



моделирования в сложных системах: Труды XIV Международной конференции – Самара: СНЦ РАН, 2008. С. 164-171

И.М. Куликовских, С.А. Прохоров

НЕЯВНАЯ РЕГУЛЯРИЗАЦИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ НА ОСНОВЕ КОНТРОЛЯ ДИНАМИЧЕСКИХ КОГНИТИВНЫХ КАРТ

(Самарский национальный исследовательский университет
им. академика С.П. Королева, Самара, Россия)

В области активно развивающихся алгоритмов глубинного обучения традиционные методы оценки сложности алгоритмов являются нецелесообразными, так как не объясняют исключительной обобщающей способности алгоритмов глубинного обучения. Последние исследования оценки качества таких алгоритмов вводят понятие неявной регуляризации как разновидности метода обучения. Данный вид регуляризации находит решение в виде модели наименьшей сложности, но не накладывает штрафа на модель и не модифицирует напрямую метод оптимизации. Согласно результатам Б. Нейшабура, Р. Салахутдинова, Н. Сребро и др. [1-3], обобщающая способность контролируется геометрией пространства параметров обучаемого алгоритма и эмпирической оптимизацией, адаптированной с учетом данной геометрии. Качество алгоритмов, в данном случае, будет обуславливаться их способностью хранить и извлекать структуру данных в течение длительных интервалов времени, компенсируя недостаток внешней памяти.

Постановка задачи

Данное исследование нацелено на создание метода неявной регуляризации обучаемых алгоритмов, который для повышения их обобщающей способности моделирует процессы кратковременной и долговременной памяти. Основной акцент работы сделан на частной проблеме формирования и контроля так называемых *когнитивных карт*, впервые предложенных психологом Э. Толменом [4].

Когнитивные карты создаются и видоизменяются в результате активного взаимодействия объекта обучения со средой окружения и могут обладать различной степенью общности, масштаба и организации в зависимости от полноты представленности пространственных отношений и выбранной точки отсчета. В рамках данного исследования рассматривается возможность формирования когнитивной карты с системой координат «угадывание-забывание», которая отражает поведение обучаемой алгоритма через последовательное представление его связи с другими алгоритмами и средой окружения.

Для формирования системы координат карты при переходе от процессов кратковременной памяти к процессам долговременной взаимодействие объекта со средой окружения рассматривается для двух задач:



1. Теории цифровой фильтрации, предлагающей методы синтеза фильтров с бесконечной импульсной характеристикой (БИХ-фильтров), которые позволяют моделировать постепенное накопление и исчезновение информации из памяти фильтра на основе оптимизации его полюсов.
2. Теории тестирования, описывающей эффект забывания, вызванный извлечением информации (Retrieval-Induced Forgetting, RIF) [5]. Данный эффект возникает как следствие сознательного запоминания тестируемой информации через ее явное извлечение, что приводит к неявному забыванию нетестируемой информации на бессознательном уровне, что позволяет моделировать процессы стимулирования памяти при тестировании с множественным выбором.

Таким образом, проводя аналогию между моделями памяти *фильтра* и *человека* в условиях фильтрации в первом случае и тестирования во втором случае, предлагаемый метод обучения позволяет моделировать взаимодействие объекта обучения со средой окружения с помощью динамических когнитивных карт, повышая обобщающую способность модели обучения. При этом рассматриваемые задачи позволяют представить модель обучения в виде обобщенной регрессионной модели и требуют ее регуляризации для понижения сложности банка фильтров и банка тестовых заданий.

Формализация задачи

Представим задачу оценивания параметров обобщенной модели регрессии, учитывающей процессы кратковременной и долговременной памяти в цифровой фильтрации, обучении моделей и тестировании, и ее решение в следующем виде:

$$\|f - Bv\|_2^2 + \sigma^2 v^T K(C)^{-1} v \rightarrow \min_v, \quad v = K(C) B^T (BK(C) B^T + \sigma^2 I)^{-1} f, \quad (1)$$

где

$f = \{H, y, y_s\}$ – обобщенное множество допустимых ответов, включающее: H – множество значений передаточной функции в непрерывной и дискретной областях (теория цифровой фильтрации), y – множество ответов алгоритма (теория обучающих систем), y_s – множество ответов человека (теория тестирования);

$B = \{D, X, X_s\}$ – обобщенное множество объектов, которое включает: $D = \{\Lambda, G\}$ – банк цифровых фильтров в непрерывной Λ и дискретной G областях, X – объекты обучающей и тестовой выборки, X_s – банк тестовых заданий с множественным выбором;

$v = \{\beta, \theta, \theta_s\}$ – обобщенное множество допустимых параметров модели, включающее: β – множество параметров цифрового фильтра в непрерывной и дискретной областях, θ – множество параметров модели обучения, θ_s – множество параметров тестового задания;

$K(C)$ – ядро, задающее множество когнитивных карт $C = \{c^{\phi, \alpha}, c^{g, f}, c_s^{g, f}\}$: $c^{\phi, \alpha}$ – когнитивная карта с системой координат в теории цифровой фильтрации, за-



дающая положение полюса фильтра $\phi \in \Phi$, $\Phi \in \{\Gamma, \Xi\}$ в непрерывной $\gamma \in \Gamma$ и дискретной $\xi \in \Xi$ областях, и параметр забывания (затухания) $\alpha \in A$; $c^{g,f}$ – когнитивная карта с системой координат в теории обучаемых систем, задающая параметр угадывания c^g и забывания c^f ; $c_s^{g,f}$ – когнитивная карта с системой координат в теории тестирования, задающая параметр угадывания c_s^g и забывания c_s^f в терминах эффекта RIF; σ^2 – дисперсия случайной составляющей множества ответов.

При этом под неявной регуляризацией регрессионной модели из соотношения (1) будем понимать задачу формирования и контроля когнитивной карты $c \in C$ согласно

$$\|f - Bv\|_2^2 + \sigma^2 v^T K(c)^{-1} v \xrightarrow{c \in C} \min.$$

На основе введенных понятий разработана модель перехода от процессов кратковременной памяти к процессам долговременной на основе когнитивной карты (см. рис. 1). Ниже дана интерпретация цифровых обозначений:

1. Анализ методов понижения сложности регрессионных моделей.
2. Анализ методов понижения сложности банка цифровых фильтров.
3. Анализ методов понижения сложности банка тестовых заданий.
4. Неявная регуляризация банка цифровых фильтров на основе $c^{\phi, \alpha}$, где $\alpha = 0$ (кратковременная память).
5. Неявная регуляризация банка цифровых фильтров на основе $c^{\phi, \alpha}$, где $\alpha \in \mathbb{N}$ (долговременная память).
6. Неявная регуляризация банка цифровых фильтров на основе $c^{\phi, \alpha}$, где $\alpha \in \mathbb{R}$ (долговременная память).
7. Неявная регуляризация банка тестовых заданий на основе эффекта RIF $c_s^{g,f} \in [0, 1]$ (кратковременная память).
8. Неявная регуляризация банка тестовых заданий на основе эффекта RIF $c_{Ls}^{g,f} \in [0, 1]$ при индивидуальном обучении (долговременная память).
9. Неявная регуляризация банка тестовых заданий на основе эффекта RIF $c_{Ls}^{g,f} \in [\underline{c}_{Ls}, \overline{c}_{Ls}]$ при групповом обучении (долговременная память).
10. Неявная регуляризация регрессионных моделей на основе $c^{g,f} \in [0, 1]$ при индивидуальном обучении (кратковременная память).
11. Неявная регуляризация регрессионных моделей на основе $c_L^{g,f} \in [\underline{c}_L^{g,f}, \overline{c}_L^{g,f}]$ при групповом обучении (кратковременная память).
12. Неявная регуляризация регрессионных моделей при индивидуальном и групповом обучении $c^{\phi, \alpha} \sim \{c^{g,f}, c_L^{g,f}\} \sim \{c_s^{g,f}, c_{Ls}^{g,f}\}$ (долговременная память).

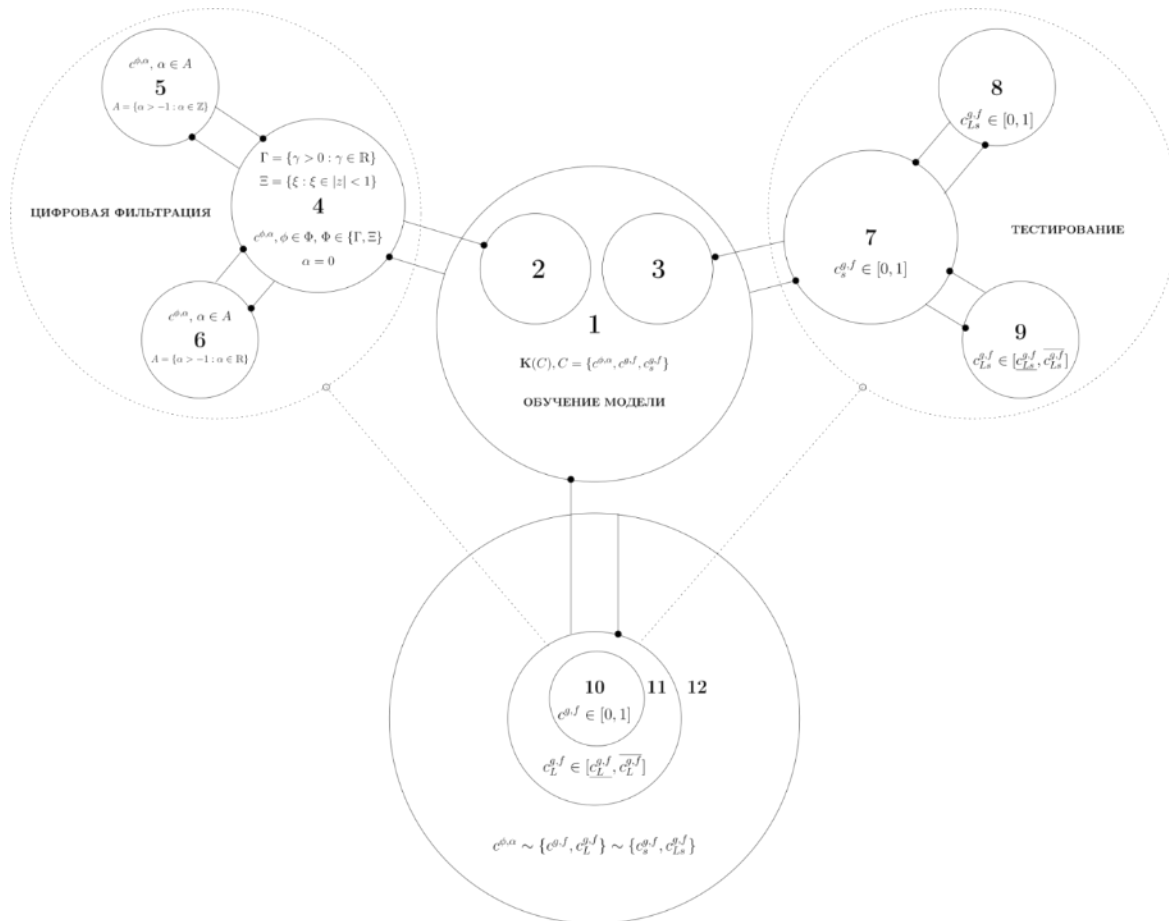


Рис. 1. Модель перехода от процессов кратковременной памяти к процессам долговременной памяти на основе динамической когнитивной карты

Благодарности

Работа выполнена при поддержке грантов Президента РФ № МК-6218.2018.9 и Минобрнауки РФ № 074-U01, а также Российского фонда фундаментальных исследований, проект № 18-37-00219.

Литература

[1] Neyshabur B., Tomioka R., Srebro N. In search of the real inductive bias: On the role of implicit regularization in deep learning. arXiv:1412.6614. - 2014.
 [2] Neyshabur B. Implicit regularization in deep learning. arXiv:1709.01953.
 [3] Neyshabur B., Tomioka R., Salakhutdinov R., Srebro N. Geometry of optimization and implicit regularization in deep learning. arXiv:1705.03071.
 [4] Tolman E.C. Cognitive maps in rats and men//Psychological Review. - 1948. - 55(4). - pp. 189-208.
 [5] Anderson M.C., Bjork R.A., Bjork E.L. Remembering can cause forgetting: Retrieval dynamics in long-term memory//Journal of Experimental Psychology: Learning, Memory, and Cognition. - 1994. - 20(5). - pp. 1063-1087.