

Разработка рекомендательной системы для подбора и конфигурирования глубоких нейросетевых моделей

С.А. Ямашкин, Е.О. Ямашкина

Мордовский государственный университет им. Н. П. Огарева, Саранск

Аннотация: В данной работе описана разработка рекомендательной системы для подбора и конфигурирования глубоких нейросетевых моделей. Ключевой областью ее работы является оценка эффективности применения нейронной сети для конкретного типа задач. Спроектированная в рамках репозитория рекомендательная система позволяет обеспечивать точный поиск, подбор и гибкую настройку нейронных сетей, в зависимости от конкретных проектно-ориентированных задач.

Ключевые слова: рекомендательная система, машинное обучение, глубокое обучение, нейросетевая модель, экспертная система, нейронная сеть, репозиторий, пространственные данные, геоинформационная система, конфигурирование, архитектура.

Введение

В настоящее время особую актуальность приобрели рекомендательные (экспертные) системы, основанные на знаниях. Такие системы играют важную роль в принятии управленческих решений в области устойчивого развития, мониторинга и управления окружающей средой, планирования землепользования, навигации и т. д.) [1]. Применение методов и принципов искусственного интеллекта в анализе пространственных данных обеспечивает большой потенциал для решения задач разработки интеллектуальных геоинформационных систем, самыми важными компонентами которых являются рекомендательные системы и глубокие нейронные сети [2,3].

Настоящее исследование посвящено описанию разработки рекомендательной системы в рамках реализации репозитория глубоких нейросетевых моделей для анализа пространственных данных. Разработка репозитория основывается на онтологической модели хранилища, которая определяет принципы систематизации глубоких моделей анализа пространственных данных по классам решаемых задач, природе и

размерности анализируемых данных, архитектуре и топологии, свойствам эффективности.

Методология и методы исследования

В рамках создания репозитория предлагается подход к обучению моделей с помощью рекомендательной системы. Ключевой областью ее работы будет оценка эффективности применения нейронной сети для конкретного типа задач. Данные из рекомендательной системы будут использоваться в качестве входных и выходных данных для нейронной сети аналогично собранным данным для определенной прикладной области. Модель репозитория нейронных сетей декомпозируется на домены моделей, задач и анализируемых данных [4]. Измеряемыми параметрами в исследованиях являются пространственно-временные данные. Задачи классификации пространственных данных с помощью нейросетевых моделей находят все большее применение в экономике, в сельском и лесном хозяйстве, в отслеживании и контроле чрезвычайных ситуаций (лесных пожаров, наводнений, природно-техногенных катастроф) и т.д. После формирования рекомендательной системы и предоставления входных данных для модели нейронной сети, полученные данные можно анализировать так же, как анализируются данные результатов других экспериментов в области анализа пространственных данных с помощью нейронных сетей.

Важной функцией репозитория глубоких нейросетевых моделей является разделение функционала для пользователей в зависимости от их роли в системе и предоставление соответствующих веб-интерфейсов с информацией о глубоких нейросетевых моделях (архитектуры, показателей эффективности, рекомендаций по настройке гиперпараметров) и примерах прикладного использования [5].

Модель распределения прав доступа пользователей в рамках репозитория нейронных сетей представлена на рис. 1. Рекомендательная система для выбора модели позволяет обеспечивать релевантный поиск, подбор и конфигурирование глубоких нейронных сетей, тонкую настройку моделей для решения конкретных проектных задач.

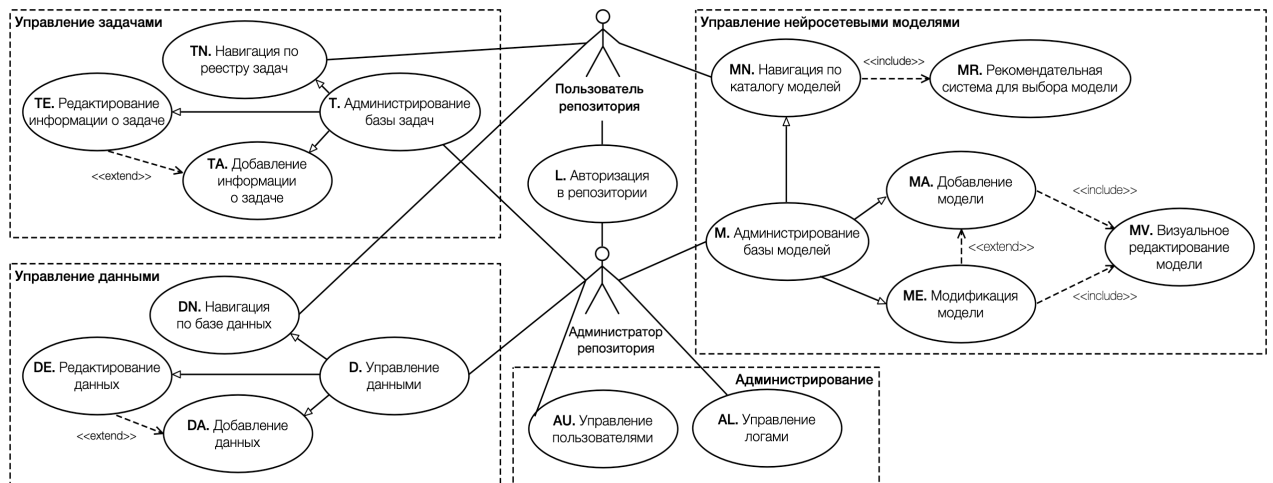


Рис. 1 – Функции рекомендательной системы. Распределение прав доступа пользователей

Рассмотрим работу рекомендательной системы в рамках репозитория нейронных сетей на примере задачи распознавания стихийных бедствий на основе данных со спутниковых изображений. В настоящее время большая часть работ по картированию стихийных бедствий выполняется вручную, что отнимает много времени и часто приводит к ошибочным результатам. Чтобы решить эти проблемы, предлагается использование сверточных нейронных сетей (CNN) для анализа данных со спутниковых изображений, которые затем будут использованы для поиска областей территорий, наиболее пострадавших от стихийных бедствий. Сверточные нейронные сети (CNN) чаще всего используются для анализа визуализированных данных и решения задач классификации изображений [6]. Основное преимущество CNN

заключается в получении наиболее полного анализа и использовании методов цифровой обработки изображений.

В рамках эксперимента по выявлению лучшей архитектуры была обучена модель, объединяющая два набора данных со спутниковых изображений высокого разрешения «Ресурс», «Landsat», «Sentinel» [7]. Далее опишем архитектуру обычной сверточной сети (CNN) и ее расширенных модификаций, таких как AlexNet, ResNet, DenseNet и MobileNet [8,9].

1) Сверточная нейронная сеть (CNN) является наиболее распространенной моделью нейронной сети, используемой для задач классификации пространственных данных [10]. Архитектура CNN состоит из чередующихся сверточных слоев и слоев объединения, за которыми следуют полносвязные слои для генерации выходных данных и слои для изучения каждого пикселя. CNN уменьшает количество свободных параметров и позволяет увеличивать глубину сети за счет использования операций свертки. Преимущества применения сверточных нейронных сетей для классификации изображений заключаются в том, что при анализе требуется меньше времени на предварительную обработку, снижаются вычислительные затраты и вероятность переобучения модели.

2) Архитектура AlexNet считается одной из самых эффективных архитектур в задачах компьютерного зрения за счет достижения практически 50-процентного снижения частоты ошибок. Основные улучшения AlexNet заключаются в реализации функции активации ReLU (Rectified Linear Units) и Dropout layers (исключения слоев) в сетевой архитектуре. Исключение слоев еще эффективнее снижает вероятность переобучения, предполагая, что хорошо выполненная модель должна обеспечивать хорошие классификации материалов космической съемки, даже если некоторые случайные активации выпадают.

3) Архитектура ResNet решает проблему снижения точности классификации, но при этом увеличивается количество слоев за счет использования блоков идентичности. Суммирование признаков, полученных из предыдущих слоев, повышает точность сети.

4) Архитектура DenseNet, подобно ResNet, также использует кратчайшие связи в сетевой структуре. Основное фундаментальное отличие состоит в том, что DenseNet использует объединенные карты объектов из всех предыдущих слоев, а не суммирование, как в ResNet. Преимущества DenseNet включают использование меньшего количества параметров для обучения и снижение вычислительных затрат. Например, сеть ResNet со 101 слоем может достичь такой же точности, как и DenseNet с 201 слоем. Однако DenseNet можно обучить почти в два раза быстрее, чем ResNet.

5) Архитектура MobileNet. Мобильные устройства составляют большую долю рынка для моделей глубокого обучения. Из-за компромисса между количеством уровней с точки зрения точности и стоимости памяти MobileNet стала популярной для развертывания на мобильном оборудовании. Основная идея MobileNet заключается в использовании разделяемых по глубине сверток вместо точечных сверток, как в других моделях CNN. В MobileNet применяется пакетная нормализация и функция ReLU после каждой свертки.

Результаты исследования

Спроектированная в рамках репозитория нейросетевых моделей рекомендательная система позволяет обеспечивать точный поиск, подбор и гибкую настройку нейронных сетей в зависимости от конкретных проектно-ориентированных задач.

Для проведения эксперимента по выявлению лучшей архитектуры был взят набор пространственных данных, содержащий 2000 образцов, который используется для обучения и тестирования каждой модели. Набор данных

представляют собой случайно выбранные изображения, где половина образцов помечена как «стихийное бедствие» на примере обнаружения наводнений, а другая половина помечена как «другие виды повреждений» и «отсутствие повреждений». Целью обучения и тестирования различных моделей с таким набором данных является предоставление двоичного классификатора для классификации того, содержит ли изображение признаки наводнения или нет. Выборку разделили на 80% образцов (1600 изображений) для обучения и 20% образцов (400 изображений) для тестирования. Количество эпох при обучении каждой модели – 30.

В результате эксперимента были получены параметры оценки точности для разных классификаторов. Обычная 4-слойная модель CNN имеет наибольший размер и наименьшую точность. Модель ResNet достигает лучшей точности (78%) и занимает относительно небольшой объем памяти. По мере увеличения количества слоев точность и размер модели также увеличивается. DenseNet обеспечивает удовлетворительную точность 75%, а также сохраняет относительно небольшой размер. AlexNet, считающаяся одной из самых эффективных моделей в Computer Vision, также имеет точность 75%, но с большим размером. Напротив, MobileNet, хотя и не имеет выдающихся показателей точности, имеет наименьший размер всего 17 мегабайт, что свидетельствует о потенциале данной архитектуры для развертывания на мобильных устройствах, встроенных системах и веб-серверах. Делаем вывод, что для задачи классификации наводнений самую высокую точность обеспечивает модель ResNet. Полученный результат будет учитываться рекомендательной системой для решения класса похожих задач классификации.

Благодарности

Работа выполнена при финансовой поддержке гранта Президента Российской Федерации (грант № МК-199.2021.1.6).

Литература

1. Касимов Н. С., Мазуров Ю. Л., Тикуннов В. С. Концепция устойчивого развития: восприятие в России // Вестник Российской академии наук. 2004. Т. 74. № 1. С. 28-36.
 2. Kahn A. B. Topological sorting of large networks Communications of the ACM. 1962. Vol. 5. № 11. pp. 558-562.
 3. Fanca A., Puscasiu A., Gota D., Valean H. Recommendation Systems with Machine Learning. 2020 21th International Carpathian Control Conference (ICCC). IEEE, 2020. DOI: 10.1109/iccc49264.2020.9257290.\
 4. Ямашкин, С. А. Ямашкина Е. О., Ямашкин А. А. Разработка прикладного программного интерфейса для обеспечения обмена данными с репозиторием нейросетевых моделей // Современные наукоемкие технологии. 2022. № 5-2. С. 226-231.
 5. Ямашкин, С. А., Ямашкина Е. О. Трансляция нейросетевых моделей в программный код на языке программирования высокого уровня // Инженерный вестник Дона. 2022. № 6. URL: ivdon.ru/ru/magazine/archive/n6y2022/774.
 6. Лиля В. Б. Алгоритм и программная реализация адаптивного метода обучения искусственных нейронных сетей // Инженерный вестник Дона. 2012. № 1. URL: ivdon.ru/magazine/archive/n1y2012/626.
 7. Ямашкин С. А., Ямашкин А. А., Ямашкина Е. О., Занозин В. В. Интеграция знаний в цифровых инфраструктурах пространственных данных // Саранск: Национальный исследовательский Мордовский государственный университет им. Н.П. Огарёва, 2021. 216 с.
 8. Bengio Y. Learning deep architectures for AI // Foundations and Trends in Machine Learning. 2009. vol. 2. № 1. pp. 1-127.
 9. Шолле Ф. Глубокое обучение на Python. – СПб.: Питер, 2018. 400 с.
-

10. LeCun Y., Bengio Y., Hinton G. Deep Learning // Nature. 2015. № 521. 436-444.

References

1. Kasimov N. S., Mazurov Ju. L., Tikunov V. S. Vestnik Rossijskoj akademii nauk. 2004. T. 74. № 1. pp. 28-36.

2. Kahn A. B. Topological sorting of large networks Communications of the ACM. 1962. Vol. 5. № 11. pp. 558-562.

3. Fanca A., Puscasiu A., Gota D., Valean H. 2020 21th International Carpathian Control Conference (ICCC). IEEE, 2020. DOI: 10.1109/iccc49264.2020.9257290.

4. Yamashkin S. A., Yamashkina E. O., Yamashkin A. A. Sovremennye naukoemkie tehnologii. 2022. № 5-2. pp. 226-231.

5. Yamashkin S. A., Yamashkina E.O. Inzhenernyj vestnik Dona, 2012. № 6. URL: ivdon.ru/ru/magazine/archive/n6y2022/774.

6. Lila V. B. Inzhenernyj vestnik Dona. 2012. № 1. URL: ivdon.ru/magazine/archive/n1y2012/626.

7. Yamashkin S. A., Yamashkin A. A., Yamashkina E. O., Zanozin V. V. Saransk: Nacional'nyj issledovatel'skij Mordovskij gosudarstvennyj universitet im. N.P. Ogareva, 2021. 216 p.

8. Bengio Y. Foundations and Trends in Machine Learning. 2009. vol. 2. № 1. pp. 1-127.

9. Sholle F. Glubokoe obuchenie na Python. [Deep Learning in Python]. SPb: Piter, 2018. 400 p.

10. LeCun Y., Bengio Y., Hinton G. Deep Learning // Nature. 2015. № 521. pp. 436-444.