Journal of Systems and Information Technology

Engineering social media driven intelligent systems through crowdsourcing: Insights from a financial news summarisation system

Martin Sykora

Engineering social media driven intelligent systems through crowdsourcing

Insights from a financial news summarisation system

Martin Sykora
School of Business and Economics, Loughborough University, Loughborough, UK

Received 2 March 2016
Revised 15 May 2016
Accepted 20 May 2016

Abstract

Purpose – The purpose of this paper is to explore implicit crowdsourcing, leveraging social media in real-time scenarios for intelligent systems.

Design/methodology/approach – A case study using an illustrative example system, which systematically used a custom social media platform for automated financial news analysis and summarisation was developed, evaluated and discussed. Literature review related to crowdsourcing and collective intelligence in intelligent systems was also conducted to provide context and to further explore the case study.

Findings – It was shown how, and that useful intelligent systems can be constructed from appropriately engineered custom social media platforms which are integrated with intelligent automated processes. A recent inter-rater agreement measure for evaluating quality of implicit crowd contributions was also explored and found to be of value.

Practical implications – This paper argues that when social media platforms are closely integrated with other automated processes into a single system, this may provide a highly worthwhile online and real-time approach to intelligent systems through implicit crowdsourcing. Key practical issues, such as achieving high-quality crowd contributions, challenges of efficient workflows and real-time crowd integration into intelligent systems, were discussed. Important ethical and related considerations were also covered.

Originality/value – A contribution to existing theory was made by proposing how social media Web platforms may benefit crowdsourcing. As opposed to traditional crowdsourcing platforms, the presented approach and example system has a set of social elements that encourages implicit crowdsourcing. Instances of crowdsourcing with existing social media, such as Twitter, often also called crowd piggybacking, have been used in the past; however, using an entirely custom-built social media system for implicit crowdsourcing is relatively novel and has several advantages. Some of the discussion in context of intelligent systems construction are novel and contribute to the existing body of literature in this field.

Keywords Natural language processing, Social media, Crowdsourcing, Crowd-Powered systems, Intelligent systems

Paper type Research paper

Journal of Systems and Information Technology
Vol. 18 No. 3, 2016
pp. 255-276

1328-7265
DOI 10.1108/JSIT-03-2016-0019

1. Introduction

Online-based crowdsourcing has opened up new and interesting applications in areas, where cognitive capabilities and the collective intelligence of the crowd allow for accurate solutions, where traditionally, individuals with expert knowledge were required to approach the tasks at hand (Brabham, 2008). Crowdsourcing platforms now provide for a scalable human intelligence processing resource that can be tapped into by researchers and system engineers alike. Especially, the fields of artificial intelligence (AI) and machine learning (ML) have benefited from the readily available crowds (on platforms such as Amazon Mechanical Turk; Paolacci *et al.*, 2010), which can now annotate huge and often complex data sets in a fraction of the usual time and costs required for annotation tasks. These crowd-annotated data sets in turn are used to train and develop better and more accurate AI/ML models. However, computer-based systems purely relying on AI and ML have not delivered truly intelligent systems, which is where a closer integration with human cognitive and reasoning capabilities, if integrated effectively, hold considerable promise. To this end, Lasecki (2014), for instance, provides several inspirational examples of crowd-driven intelligent systems. He points out, however, the challenges of seamless integration of the crowd and especially design for on-demand and real-time intelligent systems, where collecting and motivating crowd contributions in real time, is a significant challenge, which, to date, has mostly been overlooked in academic research.

In this paper, we argue for the benefits of implicit crowdsourcing by harnessing a crowd's collective intelligence and cognitive capabilities through social media. We discuss how social media-based websites can be directly used within crowd-driven intelligent systems. An example social media-based, crowd-driven intelligent system, for news analysis is presented and evaluated and related issues and insights from its development are discussed. Specifically, issues of achieving high enough quality crowd contributions and challenges of efficient workflow and real-time crowd integration into intelligent systems are considered. As opposed to traditional crowdsourcing platforms, the presented system has a set of social elements that encourages implicit crowdsourcing. Instances of crowdsourcing with existing social media, such as Twitter, often called piggybacking, have been used in the past (Grevet and Gilbert, 2015); however, using an entirely custom-built social media system for implicit crowdsourcing is relatively novel and has several advantages.

The remainder of this paper is organised as follows. Section 2 introduces some background on crowdsourcing, related intelligent systems and the approach to crowdsourcing through social media integration into intelligent systems applications. Section 3 presents the crowd-driven news analysis system and study, and Section 4 provides a discussion and limitations to the presented work. The paper is finally concluded in Section 5.

2. Related literature and theoretical background

The term crowdsourcing was coined about 10 years ago, by Howe (2006), although, by 2012, Estellés-Arolas and González-Ladrón-de-Guevara (2012) reviewed over 40 different definitions of the term. Bringing together the various definitions, the main element of their integrated definition was highlighted as:

> Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the

crowd should participate bringing their work, money, knowledge and / or experience, always entails mutual benefit.

Brabham was one of the first to systematically study crowdsourcing as a field in its own right (Brabham 2008)[1]. Aspects of crowdsourcing itself have been around for quite some time before, in one form or another. Open-source bounties and reverse open-source bounties in software feature requests have been fairly common for some time, and crowdsourcing can, in many ways, be seen as instances of existing open-source, and Web 2.0 processes or calls for participation in the classical sense of public tenders or as citizen science conducted on the Web. Prpić, *et al.* (2014) introduced the idea of crowd capital within IT-mediated crowds for collective intelligence, which is based on structure – IT used to engage crowds, content – type of desired input and process – the organisational processes used to form and structure the resources from the crowd. In this paper (Section 3), we propose to use social media to align with such work, however focusing specifically on collective intelligence potential that emerges as a by-product from social media application usage, and how this can be used in intelligent systems.

2.1 Quality of crowd contributions

The quality of contributions in crowdsourcing is naturally an issue, and often checks are embedded in tasks that allow validating contributions. For instance, in Amazon Mechanical Turk, such validation micro-tasks are commonly used, and previous user ratings based on historical crowd-contribution quality both play a key role (Paolacci *et al.*, 2010; Callison-Burch and Dredze 2010). However, in the context of social media, often such checks are not explicitly possible; yet, contributions, whether in the form of tweets, blog posts or submitted and tagged image content, are confirmed by the social network, with “up-votes”, ratings and reviews – i.e. explicit quality ratings from members of the community (Agichtein *et al.*, 2008).

Wikipedia is a useful example of crowdsourcing, as it highlights a radical experiment in trust, as even anonymous users are allowed to and in fact are encouraged to edit and re-edit this Web-based encyclopaedia in the communal hope of producing an immense and “complete” body of encyclopaedic knowledge. Critics, such as Keen (2007), point out the seemingly intrinsic problem that is such a vast text would clearly have to be riddled with inaccuracies. Quite surprisingly however Wikipedia was found to be an accurate resource and is now arguably even becoming a standard encyclopaedic reference text. A comparison with encyclopaedia Britannica (Giles, 2005) suggested a similar level of information accuracy in both encyclopaedias. Interestingly, 70-80 per cent of inaccurate edits on Wikipedia get corrected almost instantly (Adler *et al.*, 2008a, 2008b). This can be attributed to the dynamic nature and self-managing environment of collaborative participation. The NASA Clickworkers project was an early pilot study by several NASA employees to assess whether public volunteers, each working for a few minutes here and there, could perform some repetitive and routine scientific analysis tasks[2]. The work consisted in marking craters (by marking four points on a crater rim to draw a circle) in the imagery data from the Mars Viking Orbiter. A second task was also set, in which users had to categorise the age of craters (Barlow, 2000 as cited in Kanefsky *et al.*, 2001). The two sub-goals of the study were:

(1) to find whether people are interested in volunteering their free time for routine scientific tasks; and
(2) whether the public has the training and motivation to produce accurate results in a scientifically important task.

The results were reported upon in Kanefsky *et al.* (2001). In conclusion, the quality of markings showed that the computed consensus of a large number of Clickworkers was virtually indistinguishable from the inputs of a geologist with years of experience in identifying Mars craters[3]. The important element in this application was the sheer number of participants – over 85,000 users visited the site within the first six months of the sites operation. Over 1.9 million entries ensured high redundancy and averaged out any errors made by individuals, and effectively the consensus opinion of what would make out a crater on the imagery data would be collected.

This self-correcting nature of large numbers of contributors closely correlates with ideas from the field of finance, specifically prediction markets and the related efficient market hypothesis (EMH). Much of the enthusiasm for prediction markets is derived from the efficient market hypothesis or EMH, which was originally proposed in the 1960s (Fama, 1965), and states that in the market, mechanisms exist so that information will efficiently transfer into the most optimal asset price, given all known, available information about the asset. Sometimes also referred to as "information market" or also known as "event futures", they allow individuals to trade on future events, where in aggregate, these votes represent the probability of the given event. Prediction markets are speculative markets which have the sole purpose of making predictions. They can be used to predict a variety of events such as sports, politics, movies, films or the stock market. The aim of a prediction market is that, through a large collective user base, there will be a greater accuracy in predictions which will benefit those who give correct predictions through financial rewards but will detriment those who give incorrect predictions through financial or other losses. For a prediction market to be efficient however, it was found that it is not required that all individuals in a market be rational, as long as the marginal trade in the market is motivated by rational traders (Wolfers and Zitzewitz, 2004). Servan-Schreiber *et al.* (2004) have shown there not to be any substantial difference between real or play money markets either. Prediction markets were found to perform considerably better than individual human forecasters (Servan-Schreiber *et al.*, 2004). In a study by Luckner *et al.* (2007), the advantages of prediction markets were highlighted where FIFA 2006 World Cup matches were predicted with 59.4 per cent accuracy against Fifa ranking accuracy of only 46.9 per cent, over 16 matches. Finally, a good overview of the main different types of prediction markets and further literature is provided by Wolfers and Zitzewitz (2004). There are also numerous examples where systems like this have been used by companies internally to better understand their various operational aspects (Kambil and Van Heck, 2002; Kambil, 2003). These ideas extend to social media as well. For instance, Tumasjan *et al.* (2011) have used activity from Twitter to accurately forecast general election results. They further proposed that the same principles that underlie prediction/information markets are likely at play when aggregating social media activity, and effectively allow for a form of collective intelligence to emerge from the social media contributions.

2.2 Collective intelligence and systems

Models have been proposed to leverage collective intelligence; for example, the MIT Center for Collective Intelligence has recently (Nagar and Malone, 2011) proposed a model for combining human- and machine-based collective intelligence. This is a model based on prediction markets that combines predictions from groups of humans and AI agents to show that they are more robust than those from groups of humans or artificial agents alone. The idea of collective intelligence is not new; it emerged from writings by Hofstadter (1979), Russell (1983), Lévy (1994), but at an abstract level, already H. G. Wells mentioned the idea of a collective "world brain" in his essay entitled, "The Brain Organisation of the Modern World" (Wells, 1994). Recently, with the emergence of the Web, the idea has gained new momentum as it has become feasible to involve large crowds quite easily online, with numerous efforts to understand it, such as the MIT Center for Collective Intelligence[4]. However, specifically, systems that use the collective intelligence, judgements and cognitive capabilities of the crowd in real-time applications provides for a challenge. Recently, Lasecki (2014) proposed several fascinating prototype "crowd-powered" intelligent systems' examples, which illustrate the significance of crowdsourcing in systems design. An intelligent robot control crowd-powered interface (*Legion*), where the paths and decisions for a robot depended on online, real-time crowdsourcing. *Legion:Scribe*, being another system where multiple non-expert captionists collaboratively caption speech in real-time, and multiple sequence alignment algorithms subsequently merge the multiple captions into one, in real-time. *Glance* and *Chorus* are two systems that use crowd contributions to recognise activities in videos and respond to natural language text queries, respectively. Essentially, these applications leverage human understanding to facilitate intelligent systems, capable of working in real-world settings when artificial intelligence is not reliable.

Although prior work has investigated quick recruitment of members for crowdsourcing tasks, a considerable challenge is the effective recruitment of crowd workers into the specific system workflow, which may prove very problematic, especially with real-time processing demands on crowd contributors and their on-demand readiness for online tasks. This is where social media with implicit contributions from users may result in meaningful and valuable integration into systems that may be preferable to traditional specialised crowdsourcing platforms.

The value in the data contributed via social media applications, where users most often have implicit motivations to contribute, can be of significant importance. For instance, automated summarisation has been a challenge in the field of Natural Language Processing (NLP) for many years (Hahn and Mani, 2000); however, platforms such as Twitter, where messages are limited to 140 characters, provide potentially useful, real-time, on-demand crowd input into an information summarisation task (Osborne *et al.*, 2014). A variety of intelligent systems can be facilitated through close integration of crowd contributions from social media with AI techniques.

As early as 2010 in their position paper, Bermingham and Smeaton (2010) discussed the potential of tapping into the user-contributed data sets from social media crowds, on a wider and more systematically applied scale, than tends to be the practice even today. Bermingham and Smeaton related to how mobile phone messaging (SMS) and instance messaging (IM) essentially represent instantaneous chatter, however intrinsically private. This they contrast with the public nature of the social Web which allows to readily tap into a global chatter, or *"collective intelligence"*, where a wide range of online opinion and contributions in the form of posts, uploaded images/multimedia content, ratings, tags and potentially many other types of Web-based social contributions represent valuable collective intelligence, to be harnessed. Based on literature discussed to this point, the following is a useful working definition for collective intelligence:

> Collective intelligence is a shared or group intelligence that emerges from the collaboration, competition, or simply sharing of many individuals in response to some challenge or implicit goal and is essentially pattern based decision making, based on collective knowledge, where collective knowledge can be effectively collected via web 2.0 / social media systems.

Arguably, the *challenge*, or *an implicit goal*, exists on practically all social media systems, in some form or shape. The social media application can have an explicit goal (e.g. Wikipedia – the goal is to amass knowledge into a world Encyclopaedia) or an undefined generic one (e.g. YouTube – share videos for entertainment or any reason). The primary appeal of social systems being that individuals participate in these systems for their own and social enjoyment. Sykora (2009) elaborates further on this and the implicit goals in social media systems from a computational intelligence perspective, and explains the possibilities for the related computer science area of interactive evolutionary computation, which is a set of optimization heuristic algorithms directly integrating crowd contributions within evolutionary algorithms.

2.3 Custom social media for crowdsourcing and its benefits

Rogstadius introduced an intelligent system for crisis mapping and related information management, by leveraging automated text-mining algorithms and social media-based crowd contributions, in one single system. His system uses clustering of messages from Twitter and other crowd-contributed sources, specifically a custom social Web-based user interface further encourages and allows crisis-workers (i.e. the crowd) to double check and correct miss-categorisations in real time, and subsequently presents actionable intelligent system outputs (Rogstadius *et al.*, 2011). Ushahidi is another instance of such intelligent crowd-driven crisis mapping system (Ushahidi, 2015). Integrating social media platforms with specific tasks for a system that can present its outputs, through the intelligence of the crowds, can be a worthwhile approach. The advantages of exploring social media type platforms for this are several;

- Crowd input is more often readily available (i.e. on-demand collective intelligence and cognitive judgments) through the platforms, as these tend to have strong, established user-bases (Besten, 2012; Grevet and Gilbert, 2015), e.g. Flickr, Pinterest, Twitter, etc.
- Crowd is often motivated with intrinsic motivations, and a considerable body of academic research on driving motivators behind social media contributions points to two main streams of motivators, which align with altruistic and the social self-presentation reasons (Forte and Bruckman, 2005; Kuznetsov, 2006).
- Data from these sources are commonly available in real time through robust, stable and on-demand programming interfaces, e.g. the streaming Twitter API (Driscoll and Walker, 2014).

Given the vastness of the social media application landscape, it is often sufficient to leverage existing sources for crowd/collective intelligence, and up till now most studies in this research area were concerned with analysing existing social applications and related user-generated data. However, in some cases, custom *sources* of collective intelligence (henceforth referred to as CI) are needed to fill a gap in the social media application landscape. Deciding to do so may be a relatively costly initiative (in terms of time, complexity and resources); therefore, it must be carefully considered whether existing social media sources will provide the necessary CI. Using custom sources may have some significant advantages, such as provide finer-grained user-generated contributions, at a level that can be custom-build for the CI task. Because all the data are available or "owned", complicated information extraction, via APIs and HTML-page parsing, can be entirely avoided and of course the potential flexibility to accommodate a given problem domain is generally incomparably more substantial than with existing social media systems. There are unfortunately also major challenges in constructing custom-built CI sources, primarily in development of the software, its maintenance/fault-free operation, provisions for architecture/hardware and most importantly attracting and socially engaging user-participation.

Some of these issues can be addressed, for instance, by integrating an existing user workflow into a social media-based habit within the custom user interface. In the next section, the example case study of an intelligent social media type system for news summarisation, called Newsmental, that was developed and run for a couple months, is presented, evaluated and discussed in detail.

3. Newsmental: an intelligent news summarisation system

In computational finance, much research has been done on forecasting, interpreting the financial markets and a better understanding of financial events (Taylor *et al.*, 2002, Zemke, 2003). Perceived wider sentiment is detrimental to the price of assets, and it was also shown that a substantial effect on the price creation process in finance is explained by news-events sentiment (Ederington and Lee, 1993; Barberis *et al.*, 1998; Chan, 2003). Researchers have highlighted (Fung *et al.*, 2005) that in the existing literature on forecasting and trading models, there is an overwhelming tendency to focus on quantitative (macro-economic and price based) data, with very little work investigating the use of qualitative data sets in such models. Given that there is an enormous quantity of qualitative news data in the form of unstructured text, there have been numerous efforts to automatically annotate sentiment in financial news (Mittermayer and Knolmayer, 2006). Understanding or analysing news is inherently a very difficult natural language processing task, as even human experts often fail, or disagree on what particular news actually means and how it applies to various entities. Depending on the perspective, situation and background, one same news-item may appear to have different polarity (negative / positive) and impact to various individuals (Koppel and Shtrimberg, 2006); hence, a collective agreement on news is in fact highly desirable. The task of news analysis lends itself well to human-based processing; however, not so much to fully automated AI-based solutions.

A social media system that explicitly facilitates participants' collection of their opinion on financial news events in a productive, social and streamlined manner, entitled Newsmental has been designed and integrated with automated techniques, for accurate and reliable news summarisation.

3.1 System design

A system was built that would:

- Extract news articles from a number of (mostly British) financial news sources.
- News articles would be automatically analysed, pre-processed and automated entity recognition applied to the unstructured text to extract entities and some basic relationships between certain types of entities.
- Cluster similar news (using a clustering algorithm based on the extracted entities).
- Top news items would be presented on the website with a breakdown of the news and charts (based on the extracted entities and relations).
- Each news item would be available for a quick non-obtrusive evaluation by readers/visitors and the ratings would be shared amongst the entire community on the system.
- News reading history would be tracked automatically and made available to all registered users with historical views, charts and other comparisons against community ratings – effectively creating an online community of news readers.

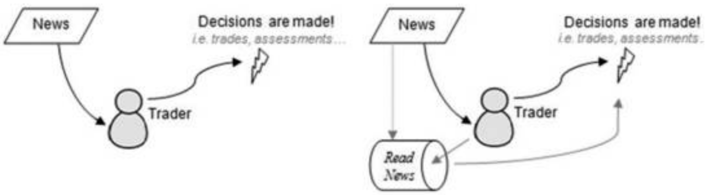
The following design choices highlight some benefits of the system to potential users:

- News articles are automatically pulled at regular intervals from multiple sources, meaning that a single place for viewing all the news may be more convenient for visitors. The news articles are clustered, so that the same news from various sources does not get repeated unnecessarily. This is essentially what many news aggregators such as *news.google.com* also do.
- Entities, facts and relationships between important entities are automatically annotated, and presented as a breakdown analysis with each news item. This provides for a useful breakdown of the main actors, and out-takes from a lengthy article, which can greatly increase the speed and efficiency at which news is read. It also makes it easier for a reader to comprehend the news and help speed up news analysis in general (Zwaan *et al.*, 1993).
- News reading and understanding is augmented via collective news analysis, in that all previous news-ratings are summarised/averaged out and presented to all subsequent visitors (i.e. shared within the community of readers). This effectively facilitates reading news "in a collaborative" manner, as chances are that other users have read and analysed the news already, which may in turn drastically speed-up news reading and the news opinion forming process.
- News reading becomes tractable. Submitting judgements for news essentially creates a footprint of all read articles and opinions held at specific times on the various topics. The news reading process has been generally intractable – even with the largest news portals. However, *our system* provides features for retrieving news-reading history by time, topic and similar views.

The central idea behind Newsmental is to use a non-intrusive manner for collecting news judgements and opinions. The system's aim has been to streamline the news reading and rating process by taking an existing need and using a social media system to aid in delivering a working solution towards this need. Figure 1 presents a scenario involving a potential use-case of an information consumer – e.g. a trader tracking financial news, who makes regular daily trading-related decisions based on a news feed.

The left illustration in Figure 1 shows a trader taking-in news information which is eventually actioned into trading decisions, with potentially many news items over time. Unfortunately, due to the effect of selective memory, it is very difficult to reconstruct the thinking process behind historical trading decisions without explicit note-taking. Traders are known to keep logs (i.e. diaries) of trades (Schwager, 1993); however, having to take note of each news-article that impacted a trading decision (e.g. into a spreadsheet) would break down the natural news reading workflow of a trader, to the point where it potentially becomes infeasible. Instead, Newsmental allows for all read news to be automatically tracked over time, with no disruptions to the workflow, as illustrated in right-hand side of Figure 1. The trader can rate, comment and highlight text excerpts from news articles in a streamlined way (using a light-weight Web 2.0 style user interface), which gets stored into database (*Read News* in Figure 1). These data are an accurate representation of a trader's opinions, perceptions of significance and sentiments over time and can be reviewed by the trader and compared to the rest of the community. Each registered user can inspect news as they were rated, in reverse chronological order, with direct links to the original news sources – effectively acting as a kind of augmented bookmarking service for news articles. Other views include: view by most recent topic or a view that compares ratings with the community, as a social element (Figure 2). For each news article, a radar-chart is generated, which shows a thin line (user opinion) and a thick line (community opinion) for each of the news-rating dimensions. For instance, in Figure 2, the yellow circle on the chart highlights the users' sentiment for the "US Government" which is visibly lower than that of the community; otherwise, the ratings are pretty much the aligned.

3.1.1 Overall architecture implementation. A relatively involved architecture had to be used to satisfy functionality requirements and minimise latency of the time-demanding long-running article extraction and text-analysis processes. RSS parsing, Restful-API, HTML processing and article extraction using *XPATH*, named entity (NE) extraction, clustering, programmatic caching and an ASP.NET AJAX and JQuery based interface were integrated to build Newsmental. To ensure that news articles were up-to-date, an independent background process running on a separate thread was responsible for processing news articles, updating the database (MySQL) and ensuring that the memory cache represented most recent states, mirroring database data. Caching was an important consideration, as the system had to achieve good



Notes: Left figure – usual work-process; Right figure – system based work-process

Figure 1.
Use-case of a trader

response times despite working with memory heavy data (i.e. large chunks of article text and text summaries).

3.1.2 News summarisation approach. The system can be split into two parts that are integrated within the single social media platform. The automated analysis and the user-led/crowd-based contributions part.

First (*1*), the automated part processes the raw summary article text through automated sentence segmentation, tokenisation, POS tagging and the actual NE extraction, which is based on gazetteers of entity names and types, with entries to handle synonyms and similar linguistic issues (Ye, 2003). In the building of the NE extraction module for our system, the gazetteers provided by a third-party API service, *Open Calais* from Clear Forest Inc.[5], were employed. A greedy clustering algorithm based on the low feature space word-vector of NEs is subsequently run to cluster similar articles together for effective presentation of news articles.

Second (*2*), this news summary, which at this point is represented through the extracted NEs, NE relationships (Figure 3[6]) and article summary clusters are augmented with a crowd-/users-based news summary. Each input panel (Figure 4) is composed of a set of horizontal and vertical slider controls, a comment text box with 120-character limit, a text-box with autosuggestions for tagging the news-article and a submit button. The submit functionality is implemented entirely as an AJAX partial page postback script, so that user experience is not negatively affected by full-page reloads.

Figure 4 shows the slider input panel in its full view, and Figure 5 illustrates a judged news item, ready for submission. The comment from the 120-character comment text-box is shared with other users of Newsmental with each article viewed, and tags are used in search functionality. It can be observed from Figures 4 and 5 that the first two vertical sliders relate to "sentiment for the US Government" and "sentiment for Standard & Poor's", and that only the former was rated negatively (sentiments are rated on a five-point ordinal scale, i.e. -2 = very bad, -1 = bad, 0 = alright/no-opinion, $+1$ = good, $+2$ = very good news). The announcement of the news that S&P has issued a debt warning for the USA is clearly bad news for the USA; however, S&P was more or less unaffected by the news. Hence, it is noteworthy the interface allows for such a more

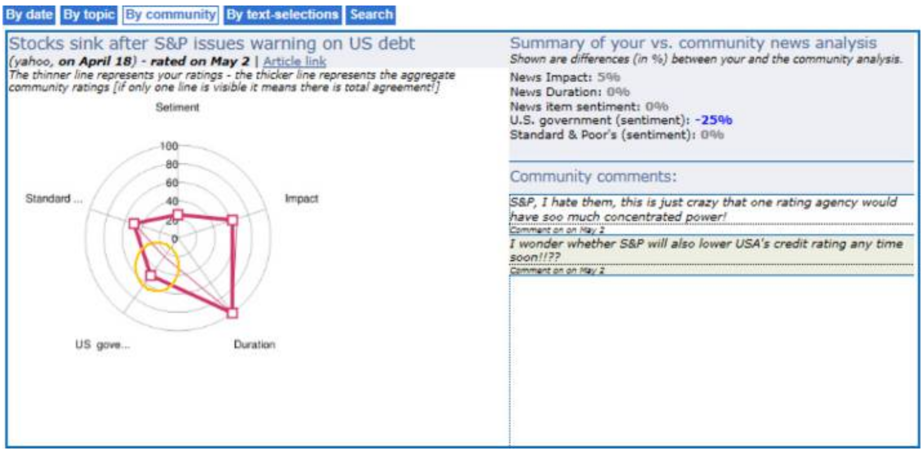


Figure 2.
Browse by community view (radar chart showing the agreement amongst ratings with the community), other views include by dates, topics, etc.

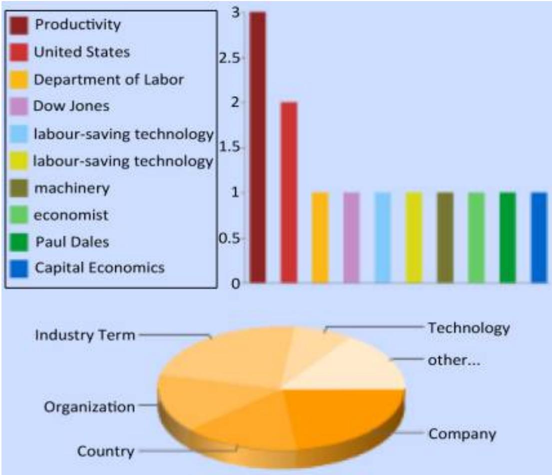
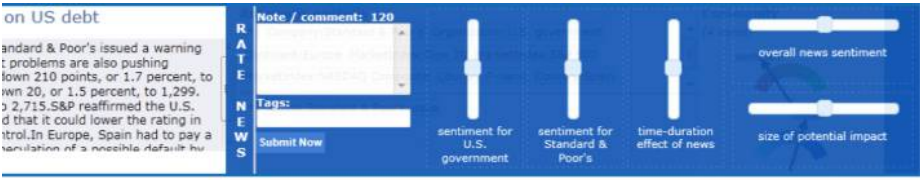


Figure 3. Visual display of the entity composition of a news-article (available for each article)



Notes: From left to right: comment [maximum, 120 char], Tags [autosuggest], sentiment for entity 1, sentiment for entity 2, time-duration, overall news sentiment and size of potential impact

Figure 4. Slide-out input panel for news article

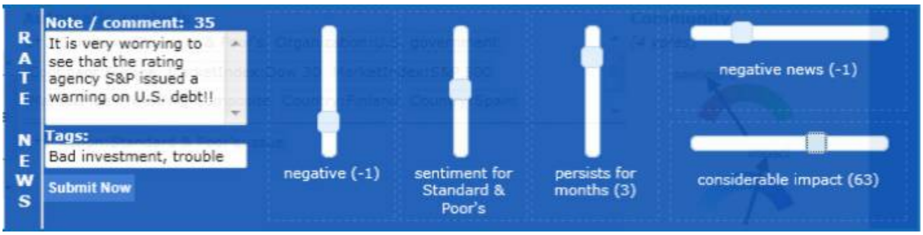


Figure 5. News input panel, ready for submission ("Submit Now" button click)

complicated sentiment/opinion to be expressed. The third vertical slider is always present in the input panel and allows a choice of five values, relating to the time-duration effect of news (minutes/hours, days, weeks, months, years). In other words, given the news, what temporal impact in terms of duration of the effect the single news-item will likely potentially have on the financial market/financial ecosystem as assessed by the user.

In the example from Figure 5, the judgement of impact is likely to be in terms of months, as the debt warning might be an indication of further troubles for the US economy that could take a few months to materialise. The two horizontal sliders relate to the "overall news sentiment" (five-point ordinal sentiment scale) and "size of potential impact" (a percentage 0-100, with five ordinal bins; no impact [0], very little impact [1-25], some impact [26-50], considerable impact [51-75] and very high impact [76-100]). In the given example, the overall news sentiment (−1) is quite bad for the USA and in fact most other economies that depend on the USA. The impact rating is also high, *considerable impact* (63), as the impact of the debt warning news will probably move the markets. However, because it is quite likely that people from different regions and especially various backgrounds will interpret certain news rather differently, each registered user was asked to provide their demographic details, i.e. age group, location, level of education (university level, pre-university level), interest (finance, politics, technology, world events), financial experience (none at all, interested, knowledgeable, expert) and news reading frequency (only sometime, once every few days, every day, every few hours). It was hoped that in aggregate, with several participants reading and rating same news, consensus opinion would emerge.

3.1.3 Community and recruitment. A top active-contributors feature was also introduced, to make the news-rating even more social. In addition to implicit benefits of use, there is some evidence to suggest that, badges, prizes or other forms of incentives within an online social environment can dramatically increase initial user engagement (Malinen, 2009). To tap into existing communities of interest, the website was advertised with 11 university student finance societies across the UK. Committee members of the relevant societies were approached and these agreed to inform their members internally in addition to social media-based (i.e. Facebook) society pages. Some word-of-mouth spread through Twitter and Facebook as the initial news judgements were starting to pour in. With the kind support of departmental administrators, undergraduate and postgraduate students across various departments (including economics and business) were also informed via departmental mailing lists at the author's institution. A mailing list of around 200 individuals from a previous study on social media, who agreed to be informed of the launch of Newsmental, were also notified about its launch.

3.2 News summarisation outputs and analysis

All in all, 2,138 ratings were submitted during the 40 days of the study being online (averaging to 54 ratings per day); however, 650 ratings were submitted by anonymous users, i.e. not logged into the system. These ratings came from 55 different individuals who rated news, with 48 users having registered an account with the system. Because it was not a requirement to be logged-in to rate news, it is likely that some users forgot to log into the system on occasion, despite reminders to do so. In summary, based on registration information, users were predominantly male; between the age of 20-39 years; 72 per cent claimed to read news daily or more frequently; over 70 per cent were interested or knowledgeable in finance; and 91 per cent of users came from the UK, USA or Europe. Out of all 2,138 ratings, there were 199 ratings where a shared/public user-note (i.e. comment) was left behind. All the ratings covered a total of 1,070 individual news articles[7]. The 35 most-rated news-item headlines are shown in the dot-plot diagram in Figure 6.

In relation to each other, the average sentiment, duration and impact of the news stories make sense and can be explained well. For example, "Kate and William give UK wine a boost" was perceived to be good news (SD, −0.45), but with a low impact (SD, 19.91) and time-duration (SD, 0.73); on the other hand, "Obama: Still differences on debt, new talks Sunday" was strongly perceived at the time as bad news with one of the lowest standard deviations for sentiment and impact (among the 35 news), however a higher standard deviation for duration. Standard deviation in this context highlights the disagreement or uncertainty associated with an average news-item judgement. A news-item that carries a lot of significance (one of the highest average durations and impact) is "Trichet says debt is global, not European problem", and the news-story "George Osborne needs a bolder plan for growth" was rated with high duration; however, the impact is lower, which seems to be a logical interpretation of the news and makes relatively good sense.

One significant advantage of this type of insight and collective crowd-driven intelligent news summarisation is its real-time and on-demand nature, as news coming out were constantly monitored and available to the crowd for review and analysis.

3.3 Evaluation and quality of contributions

Given that users rated news-items independently[8], it begs the question of the degree of agreement between them. Inter-rater agreement or reliability measures assess the agreement between two or more observers who describe the units of analysis separately from each other. These statistic measures are in frequent use in social sciences, especially in human-driven content analysis and similar methods (Krippendorff, 2004). They help to answer the question of whether ratings are the result of irreproducible human idiosyncrasies or whether they reflect properties of the phenomena of interest on which others could agree as well. Over the years, a number of measures were proposed, such as Cohen's kappa (Cohen, 1960; κ), Fleiss's kappa (Fleiss, 1971; κ) or Cronbach's alpha (Cronbach, 1951; α); however, the most appropriate type of inter-rater reliability statistic for the data in this study (multiple raters, missing values and scale of measurement) is Krippendorff's alpha (Krippendorff, 2004; Hayes and Krippendorff, 2007[9]; α). The degree to which news readers agree on the sentiment, impact and

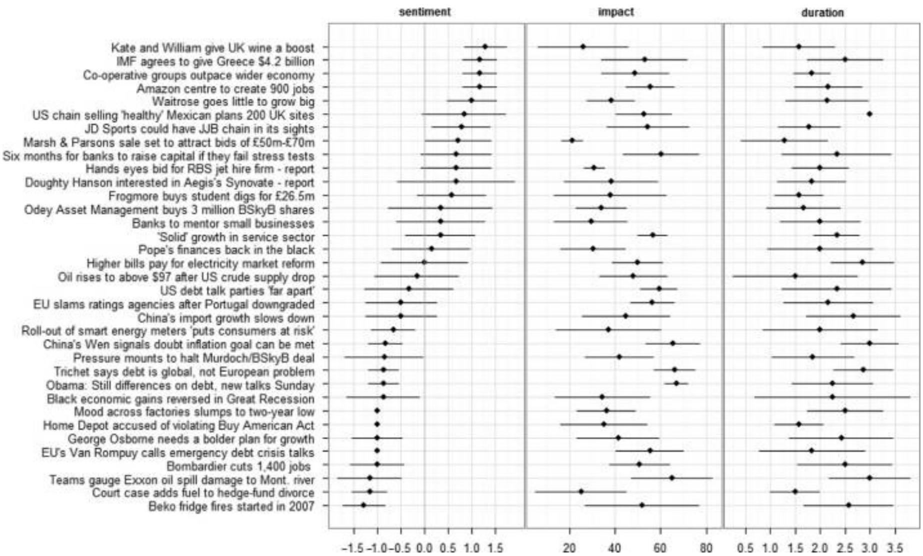


Figure 6.
Sentiment, impact and duration dot-plots (from left to right) of news stories with six or more individual ratings (top 35 headlines), ordered by decreasing average sentiment (standard deviations shown)

implied duration of news items was evaluated for the six (top) users[10]. The three alpha values with 95 per cent lower and upper limit confidence intervals (based on 10,000 bootstrap samples) were 0.6038 (0.5340, 0.6699) for sentiment, 0.2383 (0.1343, 0.3376) for impact and 0.0702 (−0367, 0.1747) for duration ratings. The value for sentiment is large enough to indicate a moderately strong and consistent agreement among raters; however, the latter two alpha values indicate poor agreement for impact and virtually no agreement for duration. Moreover, Table I reports agreements between individual users, where clearly there are users with much higher or lower agreement levels than the overall agreements. For instance, User 6 tends to agree the least with all other users and at times has a tendency to consistently disagree (indicated by a negative alpha value, e.g. last row, right most column in Table I); however, agreements between Users 1 vs 4, 1 vs 3, 3 vs 4 and 4 vs 5 tend to be consistent and high (rows 2, 3, 10, 13).

For the generic purposes of collective intelligence, it is desirable for the users' judgements to provide an estimate of the population average perception even though homogeneous subsets of users may produce a far more consistent perspective on the news data. For example, Krippendorff's statistic for Users 1, 2 and 5 shows a moderate agreement of 0.4420 (0.2514, 0.6146) for duration judgements, as opposed to virtually no agreement (0.0702) amongst all raters. Identifying homogenous subsets of users in terms of common news-interpretation is plagued with several issues and unfortunately was not further investigated in this study (see limitations section).

Overall, it would seem that individuals have at least some common interpretation of news events, especially polarity judgements were shown to be highly reliable. Judgements regarding duration and impact of news events may be treated as indicative only. A better understanding of the identities of various homogenous subsets of users, however, may provide more insight in this respect. Nevertheless, using a simple average of the judgement ratings throughout further analysis seems reasonable at this point also

Comparison	Sentiment – Krippendorf α	Impact – Krippendorf α	Duration – Krippendorf α
User 1 vs 2	0.6729 (0.4499, 0.8662)	0.0722 (−0.5041, 0.5342)	−0.0878 (−0.7236, 0.4489)
User 1 vs 3	0.8407 (0.6960, 0.9421)	0.3369 (−0.0100, 0.6488)	0.0880 (−0.1386, 0.3267)
User 1 vs 4	0.8215 (0.6568, 0.9451)	0.4810 (0.2312, 0.7101)	0.7831 (0.4925, 0.9880)
User 1 vs 5	0.8083 (0.6314, 0.9410)	0.2674 (−0.1507, 0.6189)	0.3540 (0.0131, 0.6371)
User 1 vs 6	0.3593 (−0.0349, 0.6797)	0.0446 (−0.3661, 0.3976)	−0.0632 (−0.4369, 0.2927)
User 2 vs 3	0.6360 (0.4043, 0.8180)	0.3962 (0.0534, 0.6843)	0.0940 (−0.1057, 0.3036)
User 2 vs 4	0.6339 (0.4143, 0.8243)	0.3631 (−0.0298, 0.7012)	0.0373 (−0.5389, 0.5049)
User 2 vs 5	0.5528 (0.1592, 0.8390)	−0.0079 (−0.4826, 0.4253)	0.1238 (−0.3095, 0.5142)
User 2 vs 6	0.3337 (−0.0146, 0.6214)	0.1310 (−0.3280, 0.5306)	−0.1109 (−0.6782, 0.3753)
User 3 vs 4	0.7900 (0.6100, 0.9400)	0.4876 (0.2542, 0.7098)	0.1233 (−0.1293, 0.3829)
User 3 vs 5	0.7418 (0.5159, 0.9193)	0.3573 (0.1839, 0.5408)	−0.0850 (−0.4894, 0.2502)
User 3 vs 6	0.2691 (0.0001, 0.6205)	0.0351 (−0.3824, 0.4019)	0.0124 (−0.1242, 0.1606)
User 4 vs 5	0.8809 (0.7469, 0.9702)	0.4336 (0.0891, 0.7252)	0.1681 (−0.2347, 0.5142)
User 4 vs 6	0.3304 (−0.1786, 0.7054)	−0.0272 (−0.4036, 0.3304)	−0.0038 (−0.4519, 0.3906)
User 5 vs 6	0.3558 (0.0187, 0.6704)	0.3290 (−0.0053, 0.6281)	−0.2280 (−0.6404, 0.1512)

Note: Krippendorf's alpha values, in brackets, are the LL95%, UL95% confidence intervals based on 10,000 bootstrap samples

Table I.
Sentiment, impact and duration agreement for top six Newsmental users

as relative to each other the ratings for duration and impact presented in Table I seem to provide an intuitive interpretation of events.

3.3.1 Feedback from users. Three weeks after the launch, an informal focus group was held with six users of Newsmental. After the study was concluded, a message to all registered users was also sent out asking for their feedback. The main points from both are summarised in the bullet points below.

- The functionality of being able to keep track of all rated news was very well received, and there was some indication that users appreciated the collaborative news analysis, as this specific feedback from a user illustrates:

> I'm an avid online news reader, and love keeping myself updated, but by adding an opinion/rating to an article it gives you an idea as to what the overwhelming response to an article is. The real bonus of this type of rating means that you can say more than other news websites, where all you seem to be able to do these days is "like!", and "Tweet", but this process gives the reader an opinion on the article.

The overall design of the page was otherwise positively received, an example user comment being; "The website was very simple to use, and really well structured, and by keeping it very light and lots of white space, it made for very easy viewing". Feedback also indicated that many users enjoyed news-reading on Newsmental, except in some cases where users were not quite clear on how to make best use of all the system features – despite two tutorials available on the website. As a response to this concern, a quick-intro bar on how to use Newsmental was added to the top of the home-page[11], which had a positive effect.

There was one instance in which a user complained to us about suspicion of another user's news-ratings being overly negative. This is an inevitable effect of a social Web 2.0 system with transparent user contributions. Disagreement between users can arise and as maintainers of the system, we may be approached to help resolve the situation, unless a social or automatic problem-resolution is not in place. Because it was rather interesting to investigate this complaint further, the user was subsequently contacted for their reasoning in news judgement, and explained they felt that the news were generally quite negative and a completely satisfactory explanation was given by the user.

Initially, several people felt slightly overwhelmed with the automatic analysis (i.e. presentation of entities/relations); however, in all followed-up cases (once subjects got to use the system), they deemed the presentation highly useful and conductive to reading articles. A possible improvement for the future would be to present extracted entities in a visual manner and simplify user-interface further.

4. Discussion

As opposed to traditional crowdsourcing platforms (e.g. Mechanical Turk), the presented website has a significant social element, as contributions by any user are visible to all other users and a community ranking (scoreboard) of most active users is shared, as well as features allowing users to explore the community evaluations of news through various views. These features effectively turn Newsmental into a social media platform. The explanatory value of collective intelligence in the form of news analysis from a system such as Newsmental is evident. The judgements in aggregate provide a consensus news assessment, augmented with clarification and insight from user comments. There have been significant efforts – especially in financial computation research to automate and achieve reliable news analysis (Wuthrich *et al.*, 1998; Fung *et al.*, 2005; Mittermayer and Knolmayer, 2006), however plagued with a number of problems and with varying success. Arguably several human analysts will inevitably disagree on news interpretation (Koppel and Shtrimberg, 2006), how severe a news effect will be and what it means for affected actors or entities. Replicating human judgement, experience-based induction and analytic abilities, as well as parsing an unstructured text, is exceedingly complex. A social media system such as Newsmental does not need to replicate human judgements as a function within some statistical/AI-trained model; rather, it taps into the social participation of its users at a relatively low cost. Despite the benefits provided by a system such as the one presented in this chapter, the complexity of building a Web 2.0 application and maintaining it with high levels of user participation is plagued by a number of issues.

There is also an important discussion taking place around ethical issues. It was, for instance, suggested by several academics (Petersen, 2008; Scholz, 2008; Shirky, 2010) that the treatment of individuals participating in peer production and crowdsourcing by corporations, in some instances, is equivalent to "slave labour". Petersen states that Web 2.0 represents, "an architecture of exploitation that capitalism can benefit from". The criticism put forward is serious. Corporations have been known to claim ownership over content produced by users, which is a very explicit form of exploitation. Alternatively, corporations lock user data within an interface and allow user ownership of data, but effectively, this has the same effect as direct data ownership. Some researchers (Mason and Suri, 2012) also argue that wages on paid traditional crowdsourcing platforms (e.g. Mechanical Turk) are unethical (i.e. considerably lower than the minimum US hourly wage). This criticism was avoided in this study, as users had implicit motivation to use Newsmental for their own news-reading benefits as a social media platform, of which the by-product was the CI on news interpretation. However, more generally, Petersen (2008) provides several specific examples of such, here mentioned crowd exploitation, but finds (based on interviews) that users whom he may see as being exploited, do not see themselves as such. Unfortunately, ethical issues of crowdsourcing are often overlooked and left out of the academic debate, although we believe these should form an integral part of the academic discourse, as well as wider ethical and legal questions, especially including issues around ownership of crowd contributed content on the Web (Puschmann and Burgess, 2014).

4.1 Limitations

Hayes and Krippendorff (2007) point out that to measure reliability of ratings, the data-generating process must be informed by instructions that are common to all observers who identify, categorize or describe the units of interest. Users of Newsmental were not explicitly instructed on how to judge news items, this was done on purpose and partly efforts were made to design the website so as to imply the judgement criteria implicitly. Even though judging the sentiment of, for example, a news item on S&P downgrading USA's credit rating, may well be self-explanatory, the judging of news duration and similarly news impact, will be much less so. It may not be clear whether a news reader refers to the duration people will talk about (buzz of the news), the duration of the news' effect on the markets (which would have long term effects to the debt markets, but possibly short- to mid-term on the general stock-markets), or some other "personal interpretation" of the news item duration rating. Hence, the fact that judgement criteria were implied implicitly must be taken into consideration when interpreting results. The agreement statistics (Section 3.3) were very low for duration judgements, and this low agreement may be contributed to by an artefact of the user-interface, as the news presentation interface only showed a graphic for collaborative sentiment and impact judgements but not duration. Hence, some limitations are imposed by specifics of the user-interface design, which highlights the sensitive nature of careful design of such systems and integrating crowd input most appropriately.

Pang and Lee (2008) suggested that within many Web 2.0 applications, it would be useful to provide finer-grained objects for contributions (e.g. identify and allow to rate particular aspects of a product in a review system), and further pointed out that still very few systems use this capability. In line with this, within Newsmental, the analysis of unstructured text was automated and the user was presented with the possibility to pass judgements on entities from within the news stories themselves. Unfortunately, this feature fails on occasion to identify the most sensible objects/entities of interest in the article. A possible extension could rely on the users collectively picking an entity they deem important from a choice of system identified ones, and when enough users picked the same entity, it could be suggested by default. Forte and Bruckman (2005) discuss the role of motivation and incentive and its implications for designing online communities. One of their suggestions is that regulars should be rewarded and leaders should be empowered. The empowering can be done by giving those users more rights, for example, to organise or otherwise manage and curate content in an editorial manner. Users of Newsmental, who are more involved than others, may be given higher access rights to organise, rank and otherwise tweak certain aspects of the news display. Hence, it would be the recommendation to explicitly address such community-level questions during future system designs. More broadly, Choraria (2012) looks at what motivates users to partake in social online communities repeatedly. They investigated the role of perceived sociability and usability on motivating platform usage, and suggest that especially ease of use, dynamic behaviours and interaction play a significant role. Hence, as was the case with Newsmental, platform development should generally focus on implementing these elements. Further to this, Husin *et al.* (2016) and Vuori (2012) provide a thorough discussion of issues and approaches in effectively adopting Web 2.0 platforms. Although their focus was on internal and external idea crowdsourcing for organisations, this is still highly relevant to the here presented work.

News items generally evoke some type of affective response within most individuals. The response can be quite varied for individuals with different backgrounds and the identification of homogenous subsets of users in terms of their news-interpretation tendencies would have been interesting to explore further. This was however outside of the scope of the current study, and because of the full background information, on only 48 registered users being available. The size of the overall crowd contributions within this study was nevertheless sizeable (e.g. 2,138 ratings, see Section 3.2) and comparable if not somewhat larger than some other crowdsourcing studies, such as, for instance, the work on Mechanical Turk and Crowd-Flower by Finin *et al.* (2010).

5. Conclusion

This paper investigates implicit crowdsourcing by harnessing a crowd's collective intelligence and cognitive abilities through social media platforms, for the construction of intelligent systems. Key issues such as achieving high-quality crowd contributions, challenges of efficient workflow and real-time crowd integration into intelligent systems were discussed. An illustrative example system, which systematically used a custom social media platform for automated financial news analysis and summarisation, was presented. The system has several social features and as a social media platform, provided for an online real-time processing workforce. Replicating human judgements and analytic abilities, as well as parsing unstructured text, is exceedingly complex. The presented social media system did not need to replicate human judgements as a function within some statistical/AI model, rather it tapped into the natural social participation of its users at a relatively low cost (i.e. system design and community recruitment costs). It is argued that when social media platforms are closely integrated with other automated processes into a single system, this may provide a worthwhile online and real-time approach to intelligent systems through the implicit crowdsourcing. The development of the system, its use and significant aspects of evaluation, including limitations and wider ethical issues of crowdsourcing were also discussed. Although more research is necessary, this study highlights key considerations for the presented approach. Some future work could include research on scalability related to online processing algorithms that can be efficiently used in processing varied content (e.g. images, videos, text and metadata) characteristic of social media, for purposes of intelligent systems.

Notes

1. See Brabham (2008) for a review of several interesting extrinsically motivated Web 2.0 examples. The case studies presented in Brabham (2008) include: iStockPhoto, Threadless, InnoCentive and several advertising competition campaigns, where cash incentives were used to motivate users/producers.
2. Problems chosen had properties of being time-consuming to solve, difficult to automate and scientifically important.
3. A systematic comparison of thousands of individual Clickworker inputs to the known, already catalogued craters showed the Clickworkers coming within a few pixels of the accepted catalogue positions (essentially within the precision of the catalogue itself). Accuracy could further be improved by cross-checking redundant inputs from different clickworkers. Faint craters classed as having little to no detectable "ejecta blanket" were detected with an impressive 95 per cent accuracy on a sample (Kanefsky *et al.*, 2001).
4. http://cci.mit.edu/
5. Clear Forest Inc. is now owned by Thompson Reuters; however, the technology behind it was mostly developed and coordinated by Prof Ronen Feldman, see www.clearforest.com/ and (Ye, 2003, p. 482).
6. The top ten entity types during the covered period were (counts in brackets): Person (18,230), Company (15,373), Position (15,321), IndustryTerm (14,223), Country (11,706), Organisation (10,992), City (5,103), ProvinceOrState (1,701), MarketIndex (1,661), Continent (1,642).
7. During the same period, there were 4,429 news articles collected in total, although due to news clustering only the first article belonging to a cluster would be available for rating to avoid showing duplicate stories. Hence, the 1,070 articles and their 2,138 ratings will be considered exclusively for further analysis (even though more articles during the period were available).

8. This is not entirely true, as user comments and news judgements are shared and visible to all. Hence, some tendency for bias exists, but it is expected that individuals will act in line with their existing convictions most of the time; however, research into social news selection (Westwood and Messing, 2011) showed some significance of social bias.
9. Hayes and Krippendorff (2007) provide a useful overview of inter-rater agreement measures. The custom SPSS kalpha macro written by Hayes and Krippendorff was used to compute the alpha and its confidence intervals – see www.afhayes.com/spss-sas-and-mplus-macros-and-code.html
10. This presented a sample of 35 unique news-items, as each item had to be rated by a minimum of five users.
11. This would only show to new/un-registered users.

References

Adler, B., Chatterjee, K., de Alfaro, L., Faella, M. and Pye, I. (2008a), "Assigning trust to wikipedia content", *WikiSym 2008: International Symposium on Wikis Proceedings*, *Porto*.

Adler, B., de Alfaro, L., Pye, I. and Raman, V. (2008b), "Measuring author contributions to the Wikipedia", *WikiSym 2008 – International Symposium on Wikis Proceedings*, *Porto*.

Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G. (2008), "Finding high-quality content in social media", *WSDM '08 International Conference on Web Search and Data Mining*, *Palo Alto*.

Barberis, N., Shleifer, A. and Vishny, R. (1998), "A model of investor sentiment", *Journal of Financial Economics*, Vol. 49 No. 307, pp. 307-343.

Bermingham, A. and Smeaton, A.F. (2010), "Crowdsourced real-world sensing: sentiment analysis and the real-time web", *Sentiment Analysis Workshop at Artificial Intelligence and Cognitive Science Proceedings*, *Galway*.

Besten, D.M. (2012), "Using social media to sample ideas: lessons from a Slate-Twitter contest", *Journal of Systems and Information Technology*, Vol. 14 No. 2, pp. 123-130.

Brabham, D.C. (2008), "Crowdsourcing as a model for problem solving: an introduction and cases", *International Journal of Research into New Media Technologies*, Vol. 14 No. 1, pp. 75-90.

Callison-Burch, C. and Dredze, M. (2010), "Creating speech and language data with Amazon's Mechanical Turk", *Proceeding of the NAACL workshop on Creating Speech and Language Data with Mechanical Turk (NAACL–HLT 2010)*, *Los Angeles*.

Chan, W.S. (2003), "Stock price reaction to news and no-news: drift and reversal after headlines", *Journal of Financial Economics*, Vol. 70 No. 2, pp. 223-260.

Choraria, S. (2012), "Factors determining the flow of information among the online community users", *Journal of Systems and Information Technology*, Vol. 14 No. 2, pp. 105-122.

Cohen, J. (1960), "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Vol. 20 No. 1, pp. 37-46.

Cronbach, L.J. (1951), "Coefficient alpha and the internal structure of tests", *Psychometrica*, Vol. 16 No. 3, pp. 297-334.

Driscoll, K. and Walker, S. (2014), "Working within a black box: transparency in the collection and production of Big Twitter data", *International Journal of Communication*, Vol. 8, pp. 1745-1764.

Ederington, L.H. and Lee, J.H. (1993), "How markets process information: news releases and volatility", *Journal of Finance*, Vol. 48 No. 4, pp. 1161-1191.

Estellés-Arolas, E. and González-Ladrón-de-Guevara, F. (2012), "Towards an integrated crowdsourcing definition", *Journal of Information Science*, Vol. 38 No. 2, pp. 189-200.

Fama, E. (1965), "Random walks in stock market prices", *Financial Analysts Journal*, Vol. 51 No. 5, pp. 55-59.

Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J. and Dredze, M. (2010), "Annotating named entities in Twitter data with crowdsourcing", *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data, Los Angeles*.

Fleiss, J.L. (1971), "Measuring nominal scale agreement among many raters", *Psychological Bulletin*, Vol. 76 No. 1, pp. 378-382.

Forte, A. and Bruckman, A. (2005), "Why do people write for Wikipedia? Incentives to contribute to open-content publishing", *Proceedings of GROUP 2005 Workshop: Sustaining Community, the Role and Design of Incentive Mechanisms in Online Systems, Orlando*.

Fung, G., Yu, J. and Lu, H. (2005), "The predicting power of textual information on financial markets", *IEEE Intelligent Informatics Bulletin*, Vol. 5 No. 1, pp. 1-10.

Giles, J. (2005), "Internet encyclopaedias go head to head", *Nature*, Vol. 438, pp. 900-901.

Grevet, C. and Gilbert, E. (2015), "Piggyback prototyping: using existing, large-scale social computing systems to prototype new ones", *The 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul*.

Hahn, U. and Mani, I. (2000), "The challenges of automatic summarization", *Computer*, Vol. 33 No. 11, pp. 29-36.

Hayes, A.F. and Krippendorff, K. (2007), "Answering the call for a standard reliability measure for coding data", *Communication Methods and Measures*, Vol. 1 No. 1, pp. 77-89.

Hofstadter, D. (1979), *Gödel, Escher, Bach: An Eternal Golden Braid*, Penguin Press.

Howe, J. (2006), "Crowdsourcing: a definition", available at: http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html (accessed 20 January 2015).

Husin, M.H., Evans, N., Deegan, G. (2016), "Achieving adoption and effective usage of web 2.0 among employees within Australian Government Organizations", *Journal of Systems and Information Technology*, Vol. 18 No. 1, pp. 41-63.

Kambil, A. (2003), "You can bet on idea markets", available at: http://hbswk.hbs.edu/archive/3808.html (accessed 1 April 2011).

Kambil, A. and Van Heck, E. (2002), *Making Markets: How Firms Can Design and Profit from Online Auctions and Exchanges*, Harvard Business Press.

Kanefsky, B., Barlow, N.G. and Gulick, V.C. (2001), "Can distributed volunteers accomplish massive data analysis tasks?", Technical Report, available at: http://bit.ly/9mtG1i (accessed 1 August 2010).

Keen, A. (2007), *The Cult of the Amateur: How the Democratization of the Digital World is Assaulting Our Economy, Our Culture, and Our Values*, Doubleday Currency Publishing.

Koppel, M. and Shtrimberg, I. (2006), "Computing attitude and affect in text: theory and applications", in Shanahan, J., Qu, Y. and Wiebe, J. (Eds), *Good News or Bad News? Let the Market Decide*, Springer, Vol. 20 No. 1, pp. 297-301.

Krippendorff, K. (2004), *Content Analysis: An Introduction to its Methodology*, Sage, Thousand Oaks, CA.

Kuznetsov, S. (2006), "Motivation of Contributors to Wikipedia", *ACM SIGCAS Journal of Computers and Society*, Vol. 36 No. 2, pp. 1-7.

Lasecki, W.S. (2014), "Crowd-powered intelligent systems", *Human Computation Journal*.

Lévy, P. (1994), *L'intelligence collective. Pour une anthropologie du cyberspace*, La Découverte, France.

Malinen, S. (2009), "Heuristics for supporting social interaction in online communities", *IADIS WWW/Internet Conference Proceedings, Rome*.

Mason, W. and Suri, S. (2012), "Conducting behavioral research on Amazon's Mechanical Turk", *Behavior Research Methods*, Vol. 44 No. 1, pp. 1-23.

Mittermayer, M.A. and Knolmayer, G. (2006), "Text mining systems for market response to news: a survey", Working Paper, Bern University, available at: www2.ie.iwi.unibe.ch/publikationen/berichte/resource/WP-184.pdf

Nagar, Y. and Malone, T. (2011), "Combining human and machine intelligence for making predictions", CCI Working Paper No 2011-02, available at: http://cci.mit.edu/publications/CCIwp2011-02.pdf

Osborne, M., Moran, S., McCreadie, R., Von Lunen, A., Sykora, M., Cano, E., Ireson, N., Macdonald, C., Ounis, I., He, Y., Jackson, T., Ciravegna, F. and O'Brien, A. (2014), "Real-time detection, tracking, and monitoring of automatically discovered events in social media", *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Baltimore*.

Pang, B. and Lee, L. (2008), "Opinion mining and sentiment analysis", *Foundations and Trends in Information Retrieval*, Vol. 2 Nos 1/2, pp. 1-135.

Paolacci, G., Chandler, J. and Ipeirotis, P.G. (2010), "Running experiments on Amazon Mechanical Turk", *Journal of Judgment and Decision Making*, Vol. 5 No. 5, pp. 411-419.

Petersen, S.M. (2008), "Loser generated content: from participation to exploitation", *Internet Journal - First Monday*, Vol. 13 No. 3.

Prpić, J., Jackson, P. and Nguyen, T. (2014), *A Computational Model of Crowds for Collective Intelligence, Collective Intelligence 2014*, MIT Center for Collective Intelligence.

Puschmann, C. and Burgess, J. (2014), "The politics of Twitter data", in Weller, K., Bruns, A., Burgess, J., Mahrt, M. and Puschmann, C. (Eds), *Twitter and Society*, Peter Lang, New York, NY, pp. 43-54.

Rogstadius, J., Kostakos, V., Laredo, J. and Vukovic, M. (2011), "A real-time social media aggregation tool: reflections from five large-scale events", *Proceedings of the European Conference on Computer-Supported Cooperative Work (ECSCW), Aarhus*.

Russell, P. (1983), *The Global Brain Awakens*, Global Brain Publishing.

Scholz, T. (2008), "Market ideology and the myths of Web 2.0", *Internet Journal – First Monday*, Vol. 13 No. 3.

Schwager, J.D. (1993), *Market Wizards: Interviews with Top Traders*, Collins Publishers.

Servan-Schreiber, E., Wolfers, J., Pennock, D.M. and Galebach, B. (2004), "Prediction market, does money matter?", *Electronic Markets*, Vol. 4 No. 3, pp. 243-251.

Shirky, C. (2010), *Cognitive Surplus: Creativity and Generosity in a Connected Age*, Allen Lane Publishers.

Sykora, M. (2009), "Power of Web 2.0 mass collaboration in computational intelligence and its' uses, an example from Finance", *Proceedings of the UKCI 9th Annual Workshop on Computational Intelligence, Nottingham*.

Taylor, J.G., Shadbolt, J., Adcock, C., Attew, D., Burgess, N., Hazarika, N., Larsson, S. and Towers, N. (2002), *Neural Networks and the Financial Markets: Predicting, Combining and Portfolio Optimisation*, Springer.

Tumasjan, A.M., Sprenger, T.O., Sandner, P.G. and Welpe, I.M. (2011), "Election forecasts with Twitter: how 140 characters reflect the political landscape", *Social Science Computer Review*, Vol. 29 No. 4, pp. 402-418.

Ushahidi (2015), "Your tools for changing the world", available at: www.ushahidi.com/product/ushahidi/ (accessed 21 January 2015).

Vuori, M. (2012), "Exploring uses of social media in a global corporation", *Journal of Systems and Information Technology*, Vol. 14 No. 2, pp. 155-170.

Wells, H.G. (1994), *World Brain: The Idea of A Permanent World Encyclopedia*, Alan Mayne Publishers.

Westwood, S.J. and Messing, S. (2011), "Selective exposure in the age of social media: endorsements trump partisan source affiliation when selecting Online News Media", Working Paper, Stanford University, available at: http://stanford.edu/~seanjw/papers/socialNews.pdf

Wuthrich, B., Cho, V., Leung, S., Permunetilleke, D., Sankaran, K. and Zhang, J. (1998), "Daily stock market forecast from textual web data systems", *IEEE International Conference on Man, and Cybernetics*, *San Dieg*.

Ye, N. (2003), *The Handbook of Data Mining*, Lawrence Erlbaum Associates Publishers.

Zemke, S. (2003), "Data mining for prediction: financial series case", PhD thesis, The Royal Institute of Technology, Department of Computer and Systems Sciences, Sweden.

Zwaan, R.A., Britton, B.K. and Graesser, A.C. (1993), *Toward A Model of Literary Comprehension, Models of Understanding Text*, Erlbaum, Hillsdale NJ.

Further reading

Widener, M.J. and Li, W. (2014), "Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US", *Applied Geography*, Vol. 54 No. 1, pp. 189-197.

Corresponding author

Martin Sykora can be contacted at: m.d.sykora@lboro.ac.uk