

## **Project Final Report: Measuring Financial Sentiment in Social Media**

This pilot project has now been successfully completed and we can report a number of valuable accomplishments from it. The WSI grant, together with the matched funding from our commercial partner Channel Creator, bought us a 6-month licence to Visible Technologies' social media analytics platform. Visible Technologies was the leading brand name in social media analytics throughout the period of this study and were used by most leading companies in the US for corporate brand monitoring and management. Access to the Visible Technologies platform enabled us to put together what we believe is probably the largest dataset ever assembled for an academic finance paper – we are confident that it is the largest text-based dataset ever used in an academic finance study. We obtained in the region of 100 million social media messages (350GB) linked to the 500 largest and most liquid US stocks (the constituents of the S&P500 index). So far, this has enabled us to achieve one publication. We have another paper in an advanced stage of development, a further paper at an earlier stage of development and several more planned. We have also already secured a follow-on funding grant to explore the commercial potential of this research.

To recap, last year we were approached by Channel Creator because of our expertise in text based sentiment analysis for investment selection. Channel Creator were the UK representatives for a US based leader in social media analytics called Visible Technologies. The latter had developed a business specialising in marketing and corporate brand management. Channel Creator had received approaches from London based investment managers who were interested in using the Visible Technologies platform for investment selections. The problem was that the existing platform was not set up to do this and could not be easily adapted to do this. Furthermore, Visible Technologies had no knowledge or experience of servicing the needs of investment managers as this was not a business opportunity that had arisen in the US. As essentially a re-selling agent, Channel Creator had little or no influence over Visible Technologies' strategic direction. So, Channel Creator wanted us to help them explore how to develop a new product that used the Visible Technologies platform, together with our text based investment analysis expertise and fund management industry knowledge, which could service the social media analytics needs of UK fund managers. The timing of our dialogue with Channel Creator coincided with the call for applications for the WSI grant, which is how this project came about.

It was not all smooth sailing on this project. We encountered a number of problems. At the outset, we

experienced a delay because it proved difficult to get the relevant University authorities to sign the Visible Technologies contract and then to get the agreed payment transferred to them. We eventually got these done and gained access to the Visible Technologies platform on 10 December 2014. Also, the platform was not designed for the use that we intended. In its current form, the frontend is a generic interface for search tasks, designed for interactive use with no API available. So it took Paul Gaskell a number of weeks to identify what exact data was there, in what form and then several more weeks to develop the code to download the data we needed for our study. The commercial service that Visible Technologies provides focuses on current and recent news and new social media data, as would be of interest to senior managers of companies and their PR advisors concerned about managing corporate brand images in real time. By contract, our analysis required historical data and archived historical data had to be made especially available to us. In the absence of an API, it took us time to develop a computer programme to download and sort that data from a frontend that only allowed access to a limited number of items at a time. We also ran into difficulties with Visible Technologies when our heavy download use monopolised their servers, disrupting their service to their commercial clients, which resulted in them temporarily barring us from the service. This was subsequently resolved by a lengthy negotiation process to get our access restored – which we achieved by voluntarily agreeing to batched, sequential download job requests, concentrated in off-peak periods so that we did not completely hog their bandwidth. These difficulties arose partially because the sales and customer support staff did not understand the different nature of our needs from their usual customers, our intended use or the non-standard access that we had agreed with their company to access their data. But perhaps the more important issue was that we did not understand their bandwidth limitations until they complained to us that we were hogging 100% of it.

Channel Creator had very positive interest and feedback from a number of fund managers, most notably Blackrock (the world's largest asset manager) up until this summer. During this period, we held (every 2 or 3 weeks) business lunches and afternoon-long business strategy meetings with Paul Brothers of Channel Creator in London. However, since early summer, there has been a material change in Channel Creator's relationship with Visible Technologies. The latter had been taken over by Cision Inc.. After an initial period of continuity, Cision has now integrated Visible Technologies into its core business and Channel Creator will no longer represent this business in the UK. Cision has a different, more US focused business strategy and does not wish to pursue the business that we were trying to develop in this project. As a result of this change in its relationship with Cision, Channel Creator's strategy has also changed. Without the Visible Technologies social media platform business

as a catalyst, Channel Creator 's conversations with investment managers have fizzled out and consequently so too has their appetite to pursue this project with us any further. This has necessitated a change of direction for us to turn our research into a commercial product.

A significant outcome of this project so far is that we built a cloud-server based, 350GB, Google BigQuery database of social media messages related to the stocks in the S&P500 index, which are the 500 largest and most liquid stocks list on US stock exchanges. This data is drawn from 45,516 different content sources (including general sources like Twitter, Facebook, blogs and specialist sources like Seeking Alpha and Raging Bulls), spanning 10 distinct industrial sectors (Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunications Services, Utilities), covering the period September 2013 to September 2015. Although technically this data is drawn from sources across the globe, we only focus on English language sources in this study. Consequently, we believe that most of our data originate in the English speaking countries, with the heaviest concentration being in the United States. Our database contains approximately 50 billion individual words in total. Such a volume of data requires state-of-the-art Big-Data technology such as Hadoop and noSQL MapReduce queries to analyse. We think this is the largest dataset (by volume) that has yet been assembled for an academic finance study. We are happy to share this database with other University of Southampton researchers who might be interested in making a copy of it. However, interested researchers should be aware that the University has signed a licence agreement with terms and conditions from Visible Technologies. This agreement allows us to analyse this data and publish our findings but prohibits us from disseminating the raw data publicly.

In terms of measurable successes from this project, there are several. We mentioned above the 100 million (approx.) social media message database resource we now have. To date, we have published one paper linked to this and are working on several more. Our paper “Signal Diffusion Mapping: Optimal Forecasting with Time Varying Lags” (by Paul Gaskell, Frank McGroarty and Thanassis Tiropanis) was published in the Journal of Forecasting on 1 September 2015. We have another paper in development called “Variability in Textual Sentiment-Price Relationships” (by Paul Gaskell, Sha Liu, Frank McGroarty and Thanassis Tiropanis) also based on this research, which we hope will be finished and ready for submission to a high quality journal later this year. A further paper, “Returns to Buying Online Sentiment Winners” is at an earlier stage of development. Given the uniqueness and unusual richness of our database, we anticipate many more academic publications from this research over the next few years. This data also contributed to the successful completion of Paul Gaskell's PhD thesis

(recently submitted). Finally, this pilot study gave us the validation we needed for our SDM model in order to pursue the ICURe funding grant (£50k) which we recently obtained.

Our future focus is to get our SDM technology, and the validation analysis of it (courtesy of the WSI grant), in front of potential customers. That is what the ICURe grant supports us to do and, indeed, demands of us. This will include financial sector opportunities but will also explore opportunities for SDM in other industrial sectors. In addition, Professor Frank McGroarty has had invitations to present his research to the London Quant Group and to the Institute for Quantitative Investment Research (INQUIRE). Both of these are professional associations for investment managers with an interest in statistical modelling decision support tools. We intend to use both of these opportunities to present SDM.

## **Follow-on Funding**

We have already applied for, and been awarded, a follow-on funding grant for the next stage on the path to commercialisation of our research. Our commercial partner experienced a major change in circumstances that caused us to change tack and to pursue funding in a different form from what we had originally intended. Fortunately, this change of tack has worked out well and we have now secured the next block of funding that we need on our path to commercialisation.

During the course of this project, the supplier of the social media data for our study, Visible Technologies, was taken over by a company called Cision. Up to June of this year, it was business as usual in that our relationship with Visible Technologies, and that with their UK representative Channel Creator, continued as we had originally expected. However, over the summer it became clear that Cision had a very different strategic vision for Visible Technologies' data and platform than the original management had, and that that vision did not include the social media investment selection tool that Channel Creator, Visible Technologies had been working on with us. Over the past two years, Cision have acquired a number of different marketing and marketing analytics businesses and their aim is to integrate these in order to become a big player in the marketing analytics business and will focus mainly on the US market. Furthermore, going forward, Channel Creator will not represent Cision's interests in the UK as they had Visible Technologies' interests.

While the situation set out above was undeniably disappointing and dealt a blow to our original plans to pursue Innovate UK funding, we decided to focus on the positives and to take the project forward toward commercialisation without the involvement of Visible Technologies or Channel Creator. This pilot study enabled us to establish that our innovative Signal Diffusion Mapping (SDM) technology worked on social media data in the way that we had hoped and believed that it would. In addition, close exposure to the Visible Technologies platform enabled us to identify all the component parts of their social media data platform. We have already identified alternative sources of supply (i.e. data aggregators further up the supply chain) for most of the social media data that we established were relevant for finance. Paul Gaskell and Professor Frank McGroarty were recently awarded an ICURE grant (up to £50k from Innovate UK and HEFCE) to explore the market potential for SDM. We do not believe this would have been possible without the WSI grant / Channel Creator matched funding and this pilot study which established the efficacy of SDM in a commercial application.

Looking to the future, the ICURe grant provides us with a resource of £50k for the next 6 months to go out and talk to potential customer of our SDM based service who could be based anywhere in the world. We will examine this not just in context of Finance but also in other industrial sectors in which we now realise it could have potential application. We are approaching this next phase in an open minded way. We are convinced of the commercial viability of our SDM research and are absolutely determined to bring it to commercial fruition. We do not yet know if we will take this project forward by ourselves as an independent entity (i.e. via a start-up business led by Paul Gaskell, who has now finished his PhD research) or if we will join forces with another commercial sponsor. The ICURe grant will enable us to make that decision.