Institution: University of Stirling

Unit of Assessment: B11 Computer Science and Informatics

Title of case study: Sentic Computing

1. Summary of the impact

Extracting information and meaning from natural language text is central to a wide variety of computer applications, ranging from social media opinion mining to the processing of patient health-care records. Sentic Computing, pioneered at the University of Stirling, underpins a unique set of related tools for incorporating emotion and sentiment analysis in natural language processing. These tools are being employed in commercial products, with performance improvements of up to 20% being reported in accuracy of textual analysis, matching or even exceeding human performance (Zoral Labs). Current applications include social media monitoring as part of a web content management system (Sitekit Solutions Ltd), personal photo management systems (HP Labs India) and patient opinion mining (Patient Opinion Ltd). Impact has also been achieved through direct collaboration with other commercial partners such as Microsoft Research Asia, TrustPilot and Abies Ltd. Moreover, international organisations such as the Brain Sciences Foundation and the A*Star Institute for High Performance Computing have realised major impact by drawing upon our research.

2. Underpinning research

Sentic Computing, a term first coined by Stirling scientists Amir Hussain and Erik Cambria in 2010 [1], is the analysis of sentiment from natural language text. It is based on the semantic, latent and implicit meaning of concepts, permitting open-domain sentiment analysis. The analysis of natural language is based on affective ontologies and common sense reasoning tools, enabling the analysis of text not only at a document, page or paragraph level, but also at sentence and clause level. Current keyword-based approaches can perform well on specific datasets, but accuracy drops drastically on domain change. Sentic Computing provides the capability for maintaining accuracy when switching between different domains. Sentic Computing is novel because:

- It has a unique multi-disciplinary approach – which is not only computational but also biologically-inspired and psychologically-motivated.
- It goes well beyond keyword-based approaches – employing not only word co-occurrences but also cognitive and affective information associated with a range of concepts.
- It enables fine-grained analysis of text - not only at the document, page or paragraph level, but also at a sentence and clause level.

The inter-disciplinary research underpinning Sentic Computing was carried out through an EPSRC and industrial (Sitekit Solutions Ltd) co-funded research grant (CASE studentship) from 2009 to 2012 (Hussain: PI, and E. Cambria: PhD student). This was aimed at developing a novel intelligent software engine for auto-categorising and auto-tagging documents. A particular application with Sitekit is in e-health monitoring systems [2]. The research has led to the development of a range of novel Sentic Computing models, tools, and techniques, with applications in many areas, including:

- *The Hourglass of Emotions*: a biologically-inspired and psychologically-motivated model for the representation and the analysis of human emotions.
- *AffectiveSpace*: a vector space model for reasoning by analogy on affective common sense knowledge.
- *SenticNet*: a publicly available semantic resource for opinion mining built using an ensemble of AI and Semantic Web techniques, released in 2010 [3].
- *IsaCore*: a semantic network of common and common sense knowledge for auto-categorization.

As will be detailed below, much of this work was carried out in direct collaboration with a number of companies (in addition to Sitekit), so that commercial application began before the underpinning research being carried out at Stirling was finally published. The current state-of-the-art is summarised in a book on Sentic Computing by Cambria and Hussain [4], and the research is reported in recently published articles [1,2,5-6].

3. References to the research

[1] E. Cambria, A. Hussain, C. Havasi, and C. Eckl. Sentic computing: Exploitation of common sense for the development of emotion-sensitive systems. In: Lecture Notes in Computer Science, vol. 5967, pp. 148-156, Springer (2010)
[2] E. Cambria, T. Benson, C. Eckl, and A. Hussain. Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality. Expert Systems with Applications 39(12), pp. 10533–10543 (2012)
[3] (Web) Sentic Computing website: http://sentic.net
[4] E. Cambria and A. Hussain. Sentic Computing: Techniques, Tools, and Applications. Dordrecht, Netherlands: Springer, ISBN: 978-94-007-5069-2 (2012)
[5] E. Cambria and A. Hussain. Sentic album: Content-, concept-, and context-based online personal photo management system. Cognitive Computation 4(4), pp. 477-496 (2012)
[6] E. Cambria, M. Grassi, A. Hussain, and C. Havasi. Sentic computing for social media marketing. Multimedia Tools and Applications 59(2), pp. 557-577 (2012)

Related grant:

EPSRC CASE studentship (2009-2012): "Application of Common Sense Computing for Enabling Next-generation Semantic Web Applications" PI: A. Hussain; PhD Student: E. Cambria. Co-funded by SiteKit Solutions Ltd. (Scotland) and in collaboration with MIT Media Lab (USA).

See CASE Studentship Interface article: http://interfaceonline.org/uploads/3804/Sitekit_09.pdf

4. Details of the impact

Evidence of the approach's significant commercial impact can be found in the adoption of Sentic Computing tools and techniques (as part of bigger commercial systems) by large international companies who are Stirling collaborators. These can be specifically outlined as follows:

(1) **HP Labs India** for image metadata processing, social network analysis, user profiling, social communication, and troll filtering.

(2) **Microsoft Research Asia** for text categorisation, knowledge base design and sentiment polarity detection.

(3) **Zoral Labs** for real-time monitoring and extraction of transactions from any type of unstructured data.

(4) **TrustPilot** for effective commercial machine learning technology

Three UK SME companies have also adopted our techniques, namely

(5) **Patient Opinion Ltd** for the automatic analysis of unstructured patient opinions.

(6) **Abies Ltd** for the daily measurement of patients' healthcare quality of life.

(7) **Sitekit Solutions Ltd** (a Microsoft Gold Partner company) for document and web-page auto-categorisation.

In addition, the following international institutes have drawn upon our Sentic Computing research:

(8) **The Brain Sciences Foundation** (http://www.brainsciences.org/) represents a collaboration between MIT, University of Oxford, UCLA at Irvine, Boston University and University of Paris, which is physically located in Providence, Rhode Island. It has used our Sentic computing research for metaphor detection.

(9) **The A*STAR Institute for High Performance Computing** is a research institute (http://www.ihpc.a-star.edu.sg/) supported by the Singapore Agency for Science, Technology and Research. It has employed Sentic Computing to underpin a suite of analytical tools.

The depth and reach of the impact of our research on all the organisations outlined above can be highlighted by considering the following products and tools:

(i) ***Social Media Marketing Tool*:** this is an intelligent Sentic Computing based web application that helps companies efficiently visualise and manage relevant social media information, and accordingly perform product positioning. The tool, released in 2012, is used by Stirling's industrial funding partner, **Sitekit Solutions Ltd** in the field of social media monitoring and is offered as additional functionality to **Sitekit** customers within **Sitekit**'s current commercial content management systems.

(ii) ***Sentic PROMs*:** this represents a new framework that exploits the ensemble application of standard PROMs (patient-reported outcome measures) and Sentic Computing for measuring patients' health related quality of life in a semi-structured way. The commercial development of Sentic PROMs was funded by the UK Technology Strategy Board funded grant, eCommissioning Community to Support NHS GP Consortia (£858k, TSB Grant Reference: 12074-75246), with C. Grant (CEO of **Sitekit Solutions Ltd**) as Principal Investigator and T. Benson (CEO of **Abies Ltd**) as Co-Investigator (and co-author of research reference [2] above). Sentic PROMs are now in use by **Sitekit Ltd**'s partner, **Abies Ltd**, in place of their standard PROMs as a key clinical assessment tool. Stirling played a major role during this development (which required company visits by Cambria). The CEO of **Abies Ltd** has said, "*Sentic PROMs are the next-stage of health-related quality of life measurement, and are helping substantially expand Abies healthcare business".*

(iii) SENTRA: Stirling-pioneered Sentic Computing tools and techniques have also been adopted into a commercial product by **Zoral Labs**, a world-leading company specialising in processing unstructured data. According to **Zoral Labs**, most state-of-the-art textual analysis engines are not granular (i.e. are not sentence-level based), and understand neither "context" nor "relationships". They achieve accuracy rates below 70%. However, **Zoral Labs**' novel scalable Sentic Computing based sentiment transactions system (SENTRA) (now available on the market, and which exploits Stirling published research and software) has been found in their tests to routinely deliver accuracy rates in excess of 85-90%. This is equivalent to, or even slightly in excess of, human performance. The Business Development Manager at **Zoral Labs** said "*we are proud to confirm that we have employed the sentic computing engine developed by Stirling in our commercial product, SENTRA. We always want to be updated with state-of-the-art technologies and sentic computing is surely the next stage of intelligent opinion mining systems*".

(iv) ***Sentic Album*:** this is a content, concept, and context based online personal photo management system. Sentic Album modules, released in 2012, are now used by **HP Labs India**, for photo management systems. In addition, *The Hourglass of Emotions*, a biologically-inspired and psychologically-motivated model for the representation and the analysis of human emotions, released in 2012, has been adopted for several applications by **HP Labs India**. This work was facilitated by Cambria spending six months on an invited research visit to **HP Labs**. A senior researcher at **HP Labs** has said, "*The sentic framework developed by Stirling has been applied in several applications designed here at the Labs. It's been a joy to collaborate with Erik and Amir and I am looking forward to the next SenticNet release*".

(v) ***Crowd Validation*:** this is a process for mining patient opinions that can be applied to any domain for bridging the gap between unstructured and structured data. Contacts through **Sitekit Ltd** resulted in Stirling collaborating in parallel with, **Patient Opinion Ltd**. This company is now commercially exploiting such a process, released in 2010, for automatically categorising new patient opinions and hence for improving opinion search. The CEO of Patient Opinion said: "*The growth of Patient Opinion over the past few years was greatly affected by the adoption of Cambria and Hussain's Crowd Validation technique, in which we can now easily aggregate patient-related information in a more meaningful and user-friendly way*".

(vi) ***AffectiveSpace*:** this is a vector space representation of AffectNet for reasoning by analogy on affective common sense knowledge. The tool is being used by Luminoso (spin-off of Stirling collaborators, MIT Common Sense Computing) for detecting the polarity of sentiments in natural language text. MIT acted as a project partner on the collaborative project between Stirling and **Sitekit Ltd**, with Cambria spending time at MIT, with one outcome being this application.

(vii) ***IsaCore*:** this is a semantic network of common (and common-sense) knowledge for auto-categorization built upon ConceptNet and Probase. IsaCore, released in 2012, has been adopted by Stirling's Chinese collaborator, **Microsoft Research Asia**, for the retrieval of semantically related concepts/instances in Probase as it outperforms previous probabilistic approaches to reasoning by analogy. This collaboration arose from contacts made through **Sitekit Ltd** and the Chinese Academy of Sciences which (jointly with the Royal Society of Edinburgh) funds a separate research collaboration with Stirling. Cambria spent several months at **Microsoft Research Asia**. The Head of the Web Search and Data Mining Group said "*SenticNet and Sentic Computing by Cambria and Hussain are ground-breaking technologies that are helping our team to enrich ProBase and to showcase its usefulness in many different NLP applications such as*

topic modelling and opinion mining".

(viii) ***IARPA Metaphor Project:*** The **Brain Sciences Foundation** has drawn upon our research in their IARPA (Intelligence Advanced Research Projects Activity) Metaphor project to enable deeper understanding of metaphor detection. The MIT Director of the **Brain Sciences Foundation**, involved with IARPA, said, "*I am delighted to confirm that sentic computing techniques pioneered by Erik and Amir have been widely adopted by the Brain Sciences Foundation and have been deployed as the core sentiment and opinion mining module of our existing commercial system. We believe this has added significant value to our system, and is continuing to attract the interest of various companies around the world. We have employed SenticNet in many different projects of ours including the IARPA Metaphor Project, for metaphor detection and understanding*".

(ix) ***SPICE*****:** This is a suite of analytical tools that has been developed by the **A*STAR** institute and which draws heavily on our research. A principal investigator at **A*STAR** said, "*We have adopted Sentic Computing to develop a human-in-the-loop platform with a suite of analytic tools for social media monitoring, analysis, tracking, as well as provide communication strategy recommendation that we refer to as SPICE [Strategic Public Information and Communication Enhancement]. Sentic computing techniques turned out to be key in enhancing SPICE performance for tasks such as fine-grained opinion mining and sentiment analysis, influencer network analysis, insight analysis, and social media based psychographic analysis".*

(x) ***TrustPilot Technology***: A developer at **TrustPilot** said: "*I participated to WWW13 Conference to attend Dr Cambria's tutorial on Sentic Computing. We are now using SenticNet as one of the means to enhance TrustPilot technology. Sentic Computing is taking us much further than any other machine learning technique we have applied before.*"

Sentic Computing tools are freely available for download and use through the Sentic API at sentic.net. The Sentic API is being used by a continuously growing number of researchers and companies world-wide: sentic.net was visited 600,000+ times from Oct 2012 to June 2013 (statistics available upon request from iPage.)

5. Sources to corroborate the impact (indicative maximum of 10 references)

Commercialization points of contact:

1. CEO of Sitekit Solutions Ltd. (Portree, UK) and CEO of Abies Ltd. (Berkshire, UK) - for development of Sentic PROMs (Aug 2011 - Mar 2012), collaboration within a project funded by the UK Technology Strategy Board (TSB Grant Reference: 12074-75246).
2. Business Development Manager at Zoral Labs – for Sentic Net based sentiment transactions system (SENTRA).
3. Senior Researcher at Web Access and Interaction Group, HP Labs India (Bangalore, India), within the Innovations for the Next Billion Customers Initiative, for development of Sentic Album, Sentic Corner, and Sentic Chat (Jun – Oct 2010).
4. CEO Patient Opinion Ltd (Sheffield, UK) - for Sentic computing based automatic analysis of patient opinions (Mar 2011 – Mar 2012).
5. Head of Web Search and Data Mining Group, Microsoft Research Asia (Beijing, China) – as part of the ProBase Project, for exploitation of Sentic computing techniques to knowledge representation and reasoning with ProBase (Feb – Jul 2011).
6. Director of the Brain Sciences Foundation, Director of MIT Mind Machine Project and MIT Synthetic Intelligence Project - for exploitation of Sentic computing to develop brain-inspired cognitive architectures.
7. Principal Investigator at A*STAR Institute of High Performance Computing (IHPC).
8. Developer at TrustPilot.

Weblinks:

Sentiment polarity detection demonstration: http://sentic.net/demo

SENTRA: http://zorallabs.com/products/unstructured-data-management

Probase: http://research.microsoft.com/probase