Journal of Business Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Next frontiers of research in data driven marketing: Will techniques keep up with data tsunami?

Jagdish Sheth, Charles H. Kellstadt

Goizueta Business School, Emory University, United States

ARTICLE INFO

Keywords:
Quantitative models
Data driven marketing
Video analytics
Biometric research
AI
Blockchain

ABSTRACT

The purpose of the paper is to provide a historical evolution of data driven marketing and suggest next frontiers of research opportunities as a consequence of digital tsunami as well as affordability and accessibility to conduct databased research. Earlier, techniques were in search of data but in the future data will be in search of techniques, especially in the era of social media such as Twitter, Facebook, YouTube, What's App, and Instagram.

1. Introduction

The purpose of this paper is to forecast the future of data driven marketing by analyzing its evolution since the fifties. The paper concludes that while techniques were in search of data in the early days, it will be the data which will be in search of techniques as a consequence of the shift from the structured to unstructured data and at a scale which resembles a data tsunami.

1.1. Evolution of marketing analytics

The genesis of data driven marketing goes back to the fifties. There was a conscious effort to make marketing (as well as all disciplines of business such as finance and operations) more scientific through quantitative techniques such as Markov chains and linear programming in addition to time series analysis. The key shift was from marketing as a descriptive discipline to marketing as a predictive science with a special interest in optimization. For example, marketing scholars notably at MIT, Stanford, and Purdue University, began to develop optimization of sales routes, optimal allocation of advertising budgets, location of warehouses, and inventory management through optimal distribution of products (Ronald Howard, 1963) to retail outlets (Buzzell, 1963; Bass, 1961; Massy, Montgomery, & Morrison, 1970; Bass, 1969; Massy, 1969; Kuehn, 1962).

Pricing and especially sales promotions decisions were made based on data driven heuristics and measuring elasticity of price and promotion through field experiments. The most comprehensive research was a series of field studies done by Anheuser-Busch on the Budweiser brand to measure whether advertising has any effect on sales and market share of Budweiser. More than 140 separate experiments were carried out over several years and in different geographies in the U.S. The conclusion was inconclusive due to mixed or conflicting results (Ackoff & Emshoff, 1975).

Scholars in operations research and econometrics including Alfred Kuehn, Ronald Frank, Christian Palda, John Farley, John D. C. Little, Frank Bass, Paul Green, and William Massy began to organize special conferences on analytical marketing and aligned with Management Science and INFORMS. Winer and Neslin (2014) provide an excellent history of marketing science in their edited book on marketing science.

1.2. Formation of MSI

It also led to the formation of Marketing Science Institute (MSI) in 1961 with a unique collaboration between the academic and the industry communities. Scott Paper Company President Thomas B. McCabe founded the "Institute for Marketing Science" in Philadelphia with input from leading thinkers (John A. Howard, Albert Wesley Frey and Wroe Alderson) to "stimulate increased application of scientific techniques to the understanding and solving of current marketing problems".

Twenty-nine companies responded to McCabe's membership appeal and established MSI as a nonprofit organization that would contribute to "the emergence of a definitive science or marketing" (Wikipedia, 2019).

Interest in data driven marketing was further fueled by access to data on household panels provided by Market-Research Corporation of America (MRCA) and by the Chicago Tribune. It was also enabled by access to large scale statistical packages including the BIOMED, SPSS, OSIRIS, and SAS.

Finally, emergence of desktop programs and large storage of data made quantitative research easier, faster, and cheaper. In short, democratization of data analysis through computerization led to explosive

https://doi.org/10.1016/j.jbusres.2020.04.050
Received 10 April 2020; Received in revised form 22 April 2020; Accepted 25 April 2020
0148-2963/

growth of what is referred to as marketing analytics today.

2. Evolution of different techniques

As mentioned before, econometric and operations research techniques began to search for marketing data both at the household level (micro) as well as at the aggregate market level (macro). These included MRCA and Chicago Tribune household panels and scanner data about products, brands, and stock keeping units (sku). This was during the decades of the sixties and seventies. A significant boost was provided by a large number of studies performed on the PIMS data based on the impact of market share on profitability (Buzzell, 1963; Bharadwaj & Menon, 1993).

2.1. Growth of multivariate techniques

In the seventies and the eighties, the marketing analytics shifted from econometric to psychometric techniques and from univariate to multivariate techniques (Sheth, 1977; Sheth, 1971). These included multidimensional scaling (MDS), cluster analysis, conjoint analysis, discriminant analysis, and ultimately the LISREL. Hundreds of research papers were published using these newer techniques. They required different set of data such as surveys and multi-attribute altitude data. While the multivariate revolution continued for more than three decades, there was also the emergence of biometrics such as pupil dilation and galvin skin pressure.

In the nineties, multivariate techniques started to plateau with the rise of what is known as choice models (McAlister et al., 1991). Also, as the marketing discipline began to focus on loyalty programs and relationship marketing, customer lifetime value (CLV) became a very popular technique (Venkatesan & Kumar, 2004; Kumar & Shah, 2009; Gupta, Lehmann, & Stuart, 2004; Rust, Lemon, & Zeithaml, 2004; Kumar, Shah, & Venkatesan, 2006).

2.2. Data tsunami

With the digital revolution (cell phones, internet, and World Wide Web) in the new millennium, there has been an explosion of data. Digital storage has gone through a revolutionary change. Today one can store massive data files on a flash drive. Indeed, the whole Library of Congress can be put on one or two USB (flash devices) and make the data portable.

As more and more marketing communications, market transactions, and customer feedback are online, it has generated a tsunami of unstructured data (text messages). This is further fueled by the unpreceded growth of social media including Facebook, What's App, YouTube, and Instagram. All of them generate more unstructured (non-numerical) data because of the use of text and video messages. For the first time, data is in search of techniques as opposed to techniques in search of data. Most of existing techniques are based on statistics. Therefore, they require numerical data for statistical inference and for empirical findings. In the process, inferential statistics is replaced with non-inferential statistical techniques such as natural language processing (NLP).

3. Next frontiers of research in data driven markets

Internet, World Wide Web, and Google search engines made connectivity and communication virtually universal. The old notion of six degrees of separation (you can reach someone in six sequential linkages) is now virtually two and even one degree of separation. Data storage today is cheaper than paper. However, the real revolution in data driven marketing is the popularity of social media as a way to interact, communicate, share information, influence others, and even do market transactions on a global basis (Kim & Ko, 2012; De Vries, Gensler, & Leeflang, 2012; Lipsman et al.; Tuten & Solomon, 2017).

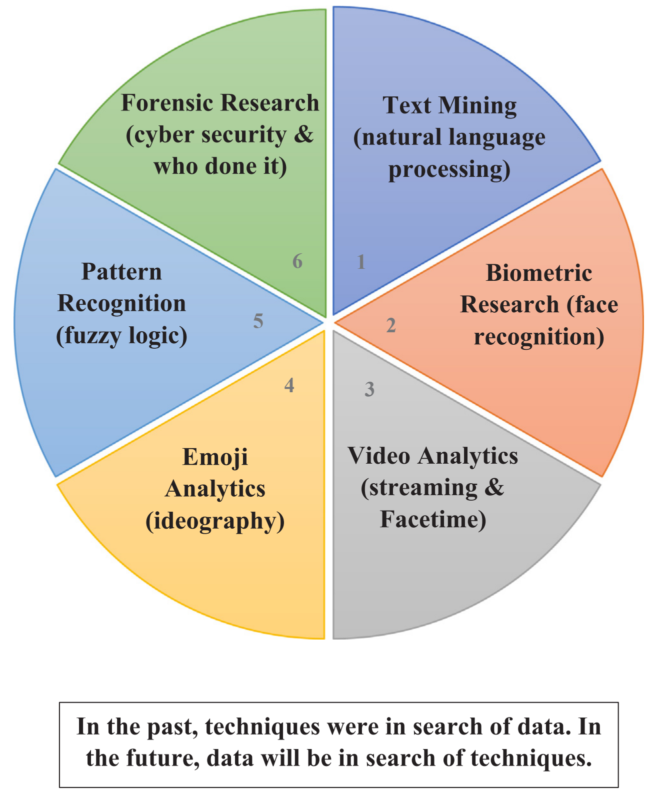


Fig. 1. Next frontiers of research in marketing.

Today, most marketers have their own website and social media presence including Facebook, Twitter, YouTube, Good, LinkedIn, Instagram, Yelp, and What's App. Today, you have celebrities who have millions of followers. Today, anybody can have a YouTube channel or a Facebook page.

In the process, the data has shifted from pre-selling to post-purchase experiences. Today, customer experience has replaced customer satisfaction as a focal interest in marketing.

In Fig. 1, I have identified six new and emerging areas of research opportunities in digital and data driven marketing.

3.1. Text mining

The first is Text Mining (e.g. Aggarwal & Zhai 2012; He, Zha, & Li, 2013; Netzer, Feldman, Goldenberg, & Fresko, 2012). Explosion of text messages on platforms such as What's App and Twitter is creating a data tsunami especially in emerging markets such as China and India (over 65 billion messages in 2018 on What's App alone). This has led to a new technique, natural language processing (NLP). While the technique emerged from cognitive psychology and linguistics, it is increasingly used in data driven marketing. NLP is not a predictive tool in a pure statistical sense. It is more a tool for typology and classification. However, once words and phrases are classified, the numerical counts can create word clouds as well as they can be subjected to either non-parametric or parametric statistics for inferential as well as predictive purposes.

The next five areas of research in marketing analytics have no agreed upon or de facto techniques. It is only a matter of time before some discipline in social sciences will develop techniques, which become de facto by their usage.

3.2. Biometric data

Biometric Research (e.g. Dunstone & Yager, 2008; Dantcheva, Elia, & Ross, 2015; Derawi, Nickel, Bours, & Busch, 2010) is most likely to be the next area of research in marketing. The biometric databases are growing exponentially both in advanced as well as in emerging economies such as China and India. For example, India now has more than one billion people with their biometric profile. It is one of the largest biometric databases in the world (Aadhar). It is already used for providing many social and economic benefits to low income people and has led to more than 200 million to have access to modern banking. It bypassed dependence on credit cards and laptops. It is anchored to cell phones and will aggregate more data as Indian consumers migrate from feature phones to smart phones.

Similarly, China has instituted a social score rating of its citizens. Every Chinese person is identified with his or her behavior including crossing the street, speeding in the car, credit card default, and use of social media and eCommerce. Each person is given a social score similar to credit score. If a person's social score is below a certain threshold, many public services are either denied or the person has to pay extra. This is like a credit score in the U.S. and other advanced countries where below a certain credit score, the customer is denied access to credit such as car loans or home mortgages. It also includes paying higher interest rates on the credit card charges.

3.3. Video analytics

A third area of growing interest is Video Analytics (e.g. Bartlett et al., 2008; Hospedales, Gong, & Xiang, 2009; He, 2013). Video content is also growing at an exponential rate. In less than fifteen years, YouTube now has more than two billion active monthly users. More than five billion videos are shared to date. Similarly, more than five billion videos are watched per day across more than 80 countries. While user generated content is still dominant, YouTube is increasingly adding content from the publishing and media companies as is the case from many government agencies and private companies.

YouTube offers unlimited channels. Therefore, it has inherent advantages over television and cable industries. It is possible for a marketer to have separate corporate channel as well as each product or brand channel on YouTube. It can, therefore, segment and target the market. As marketing expenditures shift from print and television to social media to attract and retain customers, it will become increasingly important to measure elasticity of video communication. Video analytics is in its infancy. As the old Chinese saying that one picture is worth a thousand words, the complexity of analyzing video content will be enormous.

3.4. Emoji analytics

The fourth area of research in data driven marketing is Emoji Analytics (e.g. Felbo, Mislove, Søgaard, Rahwan, & Lehmann, 2017; Wood & Ruder, 2016; Shiha & Ayvaz, 2017). As more emoji get inaugurated in the text messages, the more it is likely to generate data that have predictive powers. Fortunately, ideography is a well-accepted science and therefore data can be analyzed using ideography. Consequently, consumer insights and sentiments using emoji analytics will have rapid acceptance and utilization. This is one place where technique is in search of data, contrary to my earlier observation.

3.5. Pattern recognition

The fifth area of next generation research in data driven marketing is Pattern Recognition (e.g. Albus et al., 2012; Fukunaga, 2013; Devroye, Györfi, & Lugosi, 2013). Similar to emoji analytics, there are well known algorithms and techniques in engineering for pattern recognition. One example is fuzzy logic. In many ways, pattern recognition is ideal in a world of partial information. Therefore, detecting weak or distant signals and predicting the future is very important for marketing performance and productivity in new product introductions, pricing variations, as well as capturing the demographic and psychographic trends as market signals.

Marketing is all about anticipating the future and fuzzy logic is probably the best technique. It is embedded in most digital cameras, consumer electronics, and security-related products and services.

3.6. Forensic research

The sixth area of research in database marketing is Forensic Research (e.g. National Research Council, 2009; Houck & Siegel, 2009; Robertson, Vignaux, & Berger, 2016). It is a very useful tool in litigations such as protecting intellectual property, trademarks, and other marketing assets. It is equally important in public policy litigation such as antitrust, societal harm, deceptive marketing practices, and non-compliance of existing regulations. Examples of most recent corporate scandals are cheating on EPA mandated carbon emission by Volkswagen, and the subprime lending by banks which led to the great economic recession of 2008. Finally, forensic research in marketing will be increasingly necessary for internal governance and compliance. Just as forensic accounting is standard practice, marketing needs to develop expertise in forensic research on marketing failures of new products or advertising campaigns and pricing changes. Finally, marketing is one of the most vulnerable functions from cyber-attacks on customer databases. The recent examples of credit card data breaches as well as credit score ratings suggests that cyber security in marketing will grow in the future.

4. Challenges in next frontiers of data driven marketing

There are several challenges for marketing scholars as they embrace the six new frontiers of research.

4.1. Curating data

- Curation. The biggest challenge in the future will be data curation. There is so much information that one needs to be selective. In addition, just like in a museum, there are great fake artifacts, it is important to ensure authenticity of the data. Finally, how do you organize the data especially from diverse sources and diverse media (print, text, video, images) and integrate them so that they are ready for analysis?

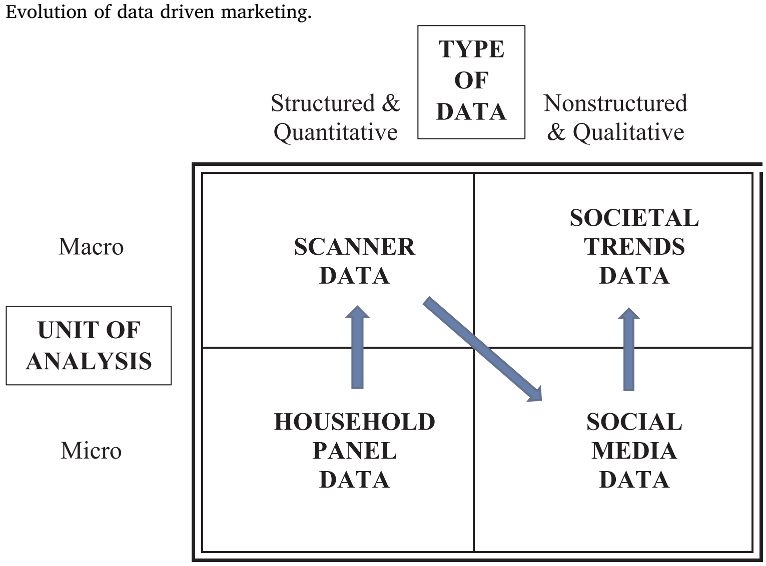
4.2. Data analysis

- Analysis. Equally important challenge will be data analysis. As mentioned before, today data is in search of a technique in contrast with earlier history where techniques were in search of data. Since there is inherent data diversity with the rise of text and video content, it will be hard to select a technique appropriate for the data.

While each type of data may generate its own technique for example, NLP for text messages, one has to decide which technique or techniques will be suitable for analysis in an integrated data environment.

More than forty years ago, Scott Armstrong (Armstrong, 1975) wrote a classic article on Tom Swift and his electronic machine. He observed that each one of us grows up with a favorite technique and we have a tendency to use it across diverse data, which may or may not meet the requirements of the technique. Examples include conjoint analysis, logit-profit models, LISREL, and more recently, latent Markov analysis. This technique bias will be a key challenge in the future because the future content is not amenable to known techniques. Obsolesce of known techniques is a real possibility. It did happen when marketing analytics made a disruptive shift from econometrics and operations research to Bayesian and multivariate statistics.

Table 1
Evolution of data driven marketing.



4.3. Insights from data

- Insights. The main objective of data driven marketing is to gain insights from the data. Insights need a perspective or a paradigm (a point of view, hypothesis, proposition, or a theory). And most of existing paradigms and perspectives are anchored to prior research and the early socialization in research scholarship. Unfortunately, they are also grounded in cultural biases. On the other hand, social media data are truly global. In my view, the largest nation in the world in not China or India, but it is what I call the Facebook Nation with more than two billion inhabitants with greater cultural diversity with respect to faith, values, and daily life than most other nations.

I believe the future will resemble like the five blind men and the elephant: one who touches the tail will interpret the reality as a rope versus the one who touches the trunk will interpret it as a snake, and the one who touches the leg will think it is a tree trunk. Each one is right from their perspective. This may require acceptance of multi-perspectives and tolerance for multiple interpretations.

4.4. Half life of knowledge

- Declining Half Life. The last challenge is to accept the fact that marketing practice is both dynamic and context bound. And as the context (ecosystem) changes and as there will be more discontinuities our time, the half-life of research will be declining and declining sharply. As it is true in software so will it be in marketing.

The declining half-life of knowledge is a serious issue. If anchored to techniques, databases, and empirical findings, we will not fully understand its impact on marketing analytics.

5. Concluding remarks

Data driven marketing has bright future. It will align and adapt to new databases and new domains. The real opportunity will lie in focusing on policy research and societal trends including poverty, emerging markets, sustainability, wellness, and education. In other words, the real contribution that data driven marketing can make is to go beyond the firm and its profit seeking motive. Table 1 summarizes the evolution of database marketing and its future trajectory.

Finally, next frontiers of research in marketing analytics will require cross disciplinary research teams. It will require collaboration between computer science, behavioral science, and quantitative science as well as with policy researchers. In other words, data driven marketing will require a programmatic research housed outside the marketing department.

Given the never-ending data tsunami, it is easier to predict that new techniques based on heuristics and algorithm will emerge from artificial intelligence, machine learning, and blockchain.

References

Ackoff, R. L., & Emshoff, J. R. (1975). Advertising research at Anheuser-Busch, Inc. (1963-68). Sloan Management Review (pre-1986), 16(2), 1.

Aggarwal, C.C., & Zhai, C., eds. (2012). Mining text data. Springer Science & Business Media.

Albus, J.E., Anderson, R. H., Brayer, J. M., DeMori, R., Feng, H., Horowitz, S. L., Moayer, B., et al. (2012). Syntactic pattern recognition, applications. Vol. 14. Springer Science & Business Media.

Armstrong, J. S. (1975). Tom Swift and his electric regression analysis machine: 1973. Marketing Papers, 37.

Bartlett, M., Littlewort, G., Vural, E., Lee, K., Cetin, M., Ercil, A., & Movellan, J. (2008). Data mining spontaneous facial behavior with automatic expression coding. Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction, pp 1-20. Springer, Berlin, Heidelberg.

Bass, F. M. (1961). Mathematical Models and Methods in Marketing. Homewood, Illinois: Richard D. Irwin, Inc.

Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science, 15*(5), 215–227.

Bharadwaj, S. G., & Menon, A. (1993). Determinants of success in service industries: A PIMS-based. *The Journal of Services Marketing, 7*(4), 19–22.

Buzzell, R. D. (1963). Is Marketing a Science? *Harvard Business Review, 41*(1), 32–40.

Dantcheva, A., Elia, P., & Ross, A. (2015). What else does your biometric data reveal? A survey on soft biometrics. *IEEE Transactions on Information Forensics and Security, 11*(3), 441–467.

Derawi, M. O., Nickel, C., Bours, P., & Busch, C. (2010). Unobtrusive user-authentication on mobile phones using biometric gait recognition. *IEEE Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing,* 306–311.

De Vries, L., Gensler, S., & Leeflang, P. S. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing, 26*(2), 83–91.

Devroye, L., Györfi, L., & Lugosi, G. (2013). A probabilistic theory of pattern recognition. Vol (31). Springer Science & Business Media.

Dunstone, T., & Yager, N. (2008). Biometric system and data analysis: Design, evaluation, and data mining. Springer Science & Business Media.

Felbo, Bjarke, Alan Mislove, Anders Søgaard, Iyad Rahwan, & Sune Lehmann. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing.

Fukunaga, K. (2013). Introduction to statistical pattern recognition. Elsevier.

Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research, 41*(1), 7–18.

He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior, 29*(1),

90–102.
He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management, 33*(3), 464–472.
Hospedales, T., Gong, S., & Xiang, T. (2009). A Markov clustering topic model for mining behavior in video. IEEE 12th International Conference on Computer Vision, 1165–1172.
Houck, M. M., & Siegel, J. A. (2009). *Fundamentals of Forensic Science.* Academic Press.
Howard, R.A. Semi-Markovian control systems. No. 3, Massachusetts Institute of Technology, Cambridge, 1963.
Kim, A. J., & Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business Research, 65*(10), 1480–1486.
Kuehn, A. A. (1962). Consumer Brand Choice as a Learning Process. *Journal of Advertising Research, 2*, 10–17.
Kumar, V., Shah, D., & Venkatesan, R. (2006). Managing retailer profitability—one customer at a time!. *Journal of Retailing, 82*(4), 277–294.
Kumar, V., & Shah, D. (2009). Expanding the role of marketing: From customer equity to market capitalization. *Journal of Marketing, 73*(6), 119–136.
Lipsman, A., Mudd, G., Rich, M., & Bruich, S. (2012). The power of "like": How brands reach (and influence) fans through social-media marketing. *Journal of Advertising Research, 52*(1), 40–52.
Massy, W. F. (1969). Forecasting the demand for new convenience products. *Journal of Marketing Research, 6*(4), 405–412.
Massy, W. F., Montgomery, D. B., & Morrison, D. G. (1970). *Stochastic Models of Buyer Behavior.* Cambridge, MA: MIT Press.
McAlister, L., Srivastava, R., Horowitz, J., Jones, M., Kamakura, W., Kulchitsky, J., et al. (1991). Incorporating choice dynamics in models of consumer behavior. *Marketing Letters, 2*(3), 241–252.
National Research Council (2009). *Strengthening forensic science in the United States: A path forward.* National Academies Press.
Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science, 31*(3), 521–543.
Robertson, B., Vignaux, G. A., & Berger, C. E. H. (2016). *Interpreting evidence: Evaluating forensic science in the courtroom.* John Wiley & Sons.
Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing, 68*(1), 109–127.
Sheth, Jagdish N. (1971). Multivariate revolution in marketing research. *Journal of Marketing, 35*(1), 13–19.
Sheth, J. N. (1977). *Multivariate Methods for Market and Survey Research.* Chicago, IL: American Marketing Association.
Shiha, M. O., & Ayvaz, S. (2017). The effects of emoji in sentiment analysis. *IJCEE, 9*(1), 360–369.
Tuten, T. L., & Solomon, M. R. (2017). *Social media marketing.* Sage.
Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing, 68*(4), 106–125.
Winer, R. S., & Neslin, S. A. (Eds.). (2014). *The History of Marketing Science*. New York, NY: World Scientific.
Wikipedia contributors (2019). Marketing Science Institute. In Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/Marketing_Science_Institute.
Wood, I. D., & Ruder, S. (2016). Emoji as emotion tags for tweets. Portorož, Slovenia. In Proceedings of the Emotion and Sentiment Analysis Workshop LREC2016, 76-79.