Identifying Key Observers to Find Popular Information in Advance (IJCAI 2016)

<u>小西 卓哉(NII)</u> 岩田 具治(NTT) 林 浩平(産総研) 河原林 健一(NII)

ウェブ上の情報共有

- アイテム(情報):
 - メッセージ (Twitter, Facebook)
 - 写真 / 動画(Instagram)
 - ウェブページ(Delicious)
 - 製品 (Amazon)
 - ホテル/レストラン (TripAdvisor, Yelp)



ウェブ上の情報共有

- アイテム(情報):
 - メッセージ (Twitter, Facebook)
 - 写真 / 動画 (Instagram)
 - ウェブページ(Delicious)
 - 製品(Amazon)
 - ホテル/レストラン(TripAdvisor, Yelp)
- 選択(アイテムの共有行動):
 - Twitterにおけるリツイート
 - ウェブページのブックマーク
 - 商品/ホテルへのレビュー





アイテムの流行予測

- 将来流行しそうなアイテムを事前に察知したい
- モチベーションは多数
 - 有名になる前においしいレストランを見つけたい
 - 将来流行りそうな研究トピックを見つけたい
 - 競合他社に先駆けたマーケティング戦略を立案したい
- 既存研究
 - [Kupavskii et al., CIKM 2012]
 - [Li et al., Social Media Processing 2014]
 - [Kong et al., Journal of Computational Information Systems 2014]



出典元: <u>https://asd-hs.wikispaces.com</u> /file/view/you-can-compete.jpg/43564 1616/you-can-compete.jpg

アイテムA



アイテムB



アイテムC







アイテムB



アイテムC





























アイテムB













アイテムB





アイテムC







流行前に選択することに成功したユーザ



流行



















流行



















観測者を経由した流行アイテムの早期検知

- 観測者: (サービスの) エキスパート, ヘビーユーザなど
- アイテム自体の予測と比べたメリット
 ユーザ推薦への応用:簡単にカスタマイズできる
 - エキスパートからの知識の受容
- どのように有用な観測者をみつけるか?
 - 流行しないアイテムを避けつつ流行アイテムを選択する観測者が良い



問題設定

- イベントログデータ $E = \{(i_e, u_e, t_e)\}_{e=1}^{|E|}$ が与えられていると仮定 – イベント (i, u, t):ユーザ $u \in U$ がアイテム $i \in I$ を時刻 $t \leq T$ で選択
- 観測者として適当な行動:
 - 最終的な選択回数が多い上位*s*%のアイテムを*m*人のユーザが選択する 前に選択



 $\mathbf{y} = \sigma(\mathbf{w}^{\top}\mathbf{x})$ 線形分類器







⇒非ゼロ要素に対応するユーザを観測者として抽出する

特徴量の設計

• 手法1: (問題設定に従い) 最初の *m* 人のユーザを使用



特徴量の設計

• 手法1: (問題設定に従い) 最初の *m* 人のユーザを使用



- データがもつ多くの情報を失ってしまう
- 閾値を超えるユーザの情報
- ユーザの重要度:初期に選択したユーザほど反応が早く有用

特徴量の設計

• 手法1: (問題設定に従い)最初の*m*人のユーザを使用



• 手法2:閾値を確率的に変動させてデータを増大





• 手法1:Non-augmented method

$$\mathbf{w}^{*} = \operatorname*{argmin}_{\substack{\hat{\mathbf{w}} \in \mathbb{R}_{\geq 0}^{|U|}, w_{0} \in \mathbb{R} \\ i = 1}} \left\{ \sum_{i=1}^{|I|} \ell\left(\mathbf{w}, \mathbf{x}_{i}, y_{i}\right) + \lambda ||\hat{\mathbf{w}}||_{1} \right\}$$

非負制約

• 手法2:Data Augmented method

$$\mathbf{w}^{*} = \operatorname*{argmin}_{\hat{\mathbf{w}} \in \mathbb{R}_{\geq 0}^{|U|}, w_{0} \in \mathbb{R}} \left\{ \sum_{i=1}^{|I|} \mathbb{E}_{p(z)} [\ell(\mathbf{w}, \mathbf{x}_{i}(z), y_{i})] + \lambda ||\hat{\mathbf{w}}||_{1} \right\}$$

Augmented損失

- p(z):閾値の確率分布

⇒ 今回はWeibull分布を使用(クロスバリデーションで推定)

- 分布を固定すれば確率最適化で学習できる(e.g. Adagrad)

観測者数の指定

- 正則化パラメータ *λ* に依存
 - 値が大きいほど、抽出される観測者数は小さくなる
- L1正則化ではパラメータの非ゼロ数を予め指定できない
 - ⇒ 解決法: Forward-backward探索

Forward-backward探索

*n*人の観測者を取得したい場合

1. Forward search:正則化パラメータの値を変化させながら探索



Forward-backward探索

n人の観測者を取得したい場合

2. Backward interpolation:より大きな観測者集合で補完



Forward-backward探索

*n*人の観測者を取得したい場合

2. Backward interpolation:より大きな観測者集合で補完





- Deliciousデータセットを用意 [Wetzker et al., ECAI 2008]
 ソーシャルブックマークサービスの履歴を記録
- ・ m=10 , s=10 に設定
 - 最初の10人までに将来上位10%に入るアイテムを検知する観測者
- 訓練データから観測者を抽出
- 抽出した観測者を利用して将来流行するアイテムを判定
 判定方法は次スライドで説明
- テストデータに対するF値で評価

判定の手順

観測者の選択したアイテムは無条件に受け取る場合を想定

- テストデータアイテムの初期10人のみ観測できたと仮定
- 少なくとも1人の観測者が初期ユーザに含まれていれば、そのアイ テムを流行すると判定





N followers

N early

Non-augmented

Augmented

結果:観測者数-F値曲線のAUCの比較

Nearly-Dataset Random **Nadoptions N**followers Nearly Nearly-pos pos/neg Non-augmented Augmented 0.315 0.051 0.168 0.200 0.172 0.252 0.133 0.305 ajax 0.344 0.034 0.135 0.174 0.156 0.293 0.178 0.332 css design 0.024 0.136 0.164 0.060 0.276 0.276 0.164 0.244 0.228 0.058 0.190 0.210 0.204 0.113 0.246 0.247 java 0.082 0.304 javascript 0.052 0.134 0.183 0.129 0.240 0.299 0.298 linux 0.056 0.171 0.194 0.195 0.266 0.102 0.299 0.306 0.374 0.376 0.085 0.163 0.247 0.224 0.327 news 0.256 0.082 0.154 0.177 0.191 0.235 0.118 0.255 opensource 0.047 0.183 0.223 0.071 0.236 0.240 0.169 0.202 photography 0.085 0.170 0.189 0.183 0.218 0.143 0.241 0.241 science 0.034 0.168 0.170 0.291 0.132 0.336 0.350 webdesign 0.143 0.295 0.055 0.157 0.192 0.179 0.256 0.131 0.291 Average

Non-augmented method

Augmented method



- ウェブ上で流行するアイテムの観測者を抽出する問題
- 特徴選択を応用した抽出手法を提案
 - Augmented method:損失関数を閾値の期待値として定義することで
 データがもつ情報を効率的に利用する
- 実データに対して有効性を確認