Text Mining of Electronic News Content for Economic Research

Panos Ipeirotis

Stern School of Business
New York University

“On the Record”: A Forum on Electronic Media and the Preservation of News
Comparative Shopping

Canon EOS 350D Digital Rebel XT Digital SLR Camera Kit
(3.0MP, 3456x2304, 3x Opt, CompactFlash Slot - SKU: 02098003)
Price Range: $639.99 - $929.59 from 51 Sellers
Description: For convenience, ease of use and no-compromise SLR performance, Canon's Digital Trinity, an 8.0 Megapixel CMOS sensor, Canon's own DIGIC II chip and a secondary small screen in the pentaprism.

PCNATION
Free FedEx 2-3 Day Shipping
$773.95
1537 Reviews

buydig.com
Free FedEx 2-3 Day Shipping
$784.99
19044 Reviews
Comparative Shopping

Canon EOS 350D Digital Rebel XT Digital SLR Camera Kit
(8.0MP, 3456x2304, 3x Opt, CompactFlash Slot - SKU: 02098003)
Price Range: $639.99 - $929.59 from 51 Sellers
Description: For convenience, ease of use and no-compromise SLR performance, Canon's Digital Trinity, an 8.0 Megapixel CMOS sensor, Canon's own DIGIC II Image Processor, 3.0 inch LCD, 3x Optical Zoom, 2.0 x Digital Zoom, 51mm Focal Length. The Canon EOS 350D Digital Rebel XT is an affordable and easy to use mid-range slr with excellent performance and a high-quality 8.0 megapixel CMOS sensor coupled with Canon's DIGIC II Image Processor. This camera is a great entry level camera for the enthusiast photographer sharing the same love for photography as you. This camera is an affordable choice for all photographers who want to upgrade from a digicam to a slr and want to capture high quality images while still keeping costs reasonable.
Are Customers Irrational?

BuyDig.com gets **Price Premium** (customers pay more than the minimum price)

$11.04 (+1.5%)
Price Premiums @ Amazon

Are Customers Irrational (?)

Number of Transactions

Price Premium
Why not Buying the Cheapest?

You buy more than a product

- Customers do not pay only for the product
- Customers also pay for a set of fulfillment characteristics
  - Delivery
  - Packaging
  - Responsiveness
  - ...

Customers care about reputation of sellers!
Example of a reputation profile

v-art
See other items from this seller

At a Glance

Feedback

Feedback Rating: ★★★★★
4.8 stars over the past 12 months (159 ratings)

<table>
<thead>
<tr>
<th>Feedback Period</th>
<th>Positive</th>
<th>Neutral</th>
<th>Negative</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 days</td>
<td>89%</td>
<td>0%</td>
<td>11%</td>
<td>19</td>
</tr>
<tr>
<td>90 days</td>
<td>91%</td>
<td>3%</td>
<td>6%</td>
<td>60</td>
</tr>
<tr>
<td>365 days</td>
<td>95%</td>
<td>2%</td>
<td>3%</td>
<td>159</td>
</tr>
<tr>
<td>Lifetime</td>
<td>95%</td>
<td>2%</td>
<td>3%</td>
<td>159</td>
</tr>
</tbody>
</table>

What do these mean?

Recent Feedback:

4 out of 5: "good"
Date: 11/9/2005 Rated by Buyer: kuniangama

5 out of 5: "Very fast shipping, perfect packaging. No complaints."
Date: 11/3/2005 Rated by Buyer: stoen

5 out of 5: "Order arrived much sooner than expected. Software works fine. Definitely very good service. Thanks!"
Date: 11/3/2005 Rated by Buyer: sengold

5 out of 5: "purchase received promptly"
Date: 11/3/2005 Rated by Buyer: tooody17

1 out of 5: "I have been waiting three weeks for an instock item. This is the worst internet experience I have ever had. I will never order from this company again."
Date: 11/2/2005 Rated by Buyer: George T.

Note: Feedback calculations only include ratings left by buyers.
<table>
<thead>
<tr>
<th>Score</th>
<th>Feedback</th>
<th>Date</th>
<th>Buyer</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>&quot;good&quot;</td>
<td>11/5/2005</td>
<td>kviramgame</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Very fast shipping, perfect packaging. No complaints.&quot;</td>
<td>11/3/2005</td>
<td>steen</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Order arrived much sooner than expected. Software works fine. Definitely very good service. Thanks!&quot;</td>
<td>11/3/2005</td>
<td>sengold</td>
</tr>
<tr>
<td>5</td>
<td>&quot;purchase received promptly&quot;</td>
<td>11/3/2005</td>
<td>twoody17</td>
</tr>
<tr>
<td>1</td>
<td>&quot;I have been waiting three weeks for an instock item. This is the worst internet experience I have ever had. I will never order from this company again.&quot;</td>
<td>11/2/2005</td>
<td>George T.</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Great seller - quick as lightning responses to all my questions -&gt; DELIVERY IN UNDER 48HRS from across the US! -&gt; could use a bit more stuffing in the box tho ;-)&quot;</td>
<td>10/27/2005</td>
<td>evan h.</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Product delivered as promised. Would buy from again.&quot;</td>
<td>10/21/2005</td>
<td>kevin_l1398</td>
</tr>
<tr>
<td>5</td>
<td>&quot;I was expecting it in two days, but it took three. Other than that, no problems.&quot;</td>
<td>10/20/2005</td>
<td>Samuel A.</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Just like I ordered A++&quot;</td>
<td>10/19/2005</td>
<td>staccoy128</td>
</tr>
<tr>
<td>5</td>
<td>&quot;speedy shipping&quot;</td>
<td>10/19/2005</td>
<td>Dayan P.</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Perfect!&quot;</td>
<td>10/15/2005</td>
<td>sperrone82</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Software was shipped promptly and arrived in less than a week. New condition. Seller provided email confirming shipment. Excellent.&quot;</td>
<td>10/14/2005</td>
<td>laser143296</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Quick delivery and exactly as advertised&quot;</td>
<td>10/14/2005</td>
<td>laser143296</td>
</tr>
<tr>
<td>2</td>
<td>&quot;Most of the Useful Books are Locked&quot;</td>
<td>10/11/2005</td>
<td>norske65</td>
</tr>
<tr>
<td>5</td>
<td>&quot;I Item was received ahead of projected delivery and thus far operates correctly.&quot;</td>
<td>10/10/2005</td>
<td>chad j.</td>
</tr>
<tr>
<td>5</td>
<td>&quot;I was in a pinch &amp; he sent it out super fast!! Quality Seller &amp; Quality Service! Thanks very much!&quot;</td>
<td>10/10/2005</td>
<td>Rudy U.</td>
</tr>
<tr>
<td>5</td>
<td>&quot;I was in a pinch &amp; he sent it out super fast!! Quality Seller &amp; Quality Service! Thanks very much!&quot;</td>
<td>10/10/2005</td>
<td>Rudy U.</td>
</tr>
<tr>
<td>5</td>
<td>&quot;package came in timely manner... thanks...&quot;</td>
<td>10/9/2005</td>
<td>rajobhan</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Delivery was pretty quick and box condition was great.. I like that!&quot;</td>
<td>10/9/2005</td>
<td>singh??</td>
</tr>
<tr>
<td>5</td>
<td>&quot;Product received missing a part, but seller resolved quickly. Excellent prices and service, will do business with again.&quot;</td>
<td>10/8/2005</td>
<td>setchmo181</td>
</tr>
<tr>
<td>2</td>
<td>&quot;OVERSELLED ITEM&quot;</td>
<td>10/6/2005</td>
<td>aldo c.</td>
</tr>
</tbody>
</table>
Conjecture:

**Price premiums** measure **reputation**

**Reputation** is captured in **text feedback**

Our contribution:

Examine how **text** affects **price premiums**
Decomposing Reputation

Is reputation just a scalar metric?

- Previous studies assumed a “monolithic” reputation
- Decompose reputation in individual components
- Sellers characterized by a set of **fulfillment characteristics** (packaging, delivery, and so on)

**What are these characteristics (valued by consumers?)**

- We think of each characteristic as a dimension, represented by a **noun, noun phrase, verb** or **verbal phrase** ("shipping", "packaging", "delivery", "arrived")
- We scan the textual feedback to discover these dimensions
Decomposing and scoring reputation

- We think of each characteristic as a dimension, represented by a noun or verb phrase (“shipping”, “packaging”, “delivery”, “arrived”)

- The sellers are rated on these dimensions by buyers using modifiers (adjectives or adverbs), *not numerical scores*
  - “Fast shipping!”
  - “Great packaging”
  - “Awesome unresponsiveness”
  - “Unbelievable delays”
  - “Unbelievable price”

How can we find out the meaning of these adjectives?
Measuring Reputation

- **Regress** textual reputation against price premiums

- Example for “delivery”:
  - **Fast** delivery vs. **Slow** delivery: +$7.95
  - So “**fast**” is better than “**slow**” by a $7.95 margin
### Some Indicative Dollar Values

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>[wonderful experience]</td>
<td>[never received]</td>
</tr>
<tr>
<td>[outstanding seller]</td>
<td>[defective product]</td>
</tr>
<tr>
<td>[excellent quality]</td>
<td>[horrible experience]</td>
</tr>
<tr>
<td>[lightning delivery]</td>
<td>[never sent]</td>
</tr>
<tr>
<td>[highly recommended]</td>
<td>[never received]</td>
</tr>
<tr>
<td>[best seller]</td>
<td>[bad experience]</td>
</tr>
<tr>
<td>[perfectly packaged]</td>
<td>[cancelled order]</td>
</tr>
<tr>
<td>[excellent condition]</td>
<td>[not as advertised]</td>
</tr>
<tr>
<td>[excellent purchase]</td>
<td>[poor packaging]</td>
</tr>
<tr>
<td>[excellent seller]</td>
<td></td>
</tr>
<tr>
<td>[excellent communication]</td>
<td></td>
</tr>
<tr>
<td>$5.86</td>
<td>-$7.56</td>
</tr>
<tr>
<td>$5.76</td>
<td>-$6.82</td>
</tr>
<tr>
<td>$5.27</td>
<td>-$6.79</td>
</tr>
<tr>
<td>$4.84</td>
<td>-$6.69</td>
</tr>
<tr>
<td>$4.15</td>
<td>-$5.29</td>
</tr>
<tr>
<td>$3.80</td>
<td>-$5.26</td>
</tr>
<tr>
<td>$3.74</td>
<td>-$5.01</td>
</tr>
<tr>
<td>$2.70</td>
<td>-$4.87</td>
</tr>
<tr>
<td>$2.38</td>
<td>-$4.39</td>
</tr>
</tbody>
</table>

**Natural** method for extracting sentiment strength and polarity

**good packaging**  
Positive? Yes  
Negative? Yes  
-$0.56

**Naturally** captures the **pragmatic** meaning within the given **context**
Product Reviews and Product Sales

- Examine changes in demand based on published product reviews (newspapers, blogs, online stores)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;excellent lens&quot;</td>
<td>+3%</td>
</tr>
<tr>
<td>&quot;poor lens&quot;</td>
<td>-1%</td>
</tr>
<tr>
<td>&quot;excellent photos&quot;</td>
<td>+6%</td>
</tr>
<tr>
<td>&quot;poor photos&quot;</td>
<td>-2%</td>
</tr>
</tbody>
</table>

- Feature “photos” is two times more important than “lens”
- “Excellent” is positive, “poor” is negative
- “Excellent” is three times stronger than “poor”
Feature Weights for Digital Cameras

Point & Shoot

SLR

- Quality
- Battery
- Resolution
- Size
- Color
- Photos
- Lens
- Screen
Show me the Money!

Broader contribution

- Economic data are affected in many contexts by text
- Economic data are affected in many contexts by news

Applications with Electronic News

- Political News and Prediction Markets
- Financial News and Stock/Option Prices
A prediction market is a market for a contract that yields payments based on the outcome of a partially uncertain future event, such as an election.

A contract pays $100 if candidate X wins the election, and $0 otherwise.

When the market price of an X contract is $60, the prediction market believes that candidate X has a 60% chance of winning the election.
Political News and Prediction Markets

Dec 2, 2007
On my blog

Hillary Clinton

...To put our money where our mouth is, the signal from the last few days shows that Hillary's market price will edge lower in the next few days/weeks...
And suddenly...

We predicted decline here

Why stop here?
An interesting sequence of emails...

Date: Mon, 14 Jan 2008 11:26:27 -0500

Subject: Excessive downloading from licensed database

We have received a complaint from XXXXXXXX about a massive number of articles (over 10,000 per session) being downloaded from their database to a system at Stern, using IP 128.122.130.34 at the times below (Eastern time).

Date: Mon, 14 Jan 2008 12:16:53 -0500

Subject: Excessive downloading from licensed database

Got a call from Jane this morning that Panos has downloaded bulk information from XXXXXXXX last Thursday 10th (2GB download) and Friday 11th (2.5GB download). This is creating a big issue with NYU libraries and XXXXX, with a threat for a bill of up to $250K...

Date: Tue, 15 Jan 2008 15:02:13 -0500

Subject: About XXXXXXX...

It is clear that the interface is meant only for humans, not to download articles for processing with computers...
XML is for humans?
management team was restructured into four main business units on October 1; marine, aviation and transport, led by Mervyn Sugden, property treaty led by David Carson, direct and facultative property led by Patrick Gage and specialty lines led by Adrian Daws. The restructure has been put into operation in time for the main renewal season later in the year.

To view the full story visit

<eLink type="webpage" reference="http://www.thereweb.com/"/>

AROUND THE MARKETS

THIS MONTH: NORTH AMERICA

A recent report from consultants Atlas Partners explores five possible scenarios for the global credit crisis and the implications they could pose for the non-life sector. The most likely scenario, according to the report, is for the economy to slip into a temporary recession. Debt would remain above the normal economic rates. Central banks would reduce interest rates to stimulate recovery.

The underlying demand for insurance would reduce across most economies. Continued price-based competition means a perpetuated soft rating cycle, with the potential for the bottom of the soft market to be reached in two or three years. The decline in demand for AIG cover will result in the allocation of capital to reinsurance operations by reinsurers. And investors' priorities will turn to reinsurance investment opportunities over direct insurance.

Financially a temporary recession would cause reduction in direct premium income and increase in reinsurance premium rates.

To view the rest of this report visit the Around The Markets section of

<eLink type="webpage" reference="http://www.thereweb.com/"/>

GUIDES

STRATEGIC RESTRUCTURING AND REDOMESTICATION

Swiss Re, Scor, Fagostone, XL and Partner Re are just some of the companies who have changed their business structures in recent times. The Review has produced this introductory guide to the opportunities and challenges associated with strategic restructuring and redomestication in Europe in association with KPMG.

To view this publication visit

<eLink type="webpage" reference="http://www.thereweb.com/"/>

The Australian Women’s Weekly © 2008 ACP Publishing Pty Ltd. All rights reserved.

<eLink type="webpage" reference="http://www.ninemsn.com.au/"/>

Politics: Women in the wings

With the US presidential election just days away, the attention of the world is on the women in the race - whether vying for office, or staunchly standing beside their men. Sharon Krum weighs up the scramble for the White House.
Some Lessons

- Unclear if commercial vendors are willing to help researcher when research can lead to competition for their own products

- Researchers increasingly use bigger data set (5-10Tb data sets routinely used by non-CS researchers)

- Need a public, comprehensive repository of archival news, accessible for machine processing

- Allow annotation and tagging from multiple parties to be part of repository

- Build reputational and usage statistics on contributed annotations (to pick the best)
Thank you!

Questions?

http://pages.stern.nyu.edu/~panos
panos@nyu.edu