

Форма рецензии статей в сборник трудов DLCP2023.

ФИО первого автора: D.N. Polyakov

Название работы: Hyper-parameter tuning of neural network for high-dimensional problems in the case of Helmholtz equation

ФИО рецензента: Mikhail A. Krinitskiy

1. Содержание работы (выбрать вариант из предложенных).

1.1 Соответствует ли статья тематике конференции

да

1.2. Отражает ли название тему работы.

да

1.3. Отражает ли аннотация основное содержание работы.

да

1.4. Сформулирована ли научно-техническая проблема, на решение которой направлена работа:

да

1.5. Представлено ли современное состояние этой проблемы (state of arts):

недостаточно четко

1.6. Сформулирована ли новизна предложенного подхода по сравнению с имеющимися и в чем его ожидаемое преимущество:

нет

1.7. Сформулирован ли результат работы, в частности, достигнут ли ожидаемый результат:

да

2. Оформление работы:

2.1. Английский язык:

удовлетворительно

2.2. Качество изложения материала, включая структуру статьи (название, аннотация, ключевые слова, введение, обзор текущего состояния, предлагаемый метод, проведенные исследования и обсуждение полученного результата, заключение, литература):

удовлетворительно

2.3. Рекомендация по объему статьи:

полная (10-12стр.)

3. Общая рекомендация:

отправить на доработку

4. Ваша уверенность в данной рекомендации. высокая

5. Комментарий для программного комитета (по желанию).

Работа актуальная, изложение довольно подробное, исследование построено методологически вполне последовательно.

Оформление следует привести к стандартам научных публикаций. Сейчас манускрипт выглядит, скорее, как черновой вариант, в спешке сверстаный в формат статьи. Больше всего этому впечатлению способствуют чересчур лаконичные и неинформативные подписи к рисункам, оформление диаграмм и таблиц.

Введение и теоретическое введение (раздел 2) написаны с некоторой небрежностью в смысле терминологии и утверждений, больше характерной для стилистики научно-популярных текстов, нежели научной статьи (см. замечания авторам).

Выводы раздела 4 приведены без интерпретации результатов и без отсылки к конкретным особенностям результатов. Некоторые выводы не подкреплены продемонстрированными результатами. Это уж точно должно быть исправлено.

После соответствующей корректировки оформления, разделов 1,2, а также выводов в разделе 4 можно рекомендовать статью к публикации.

6. Комментарий для авторов работы.

В работе рассмотрен вопрос оптимизации гиперпараметров искусственной нейронной сети, предназначенной для аппроксимации решения уравнения Гельмгольца в синтетической постановке задачи.

Проблема, рассмотренная в статье, актуальна, поскольку исследование свойств подходов оптимизации структурных и процедурных гиперпараметров нейросетевых решений – в настоящее время недостаточно полно исследованная тема.

В целом работа выполнена на приемлемом уровне, однако к некоторым положениям манускрипта есть замечания.

Весьма спорным выглядит высказывание в разделе Введение: "*The advantage of neural networks over FDM and FEM is their ability to generalize, i.e., to predict a solution in a domain outside the training sample.*" С этим сложно согласиться, поскольку, наоборот, статистические модели с высокой выразительной способностью (к коим относятся искусственные нейронные сети), напротив, по умолчанию не обладают свойством экстраполяции. То есть, за пределами распределения данных, предоставленных на обучение, качество аппроксимации зачастую существенно ниже по сравнению с областью поддержки распределения данных в тренировочной выборке. Обобщающая способность нейросетей - скорее, свойство, которого обычно следует добиваться специальными подходами. Напротив, решение ДУЧП конечно-разностным или конечно-элементным интегрированием обладает способностью к обобщению по определению этих методов, поскольку они не являются обучаемыми, и их качество не зависит ни от распределения данных в обучающей выборке, ни от наличия обучающей выборки как таковой. В случае, если авторы имели в виду какие-то специальные свойства нейросетей, в которые заложены физические законы любым их существующих способов (т.н. физически-информированные нейросети), спорное высказывание следует сформулировать более точно.

Кроме этого, общая лексика раздела Введение требует доработки в смысле менее категоричных высказываний в случаях, когда свойства нейросетей или эффект от настройки гиперпараметров описываются как нечто общепринятое, но не обоснованное строгим образом в литературе.

Например, это касается высказывания о роли гиперпараметров: "*Hyper-parameters determine the rate of convergence of the model and its maximum possible accuracy.*" Здесь следует переформулировать высказывание более аккуратно, поскольку роль гиперпараметров не ограничивается только лишь регулировкой скорости схождения оптимизации или достижимым уровнем качества. Гиперпараметрами регулируются многие свойства нейросетевых решений - от обобщающей и выразительной способности до тех, что упоминаются авторами. Кроме того, оптимизация гиперпараметров необязательно приводит к изменению скорости оптимизации.

В качестве примера также можно указать на чересчур категоричное перечисление опций для расписания шага обучения (стр. 2, раздел " 2.2. Hyper-parameters "). Вариантов расписания гораздо больше, однако высказывание сформулировано таким образом, будто других вариантов нет. Возможно, имеет смысл упомянуть, что именно в этом исследовании были рассмотрены именно такие варианты расписания LR.

Также в качестве примера можно упомянуть чересчур категоричное упоминание регуляризации как внедрения аддитивной компоненты функции потерь. Регуляризацией можно назвать любой способ ограничения выразительной способности нейросети - от внедрения компонент функции потерь до специальных подходов подачи обучающих данных алгоритму оценки градиентов и выполнения шагов градиентной оптимизации.

Стр. 3, последний абзац раздела 2.2: "*Without activation functions, the neural network turns out to be a linear regression model that is unable to represent the complex relationship between the data*". В то время как общий смысл высказывания верный, однако же, следует более аккуратно использовать терминологию. Линейная регрессия - модель, подразумевающая четыре предположения о (а) специальном параметрическом распределении (обычно нормальном, но не всегда) ошибки в порождаемых данных целевой переменной; (б) линейной связи параметра такого распределения с признаками объектов; (в) эквивариантности ошибки порождаемых данных целевой переменной и (г) идентичности распределения, порождающего эти ошибки. Из этих четырех предположений линейная нейросеть как параметрическая функция закрывает лишь одно, действительно, как композиция линейных операторов будучи эквивалентной одному линейному оператору. Остальные три предположения, закладываемые в модель линейной регрессии никак не связаны с линейным свойством преобразования линейной нейросети. Предложение следует переформулировать более строго.

Стр. 3, раздел 2.3 "*trial scheduler — stops the model training process in advance in case of poor convergence compared to previous results.*" Такое назначение расписания оптимизации гиперпараметров недостаточно точно, поскольку его назначение, в том числе, в предварительной остановке оптимизации (для некоторых из алгоритмов сэмплирования новых точек в пространстве гиперпараметров) в случае достижения нужного уровня качества до достижения количества попыток, заложенных (если они были заложены на старте оптимизации) как верхний порог итераций. Высказывание следует переформулировать более строго.

Стр. 3, раздел 2.3.1 "*The initial set of points is sampled randomly.*" и далее - следует уточнить, что за точки имеются в виду.

Стр. 5, раздел "3.4. Collocation points generation". Следует объяснить, что такое "*Collocation points*", поскольку это не является широко используемым термином в нейросетях, теории оптимизации или методах решения ДУЧП.

Рисунок 3 и подпись к нему следует скорректировать.

Подпись "*Dynamics of model training for various hyper-parameters.*" недостаточно полно и достоверно отражает содержимое диаграмм и не поясняет его.

В подписи к рисунку должно быть упомянуто, что именно отражено в диаграммах, несмотря на то, что обозначение величины, подписанной по оси ординат, описано в одном из предыдущих разделов.

Также неясно, на каких подмножествах точек оцениваются меры, отраженные на диаграммах, - на тренировочных или валидационных. Это должно быть упомянуто в подписи к рисунку или на самих панелях.

В подписи к рисунку также должно быть упомянуто, что он отражает результаты определенной стадии исследования, - чтобы у читателя не сложилось ощущение, что эти результаты применимы ко всем трем стадиям.

Переменные, упоминаемые на диаграммах (например, обозначения гиперпараметров), должны быть описаны в подписи к рисунку или должна быть ссылка на часть текста, в которой четко и последовательно описывается смысл этих переменных.

Все линии на диаграммах должны быть ясно видны, что в настоящий момент не так, например, для второй панели (с варьируемым $|N_f|$).

Все панели рисунка должны быть пронумерованы, и каждой из панелей в подписи должно быть разъяснение относительно отображаемой зависимости, тестируемого сценария, *etc.*

Согласно этим требованиям следует скорректировать и Рисунок 4 и подпись к нему.

Подписи к таблицам 4,5 также должны быть скорректированы, чтобы читателю было ясно, к какой части исследования они относятся и в каком сценарии исследования были получены приводимые результаты.

Стр. 6, раздел 4. Выводы по результатам экспертной оптимизации гиперпараметров должны быть подкреплены ссылками на конкретные диаграммы или части таблицы результатов, таким образом, чтобы было понятно, из чего именно, по мнению авторов, следует такой вывод. В настоящий момент ни один из выводов не подкреплен такой интерпретацией таблицы и диаграмм результатов (видимо, выводы построены на основании Таблицы 3 и Рисунка 3). Например, судя по диаграммам Рис. 3 сложно согласиться с выводом "*Increasing the number of collocation points ... sometimes even significantly worsens it.*"

Выводы "*Of the activation functions tested, sin outperforms tanh and greatly outperforms all other functions*" и "*The quasi-random sample generation method is superior to the random method, but the gain is small and is usually noticeable when the size of the sample is small*" не подкреплены результатами, представленными в диаграммах или таблицах. Все выводы должны быть подкреплены продемонстрированными результатами.

Основное замечание к организации исследования заключается в том, что результаты были получены без оценки неопределенностей. В частности, результатом такого подхода стали результаты, упоминаемые в разделе 5 "COMPARISON OF HPO ALGORITHMS ON THE EXAMPLE OF A 5-DIMENSIONAL PROBLEM", где упоминается, что было получено хорошее качество в одном из подходов оптимизации, который не подтверждается следующими итерациями. Это может означать неустойчивость метода оптимизации, которая может приводить к результатам с высокой неопределенностью. Эту особенность следует описать. Кроме того, следует упомянуть, что результаты оптимизации получены в подходе, не подразумевающим оценку неопределенности, что может приводить к упомянутым эффектам неустойчивого результата.

Несмотря на очевидно высокий уровень владения английским языком авторами манускрипта, в тексте замечены множественные орфографические, грамматические и стилистические недочеты. Авторам можно порекомендовать воспользоваться сервисами экспертной (эксперт-переводчик) или автоматической корректуры англоязычных текстов. Ниже приведены лишь некоторые из недочетов:

Стр. 2, раздел 2.2, "*Typical such parameters are learning rate, batch size and the optimizer*" - вероятнее всего, имелось в виду слово "Typically"

Стр. 2, раздел 2.2, "*Root mean square prop (RMSprop)*" - слово **Propagation** следует писать полностью.

Стр. 3, раздел 2.3.1, "... *avoid trapping into the local optimum*" - предположу, что уместнее использовать неопределенный артикль. В случае же, если точно известно, о каком именно локальном минимуме идет речь, как это подразумевается использованием определенного артикля, это следует описать более подробно.

Стр. 5, раздел 3.5, "*the number of neurons in the hidden layer*" - похожий недочет по использованию определенного артикля

Стр. 6, последний абзац раздела 3.5, "*Of course,*" - в научных публикациях следует избегать экспрессивных выражений, подобных "конечно же", "безусловно" и т.п.

Считаю, что после надлежащей корректировки оформления рисунков и таблиц, корректировки разделов 1,2, а также выводов в разделе 4 в соответствии с замечаниями можно рекомендовать статью к публикации.