric datasets, have shown significant improvements in Korean-based NLU tasks. This indicates that PLMs are sensitive to the language of the text data used for pretraining. Most studies regarding pretraining with Korean-specific corpora are typically conducted using datasets obtained based on different criteria; however, the performance is not comprehensively verified for each NLU benchmark, rendering the selection of pretrained specifications difficult.

In this study, we investigated the manner in which the current PLMs classify climate technology in research proposals written in the Korean language, which is a multiclass
text classification task. We fine-tuned seven PLMs using a dataset and compared the classification performance across PLMs while matching all hyperparameters. The main contributions/findings of this study are as follows:

- We compared the performance of a fine-tuned model with those of seven BERT-based models pretrained with different pretraining corpora.
- We evaluated the performance of each model when performing Korean-based multi-class text classification using a climate technology classification dataset that contained more than 200,000 research proposals in Korean spanning 45 different categories.
- The pretrained model, which was fine-tuned using the most recent novel Korean corpora, showed up to a 7% performance improvement as compared with the model with different Korean corpora.

2. Background

2.1. Transfer Learning and Pretrained Language Modeling

Transfer learning improves performance by sharing the parameters of a model that has been trained on similar task data in advance, unlike conventional machine learning that performs isolated and single-task learning. Furthermore, in transfer learning, previously learned tasks are required when learning a new task; therefore, it is more accurate, requires less training data, and its learning process is faster than that of isolated machine learning.

Transfer learning was first presented and actively applied in image classification, where pretrained network models such as VGG [10], Inception [11], and ResNet [12] were used. In natural language processing (NLP), PLMs such as GPT [13], ELMo [1], and BERT [3] were released in 2018 and achieved state-of-the-art performances in most NLP tasks. In particular, BERT and ELMo overcame the critical constraint of previous language models such as GPT, which is a unidirectional learning method. This unidirectional learning method does not consider the “context”, which is one of the most important factors of language models for a wide range of NLP tasks; therefore, it restricts the capability of the pretrained representations. To address this issue, bidirectional learning methods that are well-contextualized have been proposed in BERT and ELMo. Unlike ELMo, which is based on long short-term memory (LSTM) and applies a pretrained model to downstream tasks using a feature-based method, BERT solves the vanishing gradient problem constraint caused by recurrent neural network layers by replacing transformer-based modeling with a layer. BERT can be applied to downstream tasks by fine-tuning all pretrained parameters, as shown in Figure 1.

![Figure 1](image_url)

**Figure 1.** Architecture of fine-tuning with pretrained BERT in single-sentence classification. For classification, only final hidden vector C is used to predict labels and the other vectors are disregarded.
2.2. BERT

BERT is a deep bidirectional pretrained encoder model that operates by stacking transformers in multiple layers for language understanding. BERT is available in two sizes: BERT-BASE (containing 12 transformer layers and 768 hidden layers) and BERT-LARGE (containing 24 transformer layers and 1024 hidden layers). These two models were trained on the same datasets: BookCorpus (800M words) [14] and English Wikipedia (2500M words). BERT can be applied to downstream tasks by fine-tuning all pretrained parameters, as shown in Figure 1. As such, BERT can be easily applied to several NLP tasks by adding output layers. BERT performed the best in 11 NLP tasks and achieved state-of-the-art performances in 8 tasks of general language understanding evaluation (GLUE) [15], two areas of SQuAD v1.1 [16] and v2.0 [17] (question answering tasks), and SWAG [18] (a common-sense inference task). BERT is designed for training using self-supervised learning, wherein the model learns the context of a sentence during training by introducing a masked language model. The masked language model randomly masks tokens from the input text to predict the original vocabulary of the masked word based only on its context. This method solves the unidirectionality constraint that does not consider the context of the sentence comprehensively and allows BERT to consider a change in the meaning of words based on the context in the training method, thereby enabling the BERT language model to achieve performances comparable to those afforded by humans for language understanding tasks. In addition, the next sentence prediction task allows the relationships between sentences to be determined based on text-pair representations. By applying next sentence prediction, BERT can be utilized in many downstream tasks such as question answering as well as natural language inference by understanding the relationship between two sentences.

2.3. Multilingual Pretraining Using BERT

A few models demonstrated excellent performances on NLU tasks, such as the XLNet [19] and T5 [20]; however, BERT was still adopted as the baseline model in many studies, as it can be easily applied to fine-tuning, and its performance is comparable to those of other, newer models. In addition, RoBERTa, which affords better performances by supplementing the weaknesses of BERT, has been widely applied in conjunction with BERT in many studies. However, most of these BERT-based models are based on an English-centric design; consequently, researchers have attempted to develop models based on languages other than English.

Multilingual BERT [19] retains the model structures of BERT but replaces the pretrained corpus attributes with those that include more than 100 languages from only the existing English-centric BERT, resulting in significant performance improvements over the original BERT in natural language comprehension tasks. However, this implies that its vocabulary size is large and that its size increases inefficiently owing to the processing of more than 100 languages, which consequently restricts the memory efficiency.

Cross-lingual modeling [4] trains a model via unsupervised-learning-based pretraining, where continuous learning in English and other languages is applied simultaneously. BERT pretrained through the cross-lingual method improved the NLU accuracy in multilingual tasks as compared with the original BERT that was pretrained using a dataset containing 100 languages. This method significantly improves symbolic performance in multilingual tasks other than those in English. However, cross-lingual modeling is not comparable to well-preprocessed models in English in terms of accuracy.

Meanwhile, researchers achieved performances comparable to those of English-based models by completely replacing the pretraining corpus attributes with Korean and excluding English [8,9]. Because these studies were conducted using different methods in recent years, the pretraining corpus was determined independently and in parallel for each study, wherein the difference was indicated in only the composition of the dataset, thereby rendering performance prediction difficult. For these Korean-based models, various experiments must be conducted using new NLU task-based benchmark datasets instead.
of substantially standard task benchmarks such as NSMC [21] (text classification) and KorQuAD [22] (question and answering). Benchmark evaluation is primarily conducted through NSMC in Korean text classification, which involves the classification of two labels; in fact, it is difficult to predict performance using only the data presented in the results.

3. Robust Language Models in Korean

We selected a total of seven pretrained language models based on BERT; three in Korean, three that were multilingual, and one in English. This section discusses the criteria for selecting the PLMs and provides a brief overview of the seven PLMs. We selected PLMs that satisfied the following criteria for our task: (1) the target PLMs must be based on BERT and RoBERTa; (2) their pretraining is expected to be robust to Korean text data, i.e., they are Korean pretrained or multilingual pretrained (the pretraining details are shown in Table 1); and (3) RoBERTa (original version), which is English-pretrained and not directly related to Korean, and can be used to demonstrate the differences in performance between the models pretrained using Korean and English, separately. An overview of these models is presented below, and some of the pretraining details are shown in Table 1.

Table 1. Summary of the seven pretrained language models.

| Model                     | Number of Languages | Total Size (GB) | Korean (GB) | Pretraining Data                                      |
|---------------------------|---------------------|-----------------|-------------|-------------------------------------------------------|
| **English-pretrained baseline (baseline)**                        |                     |                |             |                                                       |
| RoBERTa                  | 1 (en)              | 160             | 0           | BookCorpus, English-Wikipedia, CC-News, OpenWebText, Stories |
| **Multilingual baseline**                                        |                     |                |             |                                                       |
| XLM-RoBERTa              | 100 (en, ko, . . .) | 2500            | 58.2        | CC-100 corpus                                        |
| Multilingual BERT (uncased) | 102 (en, ko, . . .) | Unknown        | Unknown     | Top 100 languages Wikipedia                           |
| Multilingual BERT (cased) | 104 (en, ko, . . .) | Unknown        | Unknown     | Top 100 languages Wikipedia (+Thai, Mongolian)        |
| **Korean-pretrained baseline**                                     |                     |                |             |                                                       |
| KoBERT                   | 1 (ko)              | Unknown         | Unknown     | Korean Wikipedia                                      |
| KLUE-RoBERTa             | 1 (ko)              | 62              | 62          | Modu, CC-100-Kor, NamuWiki, NewsCrawl, Petition      |
| KLUE-BERT                | 1 (ko)              | 62              | 62          | Modu, CC-100-Kor, NamuWiki, NewsCrawl, Petition      |

**Multilingual BERT (BERT-M):** BERT provides BERT-BASE-Multilingual (BERT-M) and applies multilingual pretraining using a Wikipedia corpus written in 102–104 languages. BERT-M can be applied to multilingual NLP tasks and exhibits the same model architecture and size as BERT-BASE. Furthermore, BERT-M is available in two model options for case sensitivity, i.e., BERT-M-cased and BERT-M-uncased.

**KoBERT:** Owing to the limitations of BERT-base-multilingual-cased (BERT-M-cased) in the Korean NLP task, KoBERT (the Korean BERT pretrained case) was released [7], pretrained only in the Korean corpus with BERT. Its architecture is the same as that of BERT (12 transformer encoders and 764 hidden layers), and its pretraining dataset is primarily based on the Korean Wikipedia (54M words, 5M sentences).

**RoBERTa:** RoBERTa improves BERT’s undertrained points through the following four tuning procedures: using more pretraining data, applying dynamic masking, removing NLP loss, and training on longer sequences. RoBERTa achieved state-of-the-art results on all nine GLUE tasks and outperformed recently proposed architectures, such as the XLNet [19]. We used the original RoBERTa that only considered pretraining using the English corpus as the baseline.

**XLM-RoBERTa:** XLM-RoBERTa is a transformer PLM that is typically used in multilingual tasks [23]. It is pretrained as a masked language model on 100 languages and 2.5 TB of filtered common crawl data. XLM-RoBERTa is a multilingual version of RoBERTa,
where XLM is an acronym for cross-lingual language model. RoBERTa is an unsupervised model that relies only on monolingual data, whereas the XLM is a supervised model that leverages parallel data with a new XLM model objective [4].

**KLUE-Pretrained Language Models**: KLUE-PLMs refers to a BERT (or RoBERTa) model that is pretrained with a large and diverse Korean corpus such as Modu, CC-100-Kor, NAMUWIKI, NEWSCRAWL, or PETITION [8]. Its total size is 62.65 GB.

### 4. Evaluations

We evaluated the performance of the seven models of Korean natural-language-based climate technology classification as a benchmark dataset. Some PLMs can be broadly classified into English-pretrained, multilingual-pretrained, or Korean-pretrained models. They are expected to demonstrate different levels of classification performance depending on the data properties.

We implemented a neural network model that performs a text classification downstream task by adopting a BERT-based pretrained encoder and adding an output layer for multiclass classification. Subsequently, we conducted a comparative study where we analyzed the process by which BERT performs downstream tasks and compared the performance of each model in our main task (Korean text-based multiclass classification). The seven models are as follows:

- RoBERTa-BASE (RoBERTa);
- XLM-RoBERTa-BASE (XLM-RoBERTa);
- BERT-BASE-Multilingual-uncased (BERT-M-uncased);
- BERT-BASE-Multilingual-cased (BERT-M-cased);
- KoBERT-BASE (KoBERT);
- KLUE-RoBERTa-BASE (KLUE-RoBERTa);
- KLUE-BERT-BASE (KLUE-BERT).

#### 4.1. Climate Technology Classification Dataset

We specified the natural-language-based climate technology classification corpus provided by the Green Technology Center (Seoul, Korea) as a natural language-based multiclass text classification task to evaluate the seven pretrained language models and expected robustness in Korean-based tasks through pretraining in Korean.

The dataset was organized to classify research proposals into 45 classes as category labels in Korean, based on climate technology from the National Science & Technology Information Service. The class labels were defined based on the Climate Technology Information System (CTIS) of the Green Technology Center (Table A1); the index number (1 to 45) and the name of the rightmost column (labeled “Section”) were used as labels. The dataset included approximately 170,000 training data and 43,000 test data in the form of research proposals based on Korean text. Each example comprised 13 features of the research proposal such as the title, aims, descriptions, keywords, and other trivial information. We selected three features in the dataset—the title, the aim, and the keywords—because these three features contain the most relevant information to the classes after inspecting all features manually. In addition, the size of sequence length was appropriate for the BERT-based language models when appending the three features. A detailed description of an example from the dataset is presented in Table 2.
Table 2. Descriptions of example in dataset.

| Column | Description (Korean) | Description (Translated into English) |
|--------|----------------------|---------------------------------------|
| Title  | 도로정보를 활용한 고연반 PHEV의 동력분배 제어 기술 개발 | To develop the power distribution control technology of high-fuel-efficiency PHEV using road information |
| Title  | [2차년도 개발 목표] | (Goals for second year) |
| Aims   | 실시간 운전자 성향 판단 기술 비교 평가 | o Comparative evaluation of real-time driver tendency judgment technology; technology analysis for determining driver tendencies |
| Aims   | 운전자 성향 및 도로상황을 고려한 속도 프로파일 예측 기술 | o Speed profile prediction technology considering driver propensity and road conditions. To develop speed profile prediction technology using the established driving information DB. To develop global speed profile prediction technology using driver propensity, GIS information, and traffic light situation in the absence of a driving route DB. |
| Aims   | 구축된 주행정보 DB를 활용한 속도 프로파일 예측 기술 개발 | o Establishing a foundation for the development and verification of power distribution control algorithms. To develop a path-based real-time optimal power distribution control algorithm. To develop engine operation minimization control technology based on real-time vehicle external information. To construct a human-in-the-loop simulation environment. |
| Keywords | 플러그인 하이브리드 자동차, 최적 동력분배 제어, 주행정보 예측, 운전자 운전 성향, 휴먼인더라우 | Plug-in hybrid vehicles, optimal power distribution control, driving information prediction, driving propensity of driver, and human-in-the-loop simulation. |

4.2. Preprocessing

We used the Mecab morphological analyzer [24] to perform part-of-speech tagging using text data, where certain stop words such as special characters (“{ }, “( ),” and “[ ]”) were removed. We only extracted nouns from the text data because the dataset primarily comprised text summarized from complete research proposals, and extracting nouns in Korean is a widely accepted method since Korean is an agglutinative language, which is the form of various morphemes attached to the original nouns [25–27]. An example of preprocessing performed using the morphological analyzer is shown in Figure 2.

In this section, we describe our preprocessing results based on data from the dataset. Although the dataset comprised primarily Korean text, we provide not only the Korean processing results, but also the English translation of the example data in Table 2. We applied nouns using the Mecab morphological analyzer. We selected the same example data as those in Table 2. First, we appended 3 of the 13 columns to the training set and used it as the main text data; the other columns were removed. Second, noun filtering was performed using the Mecab analyzer. As shown in Figure 2, several unique characters, numbers, and English characters in the original text remained after it was appended. These characters were removed through text filtering, as shown in Figure 2, and only nouns were extracted. Consequently, only nouns separated by spaces were used as input data. We applied the same preprocessing for the training and test datasets.
The Macro-F1 score is widely used in multiclass classification and is calculated by adding and averaging all the classification F1 scores for each label. The F1 and Macro-F1 scores are defined in Equations (1) and (2), respectively (TP: true positive, FP: false positive, and FN: false negative).

$$F_1 = 2 \cdot \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$  \quad (1)$$

$$\text{MacroF}_1 = 2 \cdot \frac{\text{Total precision} \times \text{Total recall}}{\text{Total precision} + \text{Total recall}}$$  \quad (2)$$

Furthermore, we evaluated the validation and testing performances. The validation performance was calculated using a portion of the test dataset, whereas the test performance was calculated using the remainder of the test dataset to determine whether overfitting occurred.

Hyperparameters: We fixed all the parameter conditions (loss function, optimizer, epochs, validation split rate, and sequence length) by evaluating the performances of the BERT-based encoder models. We applied Adam [28] as an optimizer and sparse categorical cross-entropy loss as a loss function, which involves converting a categorical cross-entropy (CCE) data target from a one-hot vector to an integer. The CCE loss is widely accepted in multiclass classification tasks and is defined as follows:

$$\text{CCE Loss} = - \sum_{i=1}^{\text{output size}} y_i \log \hat{y}_i$$  \quad (3)$$

where $\hat{y}_i$ is the predicted model output, and $y_i$ is the target value. Both are described as one-hot vectors. We trained our models using 10 epochs and a split validation data rate of 20%. The sequence length for BERT was fixed across all pretrained models. The hyperparameters are summarized in Table 3.
4.4. Evaluation on Climate Technology Classification

To investigate the performance of the PLMs in classifying climate technology in research proposals, we calculated the classification performance for each PLM fine-tuned using the preprocessed text (Table 4). RoBERTa is the only pretrained model in English text; therefore, it can be used as a baseline model to estimate the improvement in the classification performance of the other models. It achieved classification performances of 0.59 and 0.56 for the validation and test datasets, respectively. Meanwhile, XLM-RoBERTa, which is a cross-lingual-based pretrained model, achieved classification performances of 0.63 and 0.61 for the validation and test datasets, respectively. Compared with RoBERTa, which is a model pretrained using English text, XLM-RoBERTa demonstrated an improved performance by approximately 4%. This indicated that the XLM achieved a better classification performance than the model pretrained using English text. Meanwhile, BERT-M-cased pretrained with the multilingual method achieved classification performances of 0.65 and 0.64 for the validation and test datasets, respectively. This indicated that the multilingual model performed better than the XLM. To investigate the effect of pretraining using uncased text in the multilingual model, we calculated the classification performance of BERT-M-uncased, where 0.70 and 0.68 were obtained for the validation and test datasets, respectively. This indicated that the multilingual model pretrained using uncased characters performed better than the cased model in the classification task.

Table 4. Classification performance of climate technology classification by the seven language models.

| Model                  | Validation (Macro-F1) | Test (Macro-F1) |
|------------------------|-----------------------|-----------------|
| **English-pretrained baseline (baseline)** |                       |                 |
| RoBERTa                | 0.59                  | 0.57            |
| **Multilingual baseline** |                       |                 |
| XLM-RoBERTa            | 0.63                  | 0.61            |
| Multilingual BERT (cased) | 0.65            | 0.64            |
| Multilingual BERT (uncased) | 0.71            | 0.68            |
| **Korean-pretrained baseline** |                       |                 |
| KoBERT                 | 0.67                  | 0.67            |
| KLUE-RoBERTa           | 0.72                  | 0.70            |
| KLUE-BERT              | 0.74                  | 0.72            |

Additionally, we calculated the classification performance of the models pretrained with Korean text to investigate language-specific effects on pretraining. The classification performance of KoBERT, which was pretrained using the Korean Wikipedia only, was 0.67 for both the validation and test datasets. Meanwhile, KLUE-RoBERTa, which was pretrained with RoBERTa using the KLUE dataset, demonstrated classification levels of 0.72 and 0.70 for the validation and test datasets, respectively. The differences in the classification performance were attributable to the size and detailed refinement of the dataset used for pretraining. Finally, KLUE-BERT achieved the highest classification performance (0.74 and 0.72 for the validation and test datasets, respectively). These findings imply that matching the language, quality, and size of the text for pretraining are key in multiclass classification. In addition, the error range was approximately 0.15 between the lowest and highest values.
5. Conclusions

In this study, we investigated the applicability of PLMs in Korean text datasets and evaluated them through multiclass classification. We compared the performances of BERT-based encoder models and discovered that KLUE-BERT outperformed the other models in terms of classification; additionally, KLUE-BERT indicated an error range of up to 0.15 in terms of the Macro-F1 score and matched the language for a classification task.

We aim to show the degree of performance difference caused by the encoder model pretrained in Korean on a multiclass dataset with 45 labels, which is more challenging than the existing benchmarks of text classification. Through our findings, we imply that the size and detailed refinement of the corpus used for pretraining could be crucial factors and we suggest that applying the most appropriate pretraining language model to NLP tasks is fundamental.

As with any scientific study, this study has limitations. First, we tested the PLMs in a classification task only. Other NLP tasks, such as natural language inference or question and answering, should be tested because these tasks may be more sensitive depending on language-specific models. Second, we selected BERT-based pretrained language models only for the dataset. Other pretrained language models such as ELECTRA should be considered to investigate language model-specific effects. Third, we extracted nouns only from the dataset because the dataset consisted of summarized text in Korean. Noun extraction has been regarded as an effective way for understanding Korean, but other tags could be used in a future study. Thus, future research should consider other NLP tasks and state-of-the-art language models to verify our findings.

Author Contributions: Conceptualization, E.L., C.L. and S.A.; methodology, E.L.; validation, E.L., C.L. and S.A.; formal analysis, E.L.; writing—original draft preparation, E.L.; writing—review and editing, E.L., C.L. and S.A.; supervision, S.A.; funding acquisition, S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Kyungpook National University Research Fund, 2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A  Category Labels of Climate Technology Classification Dataset

In this section, we provide raw data for the 45 categories used to classify the climate dataset multilabel. The labels were related to research areas of climate technology such as greenhouse gas mitigation, agriculture, water management, and climate change forecasting. According to the CTIS of the Green Technology Center and the training data of the dataset, 45 class labels existed with a default label “0”; therefore, they were configured to 46 labels to perform the multiclass classification task.
Table A1. Descriptions of climate technology information system of green technology center. Index numbers of the rightmost column (Section) were used as labels for our classification task.

| Category | Division | Section (Label) |
|----------|----------|-----------------|
| **Mitigation** | **Greenhouse Gas Mitigation** | (1) Non-renewable Energy |
| | | 1. Nuclear power |
| | | 2. Nuclear fusion power |
| | | 3. Clean thermal power and efficiency |
| | | (2) Renewable Energy |
| | | 4. Hydropower |
| | | 5. Photovoltaic power |
| | | 6. Solar heat |
| | | 7. Geothermal power |
| | | 8. Wind power |
| | | 9. Ocean energy |
| | | 10. Bio energy |
| | | 11. Waste |
| | | (3) New Energy |
| | | 12. Hydrogen manufacturing |
| | | 13. Fuel cell |
| | | (4) Energy Storage |
| | | 14. Power storage |
| | | 15. Hydrogen storage |
| | | (5) Transmission and Distribution and Power IT |
| | | 16. Transmission and distribution system |
| | | 17. Intelligent electric device |
| | | 18. Transport efficiency |
| | | 19. Industrial efficiency |
| | | 20. Building efficiency |
| | | (6) Energy Demand |
| | | 21. CCUS |
| | | 22. Non-CO₂ mitigation |
| | | 23. Generic resources and genetic improvement |
| | | 24. Crop cultivation and production |
| | | 25. Livestock disease control |
| | | 26. Processing, storage and distribution |
| | | (7) Greenhouse Gas Sequestration |
| | | 27. Water system and aquatic ecosystem |
| | | 28. Water resource security and supply |
| | | 29. Water treatment |
| | | 30. Water disaster control |
| **Adaptation** | **(8) Agriculture and Livestock** | (8) Agriculture and Livestock |
| | | 31. Climate change forecast and modeling |
| | | 32. Climate information and warning system |
| | | 33. Ocean ecosystem |
| | | 34. Marine resources |
| | | 35. Offshore disaster control |
| | | 36. Contagious disease control |
| | | 37. Food safety and prevention |
| | | 38. Productive forest improvement |
| | | 39. Forest damage mitigation |
| | | 40. Ecosystem monitoring and recovery |
| | | (12) Healthcare |
| | | (13) Forest and Land Management |
| **Mitigation/Adaptation Convergence** | **(14) Mixture of Multiple Areas** | (14) Mixture of Multiple Areas |
| | | 41. New and renewable energy hybrid |
| | | 42. Low power consumption equipment |
| | | 43. Energy harvesting |
| | | 44. Artificial photosynthesis |
| | | 45. Other technologies related to climate change |

References

1. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep Contextualized Word Representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, LA, USA, 1–6 June 2018; Volume 1, pp. 2227–2237. [CrossRef]
2. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. Adv. Neural Inf. Process. Syst. 2017, 2017, 5999–6009.
3. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv 2018, arXiv:1810.04805.

4. Lample, G.; Conneau, A. Cross-Lingual Language Model Pretraining. arXiv 2019, arXiv:1901.07291.

5. Bilal, M.; Almazroii, A.A. Effectiveness of Fine-Tuned BERT Model in Classification of Helpful and Unhelpful Online Customer Reviews. Electron. Commer. Res. 2022, forthcoming.

6. Ma, K.; Tian, M.; Tan, Y.; Xie, X.; Qiu, Q. What Is This Article about? Generative Summarization with the BERT Model in the Geosciences Domain. Earth Sci. Inform. 2022, 15, 21–36. [CrossRef]

7. SKTBrain/KoBERT: Korean BERT Pre-Trained Cased (KoBERT). Available online: https://github.com/SKTBrain/KoBERT (accessed on 23 March 2022).

8. Park, S.; Moon, J.; Kim, S.; Cho, W.I.; Han, J.; Park, J.; Song, C.; Kim, J.; Song, Y.; Oh, T.; et al. KLUE: Korean Language Understanding Evaluation. arXiv 2021, arXiv:2105.09860.

9. Lee, S.; Jang, H.; Baik, Y.; Park, S.; Shin, H. KR-BERT: A Small-Scale Korean-Specific Language Model. arXiv 2020, arXiv:2008.03979.

10. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2015, arXiv:1409.1556.

11. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]

12. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

13. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. Available online: https://www.cs.ubc.ca/~amulham01/LING530/papers/radford2018improving.pdf (accessed on 23 March 2022).

14. Zhu, Y.; Kiros, R.; Zemel, R.; Salakhutdinov, R.; Urtasun, R.; Torralba, A.; Fidler, S. Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 19–27. [CrossRef]

15. Wang, A.; Singh, A.; Michael, J.; Hill, F.; Levy, O.; Bowman, S.R. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. arXiv 2019, arXiv:1804.07461.

16. Rajpurkar, P.; Zhang, J.; Lopyrev, K.; Liang, P. SQuAD: 100,000+ Questions for Machine Comprehension of Text. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–4 November 2016.

17. Rajpurkar, P.; Jia, R.; Liang, P. Know What You Don’t Know: Unanswerable Questions for SQuAD. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018.

18. Zellers, R.; Bisk, Y.; Schwartz, R.; Choi, Y.; Allen, P.G. Swag: A Large-Scale Adversarial Dataset for Grounded Commonsense Inference. arXiv 2018, arXiv:1808.05326.

19. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.; Le, Q.V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. arXiv 2019, arXiv:1906.08237.

20. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. arXiv 2019, arXiv:1910.10683.

21. Naver Sentiment Movie Corpus. Available online: https://github.com/e9t/nsmc (accessed on 23 March 2022).

22. KorQuAD. Available online: https://korquad.github.io (accessed on 23 March 2022).

23. Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; Stoyanov, V. Unsupervised Cross-Lingual Representation Learning at Scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020.

24. MeCab: Yet Another Part-of-Speech and Morphological Analyzer. Available online: https://taku910.github.io/mecab (accessed on 23 March 2022).

25. Kim, J.H.; Kwak, B.K.; Lee, S.; Lee, G.; Lee, J.H. A Corpus-Based Learning Method of Compound Noun Indexing Rules for Korean. Inf. Retr. 2001, 4, 115–132. [CrossRef]

26. Kang, B.J.; Choi, K.S. Effective Foreign Word Extraction for Korean Information Retrieval. Inf. Process. Manag. 2002, 38, 91–109. [CrossRef]

27. Yang, S.; Ko, Y. Finding Relevant Features for Korean Comparative Sentence Extraction. Pattern Recognit. Lett. 2011, 32, 293–296. [CrossRef]

28. Kingma, D.P.; Lei Ba, J. Adam: A method for stochastic optimization. arXiv 2014, arXiv:1412.6980.