МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МАТЕРИАЛЫ
VIII Международной молодежной научной
конференции
«МАТЕМАТИЧЕСКОЕ
И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ
ИНФОРМАЦИОННЫХ,
ТЕХНИЧЕСКИХ
И ЭКОНОМИЧЕСКИХ СИСТЕМ»

Томск, 26–30 мая 2021 г.

Под общей редакцией И.С. Шмырина

Томск
Издательство Томского государственного университета
2021

9. Yahia S.B., Jaoua A. Discovering Knowledge from Fuzzy Concept Lattice // Data Mining and Computational Intelligence. Studies in Fuzziness and Soft Computing. – Physica, Heidelberg. – 2001. – V. 68. – P. 167–190.
10. Krajci S. Cluster based efficient generation of fuzzy concepts // Neural Network World. – 2003. – 13. – P. 521–530.
11. Belohlavek R., Sklenar V., Zaczal J. Crisply Generated Fuzzy Concepts // B. Ganter et al. eds. ICFCA, LNAI. – 2005. – 3403. – P. 269–284.
12. Krajci S. A generalized concept lattice // Logic J. IGPL. – 2005. – 13(5). – P. 543–550.
13. Medina J., Ojeda-Aciego M., Ruiz-Calvino J. Relating Generalized Concept Lattices and Concept Lattices for Non-Commutative Conjunctions // Applied Mathematics Letters. – 2008. – 2. – P. 1296–1300.
14. Medina J., Ojeda-Aciego M., Ruiz-Calvino J. On Multi-adjoint Concept Lattices: Definition and Representation Theorem // ICFCA. LNAI. – 2007. – 4390. – P. 197–209.
15. Medina J., Ojeda-Aciego M., Ruiz-Calvino J. Formal Concept Analysis via Multi-Adjoint Concept Lattices // Fuzzy Sets and Systems. – 2009. – 160. – P. 130–144.
16. Elloumi S., Jaam J., Hasnah A., Jaoua A., Nafkha I. A Multi-Level Conceptual Data Reduction Approach Based on the Lukasiewicz Implication // Information Sciences. – 2004. – 163. – P. 253–262.
17. Shao M.W., Zhang W.X. Information Granularity Lattices // Proc. of the Sixth IEEE Int. Conf. on Machine Learning and Cybernetics. – Hong Kong, 2007.
18. Belohlavek R. A note on variable threshold concept lattices: Threshold-based operators are reducible to classical concept-forming operators // Information Sciences. – 2007. – 177. – P. 3186–3191.
19. Belohlavek R., Vychodil V. Fuzzy concept lattices constrained by hedges // Journal of Advanced Computational Intelligence and Intelligent Informatics. – 2007. – 11(6). – P. 536–545.

DOI: 10.17223/978-5-907442-42-9-2021-6

ПРИМЕНЕНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ К ЗАДАЧЕ КЛАССИФИКАЦИИ ВРЕМЕННЫХ РЯДОВ

Дуля И.С., Мурзагулов Д.А.

Томский государственный университет
idulya7@gmail.com, murzagulovdamir@gmail.com

Введение

В эпоху Индустрии 4.0 компании все больше сосредотачиваются на разработке технологических решений и стремятся максимально перестроить существующие бизнес-процессы под новую концепцию, основанную на сборе и анализе данных. Большая часть собираемых данных на предприятиях представлена в виде сигналов (данные, динамически изменяющиеся во времени). Например, электрические импульсы, данные с акселерометров, давление в устройствах и т.д. В целях компьютерной обработки исходные аналоговые сигналы дискретизируют и представляют в виде временных рядов. Далее к данным применяют статистические модели и методы машинного обучения, позволяющие извлечь ценные знания для бизнеса, которые используются в деятельности предприятия и оптимизации производства. В данном контексте классификация временных рядов была и остаётся одной из наиболее популярных задач интеллектуального анализа данных [1].

С ростом числа открытых наборов данных были предложены сотни подходов и алгоритмов решения задачи классификации временных рядов. Можно выделить три основные группы подходов к решению задачи классификации временных рядов. Первый подход самый очевидный: на вход статистической модели подаются значения временного ряда. Здесь существует ограничение на то, что временные ряды должны быть одинаковой длины, кроме того, малейшее смещение может привести к существенному искажению результата классификации. В основе второго подхода лежит использование метрических моделей (например, модель k -ближайших соседей) с «эластичной» метрикой расстояния. Третий подход предполагает переход из временного пространства в пространство признаков. В рамках подхода исследователь самостоятельно подбирает характеристики, полезные в решении конкретной задачи. Ключевая проблема состоит в том, что разные аналитики для одной задачи подберут разные наборы характеристик. Отсюда возникает проблема: невозможно определить является ли предложенный пере-

чень характеристик полным и достаточным для решения задачи. Субъективность данного этапа подталкивает к поиску такого решения, которое позволило бы систематизировать или упростить данную процедуру.

Успех моделей глубокого обучения в области компьютерного зрения послужил мотивацией к разработке подходов к классификации временных рядов на основе глубоких сверточных сетей [2]. В рамках недавнего крупного исследования [3] было оценено 18 алгоритмов классификации временных рядов на 85 наборах данных, ни один из которых не был моделью глубокого обучения. Это показывает, что научное сообщество не стремится активно изучать эффективность моделей глубокого обучения для решения задачи классификации временных рядов [4].

Преимущество применение методов глубокого обучения состоит в том, что оно позволяет справиться с несистемностью этапа подготовки данных в задаче классификации временных рядов [5]. Подходы на основе глубокого обучения применяются в связке с обратимыми преобразованиями, которые позволяют представить временной ряд в виде матрицы (например, оконное преобразование Фурье или вейвлет-преобразование). Следует заметить, что в результате таких преобразований информация практически не теряется, т.к. существует обратное преобразование, с помощью которого можно восстановить исходный временной ряд. Далее глубокие сверточные нейронные сети извлекают латентные признаки, на основе которых производится классификация с помощью полносвязной сети прямого распространения.

1. Классификация на основе непрерывного вейвлет-преобразования

Непрерывное вейвлет-преобразование (Continuous Wavelet Transform, CWT) – интегральное преобразование, которое представляет собой свертку вейвлет-функции с сигналом [6]. Вейвлет-преобразование переводит сигнал из временного представления в частотно-временное. Непрерывное вейвлет-преобразование можно описать уравнением

$$\text{cwt}(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t) \psi\left(\frac{t - \tau}{s}\right) dt,$$

где τ – сдвиг, s – масштаб, $\psi(\cdot)$ – вейвлет. Описанное преобразование основано на концепциях масштабирования и смещения. Масштабирование представляет собой растяжение или сжатие вейвлета на коэффициент масштабирования. Сдвиг представляет перемещение вейвлетов с различным масштабом от начала до конца сигнала.

В качестве основного набора данных для исследования эффективности применения непрерывного вейвлет-преобразования для извлечения признаков из временных рядов в задаче классификации временных рядов было решено взять популярный датасет Human Activity Recognition (HAR). Выбранный набор данных содержит данные с девяти акселерометров/гироскопов. Каждому набору из девяти временных рядов (данные с 9 датчиков) сопоставлены активности носителя устройства, считывающего движения.

Для решения задачи классификации временных рядов на основе вейвлет-преобразования необходимо решить две задачи:

1. Преобразовать исходные сигналы в матрицу коэффициентов вейвлет-преобразования;
2. Привести полученную матрицу в квадратную форму.

Для задачи было решено использовать вейвлет Морле с диапазоном масштабов от 1 до 65. Далее полученная матрица коэффициентов размер 64x128 преобразуется в квадратную форму (64x64) по правилу среднего. Этот шаг не является необходимым, но позволяет сэкономить вычислительные ресурсы, кроме того, сверточные нейронные сети принято обучать на квадратных изображениях.

Также необходимо определиться: как подавать сигналы с разных датчиков на вход нейронной сети? Было решено подавать сигнал с каждого отдельного датчика в виде отдельного канала (по аналогии с RGB-каналами цветного изображения). Соответственно, все зависимости между сигналами с различных датчиков могут учитываться одновременно, что очень важно.

В качестве основы свёрточной сети было решено взять архитектуру LeNet-5 и внести в неё несколько улучшений (рис. 1).

Рис. 1. Свёрточная нейронная сеть, основанная на архитектуре LeNet-5

Было решено использовать max pooling вместо average pooling, что позволит обеспечить более высокую вычислительную эффективность и извлекать признаки, полезные для обнаружения резких изменений в сигналах. Также функция активации гиперболического тангенса (Tanh) была заменена на Rectified linear unit (ReLU) для ускорения обучения и достижения лучшей производительности.

Далее была обучена и протестирована заданная модель. На рис. 2 представлена матрица несоответствий для тестового множества после 20 эпох обучения на тренировочном наборе данных.

Действительные метки	Ходьба	480	3	0	0	0	0
	Подъем по ступенькам	4	458	6	0	2	0
	Спуск по ступенькам	12	1	414	0	0	0
	Сидение	0	8	0	416	47	0
	Стояние	0	0	0	75	483	0
	Лежание	0	1	0	0	0	537
		Ходьба	Подъем по ступенькам	Спуск по ступенькам	Сидение	Стояние	Лежание
		Предсказанные метки					

Рис. 2. Матрица несоответствий

Наибольшее число ошибок связано с тем, что модель путает активности «стояние» и «сидение». Обе активности малоподвижные и сигналы с разных датчиков очень сильно похожи друг на друга, с чем и связано большое число ошибок. В целом, модель

качественно осуществляет классификацию активностей по данным с датчиков (акселерометров, гироскопов). Точность, рассчитанная на тестовом множестве, составляет 94.6%. Этого вполне достаточно для практических задач (например, учёт активности и расчёт потраченных калорий). Полученный результат следует рассматривать как доказательство того, что комбинация непрерывного вейвлет-преобразования и свёрточной сети является эффективным методом классификации многомерных нестационарных временных рядов.

2. Сравнение метода с базовыми методами классификации временных рядов

Для подтверждения качества рассмотренного метода необходимо его сравнить с другими моделями/методами классификации временных рядов на наборах данных различной природы. Помимо рассмотренного датасета Human activity recognition были рассмотрены еще четыре набора данных: Waffer [7], FordA [8], StarlightCurves [9] и синтетический набор данных с 10-тью видами нестационарных сигналов.

В качестве базовых методов классификации были взяты три метода: классификация по исходным временным рядам без извлечения признаков, классификация по признаковому представлению временных рядов и классификация на основе метрической модели (метод k-ближайших соседей) с «эластичной» метрикой расстояния DTW. Для первого и второго базового методов использовалась модель случайного леса в качестве классификатора. В табл. 1 приведены сводные результаты оценки точности моделей на тестовом множестве для различных наборов данных.

Таблица 1

Сводные результаты оценки точности моделей (Accuracy)

	HAR	Waffer	FordA	SLC	Synthetic
WT + CNN	0.924	0.984	0.922	0.951	0.986
Raw TS + RF	0.776	0.957	0.705	0.934	0.869
Feature-based TS + RF	0.858	0.954	0.745	0.978	0.937
KNN + DTW	0.605	0.860	0.500	0.911	0.922

Рассматривая значение метрики Accuracy, можно заметить, что в четырех из пяти датасетов классификация на основе скалограмм вейвлет-преобразования и свёрточной нейронной сети дала лучшие результаты. Только в рамках задачи определения положения небесного объекта (StarLightCurves) основной подход уступил классификации на основе признакового представления. Если упорядочить оставшиеся подходы, то самым мощным окажется классификация на основе признакового представления, далее идёт классификация на основе исходного временного ряда и самые плохие результаты у подхода, основанного на метрической модели с «эластичной» метрикой расстояния.

Следует заметить, что на основании этого эксперимента нельзя утверждать, что для любой задачи классификации временных рядов основной подход будет давать более высокие результаты. Это связано с тем, что предметных областей и задач, в которых применяется анализ временных рядов, крайне обширен, соответственно, временные ряды существенно отличаются между собой по свойствам и типам генеративных процессов. Это исследование лишь показывает, что во многих задачах классификация на основе скалограмм вейвлет-преобразования может дать более качественный результат, чем другие подходы.

3. Тестирование в условиях ограниченного набора обучающих данных

В задачах классификации критическую роль играет объем обучающей выборки. В настоящих прикладных задачах существуют ограничения на её размер, возникающие по различным причинам (стоимость, небольшое число наблюдаемых объектов и т.д.). Чем больше параметров содержит модель, тем больше обучающих наблюдений необходимо, чтобы найти оценки этих параметров. Нейросетевые классификационные мо-

дели содержат достаточно большое количество параметров, для глубоких свёрточных сетей их число может доходить до нескольких миллионов. С этим связано ограничение в применении нейросетевых классификаторов в реальных задачах. Часто предпочтение отдают более простым моделям.

В связи с этим, критически важно показать, что основной метод может работать и в условиях ограниченного объема обучающих данных. Для каждого из пяти наборов данных были созданы обучающие выборки с числом наблюдений от 50 до 1000. Полученные результаты приведены на рис. 3.

Рис. 3. Зависимость Accuracy от числа обучающих наблюдений

Несмотря на то, что выбранные свёрточные модели имеют большое число параметров, для их обучения требуется сравнительно небольшое число наблюдений. В среднем на различных наборах данных достаточно 250 обучающих наблюдений, для некоторых задач (Waffer) достаточно и 50 размеченных временных рядов для получения требуемой точности. Полученный результат позволяет утверждать, что подход на основе глубокого обучения можно использовать в условиях ограниченного объема данных, когда доступно всего лишь несколько сотен обучающих наблюдений.

Заключение

Появление новых открытых наборов данных, развитие моделей и методов машинного обучения, а также активное внедрение достижений четвёртой научно-технической революции подогревает интерес к задаче классификации временных рядов. Аналитики пытаются найти новые более эффективные и универсальные решения задачи. Прогресс в области глубокого обучения и задачах компьютерного зрения служит мотивацией к разработке подходов к классификации временных рядов на основе свёрточных нейронных сетей.

Подходы на основе глубокого обучения обладают существенным преимуществом перед базовыми подходами – высокой универсальностью и системностью. Если в базовых подходах качество решения существенно зависит от того, насколько удачно аналитик подобрал перечень признаков, характеризующих временные ряды, то подходы на основе глубокого обучения используют единую процедуру подготовки данных. На этапе подготовки используются обратимые преобразования (оконное преобразование Фурье, вейвлет-преобразование и др.), сохраняющие большую часть информации в результате преобразования, т.к. существует обратное преобразование, с помощью которого можно восстановить исходный временной ряд. Далее глубокие свёрточные нейрон-

ные сети извлекают латентные признаки, на основе которых производится классификация с помощью полносвязной сети прямого распространения. Также глубокое обучение позволяет работать с многомерными временными рядами. В свёрточных нейронных сетях сигнал с каждого отдельного датчика подаётся в виде отдельного канала (по аналогии с RGB-каналами цветного изображения). Соответственно, все зависимости между сигналами с различных датчиков могут учитываться одновременно, что очень важно при работе с многомерными временными рядами.

В данной работе было проведено исследование эффективности применения глубоких сверточных сетей в связке с непрерывным вейвлет-преобразованием для решения задачи классификации временных рядов. Был описан алгоритм подготовки данных и обучения моделей на примере задачи распознавания человеческой активности. Было проведено сравнение подхода с базовыми алгоритмами решения задач классификации временных рядов. Полученные результаты показывают, что глубокое обучение может обеспечить существенный прирост точности для задач из разных предметных областей по сравнению с базовыми алгоритмами решения задачи классификации временных рядов (увеличение точности в четырёх задачах из пяти). В результате исследования было обнаружено, что для обучения модели требуется относительно немного обучающих наблюдений для получения необходимой точности. На рассмотренных наборах данных требуемое число обучающих наблюдений варьируется от 50 до 250. Можно предполагать, что для обучения модели с несколькими миллионам параметров потребуется значительно больше обучающих наблюдений. Соответственно, подход, основанный на глубоком обучении и непрерывном вейвлет-преобразовании, легко можно применять в реальных задачах, когда существенно ограничен размер обучающей выборки.

В дальнейшей работе полученные результаты в области классификации временных рядов планируется перенести на задачу обнаружения точек перехода. Под точками перехода понимаются моменты времени, когда изучаемый объект меняет своё состояние. В контексте задачи распознавания человеческой активности это моменты времени, когда человек меняет одну активность на другую. Так, планируется использовать глубокую нейронную сеть для извлечения латентных признаков, на основе которых будет происходить обнаружение точек перехода между состояниями.

ЛИТЕРАТУРА

1. Yang Q., Wu X. 10 challenging problems in data mining research // *Information Technology & Decision Making*. – 2006. – № 5. – P. 597–604.
2. Wang Z., Yan W., Oates T. Time series classification from scratch with deep neural networks: a strong baseline // *International Joint Conference on Neural Networks*. – 2017. – P. 1578–1585.
3. Bagnall A., Lines J., Bostrom A., Large J., Keogh E. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances // *Data Mining and Knowledge Discovery*. – 2017. – № 31. – P. 606–660.
4. Lines J., Taylor S., Bagnall A. HIVE-COTE: The hierarchical vote collective of transformation-based ensembles for time series classification // *ACM Transactions on knowledge discovery from data*. – 2018. – № 12.
5. Li T., Zhao Z., Sun C., et al. WaveletKernelNet: an interpretable deep neural network for industrial intelligent diagnosis // *Transactions on systems, man, and cybernetics: systems*. – 2021. – P. 1–11.
6. Hurley C., Mclean J. *Wavelet, Analysis and Methods*. / Waltham Abbey: ED-Tech press, 2018. – 280 p.
7. Olszewski R., Maxilon R., Siewiorek D. Generalized feature extraction for structural pattern recognition in time-series data / USA: Carnegie mellon university, 2001. – 111 p.
8. Dataset: FordA [Электронный ресурс]: набор данных / Time-series classification repository. – URL: <http://www.timeseriesclassification.com/description.php?Dataset=FordA> (дата обращения 20.04.2021)
9. Rebbapragada U., Protopapas P., Brodley C., Alcock C. Finding anomalous periodic time series // *Machine Learning*. – 2009. – № 74. P. 281–313.