Performing opinion mining and analytical study for cashless transactions

Sonakshi Vij

Department of CSE,
Indira Gandhi Delhi Technical University for Women, India
Email: sonakshi.vij92@gmail.com

Amita Jain*

Department of CSE,
Ambedkar Institute of Advanced Communication
Technologies and Research, India
Email: amita_jain_17@yahoo.com
*Corresponding author

Devendra Tayal

Department of CSE,
Indira Gandhi Delhi Technical University for Women, India
Email: dev_tayal2001@yahoo.com

Abstract: Cashless economy and digital payments are one of the hot topics of research that are gaining enormous popularity day by day. It is primarily due to their ability to combat issues of corruption, back money, illegal transactions and counterfeit money. With the increase in number of options of digital payments, people are inclining themselves towards ease of payment. This paper focuses on two aspects: firstly, it provides an analytical study performed in the field of cashless transactions and digital payments. Secondly, it finds out the opinion of the people about cashless transactions using a hybrid approach of type 2 fuzzy logic and hesitant fuzzy sets for polarity assignment. A sentiment analysis on 'cashless economy' has been carried out using data obtained from Twitter in the form of tweets.

Keywords: cashless economy; digital payments; hesitant fuzzy sets; opinion mining; sentiment analysis; type 2 fuzzy logic.

Biographical notes: Sonakshi Vij is a Research Scholar pursuing PhD in Computer Science Engineering from the Indira Gandhi Delhi Technical University for Women, Delhi in the field of Natural Language Processing. In the very early stage of her academics career, she has published more than 30 research papers in international journals/conferences with proceedings in reputed publications like ACM, IEEE, Springer, etc. Her research areas include natural language processing, fuzzy logic, soft computing, sentiment analysis

and machine learning. She is a member of IAENG Society of Artificial Intelligence, Computer Science, Data Mining, Scientific Computing and Software Engineering.

Amita Jain is presently working as an Assistant Professor in the Department Computer Science and Engineering, Ambedkar Institute of Advanced Communication Technologies and Research, under Government of NCT of Delhi, India. She has published more than 60 research papers in well reputed international journals from ACM Transactions, Elsevier, IEEE, Springer, etc. She has 16 years of teaching and research experience. She is a member of International Association of Engineers (IAENG). Her research interests include fuzzy logic-based intelligent systems, fuzzy-based information retrieval, natural language processing, etc.

Devendra Tayal is presently working as a Professor in the Department Computer Science and Engineering, Indira Gandhi Delhi Technical University for Women, Government of Delhi, India. He has published more than 80 research papers in highly reputed international journals with high impact factor including *Fuzzy Sets and Systems*, *Journal of Uncertainty, Fuzziness and Knowledge based Systems*, *Expert Systems with Applications*, etc. His research interests include fuzzy databases, type-2 fuzzy sets, sentiment analysis, etc.

1   Introduction

The term 'cashless society' was proposed in the USA during middle of the 19th century to hint at a society that would rely on digital modes of payment and transactions that would replace the existing system of coins and notes (Batiz-Lazo and Smith, 2016; Batiz-Lazo et al., 2014; Schifrin, 1997; Manchester, 1997). The emergence of cashless economy can be observed through the various ways in which several countries around the world have started adopting it (Garba and Abu Tomma, 2015; Warwick, 2004; Goldman, 1998; Kiernan, 1994).

Figure 1   Top preferred modes of payment (see online version for colours)
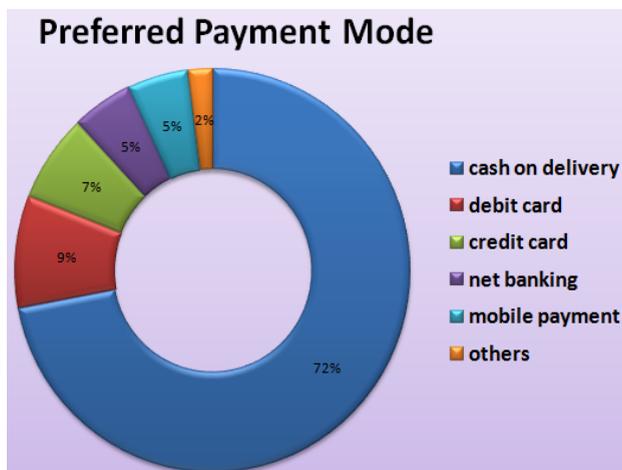
With the growth of digital payment methods like credit cards, debit cards, net banking, mobile payments, etc., India is also moving towards becoming a 'cashless society'. The preferred modes of payment for an online order[1] according to a survey conducted in India, is shown as in Figure 1. In digital payment option, people chose debit and credit cards over other options, though still a majority of people are preferring cash on delivery option for online orders.

Figure 2 highlights the growth journey of digital payments[2] from 2005 to 2015. While in 2005, 8% of the people preferred payment methods other than cash, the percentage went up to 22% in 2015. The number is expected to grow further in the coming years. As a whole, cashless transactions and digital payments are becoming a hot topic of research as well (Tavener, 2015; Rogers, 2011; Dusansky and Koc, 2009; Downes, 2004).

Figure 2   Growth in digital payments (see online version for colours)
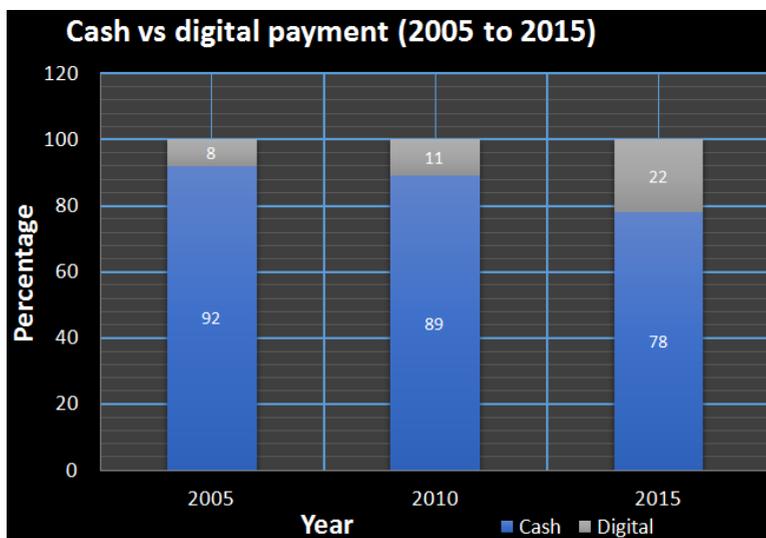
1.1 Motivation

Cashless economy has two major beneficiaries, i.e., the government and the society:

- *Benefit to government:* The motivation for analysing the various aspects of cashless transactions comes mainly from the fact that digital payment methods, like mobile payment, net banking, etc., are expected to assist in combating issues like corruption and black money as they provide an authenticated channel of payment. Illegal transactions can be tracked and halted whenever they take place. It would help the government officials to maintain transparency in various big financial deals taking place in the government agencies as well.

- *Benefit to the society:* Another beneficiary of this method of payment is the society in general. It is so because it is not possible for people to physically carry a huge amount of cash everywhere. In these scenarios, liquid money can be replaced by credit cards, debit cards or net banking options. Also, the issues of counterfeit money can be eliminated.

However, there are a few pitfalls for the same. Like the capabilities of digital payments become questionable when issues like cyber fraud, hacking and impersonation attacks creeps in. Also, not everyone is comfortable in electronic payments like the poor or lesser educated people.

Regarding this topic, people are often expressing their views on micro blogging websites, portals and social networking platforms such as Twitter, Facebook, etc. To mine the opinions of people on any topic using these social networking platforms, sentiment analysis is to be carried out. Sentiment analysis is a research topic under artificial intelligence that aims at finding the opinion of people towards a particular topic using various natural language processing techniques. In order to gain deeper insights regarding digital payments and cashless transactions, sentiment analysis has been carried out for the topic 'cashless economy' using Twitter data, i.e., tweets.

The rest of the paper is framed as follows: Section 2 highlights the recent research trends going on in the field of cashless transactions and digital payment. Section 3 performs a sentiment analysis on 'cashless economy' using data from Twitter. Section 4 illustrates the results while Section 5 concludes the work done along with relevant future scope.

2   Recent research trends

To study the latest research trends going around the world in the field of cashless transactions, Web of Science is used as the data source. Research papers are extracted from it since 1989 till 1st November 2017 corresponding to the title keywords cashless economy, cashless society, smart card, credit card, debit card, e-cash and cashless transaction. A total of 938 research publications were retrieved from it which was used to form a base for the analytical study done in this paper. The top 10 countries contributing to research publication related to the field of cashless economy and cashless transactions are listed as in Table 1.

Table 1   Top 10 countries for research publication in the concerned field

S. no.	*Country*	*Number of research publications*
1	USA	284
2	China	101
3	Taiwan	63
4	South Korea	52
5	England	49
6	India	42
7	France	34
8	Australia	32
9	Canada	31
10	Netherlands	19

It could be seen that the difference between the topmost and the second ranked country in terms of contribution is very high (almost double). The USA is amounting to the

maximum research paper publications in this field, followed by China and Taiwan. According to Table 1, India ranks sixth in this area of research.

The top 3 funding agencies around the world which funds the research paper leading to research publications in this area are shown as in Table 2. Funding details also highlight how much contribution is being made by organisations in various countries to conduct research in the concerned field.

Table 2   Top 3 funding agencies in the concerned field

S. no.	Funding agency	Number of research publications
1	National Natural Science Foundation of China	33
2	Fundamental Research Funds for the Central Universities	12
3	Deanship of Scientific Research at King Saud University	9

The top 3 organisations that are working and producing maximum research publications in the concerned area are shown as in Table 3, while the top 3 research areas corresponding to the concerned field of these 938 papers are mentioned in Table 4.

Table 3   Top 3 organisations around the world working in the concerned field

S. no.	Organisation	Number of research publications
1	Federal Reserve System USA	16
2	University of California System	14
3	Gemalto	13

Table 4   Top 3 research areas corresponding to the concerned field

S. no.	Research area	Number of research publications
1	Computer science	302
2	Business economics	229
3	Engineering	214

The title of top 5 research papers in this field (according to citation count) are as follows:

1   Examining smart card security under the threat of power analysis attacks (Messerges et al., 2002).

2   An efficient and practical solution to remote authentication: smart card (Chien et al., 2002).

3   The failure of competition in the credit card market (Ausubel, 1991).

4   Research report: richness versus parsimony in modelling technology adoption decisions-understanding merchant adoption of a smart card-based payment system (Plouffe et al., 2001).

5   Do liquidity constraints and interest rates matter for consumer behaviour? Evidence from credit card data (Gross and Souleles, 2002).

A point worth mentioning is that all the top 5 research papers are concerned with credit cards. The citation and average citation score for these research papers is summarised as in Table 5.

Table 5   Citation and average citation score for top 5 research papers

S. no.	Research paper	Citation	Average citation
1	Messerges et al. (2002)	604	37.75
2	Chien et al. (2002)	254	15.88
3	Ausubel (1991)	223	8.26
4	Plouffe et al. (2001)	216	12.71
5	Gross and Souleles (2002)	190	11.88

Also, it was found out that 98.29% of these research papers are written in English language. Other than English, the following languages are used for writing the research papers:

a   German

b   French

c   Spanish

d   Italian

e   Japanese

f   Portuguese

g   Turkish.

This shows that people are expressing their views and insights not only in English but their native language as well.

In order to gain an in depth insight in the field of cashless transactions, the control terms pertaining to them were organised. These control terms are the keywords associated with the 938 research papers under study that have maximum association with each other. The item density of these control terms is visualised using VOSviewer as shown in Figure 3.

Figure 3 helps in understanding the core concepts and issues related with cashless transactions and digital payments, in the form of control terms. It can be observed that most of the research publications deal with security issues related to these modes of payment. Security concerns such as mutual password authentication, key agreement, cryptanalysis, forgery attack, biometrics, etc. are identified as major control terms. Smart card payments and other e-payment methods need password authentication which is the best when it is done two ways. Also, key agreement seems to be mandatory for the same. Forgery attacks are rising highly in today's scenarios since access control methods like biometric are not involved with these payment methods. Hence, the occurrence of these keywords in Figure 3 seems to be well justified. This also shows that security threats are increasing day by day for these digital payment methods and corresponding growth is there in the number of research publications for the same that deals with the algorithms and methods for combating the same. A keyword strength network visualisation is shown as in Figure 4.

Figure 3 Item density visualisation for control terms (see online version for colours)
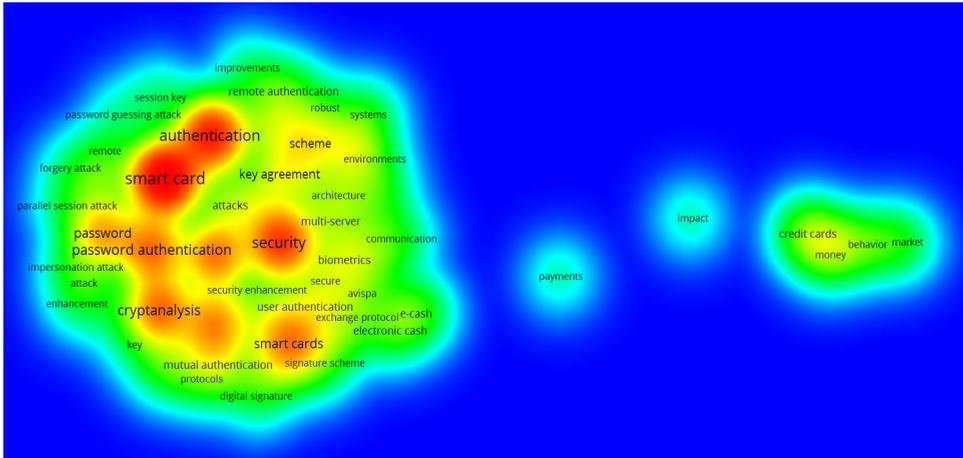
Figure 4 Keyword strength network visualisation (see online version for colours)
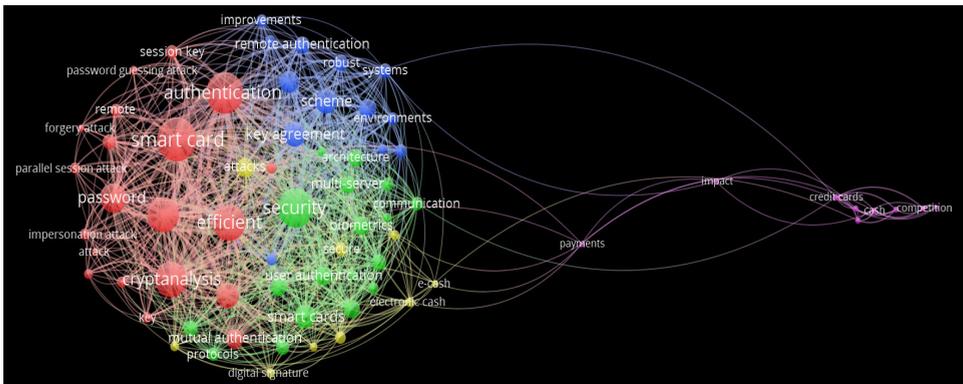
3 Sentiment analysis of 'cashless economy'

Sentiment analysis is a research topic under artificial intelligence and natural language processing that aims at finding the opinion of people towards a particular topic. This section explains the sentiment analysis for '*cashless economy*' for mining the opinion of people using Twitter data, i.e., tweets.

The sentiment analysis is carried out in a three step process as follows.

Step 1: data collection

The data used for performing sentiment analysis is collected from the widely known social networking platform Twitter, in the form of tweets. The data of 5,000 tweets was extracted using *Jefferson's GetOldTweets-python utility*, in order to bypass some of the limitations of Twitter official API which allows collecting only 100 tweets since the last

ten days. The data is extracted for the time span 9 July 2017 to 9 September 2017 (i.e., the thirrd quarter of 2017, also the current quarter).

The data was stored in mongo database. *Mongo database* (also commonly known as MongoDB) is document oriented in nature and is used to support multiple features like:

a   Handling queries that are ad hoc in nature.

b   Maintaining the availability of replica sets that consists of multiple copies of the same data.

c   Balancing huge load of input data in a consistent manner.

d   Using grid-based file system for data management.

e   Performing data aggregation using map-reduce function.

f   Maintaining data consistency and durability.

This data stored in MongoDB was ultimately converted to *pandas data frame* so that data can be explored easily. The data was further exported in the CSV format for easy execution. For the sake of simplicity, the text is then divided in individual text tokens. All sorts of punctuation marks and stop words are also removed from the same.

Step 2: negation handling

There lies a severe need of detecting negation in sentiment analysis can be illustrated by the difference in the meaning of the phrases, "This is good" vs. "This is not good."

However, it is easy to observe that the process of negation detection is not easy due to the ambiguities and unstructured nature of natural languages.

Two primary tasks that are involved in negation handling are:

• Detecting the *explicit negation cues*: for instance: nothing, have not, has not, did not, etc.

• Finding the scope of negation for the above-mentioned words.

The text received is subjected to a negation handling algorithm that uses sentence-oriented approach and relies on linguistic features of the individual tokens (Farooq et al., 2017). This method utilises the strength of the diminishers like 'hardly', 'rarely', etc. For instance, the term 'hardly', 'hardly any good' diminishes the positive polarity of the next to next word 'good'. Now, 'hard' itself is a confusing word as hard work is positive but hard job is tough and slightly negative.

Hence, we process these terms using MongoDB in python to assess their polarities in different conditions to as to simulate varied results according to the need of the scenario. These results are then generalised to suit our database by testing on a synthetic dataset.

After negation handling is well analysed, word stemming needs to be done.

Step 3: assigning polarities

Since negation handling is done successfully in the previous step, hence the polarities assigned in this step are expected to be relevant. In this step, all the collected words are assigned polarities: i.e., positive, negative or neutral (Jain and Jain, 2017). The algorithm

used for polarity detection is based on ConceptNet ontology (Jain and Jain, 2017). The words are assigned scores of polarity by using SentiWordNet (Jain and Jain, 2017).

In order to assign polarities, there exist two contemporary variants of fuzzy logic, i.e., type 2 fuzzy logic and hesitant fuzzy sets. Type 2-fuzzy logic has gained a lot of popularity among researchers in the past decade hence it is considered here for assignment (Vij et al., 2018). In this paper, a hybrid of both these techniques is adopted. The positive, negative and neutral polarity is assigned through type 2 fuzzy logic and the possible footprints of uncertainty are further taken into consideration through a hesitant fuzzy set.

The hesitant fuzzy set has values as follows:

$$H_{Positive} = \{0.49, 0.5, 0.51\}$$

$$H_{Negative} = \{0.49, 0.5, 0.51\}$$

$$H_{Neutral} = \{0.49, 0.5, 0.51\}$$

These values represent the overall nominal values. Now, type 2 fuzzy logic is applied to calculate the best selection values. Through experimental results it was observed that 0.5 gave better and more appropriate results when taken to a synthetic test dataset. Hence, it was utilised for further calculations in positive, negative and neutral cases. In the next step, final opinion mining is done.

4 Opinion of the people

This section describes the results obtained from all the 350 tweets that were obtained from Twitter for the time span 9 September 2017–9 October 2017. It can be seen that a majority of people are neither having a neutral opinion (66.8%), i.e., neither positive nor negative. But it is worth noting that the percentage of people having positive opinion (24.9%) is greater than the people who possess a negative opinion (8.3%). These results are depicted as in the radar diagram in Figure 5.

Figure 5 Results of polarity detection (see online version for colours)
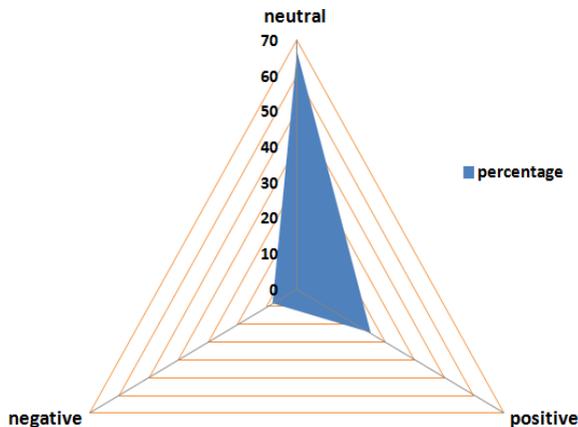
In this paper, the data is taken for just a limited amount of time. In the future, this method can further be extended using other variants of soft computing techniques and could be tested for various time intervals.

5   Conclusions and future scope

In this paper, an analytical study has been performed to investigate the recent research patterns for cashless economy and digital payments. Sentiment analysis has also been done to find out the opinion of the people regarding 'cashless economy'. It was found that people are positively inclined towards it. The method used for opinion mining is a hybrid approach of type 2 fuzzy logic and hesitant fuzzy sets for polarity assignment.

The recorded opinion during any such conducted experiment is subject to the opinions of people at any given time. These opinions vary with time and can be influenced through other surrounding conditions. In the future, these results are expected to deviate a little. For instance, if control terms such as biometric access control are dealt with in a robust manner then the corresponding results of opinion mining are expected to yield even more inclination towards positive polarity. Similarly, if forgery attacks would increase, then the results are expected to be inclined more towards negative polarity.

Considering the growth in artificial intelligence-based technologies, more convenient modes of payment are expected to be developed in the upcoming years, and hence, this inclination of people is expected to grow further in the. In the future, such an analysis can be carried out for individual digital payment methods as well. Also, state wise opinion mining on this topic can be performed.

References

Ausubel, L.M. (1991) 'The failure of competition in the credit card market', *The American Economic Review*, Vol. 81, No. 1, pp.50–81.

Batiz-Lazo, B. and Smith, A. (2016) 'The industrial organization of Hong Kong's progression toward a cashless economy (1960s–2000s)', *IEEE Annals of the History of Computing*, Vol. 38, No. 2, p.54, IEEE Computer Soc.

Batiz-Lazo, B., Haigh, T. and Stearns, D.L. (2014) 'How the future shaped the past: the case of the cashless society', *Enterprise & Society*, Vol. 15, No. 1 p.103, Oxford Univ. Press Inc., DOI: 10.1093/es/kht024.

Chien, H.Y., Jan, J.K. and Tseng, Y.M. (2002) 'An efficient and practical solution to remote authentication: smart card', *Computers & Security*, Vol. 21, No. 4, pp.372–375.

Downes, J.B. (2004) 'No cashless society coming', *Futurist*, Vol. 38, No. 5 p.4, World Future Society.

Dusansky, R. and Koc, C (2009) 'Demand for cash balances in a cashless economy', *International Journal of Economic Theory*, Vol. 5, No. 3 p.301, Wiley-Blackwell, DOI: 10.1111/j.1742-7363.2009.00111.x.

Farooq, U., Mansoor, H., Nongaillard, A., Ouzrout, Y. and Qadir, M.A. (2017) 'Negation handling in sentiment analysis at sentence level', *JCP*, Vol. 12, No. 5, pp.470–478.

Garba, A. and Abu Tomma, M.M.A. (2015) 'Acceptability of cashless economy policy by Nigerian people', *African Journal of Accounting Auditing and Finance*, Vol. 4, No. 3 p.177, Inderscience Enterprises Ltd., DOI: 10.1504/AJAAF.2015.073474.

Goldman, M.I. (1998) 'The cashless society', *Current History*, Vol. 97, No. 621 p.319, Current Hist. Inc.

Gross, D.B. and Souleles, N.S. (2002) 'Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data', *The Quarterly Journal of Economics*, Vol. 117, No. 1, pp.149–185.

Jain, A. and Jain, M. (2017) 'Location based Twitter opinion mining using common-sense information', *Global Journal of Enterprise Information System*, Vol. 9, No. 2, pp.28–32.

Kiernan, V. (1994) 'Poor join the cashless society', *New Scientist*, Vol. 142, No. 1929, p.7, New Scientist Publ. Expediting Inc.

Manchester, D. (1997) 'Smart cards: Key to cashless economy?', *Futurist*, Vol. 31, No. 1, p.29, World Future Society.

Messerges, T.S., Dabbish, E.A. and Sloan, R.H. (2002) 'Examining smart-card security under the threat of power analysis attacks', *IEEE Transactions on Computers*, Vol. 51, No. 5, pp.541–552.

Plouffe, C.R., Hulland, J.S. and Vandenbosch, M. (2001) 'Richness versus parsimony in modeling technology adoption decisions – understanding merchant adoption of a smart card-based payment system', *Information Systems Research*, Vol. 12, No. 2, pp.208–222.

Rogers, C. (2011) 'The failure of Woodford's model of the channel system in the cashless economy', *Journal of Money Credit and Banking*, Vol. 43, No. 2, p.553, Wiley-Blackwell, DOI: 10.1111/j.1538-4616.2010.00385.x.

Schifrin, M. (1997) 'Who owns the cashless society?', *Forbes*, Vol. 159, No. 12, p.42, Forbes Inc.

Tavener, P. (2015) 'What is value in a cashless society?', *New Scientist*, Vol. 226, No. 3027, p.54, Reed Business Information Ltd.

Vij, S., Jain, A., Tayal, D. and Castillo, O. (2018) 'An analytical insight to investigate the research patterns in the realm of type-2 fuzzy logic', *Journal of Automation, Mobile Robotics & Intelligent Systems*, Vol. 12, No. 2, pp.3–32.

Warwick, D.R. (2004) 'Toward a cashless society', *Futurist*, Vol. 38, No. 4, p.38, World Future Society.

Notes

1   http://www.fsstech.com/india-digital-payments-growth-trends/.

2   http://image-src.bcg.com/BCG_COM/BCG-Google%20Digital%20Payments%202020-July%202016_tcm21-39245.pdf.