

Графические модели для предсказания CTR рекламных объявлений на поиске

Вячеслав Алипов

Яндекс

11 марта 2014 г.

Обзор

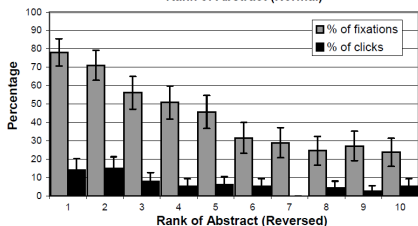
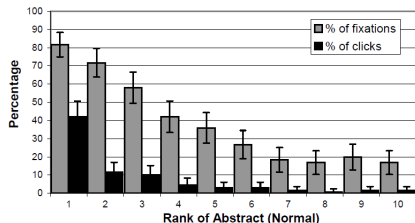
- 1 Трудности при оценке CTR
- 2 Examination Hypothesis
 - User Browsing Model
- 3 Cascade Hypothesis
 - General Click Model
- 4 Учет взаимного влияния объявлений
 - Conditional Random Field
- 5 Корректировка предсказаний с помощью CRF моделей
 - Постановка задачи
 - Позиционная модель
 - Квадратичная модель
 - Обобщенная линейная модель
- 6 Результаты
- 7 Планы

План

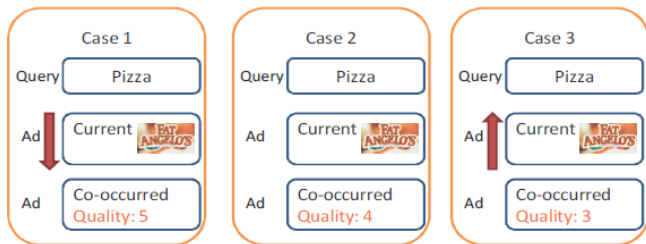
- 1 Трудности при оценке CTR
- 2 Examination Hypothesis
 - User Browsing Model
- 3 Cascade Hypothesis
 - General Click Model
- 4 Учет взаимного влияния объявлений
 - Conditional Random Field
- 5 Корректировка предсказаний с помощью CRF моделей
 - Постановка задачи
 - Позиционная модель
 - Квадратичная модель
 - Обобщенная линейная модель
- 6 Результаты
- 7 Планы

Позиционный эффект

- ▶ Click-Through Rate (CTR) на поисковой выдаче
- ▶ Т. Joachims et al. (2002)



Взаимное влияние объявлений



► Xin et al., 2012

Кликовые модели

- ▶ Моделируют поведение пользователя
- ▶ Применение
 - Предсказание CTR
 - Улучшение NDCG
 - Выбор рекламных объявлений
 - ...
 - Оценка релевантности документов
 - Замена пользовательских оценок
 - Факторы для ранжирования
 - ...

План

- 1 Трудности при оценке CTR
- 2 Examination Hypothesis**
 - User Browsing Model
- 3 Cascade Hypothesis
 - General Click Model
- 4 Учет взаимного влияния объявлений
 - Conditional Random Field
- 5 Корректировка предсказаний с помощью CRF моделей
 - Постановка задачи
 - Позиционная модель
 - Квадратичная модель
 - Обобщенная линейная модель
- 6 Результаты
- 7 Планы

Examination Hypothesis

- ▶ Click \iff Examined AND Relevant

$$P(C = 1|q, u, i) = \underbrace{P(C = 1|u, q, E = 1)}_{r_{u,q}} \cdot \underbrace{P(E = 1|i)}_{x_i}$$

Examination Hypothesis

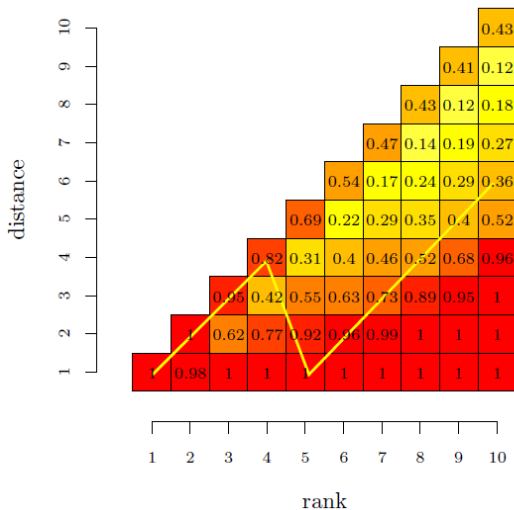
- ▶ Click \iff Examined AND Relevant

$$P(C = 1|q, u, i) = \underbrace{P(C = 1|u, q, E = 1)}_{r_{u,q}} \cdot \underbrace{P(E = 1|i)}_{x_i}$$

- ▶ UBM (User Browsing Model) (G. Dupret et al. 2008)
- ▶ Учитывает предыдущую кликнутую позицию l

$$P(C = 1|q, u, i, l) = \underbrace{P(C = 1|u, q, E = 1)}_{r_{u,q}} \cdot \underbrace{P(E = 1|i, l)}_{x_{i,l}}$$

$$P(E = 1 | i, l)$$



План

- 1 Трудности при оценке CTR
- 2 Examination Hypothesis
 - User Browsing Model
- 3 Cascade Hypothesis**
 - General Click Model
- 4 Учет взаимного влияния объявлений
 - Conditional Random Field
- 5 Корректировка предсказаний с помощью CRF моделей
 - Постановка задачи
 - Позиционная модель
 - Квадратичная модель
 - Обобщенная линейная модель
- 6 Результаты
- 7 Планы

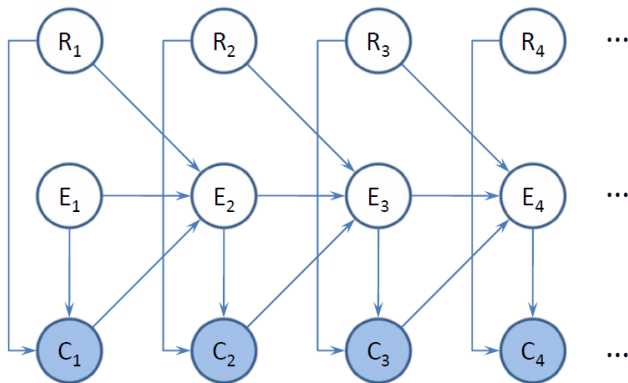
Cascade Model

- ▶ N. Craswell et al. (2008)
- ▶ Пользователь просматривает ссылки сверху вниз
- ▶ После клика сессия завершается
 - $P(E_1 = 1) = 1$
 - $P(E_{i+1} = 1 | E_i = 0) = 0$
 - $P(E_{i+1} = 1 | E_i = 1, C_i) = 1 - C_i$
 - $P(C_i = 1 | E_i = 1) = r_{u_i, q}$

Click Chain Model (CCM)

- ▶ F. Guo et al. (2009)
- ▶ Расширение Cascade Model
 - $P(E_1 = 1) = 1$
 - $P(E_{i+1} = 1 | E_i = 0) = 0$
 - $P(E_{i+1} = 1 | E_i = 1, C_i) = 1 - C_i$
 - $P(E_{i+1} = 1 | E_i = 1, C_i = 0) = \alpha_1$
 - $P(E_{i+1} = 1 | E_i = 1, C_i = 1) = \alpha_2(1 - r_{u_i,q}) + \alpha_3 r_{u_i,q}$
 - $P(C_i = 1 | E_i = 1) = r_{u_i,q}$

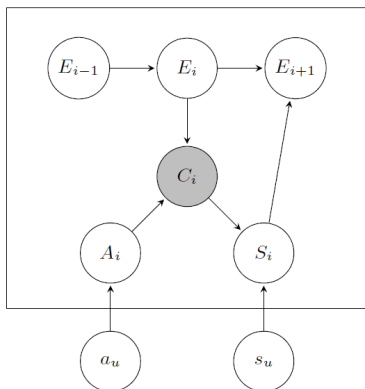
Click Chain Model (CCM)



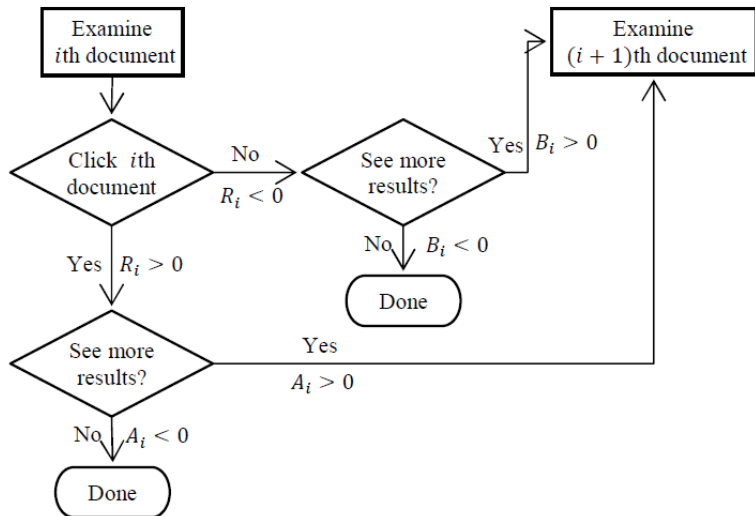
Dynamic Bayesian Network (DBN)

- ▶ O. Chapelle, Y. Zhang (2009)
- ▶ Расширение Cascade Model
 - $P(E_1 = 1) = 1$
 - $P(E_{i+1} = 1 | E_i = 0) = 0$
 - $P(E_{i+1} = 1 | E_i = 1, C_i) = 1 - C_i$
 - $P(E_{i+1} = 1 | E_i = 1, C_i = 0) = \gamma$
 - $P(E_{i+1} = 1 | E_i = 1, C_i = 1) = \gamma(1 - s_{u_i, q})$
 - $P(C_i = 1 | E_i = 1) = r_{u_i, q}$

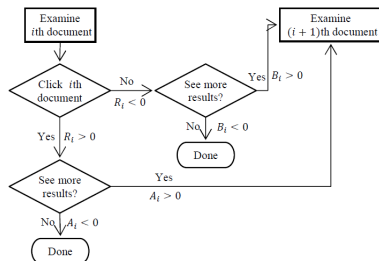
Dynamic Bayesian Network (DBN)



General Click Model (GCM) — Внешняя модель



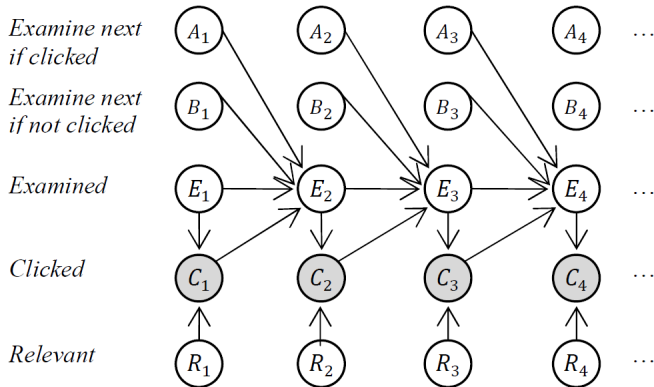
GCM — Внешняя модель



- Вероятности перехода зависят от непрерывных случайных величин

- $P(E_1 = 1) = 1$
- $P(E_{i+1} = 1 | E_i = 0) = 0$
- $P(E_{i+1} = 1 | E_i = 1, C_i = 0, B_i) = I(B_i > 0)$
- $P(E_{i+1} = 1 | E_i = 1, C_i = 1, A_i) = I(A_i > 0)$
- $P(C_i = 1 | E_i = 1, R_i) = I(R_i > 0)$

GCM — Внешняя модель



GCM — Внутренняя модель

▶ Факторы пользователя:

- $f_1^{\text{user}}, f_2^{\text{user}}, \dots, f_s^{\text{user}}$
- запрос
- регион
- user-agent
- местное время
- IP-адрес
- длина запроса
- ...

▶ Факторы URL-а на позиции i :

- $f_{i,1}^{\text{url}}, f_{i,2}^{\text{url}}, \dots, f_{i,t}^{\text{url}}$
- URL
- категория URL-а
- ключевая фраза
- ...

GCM — Внутренняя модель

- Каждому значению f каждого из факторов соответствует три нормальных случайных величины θ_f^A , θ_f^B и θ_f^R

$$\bullet A_i = \sum_{j=1}^s \theta_{f_j^{\text{user}}}^A + \sum_{j=1}^t \theta_{f_{i,j}^{\text{url}}}^A + err$$

$$\bullet B_i = \sum_{j=1}^s \theta_{f_j^{\text{user}}}^B + \sum_{j=1}^t \theta_{f_{i,j}^{\text{url}}}^B + err$$

$$\bullet R_i = \sum_{j=1}^s \theta_{f_j^{\text{user}}}^R + \sum_{j=1}^t \theta_{f_{i,j}^{\text{url}}}^R + err$$

Обучение



- ▶ Online one-pass обучение
- ▶ Байесовский вывод с помощью Expectation Propagation (Tom Minka)
- ▶ Реализован в фреймворке Infer.NET от Microsoft Research

План

- 1 Трудности при оценке CTR
- 2 Examination Hypothesis
 - User Browsing Model
- 3 Cascade Hypothesis
 - General Click Model
- 4 Учет взаимного влияния объявлений**
 - Conditional Random Field
- 5 Корректировка предсказаний с помощью CRF моделей
 - Постановка задачи
 - Позиционная модель
 - Квадратичная модель
 - Обобщенная линейная модель
- 6 Результаты
- 7 Планы

Схожесть текстов

iTunes ® Official Store

Download the Latest iTunes Music,
Movies & More from the iTunes Store
www.Apple.com/iTunes CTR=0.26

Ask Tech Support Now

18 Tech Support Reps Are Online.
Ask a Question, Get an Answer ASAP.
Tech-Support.JustAnswer.com

iTunes ® Official Store

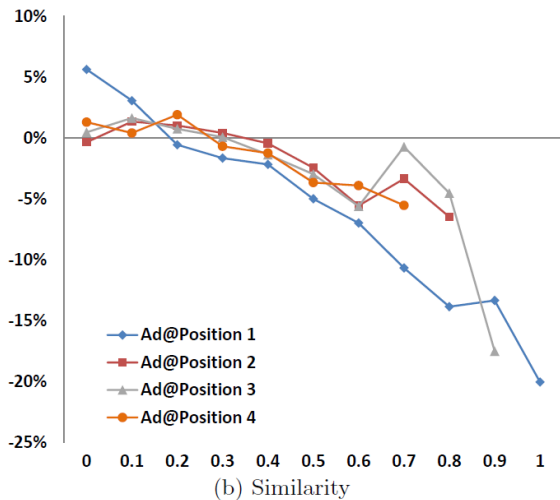
Download the Latest iTunes Music,
Movies & More from the iTunes Store
www.Apple.com/iTunes CTR=0.18

Apple iTunes® Downloads

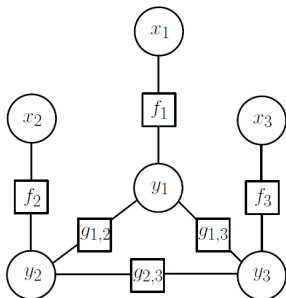
Official iTunes Downloads Music,
Movies, TV-Shows For iPod-iPad-
iPhone
www.AppleITunesDownloads.com

- ▶ CTR зависит от схожести на объявления-соседи

Изменение CTR от схожести на соседей



Conditional Random Field



- Y — вектор CTR'ов/log odds CTR'ов ($\text{logodds}(p) = \log\left(\frac{p}{1-p}\right)$)

$$P(Y|X) = \frac{1}{Z(X)} \exp \left[\sum_i h(y_i, X; w) + \sum_{j>i} \beta g(y_i, y_j, X) \right]$$

Conditional Random Field

- ▶ $h(y_i, X; w) = -(y_i - f(x_i; w))^2 = -(y_i - x_i^T w)^2$
- ▶ $g(y_i, y_j, X) = -s_{ij}(y_i + y_j)$,
где s_{ij} — похожесть текстов объявлений i и j

$$P(Y|X) = \frac{1}{Z(X)} \exp \left[\sum_i -(y_i - x_i^T w)^2 + \sum_{j>i} -\beta s_{ij}(y_i + y_j) \right]$$

Обучение и предсказание

- ▶ оценка w и β с помощью MLE
- ▶ оптимизация правдоподобия градиентным спуском
- ▶ $Y^* = \arg \max_Y P(Y|X)$

План

- 1 Трудности при оценке CTR
- 2 Examination Hypothesis
 - User Browsing Model
- 3 Cascade Hypothesis
 - General Click Model
- 4 Учет взаимного влияния объявлений
 - Conditional Random Field
- 5 Корректировка предсказаний с помощью CRF моделей**
 - Постановка задачи
 - Позиционная модель
 - Квадратичная модель
 - Обобщенная линейная модель
- 6 Результаты
- 7 Планы

Постановка задачи

- ▶ Пусть имеется алгоритм предсказания CTR для каждого объявления независимо

$$CTR \stackrel{\text{def}}{=} P(C = 1 | \text{ad}, \text{user}, \text{query})$$

Постановка задачи

- ▶ Пусть имеется алгоритм предсказания CTR для каждого объявления независимо

$$CTR \stackrel{\text{def}}{=} P(C = 1 | \text{ad}, \text{user}, \text{query})$$

- ▶ Для показа отобрали 3 конкретных объявления с предсказаниями:

$$X = \begin{bmatrix} P(C = 1 | \text{ad}_1, \text{user}, \text{query}) \\ P(C = 1 | \text{ad}_2, \text{user}, \text{query}) \\ P(C = 1 | \text{ad}_3, \text{user}, \text{query}) \end{bmatrix}$$

Постановка задачи

- ▶ Пусть имеется алгоритм предсказания CTR для каждого объявления независимо

$$CTR \stackrel{\text{def}}{=} P(C = 1 | \text{ad}, \text{user}, \text{query})$$

- ▶ Для показа отобрали 3 конкретных объявления с предсказаниями:

$$X = \begin{bmatrix} P(C = 1 | \text{ad}_1, \text{user}, \text{query}) \\ P(C = 1 | \text{ad}_2, \text{user}, \text{query}) \\ P(C = 1 | \text{ad}_3, \text{user}, \text{query}) \end{bmatrix}$$

- ▶ Можно ли уточнить прогноз, зная только X ?

Постановка задачи

- ▶ Пусть имеется алгоритм предсказания CTR для каждого объявления независимо

$$CTR \stackrel{\text{def}}{=} P(C = 1 | \text{ad}, \text{user}, \text{query})$$

- ▶ Для показа отобрали 3 конкретных объявления с предсказаниями:

$$X = \begin{bmatrix} P(C = 1 | \text{ad}_1, \text{user}, \text{query}) \\ P(C = 1 | \text{ad}_2, \text{user}, \text{query}) \\ P(C = 1 | \text{ad}_3, \text{user}, \text{query}) \end{bmatrix}$$

- ▶ Можно ли уточнить прогноз, зная только X ?
- ▶ ... зная другие характеристики объявлений?

Постановка задачи

- ▶ Y — вектор CTR'ов/log odds CTR'ов
- ▶ Моделируем условное распределение

$$P(Y|X) = \frac{1}{Z(X)} \exp[-E(Y|X)]$$

- ▶ $(X_q, Y_q)_{q=1}^n$ — обучающая выборка
- ▶ Y_q — клики/CTR/logodds(CTR)

Позиционная модель

- ▶ Добавляем унарные потенциалы

$$E(Y|X) = (Y - X)^T(Y - X) + W^T Y$$

Позиционная модель

- ▶ Добавляем унарные потенциалы

$$E(Y|X) = (Y - X)^T(Y - X) + W^T Y$$

- ▶ W находится как MLE
- ▶ Оптимальное предсказание

$$Y^* = X - \frac{1}{2}W = X - \frac{1}{n} \sum_{q=1}^n (X_q - Y_q)$$

Тест

- ▶ Выборка из показов за февраль 2014
- ▶ Обучающая выборка: 5 млн. показов
- ▶ Тестовая выборка: 2 млн. показов

Модель	LL	RMSE	AUC	Precision@1
Позиционная	+0.11%	-0.04%	+0.38%	+2.36%

Квадратичная модель

- ▶ Учим бинарные потенциалы вида $w_{ij}y_iy_j$

$$E(Y|X) = (Y - X)^T(Y - X) + Y^T W Y$$

Квадратичная модель

- ▶ Учим бинарные потенциалы вида $w_{ij}y_iy_j$

$$E(Y|X) = (Y - X)^T(Y - X) + Y^T W Y$$

- ▶ W находится как MLE
- ▶ Оптимальное предсказание

$$Y^* = [I + W]^{-1} X$$

- ▶ Можно выучить линейные члены, добавив измерение к X и Y

Тест

Модель	LL	RMSE	AUC	Precision@1
Позиционная	+0.11%	-0.04%	+0.38%	+2.36%
Квадратичная	+0.10%	-0.03%	+0.34%	+2.21%

Тест

Модель	LL	RMSE	AUC	Precision@1
Позиционная	+0.11%	-0.04%	+0.38%	+2.36%
Квадратичная	+0.10%	-0.03%	+0.34%	+2.21%

- ▶ В 40% случаев порядок сортировки по Y отличается от порядка по X !

Обобщенная линейная модель

- ▶ Учим веса при сумме CTR'ов двух объявлений как линейные функции от факторов пары объявлений

$$E(Y|X) =$$

$$(Y - X)^T(Y - X) + \sum_{i < j} f(x_i, x_j; w) \cdot (y_i + y_j) =$$

$$(Y - X)^T(Y - X) + \sum_{i < j} (w_1 f_1(x_i, x_j) + \dots + w_m f_m(x_i, x_j)) \cdot (y_i + y_j) =$$

$$(Y - X)^T(Y - X) + W^T A^T(X) Y$$

$$a_{ik}(X) = \sum_{j \neq i} f_k(x_i, x_j)$$

Обобщенная линейная модель

- ▶ Учим веса при сумме CTR'ов двух объявлений как линейные функции от факторов пары объявлений

$$E(Y|X) =$$

$$(Y - X)^T(Y - X) + \sum_{i < j} f(x_i, x_j; w) \cdot (y_i + y_j) =$$

$$(Y - X)^T(Y - X) + \sum_{i < j} (w_1 f_1(x_i, x_j) + \dots + w_m f_m(x_i, x_j)) \cdot (y_i + y_j) =$$

$$(Y - X)^T(Y - X) + W^T A^T(X) Y$$

$$a_{ik}(X) = \sum_{j \neq i} f_k(x_i, x_j)$$

- ▶ W находится как MLE
- ▶ Оптимальное предсказание

$$Y^* = X - \frac{1}{2} A(X) W$$

Обобщенная линейная модель

- ▶ В модель легко добавляется учет позиционности

$$E(Y|X) = (Y - X)^T(Y - X) + [A(X)W_1 + W_2]^T Y$$

$$Y^* = X - \frac{1}{2}[A(X)W_1 + W_2]$$

Тест

Модель	LL	RMSE	AUC	Precision@1
Позиционная	+0.11%	-0.04%	+0.38%	+2.36%
Квадратичная	+0.10%	-0.03%	+0.34%	+2.21%
Обобщенная линейная	+0.12%	-0.04%	+0.38%	+2.37%

- ▶ В качестве фичей пока — функции от предсказанных CTR'ов x_i

Обобщенная линейная модель

- ▶ X — матрица «объект-признак»

$$E(Y|X) = (Y - XU)^T(Y - XU) + W^T A^T(X)Y$$

Обобщенная линейная модель

- ▶ X — матрица «объект-признак»

$$E(Y|X) = (Y - XU)^T(Y - XU) + W^T A^T(X)Y$$

- ▶ Модель похожести текстов — частный случай

$$\sum_i (y_i - x_i^T w)^2 + \sum_{i < j} \beta s_{ij} (y_i + y_j) = (Y - XW)^T(Y - XW) + \beta^T S^T(X)Y$$

Обобщенная линейная модель

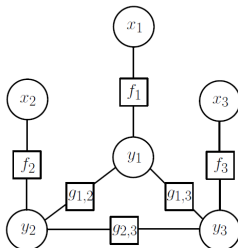
- ▶ X — матрица «объект-признак»

$$E(Y|X) = (Y - XU)^T(Y - XU) + W^T A^T(X)Y$$

- ▶ Модель похожести текстов — частный случай

$$\sum_i (y_i - x_i^T w)^2 + \sum_{i < j} \beta s_{ij} (y_i + y_j) = (Y - XW)^T(Y - XW) + \beta^T S^T(X)Y$$

- ▶ Но тогда модель из статьи неверна!



Обобщенная линейная модель

- ▶ При наблюдаемом X распределение факторизуется по y_i

$$\begin{aligned} E(Y|X) &= (Y - XU)^T(Y - XU) + W^T A^T(X)Y = \\ &= (Y - XU)^T(Y - XU) + V(X)^T Y = \\ &= \sum_{i=1}^k (y_i - x_i^T u_i)^2 + \sum_{i=1}^k v_i(X) y_i = \\ &= \sum_{i=1}^k [(y_i - x_i^T u_i)^2 + v_i(X) y_i] = \\ &= \sum_{i=1}^k \log \Phi(y_i|X) \end{aligned}$$

Обобщенная линейная модель

- ▶ Задача обучения сводится к МНК и решается аналитически

$$\begin{aligned} P(Y|X) &= \\ &= \frac{1}{Z(X)} \exp[-(Y - XU)^T(Y - XU) - W^T A^T(X)Y] = \\ &= \frac{1}{Z'(X)} \exp \left[- \left(Y - \left(XU - \frac{1}{2} A(X)W \right) \right)^T \left(Y - \left(XU - \frac{1}{2} A(X)W \right) \right) \right] = \\ &= \frac{1}{Z'(X)} \exp[-(Y - A'(X)W')^T(Y - A'(X)W')] \end{aligned}$$

Обобщенная линейная модель

- ▶ Задача обучения сводится к МНК и решается аналитически

$$\begin{aligned}
 P(Y|X) &= \\
 &= \frac{1}{Z(X)} \exp[-(Y - XU)^T(Y - XU) - W^T A^T(X)Y] = \\
 &= \frac{1}{Z'(X)} \exp \left[- \left(Y - \left(XU - \frac{1}{2} A(X)W \right) \right)^T \left(Y - \left(XU - \frac{1}{2} A(X)W \right) \right) \right] = \\
 &= \frac{1}{Z'(X)} \exp[-(Y - A'(X)W')^T(Y - A'(X)W')]
 \end{aligned}$$

- ▶ Или обучаем любой другой классификатор с новыми парными фичами

$$y_i^* = F(x_i, a_i(X); \Theta)$$

План

- 1 Трудности при оценке CTR
- 2 Examination Hypothesis
 - User Browsing Model
- 3 Cascade Hypothesis
 - General Click Model
- 4 Учет взаимного влияния объявлений
 - Conditional Random Field
- 5 Корректировка предсказаний с помощью CRF моделей
 - Постановка задачи
 - Позиционная модель
 - Квадратичная модель
 - Обобщенная линейная модель
- 6 Результаты**
- 7 Планы

Результаты

- ▶ Смоделирован позиционный эффект с помощью CRF
- ▶ Составлена и решена модель с мультипликативными взаимодействиями между CTR объявлений блока
- ▶ Формализован класс линейных моделей с попарными взаимодействиями
- ▶ Найдено аналитическое решение для этого класса моделей

План

- 1 Трудности при оценке CTR
- 2 Examination Hypothesis
 - User Browsing Model
- 3 Cascade Hypothesis
 - General Click Model
- 4 Учет взаимного влияния объявлений
 - Conditional Random Field
- 5 Корректировка предсказаний с помощью CRF моделей
 - Постановка задачи
 - Позиционная модель
 - Квадратичная модель
 - Обобщенная линейная модель
- 6 Результаты
- 7 Планы

Планы

- ▶ Найти информативное множество признаков для обобщенной линейной модели
 - функции от существующих признаков
 - принципиально новые функции от пары объявлений
- ▶ Исследовать другие штрафы за отклонения от прогноза
- ▶ Обучить свой унарный потенциал вместо прогноза базового алгоритма
- ▶ Сравнить результаты с обучением отдельного классификатора с новыми парными фичами
- ▶ Исследовать мультипликативные парные потенциалы
- ▶ Исследовать более сложные k -нарные потенциалы

Спасибо за внимание!
Вопросы?