TriVis: Visualising Multivariate Data from Sentiment Analysis

Maryanne Doyle
University College Dublin*
Dublin 4

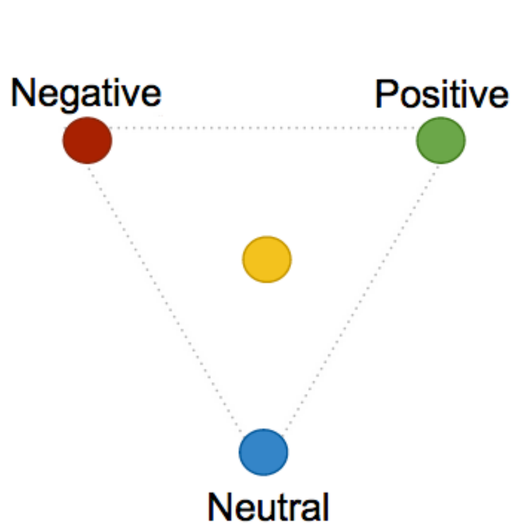
Alan F Smeaton, Adam Bermingham
Dublin City University
Dublin 9

Centre for Data Analytics

Insight

What's the problem?

In a time when a single sporting event can elicit millions of Tweets no individual can keep up with all of the expressions of sentiment produced on platforms such as Facebook and Twitter. Using a computational approach to collect and summarise this data into a more understandable form can offer insights into the reactions of millions of people to any event. Our solution is summarise the data according to sentiment labels and show the results using a novel design.



Labelling sentiment towards golf

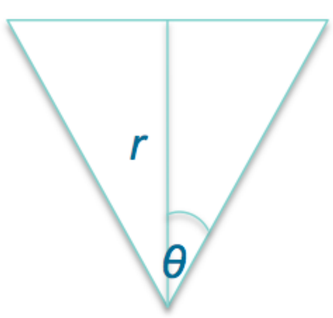
Sentiment is expressed in speech or text that is subjective, opinionated or emotional and includes evaluation or speculation. In our project we collected Tweets about golfers during the 2013 Open and labelled sentiment using a supervised machine learning algorithm.

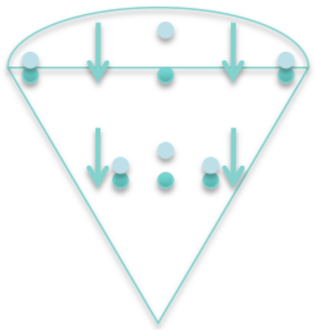
Positive: "Excellent birdie by Phil Mickelson"
Neutral: "Ian Poulter tees off next"
Negative: "Tiger Woods misses an easy shot"

Sentiment was labelled using three classifications (positive, neutral or negative) and the relative frequency of positive, neutral and negative labels were used to calculate an overall sentiment score towards each golfer for a given time period.

Graphing sentiment

First the raw frequency scores for positive, neutral and negative sentiment were normalised so that the sum of *pos*, *neu* and *neg* was always equal to one. These normalised values were converted to a plotable form using polar coordinates (r, θ) to map to a conceptual wedge shape. The horizontal placement was determined by calculating θ using a ratio of negative to positive sentiment scores. The vertical positioning was calculated by mapping the neutral value to r so that a value composed of neutral sentiment only would appear in the bottom vertex and a value with no neutral component would appear along the upper line.





Finally the value for r was scaled using $cos\theta$ so that all points fall within the triangular plot rather than a wedge shape. This prevents points with the same value for r and varying values for θ appearing to have different vertical heights.

The volume of Tweets is conveyed using the area of the plotted point and time is shown as a series of lines which connect plotted points and build up to a footprint or trail over time.

* work carried out at DCU

This project was funded by Science Foundation Ireland under grant number SFI/12/RC/2289 and Enterprise Ireland under grant number CF/2012/2618.

Using TriVis

The sentiment score representing Rory McIlroy (shown in orange) veers towards negative as he plays poorly on days one and two and fails to make the cut to progress to the weekend.

Sentiment towards Phil Mickelson (shown in blue) hovers on the positive side of neutral before surging towards positive and growing dramatically in volume after he wins the tournament on day four.

Sentiment towards Tiger Woods (shown in purple) was composed predominantly of positive and negative values which indicates that opinions of his performance were polarised and had a very low number of neutral Tweets. When plotted on a scale of negative to positive this type of score would appear in the middle and give the impression of neutrality rather than polarisation which is the case. Using TriVis polarised sentiment is clearly distinguishable from neutral sentiment.

The volume of Tweets about Tiger Woods is visibly larger than that of Ian Poulter (shown in green) until the final day when Poulter rose to third place and became the highest ranking Briton and clear home crowd favourite in the final hours of the game.

TriVis represents the volume and composition of sentiment towards multiple topics over a given interval of time in one graph. TriVis can be read at a glance by a spectator of a live event and provides an understandable insight into a complex data set in real time.

Applications

Possible applications of this design include use as a symbol which could be overlayed on live tv or as an addition to a leaderboard or scoreboard. TriVis could also be used as part of a drill down interface to allow the user to zoom or filter to explore specific times or keywords. We have used TriVis to construct a 3D model of time footprints using the z-axis to convey a timeline.

ENTERPRISE IRELAND

Ireland's EU Structural Funds Programmes 2007 - 2013
Co-funded by the Irish Government and the European Union

EUROPEAN REGIONAL DEVELOPMENT FUND

DCU NUI Galway OÉ Gaillimh UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh UCD DUBLIN

sfi Science Foundation Ireland