Causal Impact for App Store Analysis

What does it do?

Measures the impact of an event (intervention) on a metric over time

Impact significant or not?

Confidence interval?

Google uses it for measuring the success of ad campaigns
What about correlation analysis?

**Correlation analysis**

Looks at snapshot of data

Tells us relationship between vectors (+ve or -ve correlation, or no correlation)

**Causal impact analysis**

Looks at time series of data

Tells us how significant an event was
How does it do it?

Trains a predictor (prior time period)

Makes set of predictions (posterior time period)

Compares the observed vector with the predicted vector
Input Vectors

Number of ratings

Target

Compare projection with observed

Controls

Week

App x1

App x2

App x\text{\textit{n}}

Release event
Predictor Model Components
Predictor Model Components

Local trend

\[
\begin{align*}
\mu_{t+1} &= \mu_t + \delta_t + \eta_{\mu,t}, \\
\delta_{t+1} &= \delta_t + \eta_{\delta,t},
\end{align*}
\]

- Local trend value
- Expected increase
- Noise sampled from Normal distribution
Predictor Model Components

Local trend

- Local trend value
- Expected increase

\[ \mu_{t+1} = \mu_t + \delta_t + \eta_{\mu,t}, \]
\[ \delta_{t+1} = \delta_t + \eta_{\delta,t}, \]

Seasonal variance

- Adds seasonal component
- Set length and no. seasons

Noise sampled from Normal distribution
Predictor Model Components

Local trend

\[ \mu_{t+1} = \mu_t + \delta_t + \eta_{\mu,t}, \]
\[ \delta_{t+1} = \delta_t + \eta_{\delta,t}, \]

Seasonal variance

Adds seasonal component
Set length and no. seasons

Control variance

Spike and slab prior
zero coefficients small (equal) coefficients

expected increase

local trend value

noise sampled from Normal distribution
What does it do?

Causal Assumptions

External events that are not accounted for by variances do not apply

Meaning external events must do one of the following:
  - Happen globally
  - Happen in the prior time period
Causal Assumptions

The control data vectors are unaffected by the event (release)
  Non-releasing apps = control set

The relationship between the target and control data vectors is unchanged in the series
  Control set must not contain app or derivatives
Input Metrics

Number of ratings
Number of ratings / week
Rank of Downloads
Rating

Obtain: p-value for each metric, for each release
Results - Scribblenauts Remix

Posterior tail-area probability $p$: 0.00111

The blue region indicates prediction with 95% confidence interval
Apps often have rapid / agile release cycles

McIlroy et al. found that 14% of 10,713 apps updated within 2 weeks
Apps often have rapid / agile release cycles

McIlroy et al. found that 14% of 10,713 apps updated within 2 weeks

Do releases correlate with good performance?

Do releases affect performance?
Dataset

July 2014 - July 2015

Recorded apps that are consistently (every week) in the most popular free or paid lists:

Google Play apps: 307 releases: 1,570
Windows Phone apps: 726 releases: 1,617
### Metrics

#### Developer controlled factors:
- **P** - price
- **RT** - release text

### Performance metrics:
- **R** - rating
- **D** - download rank
- **N** - number of ratings
- **NW** - number of ratings in last week
Do app metrics change over time?
Do app metrics change over time?

D, N and NW have a high standard deviation over 12 months. D, N and NW are likely to change.

R has very small standard deviation. So rating is very stable, unlikely to change.
Do release statistics have a correlation with app performance?
Do release statistics have a correlation with app performance?

No strong correlations are observed

<table>
<thead>
<tr>
<th>Method</th>
<th>R</th>
<th>ΔR</th>
<th>D</th>
<th>ΔD</th>
<th>N</th>
<th>ΔN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.13</td>
</tr>
<tr>
<td>Pearson</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Google

<table>
<thead>
<tr>
<th>Method</th>
<th>R</th>
<th>ΔR</th>
<th>D</th>
<th>ΔD</th>
<th>N</th>
<th>ΔN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman</td>
<td>0.20</td>
<td>-0.17</td>
<td>-</td>
<td>0.32</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>Pearson</td>
<td>0.16</td>
<td>-0.16</td>
<td>-0.09</td>
<td>0.27</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Windows

<table>
<thead>
<tr>
<th>Method</th>
<th>R</th>
<th>ΔR</th>
<th>D</th>
<th>ΔD</th>
<th>N</th>
<th>ΔN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.12</td>
<td>0.13</td>
<td>-</td>
</tr>
<tr>
<td>Pearson</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>R</th>
<th>ΔR</th>
<th>D</th>
<th>ΔD</th>
<th>N</th>
<th>ΔN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman</td>
<td>-</td>
<td>-</td>
<td>0.16</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pearson</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Do releases impact app performance?
Do releases impact app performance?

40% of releases impact performance in Google apps

55% of releases impact performance in Windows apps
What characterises impactful releases?
What characterises impactful releases?

RT - release text
   content
   size
   change in size
P - price
Day - day of release
What characterises impactful releases?

RT - release text
content
size
change in size
P - price
Day - day of release

(new, feature) better than (bug, fix)

Releases that mention
(new, feature) are more likely to be impactful, and to positively affect Rating compared with releases that mention (bug, fix)
What characterises impactful releases?

RT - release text
content
size
change in size
P - price
Day - day of release

Releases with longer release text are more likely to positively impact Rating

(new, feature) better than (bug, fix)

more descriptive release text

Graphs showing distribution of release text size and price for Google and Windows.
What characterises impactful releases?

RT - release text
content
size
change in size
P - price
Day - day of release

(new, feature) better than (bug, fix)
more descriptive release text
higher prices

Releases with higher prices are more likely to positively impact Rating

---

CREST Open Workshop 23/11/15
William Martin
What characterises impactful releases?

RT - release text
- content
- size
- change in size

P - price

Day - day of release

(new, feature) better than (bug, fix)

more descriptive release text

higher prices

Saturday to Tuesday

Releases from Saturday to Tuesday are more likely to be impactful
Conclusions

Causal Impact Analysis can point to significant changes

We look at groups of significant releases to minimise risk of external factors

Useful developer guidelines found that apply to multiple platforms
What about correlation analysis?

<table>
<thead>
<tr>
<th>Correlation analysis</th>
<th>Causal impact analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Looks at snapshot of data</td>
<td>Looks at time series of data</td>
</tr>
<tr>
<td>Tells us relationship between vectors (+ve or -ve correlation, or no correlation)</td>
<td>Tells us how significant an event was</td>
</tr>
</tbody>
</table>

Predictor Model Components

- **Local trend**
  - Expected increase
  - Noise sampled from Normal distribution

- **Seasonal variance**
  - Adds seasonal component
  - Set length and no. seasons

- **Control variance**
  - Spike and slab prior
  - Zero coefficients vs. small (equal) coefficients

What characterises impactful releases?

- **RT** - release text content
  - More descriptive release text
  - (new, feature) better than (bug, fix)

- **P** - price
  - Higher prices

- **App** - number of ratings
  - Controls

- **Week**
  - Comparing projection with observed

- **Day** - day of release
  - Saturday to Tuesday

- **Releases**
  - From Saturday to Tuesday are more likely to be impactful

http://google.github.io/CausallImpact/CausallImpact.html