ECE1786H Project Final Report LingoEtiquette

Word Count: 1947, excluding table of content
Penalty: 0%
Haizheng Sun (1009696321)
Shihao Piao (1009735841)
Date: Dec 12th, 2023 @ 6 pm

<u>Permissions</u>

Haizheng Sun

● Permission to post video: yes

● Permission to post final report: yes

● Permission to post source code: no

Shihao Piao

● Permission to post video: yes

● Permission to post final report: yes

● Permission to post source code: no

<u>Table of Content</u>

Introduction

As Non-native English speakers, we are accustomed to typing the mother language into a translator to get the translated English sentences, then doing style transfer manually, and finally using the sentences in a suitable situation.

We thought style transfer was important because in Chinese and Korean, it is important to use both formal and informal words correctly in the appropriate situation. Wrong use will have a bad effect on our work and life. **Formal language** is characterized by a more severe tone, often adhering to conventional or established grammatical rules and used in contexts requiring respect or professionalism. **Informal language** is more casual and conversational. It is often used in everyday communication among friends or in relaxed social settings.

This traditional process (Fig.1) takes a long time, easily makes mistakes, and the sentences we use may be strange to native speakers. To help people communicate using English accurately and conveniently, especially for beginners, we do such a project. By using our model, the process becomes simple. Just by waiting a few seconds, the user can get informal and formal expressions.
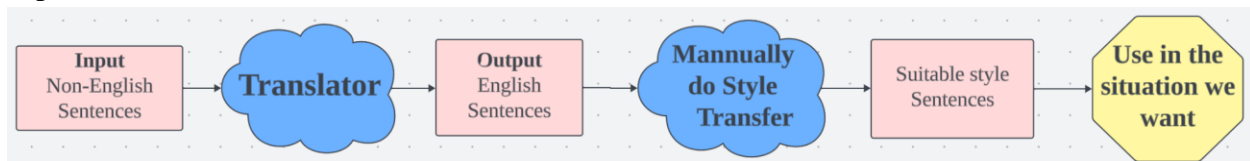
Fig.1 Traditional Process

Our model can be divided into a translation model and a style-transfer model (Fig.2). We use machine learning algorithms because they can process and learn from large volumes of bilingual text data. This capacity allows the model to identify patterns and correlations in language pairs, leading to more accurate translations. Because the training data we used in the translation model is **formal**, we **assume** it outputs English sentences in a **formal** style.

Illustration/Figure

The following picture shows the architecture of Lingo Etiquette.

Fig.2 Whole Model Structure of Lingo Etiquette

Background & Related Work

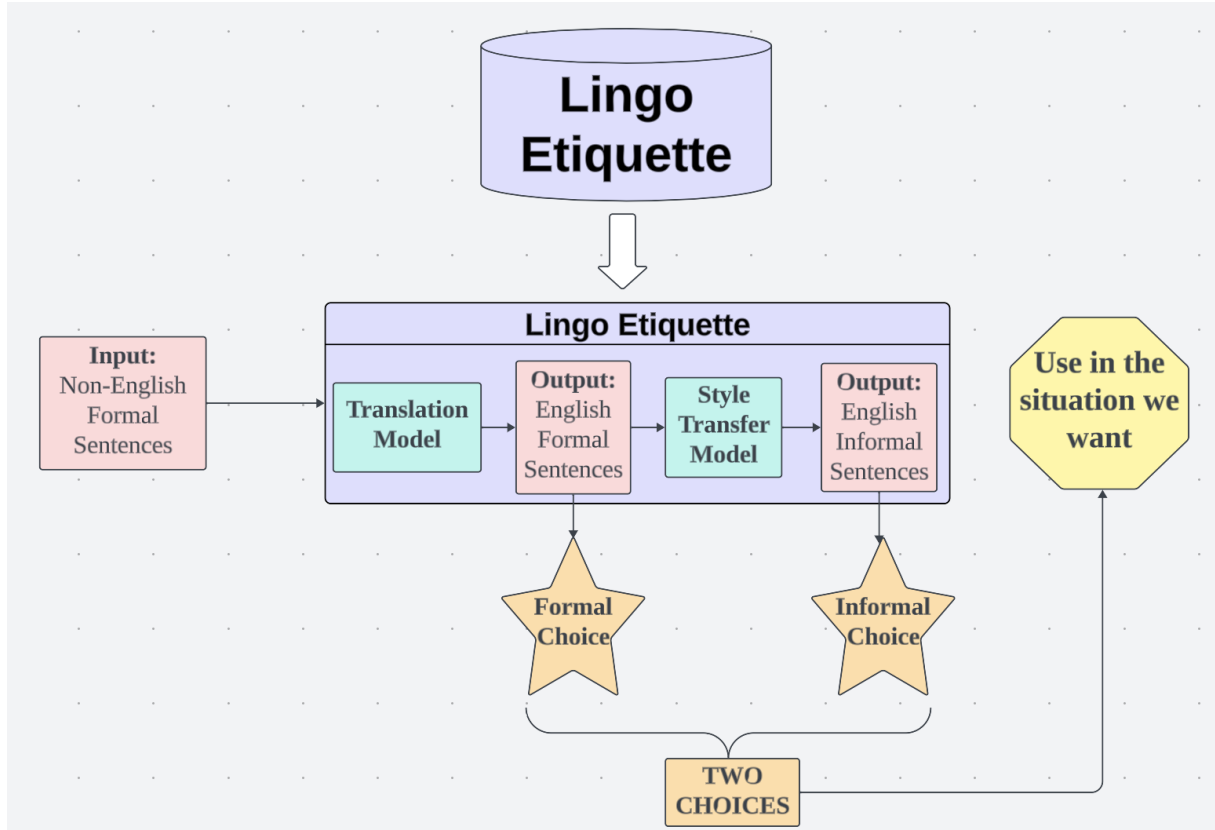Transformer model (Vaswani et al., 2017[1]), a novel neural network architecture that eschews recurrent layers for an attention mechanism. The Transformer model revolutionized machine translation by enabling more parallelization in training, which significantly increased efficiency and effectiveness in handling long-range dependencies in text. This work laid the foundation for subsequent developments in machine translation and natural language processing. The GYAFC (Grammarly's Yahoo Answers Formality Corpus) dataset (Rao & Tetreault, 2018[2]) significantly contributes to computational linguistics, particularly in formality style transfer. GYAFC is composed of parallel sentences in both formal and informal styles. These sentences are derived from Yahoo Answers, providing a rich source of naturally occurring language that reflects real-world informal and conversational text. The primary goal of the dataset is to aid in developing and benchmarking algorithms capable of converting text between formal and informal styles.

Data and Data Processing

We only used Multi30K for the translation model and GYAFC for the style transfer model at first, but the performance was bad. This is because Multi30K is not large enough, and GYAFC does not cover enough fields. To improve, we combine WMT16 with Multi30K. We dropped

duplications and cleaned the data, including removing punctuations and changing the uppercase into lowercase. The dataset, after processing, has two columns, with one column for Formal German Sentences and one column for translated Formal English Sentences (Fig.4). We split the whole dataset into a training set (60000), validation set (2000), and test set (1000) (Fig.3).

```
Train: (60001, 2)
Validation: (2001, 2)
Test: (1000, 2)
```

Fig.3 Train/Validation/Test Split

	English	Ger
0	Two young White males are outside near many b...	Zwei junge weiße Männer sind im Freien in der ...
1	Several men in hard hats are operating a giant...	Mehrere Männer mit Schutzhelmen bedienen ein A...
2	A little girl climbing into a wooden playhouse.	Ein kleines Mädchen klettert in ein Spielhaus ...
3	A man in a blue shirt is standing on a ladder ...	Ein Mann in einem blauen Hemd steht auf einer ...
4	Two men are at the stove preparing food.	Zwei Männer stehen am Herd und bereiten Essen zu.
...	...	...
59996	It is also proposed to end state aid to airlines.	Vorgeschlagen wird weiterhin, die staatlichen ...
59997	This, however, will prove disastrous for air c...	Dies wird jedoch fatale Konsequenzen für die F...
59998	Greece is a typical example.	Ein in jeder Hinsicht typisches Beispiel dafür...
59999	These are the reasons why we shall be voting a...	Aus den genannten Gründen werden wir gegen den...
60000	We agree with these objectives, but we cannot ...	Mit derartigen Zielen sind wir einverstanden, ...

60001 rows × 2 columns

Fig.4 Translation Model Processed Dataset Training Set

	English	Ger
0	A group of men are loading cotton onto a truck	Eine Gruppe von Männern lädt Baumwolle auf ein...
1	A man sleeping in a green room on a couch.	Ein Mann schläft in einem grünen Raum auf eine...
2	A boy wearing headphones sits on a woman's sho...	Ein Junge mit Kopfhörern sitzt auf den Schulte...
3	Two men setting up a blue ice fishing hut on a...	Zwei Männer bauen eine blaue Eisfischerhütte a...
4	A balding man wearing a red life jacket is sit...	Ein Mann mit beginnender Glatze der eine rote...
...	...	...
1996	Until the day when, like the man hoist with hi...	Bis zu dem Tag, an dem – als Veranschaulichung...
1997	In fact today, like yesterday, the issue is po...	In Wirklichkeit stehen wir heute wie gestern v...
1998	Lucas Casarini, leader of the White Overalls m...	Der Anführer der "Tute Bianche " Lucas Casarin...
1999	Negri, imprisoned for armed insurrection again...	Auch Negri, der wegen bewaffnetem Aufstand geg...
2000	These people are no longer socialist internati...	Diese Leute sind keine internationalen Soziali...

2001 rows × 2 columns

Fig.5 Translation Model Processed Dataset Validation Set

	English	Ger
0	A man in an orange hat starring at something.	Ein Mann mit einem orangefarbenen Hut der etw...
1	A Boston Terrier is running on lush green gras...	Ein Boston Terrier läuft über saftig–grünes Gr...
2	A girl in karate uniform breaking a stick with...	Ein Mädchen in einem Karateanzug bricht ein Br...
3	Five people wearing winter jackets and helmets...	Fünf Leute in Winterjacken und mit Helmen steh...
4	People are fixing the roof of a house.	Leute Reparieren das Dach eines Hauses.
...	...	...
995	Marathon runners are racing on a city street ...	Marathonläuferinnen laufen auf einer städtisch...
996	Asian woman wearing a sunhat while riding a bike.	Asiatische Frau trägt einen Sonnenhut beim Fah...
997	Some children are outside playing in the dirt ...	Ein paar Kinder sind im Freien und spielen auf...
998	An older man is playing a video arcade game.	Ein älterer Mann spielt ein Videospiel.
999	A girl at the shore of a beach with a mountain...	Ein Mädchen an einer Küste mit einem Berg im H...

1000 rows × 2 columns

Fig.6 Translation Model Processed Dataset Test Set

To improve the performance of the style transfer model, we combine the dataset generated from GPT-4 with GYAFC. After combining separate files into one with one column **formal** and one column **informal**. We split the whole dataset into a training set (60000), shown in Fig.7, and a validation set (2000), shown in Fig.8.

	formal	informal
0	The In–Laws movie isn't a holiday movie, but i...	the movie The In–Laws not exactly a holiday mo...
1	I don't think that page gave me viruses.	that page did not give me viroses(i think)
2	I watch it everyday, my favorite charachter is...	of corse i be wachin it evry day, my fav chara...
3	Funbrain.com and runescape.com are great for f...	runescape.com (my kids love it) & funbrain.com...
4	He was on the Late Night show with Conan O'Bri...	Is he gay?He was on Late Night with Conan O'Br...
...	...	...
59995	To get her to do what you want her to do, you ...	she is gonna do what she wants no matter what ...
59996	You get excited when you talk about him.	you sound like you like him a lot.
59997	I am a single woman.	im single, but i'm a girl sry lol
59998	Just because he says it, doesn't mean that he ...	But, that does not mean he is in love, or love...
59999	I hope that you get it resolved.	Hope you get it worked out.

60000 rows × 2 columns

Fig.7 Style Transfer Training Dataset

	formal	informal
0	Kind of like Desperate Housewives (not that th...	Kind of like Desperate Housewives (not that th...
1	The Vowels: A, E, I, O, U, and sometimes Y.	A, E, I, O, U, and sometimes Y
2	Capricorn is in no way compatible, don't ever ...	Dont fool youself, Capricorn is in no way comp...
3	Lol. They looks like they are humping each oth...	hehehe....I think they look like they are hump...
4	Sorry, but I don't think anybody will know.	I don't think anybody will know...Sorry!
...	...	...
1995	This is an acceptable program but CSI is highe...	is a good show, but csi is better!
1996	There are numerous online options such as Walm...	you can go to a couple of websites like walmar...
1997	I believe it is Aries and Scorpio.	In my opinion, Aries and Scorpio!
1998	Panic at the Disco is classified as dance rock.	At the Disco is... Dance Rock.
1999	I agree with clothing in this situation.	i have to agree with clothing on this one!

2000 rows × 2 columns

Fig.8 Style Transfer Validation Dataset

Architecture and Software

LingoEtiquette can be divided into a translation model and a style transfer model as shown in Fig. 2. After getting the non-English formal sentences, the translation model will output the translated English formal sentences, which is the formal choice of our model. Then, the translated English formal sentences will be put into the style transfer model to get the informal

choice. Finally, users can get formal and informal expressions and then use them in their desired situations.

In the modeling aspect, we constructed a transformer model from scratch for translation tasks, while for style-transfer tasks, we fine-tuned the T5 model(Raffel et al., 2020[3]). The two models are both transformer models. The Transformer follows an encoder-decoder structure. The encoder and decoder comprise a stack of identical layers containing two main sub-layers; we used three layers in our model. Since the model lacks recurrence and does not process data in order, it uses positional encodings to account for the sequence of the data. These encodings are added to the input embeddings to provide information about the words' position in the sequence. The Transformer utilizes multi-head attention, which runs the attention mechanism multiple times in parallel using eight heads in our model. In the decoder, after the multi-head attention and feed-forward network, there is a final linear layer followed by a softmax layer, which generates the output sequence.

Baseline Model or Comparison

For the translation component of our study, we utilized GPT-4 and MarianMT as our baseline models. GPT-4, a highly advanced model from OpenAI, is renowned for its exceptional language understanding and generation capabilities, which include translation tasks. MarianMT, on the other hand, is a leading neural machine translation framework specifically designed for high-performance machine translation tasks. Known for its efficiency and adaptability, MarianMT is often used in research and production environments where fast, accurate translations are required. By including GPT-4 and MarianMT as baselines, we aimed to cover a broad spectrum of translation capabilities, from highly sophisticated language models to specialized translation-focused frameworks.

For the style-transfer aspect of our research, the OpenAI API served as the baseline model, especially the GPT-4 API. This decision was bolstered by thoroughly comparing manual evaluation results and those provided by the GPT-4 API. Remarkably, we observed a high degree of correlation between the two, with the API's assessments consistently mirroring the judgments rendered by human evaluators. Given this parity, we determined that the GPT-4 API could serve as a reliable and objective metric for our evaluation needs.
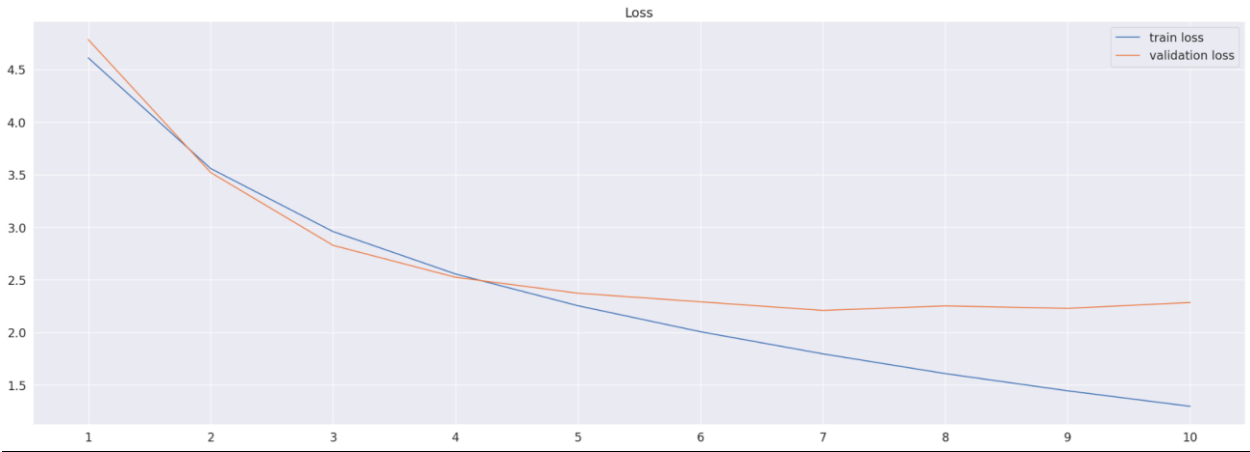
Quantitative Results

Fig.9 Loss Curve of Translation Model

Fig.9 shows the progression of model training over ten epochs. The blue line, representing training loss, and the orange line, representing validation loss, both offer a downward trend, indicating an improvement in model performance as the number of epochs increases. Initially, the training and validation losses decrease sharply, suggesting that the model is rapidly learning from the training data. As the epochs progress, the training loss reduces, which is typical as the model begins to converge and incremental learning gains become smaller. There's a notable point around epoch five where the validation loss plateaus slightly increases, suggesting the onset of overfitting or that the model has reached its learning capacity on the current dataset and architecture.

BLEU	LingoEtiquette	GPT4	MarianMT
Score	0.23	0.34	0.32

Fig.10 BLEU Score of Model

We chose BLEU as our metric; Fig.10 shows the BLEU score of our model versus baseline models, with 1 representing a perfect match. It can be seen from the results in the table that our model's BLEU is 0.23, while GPT-4 and MarianMT are 0.34 and 0.32, respectively. The performance of these models is not good, but this is because the translation of sentences may have multiple translation results, and changing words or phrases cannot mean that the translation result is bad, so this is a disadvantage of BLEU. GPT4 and MarianMT, as relatively good models, had almost no problems in their translation results. Therefore, we compared our scores with those of other models, conducted manual checks, and found that our model's performance was acceptable.
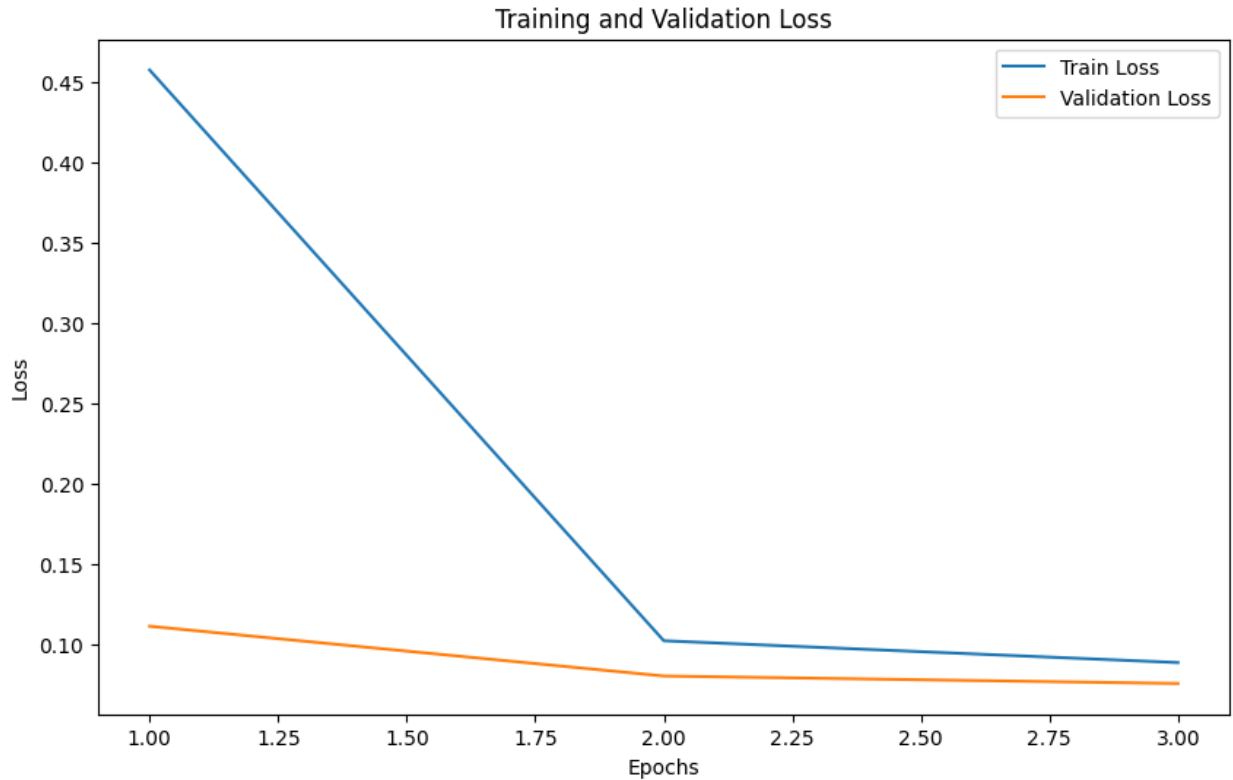
Fig.11 Loss Curve of Style-Transfer Model

Fig.11 displays the training and validation loss over three epochs. The training loss, represented by the blue line, shows a sharp decline, indicating that the model quickly learns from the training dataset. This rapid decrease suggests that the model's predictions are becoming increasingly accurate concerning the training data. By the end of the third epoch, the training loss levels off, implying that the model has begun to converge and there is less room for improvement on the training dataset.

```
Output: a guy is working on a building
Informal


Output: a guy in a vest sitting on a chair and holding the forest sector
Informal
```

Fig.12 Classified Results of GPT-4 API

For the evaluation method for the Style-Transfer part, we used the GPT-4 API to create a classifier, run the examples, informal sentences as input, and whether this is informal as output.

Some processing outputs are shown in Fig.12. With this method, we got an accuracy of 0.86, a bit higher than we expected.

Qualitative Results

Fig.13 Example of Translation Model Output

Fig.13 is an example from our translation module, which displays a German sentence at the top, followed by its English translation. After inserting this German sentence into our model, the resulting translations demonstrate a remarkable level of precision, successfully capturing the core and significance of the original text in most aspects. However, a slight error is observed towards the end of the translation. Additionally, there are instances where the model generates 'unknown' tokens, indicating a gap in its ability to recognize or translate certain words or phrases. Furthermore, the model occasionally produces a surplus of words that may be deemed unnecessary or irrelevant to the core meaning of the sentence. These 'useless words' can affect the conciseness and clarity of the translation.

Fig.14 Example of Style-Transfer Model Output

Fig.14 presents an informal version of the output sentence derived from the same example discussed earlier. A noticeable aspect of this output is the presence of spelling mistakes and word substitutions. For instance, the word 'beer' is transformed into 'booze,' a change that fits within our expectations of everyday language usage. The occurrence of spelling errors in the informal dataset is not unusual; it can be considered a symbol of informality in language. Despite these irregularities, the resultant informal sentence can be deemed acceptable within the scope of our informal language transformation objectives. However, there are instances where the output of the informal transformation remains unchanged from the original formal sentence. This lack of change suggests that the model may sometimes need help identifying and applying the appropriate informal characteristics to particular sentences.

Discussion and Learnings

The results are sensible, as we constructed the model from scratch, leading to performance that is not as good as others. Second, preprocessing is crucial, particularly due to the presence of excessively long sentences in the dataset compared to others. Such data can degrade performance and slow down the training process. Finally, to ensure consistency, running the code on a single platform is important. We observed varying results when using Colab and a local machine, so choosing one platform for all executions is recommended.

Individual Contributions

Content of Task	Person in charge
Train the Translation Model	Shihao Piao
Collect the Translation Dataset	Shihao Piao
Fine-tune the Style-Transfer Model	Haizheng Sun
Collect the Style-Transfer Dataset	GYAFC - Shihao Piao GPT-4 - Haizheng Sun
Use Baseline Models to Compare the Results (Translation Part)	Shihao Piao
Use Baseline Models to Compare the Results (Style-Transfer Part)	Haizheng Sun
Create a Gradio Interface for User Interaction	Haizheng Sun

References

[1]A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," in Adv. Neural Inf. Process. Syst., vol. 30, 2017.

[2] S. Rao and J. Tetreault, "Dear sir or madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer," *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018. doi:10.18653/v1/n18-1012

[3] C. Raffel, N. Shazeer, A. Roberts, et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," J. Mach. Learn. Res., vol. 21, no. 1, pp. 5485-5551, 2020