

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра математической кибернетики

КЛАДНЕВ

Александр Сергеевич

ГЕНЕРАТОР СИМВОЛОВ
ДЛЯ ОБУЧЕНИЯ НЕЙРОСЕТЕЙ

Дипломная работа

Научный руководитель:
кандидат физико-
математических наук,
старший преподаватель
С.Е. Бухтояров

Допущен к защите

«__» _____ 2022 г.

Заведующий кафедрой математической кибернетики,
доктор физико-математических наук, профессор А. Л. Гладков

Минск, 2022

РЕФЕРАТ

Дипломная работа: 44 страниц, 5+12+7 иллюстраций, 10 источников.

Ключевые слова: глубокое обучение, искусственные нейронные сети, классификация изображений, классификация рукописных символов, MNIST, аугментации.

Объект исследования: сверточные нейронные сети; изображения классифицируемых символов.

Цель исследования: провести сравнение разных архитектур (конфигураций) искусственных нейронных сетей в задаче классификации рукописных символов; исследовать модификации обучающих данных в повышении эффективности распознавания символов.

Методы исследования: ограниченный перебор различных конфигураций нейронных сетей и выбор лучшей на основе точности, функции ошибок, затрачиваемом на обучение времени; модификация обучающих данных и анализ точности предсказаний выбранных конфигураций нейронных сетей, обученных на новой выборке обучающих данных, и выбор лучшего [для классификации рукописных символов] подхода модификации обучающих данных.

В ходе выполнения работы в задаче классификации рукописных символов были исследованы различные архитектуры и конфигурации искусственных нейронных сетей, выбраны лучшие для дальнейшего изучения влияния модификации обучающих данных. По результатам исследования некоторых техник модификации обучающих данных были выявлены лучшие.

Область применения: задача классификации символов; задача детекции символов и текста; транскрибация рукописей, манускриптов; выявление нейродегенеративных заболеваний на основе почерка.

РЭФЕРАТ

Дыпломная праца: 44 старонак, 5+12+7 ілюстрацый, 10 крыніц.

Ключавыя словы: глыбокае навучанне, штучныя нейронныя сеткі, класіфікацыя выяў, класіфікацыя рукапісных сімвалаў, MNIST, аўгментацыі.

Аб'ект даследавання: згорткавыя нейронныя сеткі; выявы сімвалаў для класіфікацыі.

Мэта даследавання: правесці параўнанне розных архітэктур (канфігурацый) штучных нейронных сетак ў задачы класіфікацыі рукапісных сімвалаў; даследаваць мадыфікацыі навучальных даных у павышэнні эфектыўнасці распазнавання сімвалаў.

Метады даследавання: абмежаваны перабор розных канфігурацый нейронных сетак і выбар лепшай на аснове дакладнасці, функцыі памылак, выдаткаваным на навучанне часу; мадыфікацыя навучальных даных і аналіз дакладнасці прадказанняў выбраных канфігурацый нейронных сетак, навучаных на новай выбарке навучальных даных, і выбар лепшага [для класіфікацыі рукапісных сімвалаў] падыхода мадыфікацыі навучальных даных.

У ходзе выканання работы у задачы класіфікацыі рукапісных сімвалаў былі даследаваны розныя архітэктурныя і канфігурацыйныя штучных нейронных сетак, выбраны лепшыя для далейшага доследу ўплыву мадыфікацыі навучальных даных. Па выніках даследавання некаторых тэхнік мадыфікацыі навучальных даных былі выяўлены найлепшыя ў задачы класіфікацыі рукапісных сімвалаў.

Вобласць прымянення: задача класіфікацыі сімвалаў; задачача дэтэкцыі сімвалаў і тэксту; транскрыбацыя рукапісаў, манускрыптаў; выяўленне нейрадэгенератыўных захворванняў на аснове змены почырку.

ABSTRACT

Graduation paper: 44 pages, 5+12+7 illustrations, 10 sources.

Keywords: deep learning, artificial neural networks, image classification, handwritten symbols classification, MNIST, augmentations.

Object of research: convolutional neural networks; images of symbols to be classified.

Purpose of research: to compare different architectures (configurations) of artificial neural networks considering the problem of handwritten symbols classification; to research train data modifications with respect to achieving better classification accuracy.

Methods of research: iterating through limited enumeration of neural networks configurations; choosing the best one with the respect to loss function, accuracy, training time; different modifications of training data & analysis of accuracy of neural networks with such configurations, trained on modified training data; choosing the best one with respect to handwritten symbols classification problem.

In the course of the paper, different architectures of artificial neural networks were researched with respect to handwritten symbols classification, the best ones were chosen to be researched with respect to training data modifications. As a result of researching several training data modifications techniques, the best ones were chosen with respect to the problem.

Area of appliance: symbols classification problem; symbols and text; transcribing handwritten papers and ancient manuscripts; neurodegenerative diseases intelligence based on person's writing.