
Serge DAROLLES 1,2 Gaëlle LE FOL 1,2 Yoann LE NY1,3 Guillaume MONARCHA3

1PSL Research University - Université Paris Dauphine CREST, Paris, France
2INSEE-CREST, Paris, France
3Orion Financial Partners, Paris, France

Objectives
This paper aims to exhibit the importance of user’s influence on social-network content sentiment extraction. We use a dynamic measure of influence to build dynamic networks of communities. Our final objective is to improve the size of the community at each processing phase including text-analysis and Influence community.

Text Analysis
Models:
- Target Extraction: POS tagger
- Target Classification: Conv-Net Based on Word2Vec features
- Sentiment Analysis: Naive-Bayesian classifier trained on a large dataset of 1.6 Million labeled tweets.

Influence Model
With use parametrized Hawkes self-exciting point-process defined with the counting process \( N(t) \) its intensity following:
\[
\lambda(t) = \mu + \int_0^t a(t-u)dN_u
\]
Which, using \( a(t) = ae^{-\beta t} \), leads to:
\[
\lambda_t = \mu + \int_0^t a e^{-\beta(t-u)}dN_u = \mu + \sum_{t-i} e^{-\beta(t-i)}
\]
We estimate parameters by maximizing the Log-likelihood function.

Methodology
The general process of this work involves first a processing phase including text-analysis and Influence modeling then a testing phase to prove the indicator enhanced predictive power.

Results

### Linear Estimation

\[ R_{t+1} = \alpha + \beta \text{VIX}_t + \epsilon_t \]

<table>
<thead>
<tr>
<th>Sentiment Measure</th>
<th>VIX</th>
<th>InSent</th>
<th>( \Delta \text{VIX} )</th>
<th>( \Delta \text{InSent} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>( p\text{-value} )</td>
<td>0.66</td>
<td>0.04</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
<td>0.06</td>
</tr>
</tbody>
</table>

### Two-Factors Model

\[ R_{t+1} = \alpha + \beta_1 \text{InSent}_t + \beta_2 \text{VIX}_t + \epsilon_t, \]

#### SP500 Idx \((t+1)\) T-Stat \( p\)-Value

- Intercept: 0.0000, 0.00, 1.00
- \( \Delta \text{VIX} \): 0.0005, 2.29, 0.02
- \( \Delta \text{InSent} \): 0.0006, 2.76, 0.01

### VAR

\[ Y_{t+1} = c + A_{11} Y_{t+1} + A_{12} Y_{t+1} + \epsilon_{t+1} \]

#### Panel A: \( \Delta \text{VIX} \) T-Stat \( p\)-Value

- Intercept: 2.111, -4.375, -0.126
- \( p\text{-value} \): 0.035, 0.000, 0.900
- R-squared: 0.09

#### Panel B: \( \Delta \text{InSent} \) T-Stat \( p\)-Value

- Intercept: -16.949, -2.355, -0.160
- \( p\text{-value} \): 0.000, 0.019, 0.873
- R-squared: 0.21

Conclusion
- We use a Hawkes-Process Modeling to provide a dynamic measure of social-network users’ influence
- We showed that this Influence metric enhances the significance of the Sentiment signal to predict stock-markets returns
- Provide proofs that this measures provides a supplementary information to predicting market returns

Data Table: Database descriptive statistics

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Users</th>
<th>Total Tweets</th>
<th>Average Daily Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>31-12-2009</td>
<td>192,486</td>
<td>105,430,254</td>
<td>48,141</td>
</tr>
<tr>
<td>31-12-2016</td>
<td>192,486</td>
<td>105,430,254</td>
<td>48,141</td>
</tr>
</tbody>
</table>

Figure: Content volume from dec-2009 to dec-2016.

Figure: Hawkes-Process modeling of Audience-Feature ‘IN-OUT’ for user @Carl_C__Lach from 31-dec-2009 to 31-dec-2016.

Figure: Sentiment-Index (SentIn) and VIX Index from 31-dec-2009 to 31-dec-2016.