

УДК 004.032.26

УЧЕТ СТРУКТУРЫ ТЕКСТА ПРИ РЕШЕНИИ ЗАДАЧ ОПРЕДЕЛЕНИЯ ПОЛА АВТОРА ТЕКСТА ГЛУБОКИМИ НЕЙРОННЫМИ СЕТЯМИ С ЭВОЛЮЦИОННЫМ ПОДБОРОМ ИХ КОНФИГУРАЦИИ^{1,2}

© 2020 г. А. Г. Сбоев^{1,2,*}, А. А. Селиванов¹, Р. Б. Рыбка¹, И. А. Молошников¹

¹ НИЦ “Курчатовский институт”, Москва, 123182 Россия

² Национальный исследовательский ядерный университет “МИФИ”, Москва, 115409 Россия

*e-mail: sag111@mail.ru

Поступила в редакцию 30.09.2020 г.

После доработки 30.09.2020 г.

Принята к публикации 10.11.2020 г.

В работе рассматриваются результаты исследования по влиянию структурных признаков текста на решение задачи классификации пола автора. В качестве признаков исследуется линейная структура, синтаксическая, а также структура, вычисляемая во внутренних слоях языковой модели. В статье предлагается развитие существующих методов учета синтаксических признаков, имеющих высокую вычислительную сложность и время работы за счет анализа синтаксических путей для каждого слова каждого предложения, либо последовательного анализа структур каждого из предложений. Предложенное развитие основано на использовании в рамках нейросетевой архитектуры графовых слоев внимания, принимающих на вход матрицу синтаксической связности всех слов текста. К входной матрице признаков каждого текста добавляется искусственно созданный вектор, который аккумулирует активности всех слов текста и используется для характеристики текста и его классификации. Для предложенной архитектуры сети реализован метод эволюционного подбора гиперпараметров на базе метода tree parzen estimator. Полученные результаты показывают, что учет синтаксической структуры текста для рассматриваемой задачи определения пола автора текста при расчетах на открытых корпусах текстов RusPersonality и Gender Imitation Crowdsorce “a” повышает точность на 2 и 5% соответственно по метрике f1-score со взвешенным усреднением по классам.

Ключевые слова: машинное обучение, искусственные нейронные сети, обработка естественного языка, автоматизированный анализ текстов, графовые нейронные сети, авторское профилирование, эволюционное обучение, определение пола автора текста

DOI: 10.1134/S2304487X20050132

1. ВВЕДЕНИЕ

Для ряда важных задач анализа естественного языка, таких как автороведческая экспертиза, маркетинговые исследования, изучение социальной обстановки с учетом социальной стратификации, эффективное решение задачи определения пола и возраста автора весьма актуально. В предыдущих работах [1–3] показаны успешные алгоритмы ее решения, преимущественно основанные на глубоких нейронных сетях. Прогресс последних лет дает возможность получения высокоуровневого представления текста в виде струк-

тур, примерами которых являются: статистические характеристики текста, например частоты биграмм слов; информация о структуре языка из предобученных на больших корпусах языковых моделей; синтаксические особенности предложений. Применение таких признаков для задач классификации текстов изучено слабо, особенно в контексте русского языка с синтаксическими конструкциями в виде деревьев зависимости. В связи с этим целью данной работы является разработка и исследование подходов, учитывающих расширение признакового пространства за счет введения структурных признаков, на примере задачи определения пола автора русскоязычного текста, где структура выражена синтаксическими деревьями. В качестве методов используется аппарат глубоких нейронных сетей, для которых число свободно настраиваемых параметров, заданных топологией модели, является важным

¹ Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-29-10084 “мк”.

² Работа была выполнена с использованием оборудования центра коллективного пользования “Комплекс моделирования и обработки данных исследовательских установок мега-класса” НИЦ “Курчатовский институт”, <http://ckp.nrcki.ru/>.

фактором, влияющим на процесс обучения и его результат. В условиях обширной области возможных комбинаций этих параметров представляется эффективным использование единого набора алгоритмов, которые позволяют автоматизированно подобрать их комбинацию, наиболее эффективную для рассматриваемой задачи. Проводимое в работе исследование также показывает эффективность применения алгоритма выбора комбинации параметров на основе сравнения метода *tree-structured parzen estimator* [4, 5] по отношению к ручному способу настройки.

2. ПРЕДЫДУЩИЕ РАБОТЫ

Описанные в литературе решения автоматического определения признаков авторского профиля, в частности пола, в основном опираются на применение нейросетевых моделей гибридных топологий, например топология на основе совокупности свёрточных и рекуррентных слоев [1]. Наши предыдущие исследования показали, что особенно актуальным является создание контекстно независимых моделей, которые не используют в явном виде информацию о словоформах, т.к. корпуса достаточного объема, покрывающего в равной степени все тематики с метками пола автором, пока не создано. В связи с этим в данной работе для создания подходов, учитывающих структурные признаки текста, будут использоваться морфологические признаки отдельных слов и синтаксические связи между словами в предложении.

Учет структурных особенностей анализируемых объектов с помощью нейронных сетей производится на основе графовых нейросетевых моделей. В части задач текстового анализа наиболее релевантными для задач исследования, проводимого в данной работе, являются методы, которые учитывают синтаксическую структуру: *TreeLSTM* [6], *TreeLSTM* с механизмом внимания [7].

У данного подхода существует несколько недостатков: во-первых, структура в виде синтаксических деревьев — лишь один из вариантов представления структурной информации в тексте, при этом другие варианты могут быть выражены не ациклическими графами. Помимо этого, каждому входному примеру в данных моделях должно соответствовать одно дерево — в то время, как текст может содержать несколько предложений (а значит несколько деревьев синтаксической структуры).

Существуют модели, которые работают с последовательностью синтаксических деревьев для того, чтобы получить векторное представление текстов: *Sequence Graph LSTM* [8], *Sequence Graph Attention with biLSTM* [9]. В работах [8, 9] показана перспективность этих методов для ре-

шения задачи определения пола. В подходе *Sequence Graph LSTM* учет синтаксических структур осуществляется за счет анализа совокупности путей по синтаксическому дереву для каждого слова в тексте, что увеличивает вычислительную сложность и время работы алгоритма. В основе *Sequence Graph Attention with biLSTM* для учета синтаксических структур используются слои типа *Graph Attention (GAT)* [10] с последующим усреднением векторов признаков слов в каждом предложении. Для анализа последовательности предложений используется *BiLSTM*-слой. Такая процедура также является вычислительно затратной. В данной работе проводится развитие подхода на базе метода *GAT* с устранением необходимости последовательного анализа предложений за счет создания матрицы синтаксической связности всего текста.

В виду сильного влияния значений гиперпараметров, определяющих точность получаемых в результате обучения моделей, предусмотрена процедура их автоматического выбора. Существующие алгоритмы [11–13] опираются в большей степени на построение архитектуры нейронной сети на основе перебора (и поиска) отдельных блоков топологии и связей между ними. Таким образом на каждой итерации работы алгоритмов производятся вычисления для нескольких вариантов сети из структурных элементов разных типов, что обладает значительной вычислительной сложностью и следовательно требует больших затрат времени. В то же время в ряде случаев, и в частности в нашем, топология сети уже представляется известной, поэтому альтернативой является подбор гиперпараметров для известной архитектуры сети на основе метода *tree-structured parzen estimator* [4, 5], реализованного в библиотеке *Hyperopt*. В статье [14] показано, что поиск на основе данного алгоритма позволяет получить большую точность, чем ручной подбор, за меньшее число итераций, чем алгоритмы случайного поиска (*Random Search*) и перебора (*Grid Search*).

3. МЕТОДЫ И ПОДХОДЫ

3.1. *Stacked Graph Attention Network*

Данная модель представляет собой последовательность графовых свёрточных слоев с механизмом внимания. К входной матрице признаков для каждого текста добавляется искусственно созданный вектор, который аккумулирует активности всех элементов входящей последовательности слов и в дальнейшем используется для характеристики всего текста и его классификации. Данный подход используется в архитектуре *Transformer* [15], на основе которой построены контекстуальные языковые модели типа *BERT* [16].

Для данной архитектуры пространство поиска содержало следующие гиперпараметры:

- число блоков attention;
- число нейронов в блоке attention;
- число нейронов в слое graph attention;
- число блоков graph attention;
- коэффициент dropout [17];
- использовать ли дополнение векторов признаков позиционным кодированием;
- использовать ли сквозные связи (residual) в блоках graph attention [18];
- размер батча.

Помимо исходной модели, рассматривалась ее модификация Stacked Graph Attention Network with LSTM.

Graph Attention Network with LSTM создает представления текста на основе обработки последовательности векторов признаков узлов с помощью слоя LSTM.

Для данной архитектуры дополнительно в пространство поиска были включены следующие параметры:

- число нейронов слоя LSTM.

3.2. Convolutional LSTM

Данная топология описана в статье [1]. Топология включает совокупность свёрточных слоев и слой LSTM, что позволяет учитывать линейную структуру текста. Опубликованная точность данной модели используется в качестве базы для сравнения, а также для демонстрации эффекта от подбора гиперпараметров эволюционным методом.

Для данной топологии проводился поиск комбинации следующих параметров:

- число нейронов свёрточных слоев;
- число свёрточных слоев;
- размер окна свёртки;
- размер окна пулинга;
- число нейронов слоя LSTM;
- коэффициент dropout [Dropout];
- размер батча.

3.3. Алгоритм поиска комбинации параметров сети tree parzen estimator

Алгоритм включает в себя следующие шаги:

- 1) определение пространства поиска для каждого из гиперпараметров рассматриваемой архитектуры сети;
- 2) определение целевой функции, которая будет оптимизироваться в процессе поиска значений параметров;

3) проведение расчетов для нескольких случайно выбранных комбинаций параметров;

4) сортировка полученных результатов на основе значений целевой функции и разделение по порогу на две группы: та, которая содержит лучшие результаты ($\times 1$), и все остальные ($\times 2$);

5) расчет плотностей распределения $l(\times 1)$ и $g(\times 2)$ на основе ядерной оценки плотности (метода окна Парзена–Розенблатта);

6) случайный выбор множества конфигураций из распределения $l(\times 1)$, оценка их с точки зрения $l(\times 1)/g(\times 2)$, выбор конфигурации, которая соответствует наибольшему ожидаемому улучшению значения целевой функции, расчет целевой функции на основе текущей конфигурации;

7) добавление новых результатов в список результатов из пункта 3;

8) шаги 4–7 повторяется фиксированное число итераций или пока не будет достигнут лимит по времени.

3.4. Данные

Рассматриваемые подходы были проверены на двух наборах данных.

1) RusPersonality. Это представительный и валидированный лингвистами набор текстов с разметкой пола, возраста, стиля и других автороведческих параметров. Корпус содержит 1549 текстов-эссе по двум темам: “письмо другу” и “описание картины”. Из них 575 текстов, где авторы мужского пола, и 974 – женского. Тексты RusPersonality были предварительно сбалансированы по классам, итоговый размер выборки составил 1150 текстов.

2) Gender imitation crowdsorce “a” (GI cs “a”) – корпус текстов, содержащий различную информацию об авторах. Собран средствами краудсорсинга с использованием заданий, составленных. Общее число текстов в корпусе GI cs – 5150. В данной работе мы использовали часть “a” из 1716 текстов с информацией о поле авторов. Тексты GI cs “a” были предварительно сбалансированы по классам, итоговый размер выборки составил 1664 текста.

Для получения векторов морфологических признаков и синтаксических деревьев предложений был использован автоматический разборщик UDPipe [19], который имеет наибольшую точность построения синтаксических деревьев для русского языка (~87.6% по метрике UAS на корпусе SynTagRus).

Для использования в рамках архитектуры graph attention синтаксические деревья предложений каждого текста объединялись посредством соединения верхнего виртуального узла предложения с виртуальным узлом всего текста. Таким

Таблица 1. Результаты вычислительных экспериментов без эволюционного подбора гиперпараметров

Модель	Использование структур	RusPersonality	GI cs “a”
ConvLSTM [1]	Линейная структура*	81 ± 4	77 ± 4
Sequential GraphLSTM [8]	Синтаксическая	83 ± 2	83 ± 2
Sequential GAT biLSTM [9]	Синтаксическая	83 ± 2	80 ± 1
BERT (SotA) [16]	Вычисляется в слоях внимания**	85 ± 2	86 ± 1

*ConvLSTM [1] косвенно учитывает связи в последовательностях слов, благодаря использованию LSTM. **Входные данные сети не содержат структурных признаков, но сформированные в процессе обучения веса сети и рассчитываемые активности могут быть интерпретированы, как связи между входными векторами, то есть как извлекаемая в процессе работы сети структура входных данных [20].

образом получалась матрица смежности для текста.

4. ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

Были проведены следующие серии экспериментов с моделями Stacked Graph Attention Network и Convolutional LSTM, не имеющей структурных признаков, с использованием алгоритма эволюционного обучения, а также с BERT. Это позволяет оценить эффективность настройки гиперпараметров в эксперименте с помощью итеративного метода tree-structured parzen estimator [4, 5] по отношению к ручному способу настройки, а также различных способов описания структурных признаков.

Результаты вычислительных экспериментов были получены с использованием стратифицированной кросс-валидации (Stratified Shuffle Split) с разделением исходной выборки на пять частей. На каждой итерации кросс-валидации одна часть принималась за тестировочную выборку, а оставшиеся четыре разделялись на тренировочную и валидационную. Таким образом, на каждой итерации кросс-валидации было получено следующее разбиение: 72% – тренировочная выборка, 8% – валидационная выборка, 20% – тестировочная выборка.

При этом тексты тренировочной и тестировочной выборок принадлежали разным авторам. Разбиения данных для разных экспериментов фиксировались, чтобы сделать результаты сравнимыми.

Поиск комбинации параметров в сети производился на основе первой части кросс-валидации без использования тестовых данных. В качестве метрики для оптимизации используется f1-мера с взвешенным усреднением по классам (f1-score weighted), средняя по трем запускам модели.

Алгоритм поиска имел следующие параметры:

- 200 итераций алгоритма (или трое суток работы);
- 100 эпох обучения каждого варианта нейронной сети;

- ранний останов сети: в случае, если в течение 15 эпох значение метрики на валидационном множестве не улучшалось, процесс обучения заканчивался и загружались веса сети на момент с наилучшим значением метрики.

В качестве исходного представления данных были использованы бинарные вектора морфологических признаков (размерность – 58).

Языковая контекстная модель BERT была предварительно обучена на мультязычном корпусе ресурса wikipedia (все тексты для 100 языков). Размерность выходного векторного представления – 768.

В табл. 1 сопоставляются результаты вычислительных экспериментов на корпусах RusPersonality и GI cs “a” без использования эволюционного подбора гиперпараметров, включая как результаты этой работы, так и результаты предыдущих работ.

В табл. 2 приводятся точности для моделей, топология которых основана на подобранных на основе эволюционного алгоритма комбинациях гиперпараметров нейронной сети. Все представленные результаты получены с использованием ансамбля из трех моделей с подобранной топологией, обучаемых независимо.

По данным табл. 1 можно сделать вывод, что учет синтаксической структуры текста для рассматриваемой задачи в определенной степени повышает точность определения пола автора текста, хотя эффект этот маскируется погрешностью ссылочных данных ConvLSTM [1]. Сравнение с контекстуальной языковой моделью BERT, предварительно обученной на текстах ресурса “википедия” на ста разных языках и учитывающей семантику языка, демонстрирует, что разработанные архитектуры Convolutional LSTM и Stacked Graph Attention Network достигают точностей, сопоставимых с BERT, используя меньшее признаковое пространство без ресурсоемкого предварительного обучения на больших корпусах данных.

Из данных табл. 2 следует, что автоматизированная настройка гиперпараметров сетей не сильно изменяет результаты сопоставления расчетов, хотя для ConvLSTM [1] GI_cs “a” можно

Таблица 2. Результаты вычислительных экспериментов с эволюционным подбором гиперпараметров

Модель	Использование синтаксических структур	RusPersonality	GI cs “a”
ConvLSTM [1]	нет	81 ± 3	80 ± 1.5
Stacked Graph Attention Network	да	83 ± 2	82 ± 2
Stacked Graph Attention Network with LSTM	да	82 ± 2	82 ± 1

говорить о выигрыше в точности. В перспективе планируется проведение анализа точности Convolutional LSTM и Graph Attention Network на других текстовых задачах.

5. ЗАКЛЮЧЕНИЕ

В статье рассмотрены нейросетевые подходы учета структур текстов при решении задачи определения пола автора, как частный случай задачи авторского профилирования. Предложено развитие существующих методов работы со структурными признаками текста, выраженными в виде синтаксических деревьев зависимости, с анализом на базе Graph Attention слоев нейронной сети с различными выходными компонентами. Проведенные исследования выбранных подходов с применением метода эволюционного подбора параметров показывают перспективность их использования для задачи определения пола автора с точностью сравнимой со сложными моделями нейросетей, учитывающих контексты и обученных на больших размеченных корпусах по принципу языковых моделей. Планируется дальнейшее исследование методов, учитывающих структурные особенности текстов, для других задач текстовой классификации.

СПИСОК ЛИТЕРАТУРЫ

1. *Sboev A., Moloshnikov I., Gudovskikh D., Selivanov A., Rybka R., Litvinova T.* Deep learning neural nets versus traditional machine learning in gender identification of authors of RusProfiling texts // *Procedia Comput. Sci.* 2018, V. 123. P. 424–431.
2. *Sboev A., Moloshnikov I., Gudovskikh D., Selivanov A., Rybka R., Litvinova T.* Automatic gender identification of author of Russian text by machine learning and neural net algorithms in case of gender deception // *Procedia Computer Science.* 2018. V. 123. P. 417–423, ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.064>
3. *Sboev A., Moloshnikov I., Gudovskikh D., Rybka R.* A comparison of Data Driven models of solving the task of gender identification of author in Russian language texts for cases without and with the gender deception // *IOP Publishing, Journal of Physics: Conf. Ser.* 2017. V. 937. P. 1742–6588.
4. *Bergstra J.S., Yamins D., Cox D.* Algorithms for hyperparameter optimization // *Advances in Neural Information Processing Systems.* 2011.
5. *Bergstra J.S., Yamins D., Cox D.* Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures / *Proceedings of the 30th International Conference on Machine Learning.* 2013.
6. *Tai K.S., Socher R., Manning C.D.* Improved semantic representations from tree-structured long short-term memory networks. *arXiv preprint arXiv:1503.00075*, 2015.
7. *Miyazaki R., Komachi M.* Japanese sentiment classification using a tree-structured long short-term memory with attention. *arXiv preprint arXiv:1704.00924.* 2017.
8. *Sboev A., Selivanov A., Rybka R., Moloshnikov I., Bogachev D.* A Neural Network Model to Include Textual Dependency Tree Structure in Gender Classification of Russian Text Author. In: *Advanced Technologies in Robotics and Intelligent Systems. Mechanisms and Machine Science.* Springer, Cham, 2020. V. 80. https://doi.org/10.1007/978-3-030-33491-8_48
9. *Sboev A., Bogachev D., Selivanov A., Moloshnikov I., Rybka R.* Graph Convolution Network with Attention to Include Syntax Trees into Text Author’s Gender Identification Task (in press).
10. *Veličković P., Cucurull G., Casanova A., Romero A., Lio P., Bengio Y.* Graph Attention Networks. 2018. Published as a conference paper at ICLR 2018.
11. *Hanfeng Jin, Qingquan Song, Xia Hu.* Auto-keras: An efficient neural architecture search system / *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* ACM, 2019.
12. *Weill C., Gonzalvo J., Kuznetsov V., Yang S., Yak S., Mazzawi H., Hotaj E., Jerfel G., Macko V., Adlam B., Mohri M.* AdaNet: A Scalable and Flexible Framework for Automatically Learning Ensembles. *arXiv preprint arXiv:1905.00080.* 2019.
13. *Miikkulainen R., Liang J., Meyerson E., Rawal A., Fink D., Francon O., Raju B., Shahrzad H., Navruzyan A., Duffy N., Hodjat B.* Evolving deep neural networks. In: *Artificial Intelligence in the Age of Neural Networks and Brain Computing.* Academic Press, 2019. P. 293–312.
14. *James B., Yamins D., Cox D.* Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures / *Proceedings of the 30th International Conference on Machine Learning.* 2013.
15. *Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I.* Attention is all you need / In: *Advances in Neural Information Processing Systems.* 2017. P. 5998–6008.

16. Devlin J., Chang M.W., Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018.
17. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting // Journal of Machine Learning Research. 2014. V. 15. № 1. P. 1929–1958.
18. Mosella-Montoro A., Ruiz-Hidalgo J. Residual Attention Graph Convolutional Network for Geometric 3D Scene Classification / In: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019.
19. Straka M., Straková J. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe / Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 2017. P. 88–99.
20. Vig J. A multiscale visualization of attention in the transformer model. arXiv preprint arXiv:1906.05714. 2019. no. 12.

Vestnik Nacional'nogo Issledovatel'skogo Yadernogo Universiteta "MIFI", 2020, vol. 9, no. 6, pp. 554–560

Evolutionary Configuration Selection of Deep Neural Networks Accounting Text Structures in a Task of Author's Gender Profiling

A. G. Sboev^{a,b,#}, A. A. Selivanov^a, I. A. Moloshnikov^a, and R. B. Rybka^a

^a National Research Center Kurchatov Institute, Moscow, 123182 Russia

^b National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, 115409 Russia

[#]e-mail: sag111@mail.ru

Received September 30, 2020; revised September 30, 2020; accepted November 10, 2020

Abstract—The influence of structural features of the text on the author gender classification has been studied. Considered structural features include a linear structure, a syntactic structure, and a structure calculated in the hidden layers of the language model. Existing syntax-accounting methods that have high computational complexity and working time have been developed by analyzing syntactic paths for each word of each sentence, or sequential analysis of sentences structures. The proposed development is based on the use of attention graph layers (GAT) within the neural network architecture, whose input is the matrix of syntax connectivity of all words of the text. An artificially created vector is added to the input feature matrix of each text, which accumulates the activities of all words in the text and is used to characterize the text and classify it. For the proposed network architecture, the method is implemented for evolutionary selection of hyperparameters based on the tree parzen estimator. The results obtained show that the syntax structure of the text for the considered task of author's gender identification on the open corpora RusPersonality and Gender Imitation Crowdsourcе “a” increases the accuracy by 2 and 5%, respectively, according to the f1-score metric with weighted averaging over classes.

Keywords: machine learning, artificial neural networks, natural language processing, automated text analysis, graph neural networks, author profiling, evolution learning, author gender identification

DOI: 10.1134/S2304487X20050132

REFERENCES

1. Sboev A., Moloshnikov I., Gudovskikh D., Selivanov A., Rybka R., Litvinova T., Deep learning neural nets versus traditional machine learning in gender identification of authors of RusProfiling texts, *Procedia Comput. Sci.*, 2018, vol. 123, pp. 424–431.
2. Sboev A., Moloshnikov I., Gudovskikh D., Selivanov A., Rybka R., Litvinova T., Automatic gender identification of author of Russian text by machine learning and neural net algorithms in case of gender deception, *Procedia Computer Science*, 2018, vol. 123, pp. 417–423, ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.01.064>.
3. Sboev A., Moloshnikov I., Gudovskikh D., Rybka R. A comparison of Data Driven models of solving the task of gender identification of author in Russian language texts for cases without and with the gender deception, *IOP Publishing, Journal of Physics: Conf. Ser.*, 2017, vol. 937, pp. 1742–6588.
4. Bergstra J.S., Yamins D., Cox D., Algorithms for hyper-parameter optimization, *Advances in Neural Information Processing Systems*, 2011.
5. Bergstra J.S., Yamins D., Cox D., Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures, *Proceedings of the 30th International Conference on Machine Learning*, 2013.

6. Tai K.S., Socher R., Manning C.D., Improved semantic representations from tree-structured long short-term memory networks. arXiv preprint arXiv:1503.00075, 2015.
7. Miyazaki R., Komachi M., Japanese sentiment classification using a tree-structured long short-term memory with attention, arXiv preprint arXiv:1704.00924. 2017.
8. Sboev A., Selivanov A., Rybka R., Moloshnikov I., Bogachev D., A Neural Network Model to Include Textual Dependency Tree Structure in Gender Classification of Russian Text Author, in *Advanced Technologies in Robotics and Intelligent Systems, Mechanisms and Machine Science*, Springer, Cham, 2020, vol. 80. https://doi.org/10.1007/978-3-030-33491-8_48.
9. Sboev A., Bogachev D., Selivanov A., Moloshnikov I., Rybka R., Graph Convolution Network with Attention to Include Syntax Trees into Text Author's Gender Identification Task (in press).
10. Veličković P., Cucurull G., Casanova A., Romero A., Lio P., Bengio Y., Graph Attention Networks, 2018, Published as a conference paper at ICLR 2018.
11. Haifeng Jin, Qingquan Song, Xia Hu, Auto-keras: An efficient neural architecture search system, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, 2019.
12. Weill C., Gonzalvo J., Kuznetsov V., Yang S., Yak S., Mazzawi H., Hotaj E., Jerfel G., Macko V., Adlam B., Mohri M., AdaNet: A Scalable and Flexible Framework for Automatically Learning Ensembles. arXiv preprint arXiv:1905.00080. 2019.
13. Miikkulainen R., Liang J., Meyerson E., Rawal A., Fink D., Francon O., Raju B., Shahrzad H., Navruzyan A., Duffy N., Hodjat B., Evolving deep neural networks, In *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, Academic Press, 2019, pp. 293–312.
14. James B., Yamins D., Cox D., Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures, *Proceedings of the 30th International Conference on Machine Learning*, 2013.
15. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I., Attention is all you need, in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.
16. Devlin J., Chang M.W., Lee K., Toutanova K., Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805. 2018.
17. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., Dropout: a simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, 2014 vol. 15, no. 1, pp. 1929–1958.
18. Mosella-Montoro A., Ruiz-Hidalgo J., Residual Attention Graph Convolutional Network for Geometric 3D Scene Classification, in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019.
19. Straka M., Straková J., Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe, *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 2017, pp. 88–99.
20. Vig J., A multiscale visualization of attention in the transformer model, arXiv preprint arXiv:1906.05714. 2019. no. 12.