

8章 価格モデルの構築(3)

サ ミンソン

8. 8実データの利用 — eBay API

- eBay - オンラインオークションサイト
 - フリーのXMLベースのAPIが存在し、検索やアイテム詳細情報の取得などに使われる
 - 本節で学ぶもの
 - 価格データを取得する方法
 - データを変換して価格推測に利用する方法

8. 8. 1 ディベロッパキーの取得

- eBayのAPIにアクセスするには…
Quick Start Guideを利用する
(<http://developer.ebay.com/quickstartguide>)

↑で開発者アカウントを作ると

- ・ディベロッパキー
- ・アプリケーションキー
- ・証明書キー
- ・認証トークン

が得られる

8. 8. 2コネクションのセットアップ

- キーが得られたら、eBay APIへのコネクションをセットアップする。
 - ①キーやコール名など、数多くのヘッダを渡す
getHeader関数を作る。
 - ②リクエストのパラメータを入れたXMLを送る
sendrequest関数を作る。

8. 8. 3検索する

- APIコールに渡すパラメータは二つだけ考慮する
- Query ー検索語を含む文字列
- CategoryID ー検索したいカテゴリーを指定する数値

doSearch関数はこの二つのパラメータを取って検索を実行する関数

実行結果

```
>>> import ebaypredict
>>> laptops=ebaypredict.doSearch('laptop')
>>> laptops[0:5]
[(u'170344824814', u'1 DELL LATITUDE D610 DVD P4 M WiFi XP-2
  WI FI LAPTOP NR', u'275.05', u'2009-06-16T10:35:26.000Z'),
(u'170344863803', u'1 DELL LATITUDE CDRW DVD P4 M WiFi XP-2
  WI FI LAPTOP NR', u'265.0', u'2009-06-16T13:08:41.000Z'),
(u'120435171738', u'1 DELL LATITUDE CDRW DVD P4 M WiFi XP-2
  WI FI LAPTOP NR', u'259.54', u'2009-06-16T13:20:40.000Z'),
(u'170344872176', u'1 DELL LATITUDE D600 DVD P4 M WiFi XP-2 WI
  FI LAPTOP NR', u'232.5', u'2009-06-16T13:38:41.000Z'),
(u'310148926718', u'DELL LATITUDE 1 DVD CDRW P4 M WiFi XP-2
  WI FI LAPTOP NR', u'220.5', u'2009-06-16T13:40:54.000Z')]
```

実行結果(2)

```
>>> ebaypredict.getCategory('computers')
58058 Computers & Networking
>>> ebaypredict.getCategory('laptops',parentID=58058)
111422 Apple Laptops & Notebooks
177 Laptops & Notebooks
158885 For Laptops & Notebooks
168302 Laptops & Notebooks for Repair
40177 Laptops & Notebooks
>>> laptops=ebaypredict.doSearch('laptop',categoryID=177)
>>> laptops[0:5]
[(u'170344824814', u'1 DELL LATITUDE D610 DVD P4 M WiFi XP-2 WI FI LAPTOP NR', u'275.05',
  u'2009-06-16T10:35:26.000Z'),
 (u'170344863803', u'1 DELL LATITUDE CDRW DVD P4 M WiFi XP-2 WI FI LAPTOP NR' u'265.0',
  u'2009-06-16T13:08:41.000Z'),
 (u'120435171738', u'1 DELL LATITUDE CDRW DVD P4 M WiFi XP-2 WI FI LAPTOP NR', u'259.54',
  u'2009-06-16T13:20:40.000Z'),
 (u'170344872176', u'1 DELL LATITUDE D600 DVD P4 M WiFi XP-2 WI FI LAPTOP NR', u'232.5',
  u'2009-06-16T13:38:41.000Z'),
 (u'310148926718', u'DELL LATITUDE 1 DVD CDRW P4 M WiFi XP-2 WI FI LAPTOP NR', u'220.5',
  u'2009-06-16T13:40:54.000Z')]
```

8. 8. 4 アイテムの詳細を得る

- 検索結果のリストからはタイトルと価格が与えられる。
- eBayはアイテムタイプ特有の属性も提供している
 - laptop:プロセッサ形式、メモリ、ハードディスクの容量など
- こうした詳細を得るには・・・
 - 検索結果が返したアイテムIDを渡してeBay APIのGetItemをコールする

実行結果

```
>>> ebaypredict.getItem(laptops[7][0])
{'attributes': {
  u'13': u'Windows XP Professional', u'12': u'1000', u'14':
    u'Dell',
  u'26445': u'CD-RW/DVD Combo', u'26444': u'1600',
    u'26446': u'14.0',
  u'26443': u'Intel Core Duo', u'183': u'-', u'25710': u'80.0',
    u'42213': u'Dual Core',
  u'10244': u'Refurbished'},
'price': u'202.5', 'bids': u'17', 'feedback': u'51797',
'title': u'1 DELL LATITUDE CDRW DVD P4 M WiFi XP-2 WI
FI LAPTOP NR'}
```

8. 8. 5価格予測器の構築

- eBayから一連のアイテムを取ってきた上で、クロス評価関数に渡せるよう数値のリストに直す
- まず、ラップトップのリストを取得し、それから個別のリクエストをかけるものとする。個々のリクエスト結果について、予測に使える数値リストを属性により生成し、データをK近傍法の関数群に適した構造体にまとめる。

RAM容量、プロセッサ速度、画面サイズ、 フィードバックスコアで価格をK近傍法で推測

```
>>> set1=ebaypredict.makeLaptopDataset()  
...  
>>> numpredict.knnestimate(set1,(512,1000,14,40,1000))  
121.16400000000002  
>>> numpredict.knnestimate(set1,(1024,1000,14,40,1000))  
131.75399999999999  
>>> numpredict.knnestimate(set1,(1024,1000,14,60,0))  
138.49600000000001  
>>> numpredict.knnestimate(set1,(1024,2000,14,60,0))  
216.00799999999998  
>>> numpredict.knnestimate(set1,(2024,2000,14,60,0))  
362.97399999999999  
>>> numpredict.knnestimate(set1,(2024,2000,14,60,1000))  
388.37399999999997
```


HP HDX18t 2.53 QX9300/4/640/HDTV/1080p//Blu ray/V.Ulti.

HD TV Tuner/Bluetooth/802.11N/HPWarranty \$1849

Item number: 26042191

Buyer or seller of this item? [Sign in](#) for your status

[Watch this item](#) in My



FREE shipping

Buy It Now price: **US \$1,849.99** [Buy It Now >](#)

[Immediate payment required](#)

Make no payments for 3 months on this purchase - [eBay MasterCard](#)

End time: **Jul-02-09 17:21:32 PDT (15 days 14 hours)**

Shipping: **FREE shipping**
UPS Ground
Service to [United States](#)
[\(more services\)](#)

Ships to: **Worldwide**

Item location: **Nashua, New Hampshire, United States**

You can also: [Watch This Item](#)
Get [SMS](#) or [IM](#) alerts | [Email to a friend](#)

[View larger picture](#)

Listing and payment details: [Show](#)

Meet the seller

Seller: [driven.to.success](#) (1645 ★) 

Feedback: **99.4 % Positive**

Member: since Mar-14-98 in United States

- [See detailed feedback](#)
- [Ask seller a question](#)
- [Add to Favorite Sellers](#)
- [View seller's other items](#)

Buy safely

1. Check the seller's reputation

Score: 1645 | 99.4% Positive

[See detailed feedback](#)

2. Check how you're protected

Pay with **PayPal** and your [full purchase price](#) is covered | [See terms](#)

Description [\(revised\)](#)

Item Specifics - PC Laptops

Condition: --
Brand: **HP**

Screen Size: **18.4 inch**

Operating System: **Windows Vista Ultimate**

Primary Drive: **Blu-ray**

Processor Type: **Intel Core 2 Quad**

Processor Speed: **2.53 GHz**

Processor Configuration: **Quad Core**

Memory (RAM): **4 GB**

Hard Drive Capacity: **640 GB**

Hard Drive Capacity: **640 GB**

Hard Drive Capacity: **640 GB**

Hard Drive Capacity: **640 GB**

Hard Drive Capacity: **640 GB**

- `>>> numpredict.knnestimate(set1,(4048,2530,18.4,640,1000))`
798.36199999999997

考察

①アイテムの価格を決める全ての属性を考慮すべき

-ex)laptop:processor type,
VGA,ODD, etc.

②オークションのデータでいいのか

—投稿者(一般人)が価格を決める

—(業者でも)条件付でアイテムを安くする

8. 9K近傍法はどこで使うべきか

- K近傍法の欠点:

- 予測器の構築は非常に計算集約的な作業
- 大きなデータセットで良好な解を見つけるには非常に長い時間が掛かる

- K近傍法の利点:

- 新しい観測値の追加には計算力は必要としない
- 予測において重み付けした他の観測値を利用するので、厳密に何が起きているかを解釈するのが容易