

А.Н. ЧЕРНОДУБ, Д.В. НОВИЦКИЙ, Д.А. ДЗЮБА

Институт Проблем Математических Машин и Систем НАН Украины, г. Киев
a.chernodub@gmail.com

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ ОДИНОЧНЫХ НЕЙРОННЫХ СЕТЕЙ И КОМИТЕТОВ НЕЙРОННЫХ СЕТЕЙ: СРАВНИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ*

Рассматривается задача прогнозирования временных рядов на основе нейросетевых методов. Предлагается метод формирования комитетов нейросетей на основе оценки степени линейной зависимости выходов одиночных нейросетей, входящих в комитет. Проводится экспериментальное сравнение различных типов одиночных нейросетей и комитетов нейросетей на задачах прогнозирования чисел Вольфа и процесса Маккея-Гласса.

Ключевые слова: комитет нейронных сетей, прогнозирование, числа Вольфа, процесс Маккея-Гласса

Введение

Прогнозирование временных рядов является актуальной научной проблемой, имеющей обширные приложения в экономике, физике и других отраслях. Нейронную сеть можно обучить решать задачу прогнозирования временного ряда на известных примерах, при этом задачу обучения нейронной сети можно рассматривать как задачу моделирования некоторой динамической системы, порождающей этот временной ряд.

В последнее время, популярным способом повышения качества прогнозирования является использование комитетов нейронных сетей [1-2]. В этом случае, система прогнозирования состоит из нескольких нейросетей, на основе выходов которых выполняется принятие финального решения специальным обобщающим модулем. Однако, всегда ли оправдано такое усложнение?

Мы провели экспериментальное исследование использования различных архитектур как одиночных нейросетей, так и комитетов нейросетей, обучив в общей сложности более 10000 нейросетей. Эксперименты проводились на двух известных временных рядах небольшой длины, похожих по своей форме: числах Вольфа и реализации процесса Маккея-Гласса.

Постановка задачи

Временной ряд «Числа Вольфа» представляет собой запись наблюдений количества солнечных пятен за год, начиная с 1700 года. Временной ряд «Процесс Маккея-Гласса» представляет решение нелинейного дифференциального уравнения с временной задержкой:

$$\frac{dy(t)}{dt} = \frac{ay(t-\tau)}{1+y(t-\tau)^{10}} - by(t) \quad (1)$$

Для вычисления временного ряда были использованы следующие значения параметров: $a = 0.2$, $b = 0.1$, $\tau = 17$.

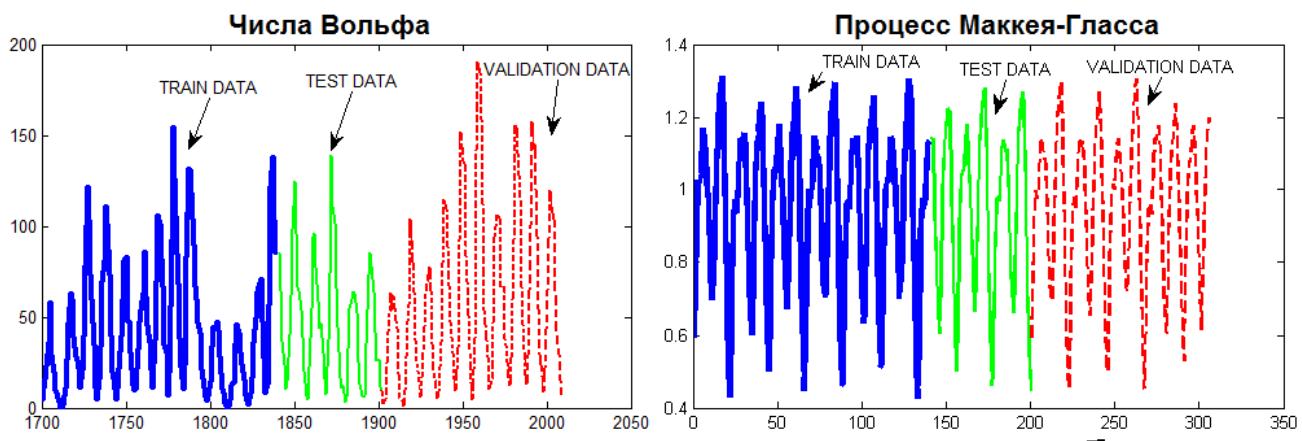


Рис. 1. Временной ряд «Числа Вольфа» (слева), временной ряд «Процесс Маккея-Гласса» (справа).

* Данная работа выполнена при поддержке международного гранта УНТЦ Р-357 (EOARD 088008).

Оба временных ряда содержат 307 значений, они были нормированы на интервале [-1;1] и разделены на тренировочную, тестовую и экзаменационную части в соотношении 45:20:35.

Задача для прогнозирования формулируется следующим образом: по известному текущему значению последовательности $y(k)$ и некоторой предыстории $y(k-1), y(k-2), \dots, y(k-m)$ дать оценку следующего значения $y(k+1)$. Процесс повторяется для всех значений последовательности, у которых есть следующие значения. Длина предыстории m , которая используется для прогнозирования, варьируется. Нейронные сети обучаются на тренировочной части последовательности, далее на тестовой части проводится отбор наилучшей одиночной нейросети или комитета по среднеквадратичному критерию (Mean Squared Error, MSE):

$$MSE = \frac{1}{N} \sum_{n=1}^N (y(n) - y^*(n))^2 \quad (2)$$

Здесь: N – количество значений процесса, $y(n)$ – реальные значения, $y^*(n)$ – прогнозируемые значения. Одиночная нейросеть или комитет с лучшим результатом на teste проходит дополнительные испытания на экзаменационной последовательности. Экзаменационный результат считается окончательным.

Одиночные нейронные сети

В нашей работе мы использовали 5 типов одиночных нейронных сетей: LNN, MLP, DLNN, DMLP, RMLP. LNN (Linear Neural Network) представляет собой статический однослойный персептрон с линейной активационной функцией [3], MLP (Multilayer Perceptron) [4] это двухслойный персептрон с сигмоидальной активационной функцией в скрытом слое и линейной функцией в выходном слое. Сеть DLNN (Dynamic Linear Network) [5] является «динанизированной» версией линейной нейросети LNN, к ее входу добавлена линия задержки, которая, кроме текущего элемента временного ряда подает на вход нейросети некоторое количество задержанных прошлых значений. Сеть DMLP (Dynamic Multilayer Perceptron) [6] представляет аналогичным образом динанизированный многослойный персептрон. Рекуррентная сеть RMLP (Recurrent Multilayer Perceptron) [7] представляет собой сеть DMLP с добавленными рекуррентными связями с выхода нейросети на ее вход. Сети LNN и MLP являются статическими нейросетями, поскольку используют для прогнозирования только текущего значение временного ряда, в случае одиночных нейросетей они рассматриваются в основном из сравнительных мотивов. Сети DLNN, DMLP и RMLP являются динамическими нейросетями, благодаря встроенным линиям задержек.

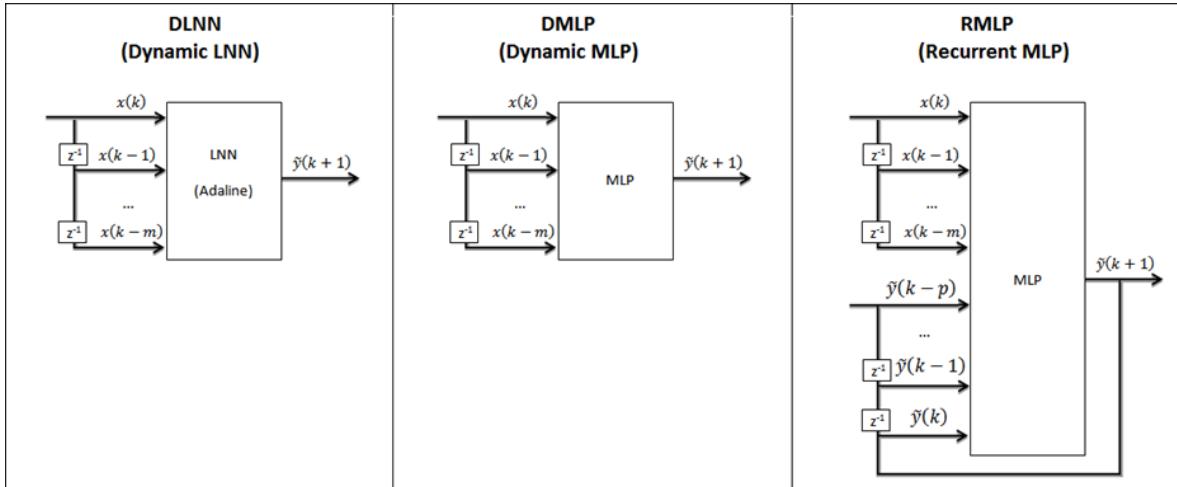


Рис. 2. Архитектуры одиночных динамических нейронных сетей.

Во время экспериментов, параметры нейросетей варьировались. В сетях MLP, DMLP, RMLP количество нейронов скрытого слоя изменялось от 5 до 15, обучение велось по методу Левенберга-Марквардта, для каждого типа архитектуры было выполнено 10 попыток обучения. В рекуррентных сетях RMLP для коррекции весов использовался механизм обратного распространения во времени ВРТ. Количество задержек на входе для сетей DLNN, DMLP и RMLP варьировалось от 1 до 10, количество рекуррентных задержек на выходе для сетей RMLP бралось от 1 до 5. Всего для каждого из временных рядов была обучена 1 сеть LNN, 110 сетей MLP, 10 сетей DLNN, 1100 сетей DMLP, 2750 сетей RMLP. Среднеквадратичные ошибки лучших на teste одиночных нейросетей для каждой из опробуемых нейросетевых парадигм представлены в табл. 1-2.

Таблица 1.

Среднеквадратичные ошибки лучших одиночных нейросетей на задаче «Числа Вольфа»

Тип нейросети	TRAIN	TEST	VALIDATION
Single LNN	0.01134	0.01042	0.02276
Single MLP	0.01063	0.01032	0.02468
Single DLNN	0.02468	0.00592	0.01166
Single DMLP	0.00204	0.00400	0.01181
Single RMLP	0.00214	0.00370	0.01300

Таблица 2.

Среднеквадратичные ошибки лучших одиночных нейросетей на задаче «Процесс Маккея-Гласса»

Тип нейросети	TRAIN	TEST	VALIDATION
Single LNN	0.018738	0.019297	0.018598
Single MLP	0.017549	0.01792	0.018866
Single DLNN	0.00369	0.003339	0.0035374
Single DMLP	4.13E-06	6.81E-06	1.27E-05
Single RMLP	1.04E-05	7.69E-06	1.04E-05

Как видно из результатов одиночных нейросетей на экзаменационной выборке, для задачи «Числа Вольфа» лучшей архитектурой оказалась линейная нейросеть с задержками DLNN, которая опередила другие, более сложные динамические сети. Для задачи «Процесс Маккея-Гласса» лучший результат обеспечила рекуррентная динамическая сеть RMLP, обладающая наиболее сложной среди рассматриваемых динамических нейросетей архитектурой.

Комитеты нейронных сетей

В нашей работе, комитеты нейронных сетей представляют собой ансамбли из нескольких одиночных нейросетей, независимо решающих задачу прогнозирования (рис. 3). Частные решения одиночных нейросетей $\{\tilde{y}_1(k+1); \tilde{y}_2(k+1); \dots; \tilde{y}_p(k+1)\}$ поступают на обобщающий модуль (Gating module), который выдает окончательное решение $\tilde{y}(k+1)$.

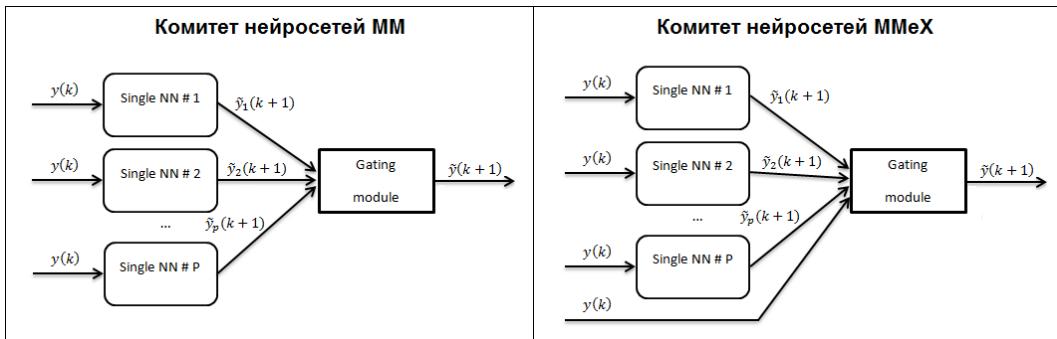


Рис. 3. Комитет нейросетей MM (слева), комитет нейросетей MMeX (справа).

Комитеты нейросетей делятся на два типа: MM (“Multi-Modular”), на вход которых поступают только частные решения одиночных нейросетей, и MMeX (“Multi-Modular with eXternal inputs”), на вход которых поступают частные решения и элементы исходного временного ряда. В качестве одиночных нейронных сетей использовались взятые из числа обученных для экспериментов с одиночными нейросетями динамические рекуррентные сети типа RMLP. В качестве обобщающих модулей были опробованы все 5 типов нейронных сетей, описанных в предыдущем разделе. Обобщающие модули обучались прогнозированию с использованием той же обучающей, тестовой и экзаменационной выборки, что и одиночные нейронные сети. Для комитетов типа MM в качестве обобщающих модулей были дополнительно опробованы операции усреднения (“Average”) и вычисления медианного значения (“Median”).

Выбор одиночных рекуррентных нейросетей для включения их в комитеты проводился в два этапа: отбор 20-ти лучших из 2750 натренированных для каждого временного ряда одиночных нейросетей, и формирование множества из 5-ти нейросетей, имеющих наименьшую степень взаимной линейной зависимости по выходу. Такой способ отбора нейросетей для комитета обеспечивает выбор набора нейросетей с достаточно хорошим качеством прогнозирования, дающих ошибки на разных элементах временного ряда, что может быть сглажено обобщающим модулем.

Для оценки линейной зависимости членов комитета для каждой из $C_{20}^5 = 15504$ возможных комбинаций значений их выходов на массиве данных TRAIN DATA было рассчитано число обусловленности (condition value)

$$cond = \frac{\alpha_{\max}}{\alpha_{\min}},$$
 где α – сингулярные числа, которые являются квадратными корнями соответствующих собственных значений корреляционной матрицы прогнозируемых траекторий. Чем меньше число обусловленности для набора векторов, тем меньше степень их линейной зависимости. Для полностью линейно зависимых векторов число обусловленности равняется бесконечности, для полностью линейно независимых оно равняется нулю.

В табл. 3-4 приведены варианты наборов одиночных нейросетей для комитетов. Значения в столбце “Rating No.” отвечают номеру нейросети в рейтинге точности прогнозирования среди Top-20 лучших одиночных рекуррентных нейросетей.

Таблица 3.

Наборы одиночных нейросетей для комитетов на задаче «Числа Вольфа»

Variant No.	Condition value	NN#1 Rating No.	NN#2 Rating No.	NN#3 Rating No.	NN#4 Rating No.	NN#5 Rating No.
1	16.76	2	4	13	15	16
2	16.82	2	7	12	13	15
3	16.91	2	12	13	15	16
4	16.94	2	5	12	13	15

Таблица 4.

Наборы одиночных нейросетей для комитетов на задаче «Процесс Маккея-Глассса»

Variant No.	Condition value	NN#1 Rating No.	NN#2 Rating No.	NN#3 Rating No.	NN#4 Rating No.	NN#5 Rating No.
1	960.45	5	10	11	17	20
2	965.37	7	15	17	19	20
3	965.70	10	11	17	19	20
4	966.74	5	10	13	18	20

Из табл. 3-4 видно, что ни один из отобранных таким способом наборов нейросетей для комитетов не вошла первая в рейтинге качества прогнозирования одиночная нейросеть. Вместо этого, каждый вариант комитета имеет в своем составе нейросети из последней пятерки рейтинга наилучших 20-ти нейросетей. Это может быть объяснено тем, что именно эти сети дают небольшое количество относительно больших ошибок на разных элементах временного ряда, что согласуется с нашими целями.

После отбора одиночных нейросетей для комитета, выполнялось обучение обобщающих модулей. Сравнительно малый объем обучающей выборки не позволил применить технику бустинга [8-9]. В комитетах MM MLP, MM DMLP, MM RMLP, MMeX MLP, MMeX DMLP, MMeX RMLP количество нейронов скрытого слоя изменялось от 1 до 5, обучение велось по методу Левенберга-Марквардта, для каждого типа архитектуры было выполнено 2 попытки обучения. В комитетах MM RMLP и MMeX RMLP для коррекции весов использовался механизм обратного распространения во времени BPTT. Количество задержек на входе для комитетов MM DLNN, MM DMLP, MM RMLP, MMeX DLNN, MMeX DMLP и MMeX RMLP варьировалось от 1 до 10, количество рекуррентных задержек на выходе для сетей MM RMLP и MMeX RMLP изменилось от 1 до 5. Всего для каждого из временных рядов было обучено 2 комитета с сетью LNN в качестве обобщающего модуля, 20 комитетов с сетью MLP, 20 комитетов с сетью DLNN, 200 комитетов с сетью DMLP, 2000 комитетов с сетью RMLP. Результаты лучших на тесте комитетов для каждого типа смещающих модулей представлены в табл. 5-6.

Таблица 5.

Результаты лучших комитетов нейросетей на задаче «Числа Вольфа»

Тип комитета	TRAIN	TEST	VALIDATION
MM Average	N/A	N/A	0.01653
MM Median	N/A	N/A	0.01432
MM LNN	0.001283	0.008261	0.01930
MM MLP	0.001279	0.008341	0.01932
MM DLNN	0.001231597	0.001231	0.02209
MM DMLP	0.001580308	0.008151	0.03134
MM RMLP	0.001179236	0.006721	0.02498
MMeX LNN	0.001297	0.010035	0.04267
MMeX MLP	0.001287	0.010418	0.04297
MMeX DLNN	0.00126295	0.010341	0.04305
MMeX DMLP	0.001279644	0.008606	0.03780
MMeX RMLP	0.000895341	0.004671	0.05874

Таблица 6.

Результаты лучших комитетов нейросетей на задаче «Процесс Маккея-Гласса»

Тип комитета	TRAIN	TEST	VALIDATION
MM Average	N/A	N/A	9.12857E-06
MM Median	N/A	N/A	5.81958E-06
MM LNN	4.45E-06	1.63E-05	4.96907E-05
MM MLP	9.63401E-07	4.99903E-06	0.1279415
MM DLNN	3.85456E-06	1.40081E-05	4.25028E-05
MM DMLP	6.35888E-07	1.05279E-05	1.63934E-05
MM RMLP	1.13236E-05	9.84379E-06	1.41329E-05
MMeX LNN	5.83929E-06	8.51037E-06	5.97136E-06
MMeX MLP	1.3196E-06	4.48807E-06	3.83486E-06
MMeX DLNN	5.74409E-06	9.12733E-06	6.51437E-06
MMeX DMLP	1.04589E-06	7.49995E-06	4.60201E-06
MMeX RMLP	1.09349E-05	1.08614E-05	1.27645E-05

Как видно из результатов комитетов нейросетей на экзаменационной выборке, для задачи «Числа Вольфа» лучшим комитетом оказался комитет без внешних входов со статической линейной нейросетью в качестве обобщающего модуля, MM LNN. Более точный прогноз обеспечили операции усреднения и взятия медианного значения выходов в качестве обобщающих модулей. Результат лучшего комитета оказался почти на 10% хуже, чем результат лучшей одиночной нейросети (см. табл. 1).

Для задачи «Процесс Маккея-Гласса» лучший результат обеспечил комитет с внешними входами со статическим многослойным персептроном в качестве обобщающего модуля MMeX RMLP. Он превосходит лучшую одиночную нейросеть почти на порядок (см. табл. 2).

Выводы

Несмотря на внешнюю схожесть временных рядов чисел Вольфа и процесса Маккея-Гласса, для первого временного ряда применение комитетов нейросетей ухудшило качество прогнозирования, тогда как для второго временного ряда – напротив, позволило значительно его улучшить. Это улучшение, по-видимому, носит фундаментальный характер, так как в случае одиночных нейросетей для чисел Вольфа лучшие результаты прогнозирования показали нейросети с более простой архитектурой, а для процесса Маккея-Гласса лучшими оказались более сложные рекуррентные нейросети. Интересной задачей для дальнейших исследований является разработка критерия выявления пользы применения комитетов нейросетей для повышения качества прогнозирования. Также, требую-

щим дополнительного исследования вопросом является то, что для обоих временных рядов лучшими нейросетевыми обобщающими модулями комитетов нейросетей оказались статические, а не динамические нейросети.

Список литературы

1. Гужва А. Г., Доленко С. А., Персианцев И. Г., Шугай Ю. С. Многоступенчатый алгоритм на основе комитета нейронных сетей для прогнозирования и поиска предвестников в многомерных временных рядах // Нейроинформатика-2009. XI Всероссийская научно-техническая конференция: Сб. научн. тр. Ч. 2. С. 116–125. М.: МИФИ. 2009.
2. Галинская А.А. Модульные нейронные сети: обзор современного состояния разработок // Математические машины и системы. - 2003. - №. 3,4, с.87-102.
3. Хайкин С. Нейронные сети: полный курс, 2-е изд., испр. : Пер. с англ. — М. : Вильямс, 2006, с. 194.
4. Там же, с. 218.
5. Там же, с. 810.
6. Там же, с. 811.
7. Там же, с. 922.
8. Avnimelech R., Intrator N., Boosting Regression Estimators // Neural Computation 1999, Vol. 1, Issue 2.
9. Schapire R.E. The Boosting Approach to Machine Learning: An Overview” // MSRI Workshop on Nonlinear Estimation and Classification, 2002.