XXXI Неделя науки СПбГПУ. Материалы межвузовской научной конференции. Ч. VI: С.61-63, 2003.

УДК 621.317.08

П.В. Кобяков (асп. каф. ИИТ).

ТЕМПОРАЛЬНЫЕ НЕЙРОННЫЕ СЕТИ И ФИЛЬТР КАЛМАНА

Темпоральные нейронные сети [1,2,3] являются расширением стандартных моделей нейронных сетей, позволяющим обрабатывать динамическую информацию. Темпоральные сети отличаются более сложной архитектурой и, как следствие, более сложными алгоритмами обучения [4,5,6]. Применение теории фильтров Калмана для сетей данного класса позволяет значительно повысить качество обучения за счет эффективного использования вычислений градиента.

Фильтр Калмана.

Алгоритм фильтрации Калмана - это линейный, основанный на модели, стохастический, рекурсивный, взвешенный алгоритм оценивания, использующий метод наименьших квадратов [7]. Алгоритм позволяет оценивать состояние системы с использованием информации о входных и выходных сигналах системы. Фильтр Калмана использует модель системы, состоящую из уравнения состояния (процесса) и уравнения измерения, оба уравнения являются линейными. Процесс оценивания является рекурсивным, таким образом, фильтр может функционировать в реальном времени. Фильтр является оптимальным в среднеквадратическом смысле.

Для линейной динамической системы уравнения состояния и измерения имеют вид

$x_{k+1} = Ax_k + Bu_k + w_k$, $z_k = H_k x_k + v_k$,

где x_k - состояние системы, u_k - входное воздействие, w_k - изменение состояния для шага k, A и B - параметры системы, z_k - измеряемое состояние, v_k - ошибка измерения. Математическое ожидание $E[v_k] = E[w_k] = 0$, при этом w_k и v_k являются случайными некоррелированными величинами.

Пусть $\tilde{x}_k$ - априорная оценка состояния системы для шага k, с учетом всей значений, полученных до шага k; $\hat{x}_k$ - апостериорная оценка для шага k, учитывающая измерение z_k. Априорные и апостериорные ошибки могут быть определены следующим образом. $\tilde{e}_k = x_k - \tilde{x}_k$, $\hat{e}_k = x_k - \hat{x}_k$. Обозначим дисперсии ошибок как $\tilde{P}_k = E\left[\tilde{e}_k^2\right]$, $\hat{P}_k = E\left[\hat{e}_k^2\right]$.

Алгоритм фильтрации Калмана для случая линейных систем имеет вид [8]:

1 Выбор начальных оценок $\tilde{x}_0$ и $\tilde{P}_0$.

2 Вычисление коэффициента усиления Калмана: $S_k = R_k + H_k \tilde{P}_k H_k^T$, $K_k = \tilde{P}_k H_k^T S_k^{-1}$, где R_k - ковариационная матрица v_k.

3 Коррекция текущей оценки: $\hat{x}_k = \tilde{x}_k + K_k\left(z_k - H_k \tilde{x}_k\right)$.

4 Коррекция оценки дисперсии ошибки: $\hat{P}_k = \left(I - K_k H_k\right)\tilde{P}_k$.

5 Предсказание величины: $\tilde{x}_{k+1} = A\hat{x}_k + Bu_k$.

6 Предсказание дисперсии: $\tilde{P}_{k+1} = A\hat{P}_k A^T + Q_k$, где Q_k - ковариационная матрица w_k.

7 Переход на шаг 2.

Выбор начальных оценок не является определяющим в работе фильтра, тем не менее, хорошие оценки обеспечивают более быструю сходимость.

Фильтр Калмана для нелинейных систем.

Теория фильтров Калмана была разработана для применения к линейным системам. На практике уравнения состояния и измерения часто являются нелинейными функциями:

$x_{k+1} = f\ (x_k\ , u_k) + w_k,\ z_k = h(x_k\ ) + v_k$.

Для обеспечения возможности применения фильтра Калмана эти уравнения линеаризуются в точке последней оценки, а именно:

$$x_{k+1} = f(\hat{x}_k, u_k) + \left.\frac{\partial f}{\partial x}\right|_{\hat{x}_k} (x_k - \hat{x}_k) + w\ ,\ z_k = h(\tilde{x}_k) + \left.\frac{\partial h}{\partial x}\right|_{\tilde{x}_k} (x - \tilde{x}_k) + v_k\ .$$

Фильтр Калмана, основанный на данных уравнениях, называется *расширенным*. Следствием использования линеаризации является утрата свойства оптимальности.

Обучение темпоральных сетей с использованием фильтров Калмана.

Нейронная сеть может рассматриваться как нелинейная динамическая система, вектор состояния которой x_k определяется множеством весов сети. Уравнения состояния и измерения для нейронной сети записываются как [9]:

$x_{k+1} = x_k + w_k,\ z_k = h(x_k\ ,\ u_k\ ,\ s_k\ ) + v_k$.,

где x_k - вектор состояния, w_k - искусственный шум процесса, u_k - входной вектор, s_k - вектор текущих значений активации нейронов, $h(\ ,\ ,\ )$ - нелинейная функция преобразования сети, z_k- желаемый отклик модели, v_k - искусственный шум измерения с ковариационной матрицей R_k.

Уравнения фильтра Калмана позволяют оценивать состояние системы, то есть изменять веса сети таким образом, чтобы реальный выходной сигнал сети y_k как можно меньше отличался от желаемого сигнала z_k.

Уравнение измерения линеаризуется следующим образом:

$z_k = H_k x_k + v_k$,

где H_k - матрица измерений линеаризованной модели размерностью $p \times W$, состоящая из частных производных p выходов сети по W весам. Частные производные вычисляются в точке последней оценки. На практике эти частные производные вычисляются с помощью алгоритма обратного распространения сквозь время (BPTT) или алгоритма обучения рекуррентной сети в реальном времени (RTRL).

Заключение.

В отличие от других градиентных методов обучения темпоральных сетей, алгоритм калмановской фильтрации более эффективно использует имеющуюся информацию, что приводит к значительному увеличению качества обучения. В [10] приводятся сравнительные данные для задачи предсказания. Ошибка сети, обученная с использованием калмановской фильтрации, была на порядок ниже ошибки сети, обученной стандартным алгоритмом обратного распространения.

Повышение точности достигается за счет увеличения вычислительных затрат. Минимальные затраты по времени и памяти для DEKF соответствуют затратам одного из алгоритмов: BPTT [4] или RTRL [6], дополнительные затраты по времени и по памяти составляют $O(pW^2)$ и $O(W^2)$, соответственно. Данный факт не является серьезным препятствием при реализации сетей на современных настольных компьютерах, и вполне окупается качеством обучения. Для встроенных применений может использоваться расщепленная версия алгоритма.

ЛИТЕРАТУРА

1. Chappelier J., Gori M., Grumbach A. Time in Connectionist Models. Sequence Learning: Paradigms, Algorithms and Applications, Springer, 2001, 105-134.
2. Kremer, S.C. Spatio-temporal Connectionist Networks: A Taxonomy and Review: Neural computation.
3. Mozer, M.C. Neural Net Architectures for Temporal Sequence Processing – Time Series Prediction, Addison-Wesley, 1994, 243-264.
4. Williams, R. and Zipser, D. A learning algorithm for continually running fully recurrent neural networks - Neural Computation 1(3) 1990, 270-280.
5. Day, S. and Davenport, M. Continuous-Time Temporal Back-Propagation with Adaptive Time Delays – IEEE Transactions on Neural Networks, 1991.
6. Wan A. Time Series Prediction by Using a Connectionist Network with Internal Delay Lines. Time Series Prediction. Forecasting The Future and Understanding the Past, Addison Wesley, 1994.

7. Welsh, G., Bishop, G. An Introduction to the Kalman Filter - Department of Computer Science, University of North Carolina, TR 95-041, 2002.
8. De Schutter, J., De Geeter, J., Lefebvre, T, Bruyninckx, H. Kalman Filters: A Tutorial – 1999.
9. Leleux, D.P., Claps, R., Chen, W., Tittel, F.K., Harman, T.L, Applications of Kalman filtering to real-time trace gas concentration measurements – Applied Physcis B 74, pp. 85-93, 2002.
10. Haykin, S. Neural Networks: A Comprehensive Foundation (second edition), 1999.