

Чернодуб А.Н.<sup>1</sup>

<sup>1</sup> Институт Проблем Математических Машин и Систем НАНУ

### **Метод смешивания мнений экспертов комитета нейронных сетей на основе расширенного фильтра Калмана для задачи прогнозирования временных рядов**

**Введение.** Прогнозирование временных рядов является актуальной научной задачей, имеющей множество применений в экономике, физике, теории управления и других отраслях. Хорошим практическим методом решения задачи прогнозирования временных последовательностей являются нейронные сети. Нейронная сеть может быть обучена на известных примерах моделируемой последовательности и затем использоваться для прогнозирования на новых, ранее не виденных нейросетью примерах данных.

Популярным способом повышения качества работы нейросетей является использование комитетов экспертов [1, с. 458]. В этом случае, система прогнозирования состоит из нескольких нейронных сетей-экспертов, функционирующих параллельно и обобщающего модуля, который принимает финальное решение. В качестве обобщающего модуля могут использоваться как простые операции усреднения выходных значений экспертов или операции взятия их медианного значения, так и более сложные алгоритмы. Использование комитетов экспертов обычно дает небольшой, порядка 1-5% прирост качества работы системы, однако и такое улучшение может быть очень существенно. Например, именно использование комитетов нейронных сетей обеспечило команде из Швейцарии победу в соревновании “German Traffic Sign Competition” на конференции IJCNN’11 за счет улучшения на 0.42% по сравнению с результатом, полученном ими же на одиночной нейронной сети [2]. Также, точность прогнозирования критична для синтеза контроллеров на основе популярного метода прогнозирующего нейроуправления (Model Predictive Neurocontrol) [3, 4], поскольку в нем эволюция объекта управления итеративно прогнозируется на несколько тактов вперед, при этом ошибка прогнозирования растет экспоненциально. Это выдвигает дополнительные требования к точности синтезируемой эмпирической модели динамического процесса, в роли которой может выступить комитет нейронных сетей.

Мы предлагаем новый алгоритм смешивания мнений экспертов для комитета нейронных сетей на основе использования расширенного фильтра Калмана, который выполняет перевзвешивание весов обобщающего модуля в режиме он-лайн и проводим экспериментальное исследование этого алгоритма.

**Задача прогнозирования.** Для проведения экспериментов был использован стандартный для данного типа задач процесс Маккея-Гласса [5, 6]. Последовательность Маккея-Гласса задается следующим разностным уравнением:

$$x_{t+1} = (1-b)x_t + a \frac{x_{t-\tau}}{1 + (x_{t-\tau})^{10}}, t = \tau, \tau + 1, \dots,$$

где  $\tau \geq 1$  — целые числа. Последовательность из 3000 значений была сгенерирована с использованием следующих значений параметров:  $a = 0.2$ ,  $b = 0.1$ ,  $\tau = 30$ , как в [6].

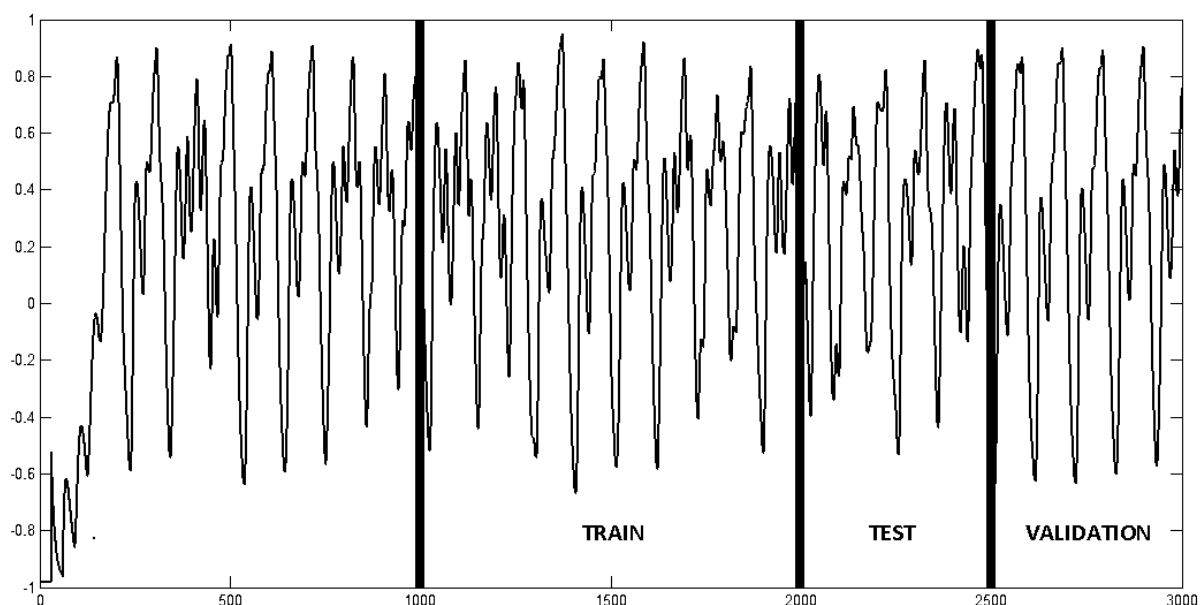


Рис. 1. Деление процесса Маккея-Гласса на обучающую (TRAIN), тестовую (TEST) и валидационную (VALIDATION) последовательности.

Первые 1000 значений сгенерированной последовательности были проигнорированы, 1000 следующих значений были использованы в качестве обучающей выборки, последующие 500 — в качестве тестовой выборки, оставшиеся 500 — в качестве валидационной выборки (рис. 1).

Задача прогнозирования ставилась следующим образом: по  $l$  известным прошлым значениям последовательности  $(x_t, x_{t-1}, \dots, x_{t-l})$  дать оценку  $y_{t+1}$  следующего значения  $x_{t+1}$ . Качество прогнозирования оценивалось по критерию нормированной среднеквадратичной ошибки NRMSE, определенной по формуле:

$$NRMSE = \frac{\sqrt{\langle (y_k - x_k)^2 \rangle}}{\sqrt{\langle (x_k - \langle x_k \rangle)^2 \rangle}},$$

где:  $y_k$  — прогнозируемые значения,  $x_k$  — целевые значения.

**Одиночные нейронные сети.** Первоначально, для решения задачи прогнозирования было обучено 1000 одиночных нейронных сетей с задержкой во времени (Time Delay Neural Networks, TDNN) [1, с. 807]. Они представляют собой многослойные персептроны с одним скрытым слоем, на вход которым подается  $l$  задержанных прошлых значений моделируемого процесса. В качестве активационных функций использовался гиперболический тангенс. Количество задержек на входе сети варьировалось от 1 до 10, количество нейронов в скрытом слое — от 3 до 12. Нейронные сети были обучены на обучающей выборке последовательности алгоритмом GEKF [6], затем из них была выбрана сеть, показавшая лучшее качество прогнозирования на тестовой выборке. Это оказалась сеть с 7 задержанными входами и 5 нейронами в скрытом слое, она показала ошибку  $NRMSE_{SINGLE} = 0.0139$  на валидационной выборке.

**Комитет нейронных сетей.** Мы использовали комитет нейронных сетей с архитектурой, показанной на рис. 1. Он представляет собой  $N$  обученных одиночных нейронных сетей-экспертов, которые подают свои оценки  $y^1(k+1), \dots, y^N(k+1)$  следующего значения процесса  $x(k+1)$  на обобщающий модуль, который по ним выдает финальную оценку  $y(k+1)$ . Алгоритм вычисления финальной оценки в обобщающем модуле может варьироваться.

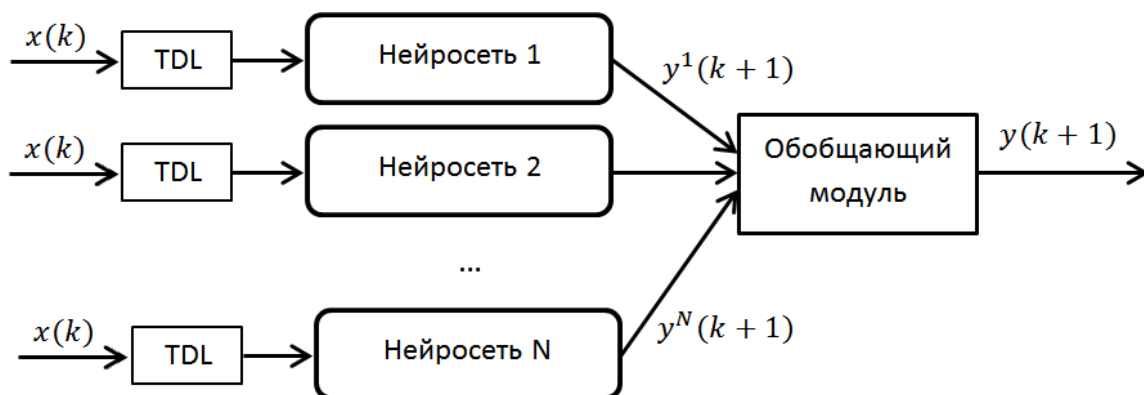


Рис. 2. Комитет нейронных сетей для задачи прогнозирования последовательностей

В нашем комитете было использовано  $N = 5$  нейронных сетей, отобранных из 1000 обученных ранее одиночных сетей.

**Метод отбора сетей в комитет.** Нейронные сети-эксперты, входящие в комитет, должны удовлетворять двум критериям: 1) обеспечивать хорошее качество прогнозирования и 2) по возможности, выдавать независимые друг от друга решения. Последнее повышает вероятность ошибок нейросетей комитета на разных элементах выборки, что может быть сглажено обобщающим модулем. Существует два популярных метода отбора сетей в ансамбль: 1) по степени линейной зависимости выходов нейросетей 2) по степени линейной зависимости ошибок на обучающей последовательности [7]. Мы воспользовались первым вариантом. Для этого, из 1000 обученных сетей было отобрано 20 лучших, отбор производился по принципу наименьшей показанной ошибки на тестовой выборке. Затем, для каждого набора сетей из  $C_{20}^5 = 15504$  возможных вариантов наборов были получены их прогнозы на обучающей выборке, эти выходы были объединены в матрицы размера  $1000 \times 5$  для каждого из наборов, и для каждой из таких матриц было рассчитано число обусловленности (condition value)  $cond = \frac{\alpha_{\max}}{\alpha_{\min}}$ , где  $\alpha$  — сингулярные числа, которые

являются соответствующими собственными значениями корреляционной матрицы значений полученных прогнозов. Чем меньше значение числа обусловленности для набора векторов, тем меньше степень их взаимной линейной зависимости. Для полностью линейно зависимых векторов число обусловленности равно бесконечности, для полностью линейно независимых векторов оно равно нулю. В наш комитет вошел набор сетей с номерами {4,10,15,17,19} как показавший наименьшее значение числа обусловленности, здесь номер сети означает ее место в рейтинге 20-ти лучших на тестовой выборке сетей. Обратите внимание, что лучшая из одиночных сетей, по определению имеющая номер 1, не вошла в отобранный для комитета набор сетей.

**Смешивание мнений экспертов на основе расширенного фильтра Калмана.** Расширенный фильтр Калмана может выступать в роли оптимизатора второго порядка, что в последнее время активно используется для обучения как сетей прямого распространения [6], так и рекуррентных сетей [8, 9]. Этот метод обеспечивает более быструю сходимость по сравнению с оптимизационными методами 1-го порядка и функционирует в режиме он-лайн, что в ряде задач выигрышно выделяет его по сравнению с другими оптимизационными с методами 2-го порядка, обычно имеющими

пакетный характер работы. Рассмотрим применение расширенного фильтра Калмана для смешивания мнений экспертов в комитете одиночных нейронных сетей. Пусть комитет состоит из  $N$  одиночных нейросетей-экспертов, которые на каждом такте работы системы независимо друг от друга генерируют вектор оценок нового значения  $g(k) = [y^1(k+1) \ y^2(k+1) \ \dots \ y^N(k+1)]$  и пусть задана весовая матрица смешивания мнений экспертов  $w(k) = [w_1 \ w_2 \ \dots \ w_N]^T$ .

На первом такте определим начальную весовую матрицу  $w(1) = [1/N \ 1/N \ \dots \ 1/N]^T$ , а также зададим корреляционную матрицу  $P$  размера  $N \times N$ , на первом шаге установим ее равной единичной матрице  $P(1) = I$ . Зададим матрицу скорости обучения  $R = \frac{1}{\eta}$ , в нашем примере  $\eta = 0.001$  и матрицу шума процесса  $Q = 10^{-4} I$ .

В ходе работы комитета, на такте  $k$  выполним следующие действия:

- 1) Найдём прогнозируемое значение следующей величины ряда:

$$y(k+1) = g(k) \times w(k).$$

- 2) Определим текущую ошибку работы комитета  $e(k)$  на основе измеренных прошлых значений процесса  $x(k)$  и прошлого выполненного прогноза комитета  $y(k)$ :

$$e(k) = x(k) - y(k).$$

При этом допускается, что поступившее прошлое значение процесса  $x(k)$  может быть зашумлено, т.е.  $x(k) = \hat{x}(k) + \varepsilon$ ,  $\varepsilon \in N(0, \sigma^2)$ .

- 3) Вычислим новые значения весов комитета  $w(k+1)$  и матрицы корреляции  $P(k+1)$  по формулам:

$$H(k) = g(k-1),$$

$$K = P(k)H(k)^T [H(k)P(k)H(k)^T + R]^{-1},$$

$$P(k+1) = P(k) - K(k)H(k)P(k) + Q,$$

$$w(k+1) = w(k) + K(k)e(k).$$

- 4) Пока не закончится последовательность, переходим к пункту 1).

Матрица наблюдений  $H$  содержит производные  $\frac{\partial y(k)}{\partial w(k)}$ , которые в нашем случае равны выходам одиночных сетей-экспертов. Обновляющаяся на каждом такте

корреляционная матрица  $P$  здесь работает как аппроксимация обратной матрицы Гессе, содержащей вторые производные функции ошибки по оптимизируемым параметрам.

**Результаты сравнительного эксперимента.** Мы провели серию экспериментов по прогнозированию последовательности Маккея-Гласса с одиночными нейросетями и комитетами нейросетей на незашумленной выборке, а также на выборке, входные значения которой были зашумлены белым гауссовским шумом с амплитудой 1%, 2%, 5% и 10%. Эксперименты были выполнены в среде MATLAB без использования пакета Neural Networks Toolbox. В таблице ниже представлены результаты прогнозирования на валидационной выборке, ошибка оценивалась по критерию NRMSE, в скобках указан выигрыш в процентах по сравнению с лучшей одиночной нейросетью:

Тип комитета	Без шума	Шум 1%	Шум 2%	Шум 5%	Шум 10%
Одиночная нейросеть	0.0139 -	0.0278 -	0.0518 -	0.1210 -	0.2386 -
Усреднение по комитету	0.0137 (+1.3%)	0.0276 (+0.9%)	0.0514 (+0.9%)	0.1195 (+1.2%)	0.2350 (+1.5%)
Взятие медианного значения	0.0138 (+0.9%)	0.0276 (+0.9%)	0.0512 (+1.3%)	0.1193 (+1.4%)	0.2338 (+2.0%)
Расширенный фильтр Калмана	0.0136 (+1.9%)	0.0272 (+2.2%)	0.0509 (+1.9%)	0.1186 (+2.0%)	0.2323 (+2.6%)

**Выводы.** Нами был предложен алгоритм перевзвешивания мнений сетей-экспертов комитета нейросетей, работающий на основе расширенного фильтра Калмана. Было экспериментально установлено, что качество прогнозирования такого комитета повышается примерно на 2% по сравнению с лучшей одиночной нейросетью из множества обученных сетей в количестве 1000 штук. Это в 1.5-2 раза лучше, чем обычные методы усреднения по комитету и взятия медианного значения. Повышение качества прогнозирования объясняется, во-первых, фактическим дообучением комитета на выборке данных в ходе его работы и, во-вторых, эффективностью алгоритма расширенного фильтра Калмана, который функционирует как оптимизационный метод второго порядка.

## Литература

1. *Хайкин С.* Нейронные сети: полный курс. — М.: Вильямс, 2006. — 1104 С.
2. *Ciresan D. C., Meier U., Masci J., Schmidhuber J.* A Committee of Neural Networks for Traffic Sign Classification. — Proceedings of IEEE International Joint Conference on Neural Networks — USA, San Francisco, July 31 – August 5, 2011 — P. 1918 – 1921.
3. *Hagan M.T., Demuth H.B.* Neural networks for control. — Proceedings of the American Control Conference – San Diego, USA, 1999. – Vol. 3. – P. 1642 – 1656.
4. *Кузнецов Б.И., Василец Т.Е., Варфоломеев А.А.* Синтез нейроконтроллера с предсказанием для двухмассовой электромеханической системы. — Электротехника и электромеханика. – 2008. – Т. 3. – С. 27 – 32.
5. *Gers F. A., Eck D., Schmidhuber J.* Applying LSTM to Time Series Predictable Through Time-Window Approaches. — Proceedings of International Conference on Artificial Neural Networks — Berlin, 2001 . – P. 669 – 676.
6. *Alessandri A., Cuneo M., Pagnan S., Sanguineti M.* On the convergence of EKF-based parameters optimization for Neural Networks. – Proceedings of the 42nd IEEE Conference on Decision and Control — Maui, Hawaii, USA, December 2001. — V. 6. — P. 6181 – 6186.
7. *Галинская А.А.* Модульные нейронные сети: обзор современного состояния разработок. — Математические Машины и Системы. — 2003. — № 3,4. — С. 87 – 102.
8. *Prokhorov D.V.* Toyota Prius HEV Neurocontrol and Diagnostics. — Neural Networks. – 2008. – N. 21. – P. 458 – 465.
9. *Cernansky M., Benuskova L.* Simple recurrent network trained by RTRL and extended Kalman filter algorithms. – Neural Network World. – 2003. – N. 13. – V. 3. – P. 223 – 234.