Cracking the Code: What customers say in their own words

Tim Macer, University of Southampton; Mark Pearson, Egg plc; Fabrizio Sebastiani, National Council of Research, Italy

Abstract

The paper describes how Egg, a large UK internet bank, in partnership with meaning ltd, a research technology consultancy, commissioned the Institute for the Science and Technology of Information of the Italian National Council of Research (ISTI-CNR) to build VCS, a novel software solution for the automatic analysis of the many thousands of verbatim comments Egg collects through event and customer experience surveys conducted online. The sheer volume of responses received had made any systematic analysis of them impossible, which is problem experienced not only by Egg but by many who conduct research online. These difficulties mean that researchers often restrict research designs to reduce or eliminate open questions from online surveys, despite the richness of the insights that can be obtained in this way.

Unlike other automated or computer-assisted software, the system developed for Egg is novel in that it applies machine learning to the problem of analysing and classifying verbatim texts from open-ended questions. With few precedents for the application of this method to verbatim textual data, which is characterised not only by great diversity of content but by frequent linguistic mistakes, a cautious experimental approach was taken in order to validate the software's accuracy and reliability in performing the task.

The results of these experiments, presented in this paper, show that the software performs well against human coders that have subject matter expertise. Overall accuracy using a range of statistical measures was broadly similar. Furthermore, it was shown that accuracy could be improved by increasing the number of training examples, thereby demonstrating that the system is inherently trainable, and its performance can be monitored and improved with respect each survey or question being analysed.

Practical experiences described here, since the software was adopted at Egg, demonstrate dramatic savings in cost and time, making comprehensive and highly systematic analysis of many thousands of verbatim responses cost effective.

It concludes by considering the implications of this technology for all engaged in research or evaluating customer feedback, and how its application could profoundly change the way research is undertaken, by effectively removing restrictions on placing open-ended questions in online surveys.

Speakers' Biographies

Tim Macer is the founder of meaning ltd, a consulting company specialising in research technology, for which he is known internationally as an independent and authoritative observer and advocate. He is Visiting Senior Fellow at the University of Southampton in marketing analytics and is a widely published writer on software and IT for research.

Mark Pearson was until recently Head of Brand and Insight at Egg and is now an independent Insight Consultant. Mark is one of the founders of Egg, where he pioneered the development of online research and consumer immersion techniques. He has presented papers at MRS and ESOMAR conferences and been a contributor to various government strategies.

Dr Fabrizio Sebastiani is a senior researcher at the Institute for the Science and Technology of Information of the Italian National Council of Research in Pisa, Italy, and a former professor of the University of Padova. He has published widely on the topics of information retrieval, machine learning, text classification, and opinion mining.