

АНАЛИЗ ВЕДУЩИХ ПОКАЗАТЕЛЕЙ НЕЙРОННОЙ СЕТИ ПО РАСПОЗНАВАНИЮ ОБРАЗОВ НА КОСМО И АЭРОФОТОСНИМКАХ ПРИ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ ГЕОПОЛЕЙ

Бирюков Н.А., Полтавский А.В.

Московский государственный университет геодезии и картографии, Москва, Россия
4096414@gmail.com, avp57avp@yandex.ru

Аннотация. Дается краткий анализ организации алгоритмов распознавания образов и рассматриваются основные группы для ведущих показателей при организации машинного обучения с использованием современных платформ, основанных на методах нейросетевых технологий. Представлены характеристики и практикуемые показатели с их функциональными возможностями для семантической сегментации геополей.

Ключевые слова: алгоритм, распознавание образов, интеллектуальная система, семантическая сегментация, показатели, функция потерь, геонинформационные системы, функция риска.

Введение

Многие алгоритмы и программы распознавания образов были созданы еще в конце 50-х и далее 60-х годов прошлого столетия, они и в настоящее время востребованы в различных областях деятельности человека. Также многие разработки из методов, моделей, алгоритмов и простых программ для нейротехнологий по распознаванию образов также примерно созданы в те же годы, когда появился первый машинный посредник (компьютеры) общения между человеческим и машинным языками. Уже тогда формировался сам облик и «первичный мостик» из средств по интеллектуализации и цифровизации информационных процессов к объектам различной природы, которые сегодня используются и интенсивно развиваются, находят свое применение в более современных информационных комплексах ИИС и всевозможных (по назначению цифровых платформ) гибридных (комбинированных) интеллектуальных информационных системах-ГИИС. Так, например, алгоритм [1] для распознавания цветовых классов объектов в информационных системах (ИИС и ГИИС) призван определить признаки из цветовых классов и дать априорное описание их на языке признаков, а также обеспечить рациональное разделение цветовых классов. В свою очередь, априорное описание здесь можно представить в виде функций из условных плотностей вероятностей признака $f_i(\chi)$. Пусть в блоках ИИС и ГИИС задан некоторый объект для распознавания темно-зеленого цвета (автомобиль, трактор, мотоцикл, здание и др.) находящийся на светло-зеленом поле. В информационной системе следует его опознать. Таким образом, к первому классу отнесем светло-зеленые цвета, а ко второму классу – темно-зеленого цвета.

В такой задаче сам подход распознавания образов в блоках сводится к статистической оценке и проверкой гипотез альтернатив Ω_1, Ω_2 . Для одномерного пространства применяется:

$$\begin{cases} Z \in \Omega_2, \chi > \chi_0 \\ Z \in \Omega_1, \chi < \chi_0 \end{cases} \quad (1)$$

где Z – образ объекта по цветовой гамме, поступающий на вход распознающего устройства; χ_0 – решающее правило для признака.

Алгоритм по распознаванию сводится к оценкам для значения решающего правила χ_0 .

Поставленную задачу можно с помощью известного критерия Байеса («байесов подход»), с помощью этого критерия производится минимизация функции риска и обеспечивается оптимальность решающего значения признака по цвету. Разделим условно значение признака χ на две части – R_1 и $R_2 \rightarrow R_1: Z \in \Omega_2, R_2: Z \in \Omega_1$. При распознавании допустимы два вида ошибок:

- ошибки первого рода – цвет для объекта, принадлежащий классу Ω_1 относится к классу Ω_2 ;
- ошибки второго рода – темно-зеленый цвет объекта, который отнесен к классу Ω_1 .

Вероятностями для дальнейшего оценивания ошибок являются:

$$P_1 = \int_{R_1} f_1(\chi) d\chi \text{ и } P_2 = \int_{R_2} f_2(\chi) d\chi. \quad (2)$$

В информационной системе имеется, как отмечено, распознающее устройство, в блоках которого ошибки учитываются в виде штрафа. Соответственно обозначим штрафы – C_1 и C_2 . Если считать штрафы за правильное решение равными нулю, то риск к принятию решения в модели определяется как:

$$\rho = C_1 P(\Omega_1) P_1 + C_2 P(\Omega_2) P_2, \quad (3)$$

где $P(\Omega_1)$, $P(\Omega_2)$ – априорные вероятности.

Учитывая, что вероятность $P_1 = \int_{R_1} f_1(\chi) d\chi$ получим формулу:

$$\rho = C_1 P_1 + \int_{R_2} [C_2 P(\Omega_2) f_2(\chi) - C_1 P(\Omega_1) f_1(\chi)] d\chi. \quad (4)$$

Решающее правило в минимизации функции риска заключается в том, что к области R_2 относят те цвета, для которых подынтегральное выражение в формуле будет отрицательным, условие для этого $C_2 P(\Omega_2) f_2(\chi) - C_1 P(\Omega_1) f_1(\chi) < 0$, т.е.:

$$\frac{f_2(\chi)}{f_1(\chi)} < \frac{C_1 P(\Omega_1)}{C_2 P(\Omega_2)}. \quad (5)$$

Схему поиска для значений признака χ_0 преобразуем в следующую формулу:

$$\frac{f_2(\chi_0)}{f_1(\chi_0)} = \frac{C_1 P(\Omega_1)}{C_2 P(\Omega_2)}. \quad (6)$$

Полагая $C_1 = C_2$, $f_2(\chi_0) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(\chi_0 - \bar{\chi}_2)^2}{2\sigma_2^2}}$, $f_1(\chi_0) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(\chi_0 - \bar{\chi}_1)^2}{2\sigma_1^2}}$, здесь символы σ_1 , σ_2 – стандартные отклонения признака χ соответствующих классов Ω_1 и Ω_2 ; символы $\bar{\chi}_1$, $\bar{\chi}_2$ – средние значения признаков, формулой для процедур является:

$$\exp - \left[\frac{(\chi_0 - \bar{\chi}_2)^2}{2\sigma_2^2} - \frac{(\chi_0 - \bar{\chi}_1)^2}{2\sigma_1^2} \right] = \frac{P(\chi_1)\sigma_2}{P(\chi_2)\sigma_1}. \quad (7)$$

Теперь, разрешив выражение относительно χ_0 , а также подставив априорные значения к вероятности цветовых классов получим алгоритм принятия решения, который обеспечивает минимальный риск в блоке распознавания $\begin{cases} Z \in \Omega_2, \chi > \chi_0^* \\ Z \in \Omega_1, \chi < \chi_0^* \end{cases}$. Так, для взятых в блоках априорных вероятностях для цветов $P(\Omega_1) = 0,12$ и $P(\Omega_2) = 0,28$ будем иметь (на выходе):

$$\begin{cases} Z \in \Omega_2, \chi > 0,458 \\ Z \in \Omega_1, \chi < 0,458 \end{cases}. \quad (8)$$

Приведенные выше рабочие формулы и сегодня необходимы для решения практических задач по распознаванию образов, например, в алгоритмах беспилотных систем, а также они представляют практический интерес для современных объектов в интеллектуальных блоках для информационных систем ИИС и блоках для ГИИС. Наряду с этими ранее известными подходами, моделями и наработками, в те времена также появились новые направления по организации «машин» и действующих алгоритмов для нейронных сетей. Одна из «машин» для того времени, обучающихся распознавать образы и объекты, называлась персептроном. Информационный процесс обучения для такой машины заключается в изменении величины взаимосвязей между этими элементами, по аналогии тому как идет обучение у самого человека. Персептрон (как и нейронная сеть) был обучен распознавать заглавные буквы, причем распознавание происходило даже тогда, когда буквы были различного размера и наклонены под различными углами. Также тогда были попытки построить вычислительную машину, распознающую речь. Отметим, что в научных лабораториях ВВС России и МО США уже изготавливались вычислительные машины, которые были способны распознавать звуки и команды человека на управления техническим объектом. Сегодня актуальны и востребованы отечественные методы и нейросетевые технологии для различных моделей информационных ИИС, ГИИС, а также с их применением в различных геоинформационных системах (ГИС) из наработанных действующих алгоритмов и программ. Эти модели, алгоритмы и программы с применением нейросетевой технологии должны обеспечивать определенные уровни для показателей качества и самим объектам распознавания и нейронной сети. Покажем это для актуальной сферы по семантической сегментации геополей к анализу объектов по космо- и аэрофотоснимкам.

1. Показатели обучения, валидации и тестирования моделей нейронных сетей при семантической сегментации

Нейронные сети – это, прежде всего, математические модели и алгоритмы, которые являются расширением известных статистических методов и построения адаптивных блоков для распознавания объектов. Модель для объекта – это набор из n параметров, которые характеризуют объект и, как правило, он представлен n -мерным вектором для признаков. В большинстве созданных моделей, а также практикуемых алгоритмов и программ применяют градиентные методы, также часто применяют и неградиентные методы. На языке машинного обучения задача к распознаванию объектов называется

классификацией, а задача выделения объектов из фона – сегментацией. Семантическая сегментация относится к задачам классификации каждого пикселя цифрового изображения. Цель для нейронной сети, алгоритма и самой программы – изучение зависимостей к информационным потокам данных, чтобы она смогла выполнить ожидаемые предсказания и распознать объект. Это делается путем обучения нейронной сети. Для задачи по обучению и классификации применяется метод контролируемого обучения (обучения с учителем), когда в информационном процессе данные «пропускаются» через слои, а выходные данные из сети сравниваются с «истиной» (эталоны). Такое сравнение в машине производится с помощью оцениваемых функций для приближений, функции потерь и др. Остановимся на показателях, которые уже применяют в интеллектуальных системах ИИС и ГИИС, а также и для различных видов систем ГИС. Сведем их в отдельные группы.

1.1. Группа 1. Показатели точности и вероятности распознавания образов

Показатель «Точность» распознавания (здесь подразумевается точность распознавания моделью нейронной сети пикселей для объектов некоторого класса на цифровом изображении) при семантической сегментации цифровых изображений — количественный показатель (и соответствующая положительная скалярная величина) отражающий долю из предсказанных нейронной сетью пикселей цифрового изображения, как относящихся к интересующему классу объектов, т.е. то, что они действительно к нему относятся [2].

Следует также отметить, что точность распознавания при семантической сегментации исходит из точности для классификации, которая представляет собой также показатель (и положительную скалярную величину), отражающий оценку доли правильного отнесения к некоторому интересующему классу. Понятие для точности классификации в данном случае конкретизируется под задачу для семантической сегментации, т.е. под задачу классификации каждого пикселя цифрового изображения, что также даёт понятие точности распознавания.

Оценка для точности распознавания, как и точности классификации, основывается на матрице ошибок (см. таблицу 1): Матрица ошибок для двух классов (класс объектов и класс «Фон», включающий в себя все оцениваемые пиксели, не относящиеся к классу объектов, что соответствует случаю бинарной классификации).

Таблица 1. Интерпретация формируемых матриц ошибок

		True condition	
		Condition positive	Condition negative
Predicted condition	Predicted condition positive	True Positive (TP)	False Positive (FP) Type I error
	Predicted condition negative	False Negative (FN) Type II error	True Negative (TN)

В таблице 1 Condition positive – истинная положительная метка класса (по эталону), Condition negative – истинная негативная метка класса (по эталону), Predicted condition positive – предсказанная нейронной сетью положительная метка, Predicted condition negative – предсказанная нейронной сетью негативная метка.

С практической точки зрения при семантической сегментации: True Positive – число правильно отнесённых нейронной сетью к интересующему классу пикселей цифрового изображения; False Positive – число неправильно отнесённых нейронной сетью к интересующему классу пикселей цифрового изображения; False Negative – число неправильно отнесённых к классу «Фон» пикселей цифрового изображения; True Negative – число правильно отнесённых нейронной сетью к классу «Фон» пикселей цифрового изображения.

Следует отметить, что единой методики и общепринятой формулы для измерения точности распознавания моделью нейронной сети объектов какого-либо класса на цифровых изображениях пока еще нет и поэтому для её оценки используется ряд из метрик, а именно: Precision, Recall, F-мера, IoU и т.д. [2].

Их запись, исходя из матрицы ошибок, следующая:

- Precision – вероятность того, что предсказание интересующего класса было проведено правильно:

$$Precision = \frac{TP}{TP+FP};$$

- Recall – доля предсказанных нейронной сетью пикселей интересующего класса из их общего

$$Recall = \frac{TP}{TP+FN};$$

- F-мера – гармоническое среднее Precision и Recall: $F_1 = 2 \frac{Precision \cdot Recall}{Precision + Recall}$.

1.2. Группа 2. Показатели Дайса и Жаккара

При работе в блоках для ИИС и ГИИС, связанных с практическим использованием методов обучения нейросети, также применяют количественные метрики – коэффициенты Дайса и Жаккара. Приведем краткое их описание. Коэффициент Дайса (разработан Л. Дайсом в 1945 г.) – это один из статистических показателей для нейронной сети, который используют для оценки сходства результатов, получаемых на выходе модели, и эталонов. Рабочая формула для расчёта: $Dice = \frac{2|X \cap Y|}{|X| + |Y|}$, в которой компоненты X и Y – два множества, вертикальная черта обозначает кардинальность для множества, то есть количество элементов в нём. По матрице ошибок коэффициент Дайса определяется как $Dice = \frac{2TP}{2TP + FP + FN}$. Данный коэффициент используется в алгоритмах семантической сегментации изображений, в частности, для сравнения выходных данных алгоритма с эталонами (представленными масками классов).

Следующим важным показателем является коэффициент Жаккара, который определяется, исходя из матрицы ошибок, формулой $IoU = \frac{TP}{TP + FP + FN}$.

Коэффициент Жаккара (более распространён в литературе как Intersection over Union или IoU) является более «строгой» метрикой для оценки показателей распознавания при семантической сегментации цифровых изображений. В общепринятой трактовке IoU – коэффициент, представляющих собой пересечение массива предсказанных значений классов для пикселей цифровых изображений и массива их истинных значений относительно интересующего класса, делённое на объединение этих массивов. Рассматривать попиксельную (как обобщённую) точность в качестве одного из показателей к точности распознавания не следует, т.к. она учитывает при вычислении и правильно отнесённые пиксели для класса «Фон», что приводит к ошибкам в оценке по точности распознавания в указанной трактовке при наличии дисбаланса для классов (т.е. ситуации, при которой на цифровом изображении число пикселей интересующего класса мало, а превалирует класс «Фон»), а именно:

$$OA = \frac{TP + TN}{TP + FP + FN + TN}, \quad (9)$$

где OA – показатель «общая точность».

Примером для указанной выше ошибки может послужить следующее:

- пусть некоторый эталон есть вектор из бинарных значений: 111000
- нейронная сеть выдаёт результатом вектор из бинарных значений 000000.

Посредством расчётов по матрице ошибок получим следующее: $OA = \frac{TP + TN}{TP + FP + FN + TN} = \frac{0 + 3}{0 + 0 + 3 + 3} = 0,5$. Показанный в данном случае пример отражает, что оценка точности распознавания при семантической сегментации цифровых изображений посредством OA будет являться неудачной с точки зрения достоверности и доверия для самой выбранной метрики.

1.3. Группа 3. Показатели времени обучения и устойчивости процесса

Экспериментально уже установлено, что при обучении моделей нейронных сетей на подвыборке для обучения из ~5000 данных размерностью 512x512x3 при 20 эпохах время обучения может составлять в среднем от 2 до 4-5 часов [2]. Однако здесь следует отметить то, что время обучения зависит от организации архитектуры нейронной сети (от количества используемых слоёв, их положения и назначения), от количества потока данных в выборке, размерности данных и того, что непосредственно используется при обучении нейронной сети – видеокарта (считается, что обучение проходит быстрее) или процессор (тогда считается, что обучение проходит медленнее). Поэтому возможны такие случаи, когда нейронная сеть обучается в указанном выше случае от 10-14 часов до одного дня. Случай, когда на указанной выше подвыборке для обучения нейронная сеть обучается день и более подразумевает использование процессора. Время и скорость для информационного процесса с использованием нейросети важны. В первом приближении оценивания информационного процесса и времени обучения можно ввести еще одну меру и дополнительный показатель (коэффициент) – количество баллов для обучающего сценария с использованием определенной модели нейронной сети при анализе семантической сегментации изображений для геополей:

$$K6 = \left(1 - \frac{T_{min} + T_{max}}{T}\right) * 100, \quad (10)$$

где T_{min} – время нахождения параметров в зоне для минимального допустимого значения; T_{max} – время нахождения искомых параметров в зоне для максимального значения; T – общее время наблюдаемого информационного процесса.

С примерами результатов информационного моделирования и вычислительного эксперимента с использованием нейронных сетей U-Net и др., которые были доложены на научном семинаре на кафедре «Информационно-измерительные системы» МГУТиК (МИИГАиК) в 2024 г. можно ознакомиться в работе [2].

2. Обсуждение

Как правило, под образом понимают множество объектов, сходных друг с другом в каком-либо отношении. Сегодня в алгоритмах ИИС, ГИИС и ГИС используют глубокие нейронные сети. Они уже показывают высокую эффективность и точность в предсказании для разработок в сфере промышленности, в авиации, в космонавтике, сельском хозяйстве и т.д. Качество компьютерного зрения для обработки космо- и аэрофотоснимков зависит во многом от носителей технических средств, бортовой аппаратуры и бортовых алгоритмов проведения мониторинга, осуществления наблюдения и обработки сигналов от объектов в различных условиях применения по функциональному назначению. Отметим также, что само моделирование как информационный процесс и метод научного анализа объектов различной природы в той или иной форме содержит элементы из распознавания образов, т.к. создание модели уже предполагает наложение на весь процесс и объект некоторой макроструктуры. Задача оценки информативности объекта и процессов – это, прежде всего нахождение оптимальной системы признаков. Как видим, приведенные выше показатели можно изначально принимать за основные, но при окончательном мнении по принятию решений их следует также дополнить блоками и автоматизированными процедурами многокритериальной оценки поддержки решений и согласования экспертных оценок [3-5].

3. Заключение

Методам и моделям интеллектуального управления в системах различной природы посвящено много работ и много дается определений. Разрабатываемые информационные системы с элементами ИИ, по существу, только приступают к интеграции формируемых моделей и методов нейронных сетей автоматизированных процедур и оценок из теоретико-множественного, логического и вероятностного подходов (как комбинированный подход). Действующие алгоритмы распознавания образов с использованием нейросетевых ресурсов информационных технологий уже находят свое широкое применение в структурах информационно-измерительных, в частности в структурах многоспектральных оптико-электронных приборов беспилотных летательных аппаратов, в которых анализируются информационные процессы по наблюдениям за объектами с оценками вероятности их правильного обнаружения и точности измерений географических координат. Изначально, как правило, во многих интеллектуальных информационных системах формируются проектные подходы, производится отбор методов и моделей для анализа, а также оцениваются первичные детерминированные, вероятностные, логические и комбинированные признаки наблюдаемых объектов. Большую роль при этом отводят современным технологиям на основе нейронных сетей, которые сегодня интенсивно развиваются и востребованы на мировом экономическом рынке.

Литература

1. Полтавский А.В. Интеллектуальные модели в информационно-измерительных и управляющих системах для распознавания образов / А.В. Полтавский, Г.Н. Ахобадзе, Е.Ю. Русяева // Приборы и системы. Управление, контроль, диагностика. – 2024. – № 8. – С. 35–41.
2. Бирюков Н.А. Семантическая сегментация геополей с использованием нейронных сетей на примере проблематики выделения зданий на космо- и аэрофотоснимках / Н.А. Бирюков, А.А. Майоров, М.П. Лапчинская // Известия ВУЗ. Геодезия и аэрофотосъемка. – 2024. – Т. 68, № 4. – С. 44–61.
3. Трахтенгерц Э.А. Компьютерные системы и методы поддержки информационного управления. – М.: СИНТЕГ, 2010. – 135 с.
4. Алчинов А.И. Методы цифровой фотограмметрии. Технология «Талка» / А.И. Алчинов, Н.Д. Беклемишев, В.Б. Кекелидзе. – М.: Московский государственный университет печати, 2007. – 259 с.
5. Bashkirov I.G. Parametric synthesis optimization models for high speed transport aerodynamic design to comply with flight safety and low environmental impact requirements / I.G. Bashkirov, S.L. Chernyshev, G.S. Veresnikov // Acta Astronautica. – 2023. – Vol. 204. – P. 720–727.