Received August 26, 2019, accepted September 19, 2019, date of publication October 2, 2019, date of current version October 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945105

Identifying Worldwide Interests in Organic Foods by Google Search Engine Data

XU LIU[1], MEIYING ZHONG[1], BIN LI[2], YE SU[3], JINGLU TAN[4], SEYED MOHAMMAD TAGHI GHARIBZAHEDI[5], YA GUO[1,4], AND JUAN LIU[6]

[1]Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi 214122, China
[2]Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China
[3]Department of Economics, University of Nebraska Kearney, Kearney, NE 68847, USA
[4]Department of Bioengineering, University of Missouri, Columbia, MO 65211, USA
[5]Young Researchers and Elites Club, Lahijan Branch, Islamic Azad University, Lahijan 44169-39515, Iran
[6]School of Foreign Languages and Tourism, Wuxi Institute of Technology, Wuxi 214121, China

Corresponding authors: Ya Guo (guoy@missouri.edu) and Juan Liu (jlrm7@mail.missouri.edu)

This project is partially supported by the National Natural Science Foundation of China (No: 51961125102, 31771680, 51961125102), Fundamental Research Funds for the Central Universities of China (No: JUSRP51730A), the Modern Agriculture Funds of Jiangsu Province (No: BE2018334), the 111 Project (B12018), and the Research Funds for New Faculty of Jiangnan University.

ABSTRACT Global interests in organic foods are of importance to researchers and the food industry. Traditional questionnaire-based methods do not provide a broad picture. To meet this need, worldwide interests in organic foods were studied by integrating query data from the Google search engine and deep learning methods. The results show that organic oil, organic milk, organic chicken, and organic apples are the most interested organic foods; people from Singapore, US, New Zealand, Australia, United Kingdom and Canada care about organic foods the most; consumers' interest in organic foods has no correlation with GDP and life expectancy but has significant correlations with other dimensions of culture such as individualism, uncertainty avoidance, and long-term orientation. A recurrent neural network (RNN) model structure is useful in predicting people's interests in major organic foods over time.

INDEX TERMS Organic food, search engine, search interest, neural network, data modeling, deep learning, consumer behavior.

I. INTRODUCTION

The development of science and technology has been continuously improving human life. In parallel with the rapid development of economy, consumers are increasingly concerned with food safety and their health. Therefore, knowing people's interest worldwide in organic foods is useful for food production and policy making (Wang & Liu, 2014). Consumers' preference for organic foods engenders an important research topic, which has often been studied with questionnaires (Briz & Ward, 2009; Honkanen et al., 2006). However, covering multiple cities and regions by questionnaires can be a big challenge to attempts for a comprehensive study. People from different locations have diverse backgrounds and thus survey results from one location at a specific time usually do not apply to other places. As a result, there is a serious need on a global scale for determining sustained interests in organic foods over time.

The associate editor coordinating the review of this manuscript and approving it for publication was Osama Sohaib.

With the development of increasingly convenient and accessible search engines, people tend to rely on search engine data (SED) to find information for preference recommendations or research data collection. People who search more times on an item tend to show more interests in it than those who search fewer times. SED have been extensively used in various studies, including detecting potential influenza outbreaks (Ginsberg et al., 2009; Pelat et al., 2009), predicting economic status (Choi & Hal, 2012), and ranking universities (Vaughan & Esteban, 2014). In this work, not only were SED used to investigate consumer interests in organic foods world widely, but deep learning techniques were applied to predict the trends in people's interests over time.

II. BACKGROUND

A. RESEARCH ON INTEREST IN ORGANIC FOODS

Many past efforts focused on consumer attitudes towards organic foods. One of the major features of these studies is the

TABLE 1. Studies on organic foods across the world.

Scholars	Study Focus	Area	Year	Major Work/Result
Davies	Purchasers of organic food	Northern Ireland	1995	The primary factor in organic food purchase is the consumer's level of personal disposable income.
Schifferstein	Comparison between organic food purchasers and nationally representative household	Netherlands	1998	Organic food buyers were best separated on the basis of education, body mass index, scores on the health locus of control scales, or interest in vegetarianism and naturopathy.
Cederberg & Mattsson	Organic food marketing share	Sweden	1999	The largest share of organic food-related products was in dairy and about 3% of milk sales in 1998 were labeled as organic.
Cederberg et al.	Organic food farming	Sweden	2000	Organic milk production is a way to reduce pesticide use and mineral surplus in agriculture but this production form also requires substantially more farmland than conventional production.
Janssen et al.	Consumers' attitudes towards governmental logo on organic foods	Germany	2014	The awareness of the new EU logo, the trust for the logo, and the understanding of the logo were supposed to be key points for EU to strengthen the organic sector.
Ramesh & Divya	Consumers' awareness, attitude and satisfaction towards organic food products	India	2015	The major reason for customers to purchase organic foods is the expectation of a healthier lifestyle and an environmentally-friendly means of production.
Basha and Mohamed	Consumers attitude towards organic food	India	2015	Food quality, environmental concern and lifestyle are the most common motives for purchasing organic foods.
Thøgersen et al.	Determinants for consumer purchasing of organic foods	China and Brazil	2015	Healthiness, taste, and environmental friendliness are major determinants for consumer purchasing of organic foods
Vittersø et al.	The tendency to purchase organic foods	Norway	2015	People in 2013, compared to 2000, had a more positive tendency to purchase organic foods.
Teng & Wang	Decisional factors driving consumption of organic foods	Taiwan	2015	Trust attitudes about organic foods and the subjective norm exert a significant influence on consumer purchasing behavior.
Dias et al.	Consumers' loyalty for organic foods	Brazil	2016	A one-dimensional validated scale was developed and used, which consists eight questions and shows high composite reliability.
Thøgersen et al.	The stability of value basis for consuming organic foods	China	2016	Consumption for organic foods increased in 2012 compared to that in 2009.
Hansen et al.	The interplay between consumer motivations and values influence organic food identity	Danish	2018	Health consciousness has a positive effect on intentional organic food behavior through organic food identity.
Nguyen et al.	Relations between consumers' personal factors and organic food purchases	Vietnam	2019	Factors affecting purchase for organic food include concerns for environment, health, and food safety and consumers' knowledge of organic food.

data collection in a local area at a specific time. Table 1 shows selected studies related to organic food around the world from 1995 to 2019.These researches were conducted in one country or two countries for comparison at a specific time.

Davies (1995) and Schifferstein (1998) found that organic food buyers considered themselves more responsible for their own health and were more likely to undertake preventive health actions than the general public. The organic food consumers were best separated on the basis of education, body mass index, scores on the health locus of control scales, or interest in vegetarianism and naturopathy. The differences between buyers and non-buyers on a multitude of measures suggest that the consumption of organic foods is a way of life, resulted from ideology and connected to a particular value system, which affects personality measures, attitudes, and consumption behavior. This research links the education level and values of individuals to the consumption of organic food. Ramesh & Divya (2015) conducted an analysis through interviews, surveys, and data collection from various journal articles to understand consumers' awareness, attitude and satisfaction towards selected organic food products in India. They found that the major reason for customers to purchase organic foods in India is the expectation of a healthier lifestyle and an environmentally-friendly means of production. This was consistent with the findings of Basha and Mohamed (2015) that food quality, environmental concern and lifestyle are the most common motivations for purchasing organic foods. For the emerging markets of China and Brazil, survey data were collected in Guangzhou of China and Porto Alegre of Brazil, and analyzed through structural equation modeling (Thøgersen et al., 2015). The similar conclusion has been drawn: beliefs about healthiness, taste, and environmental friendliness are major determinants for consumer purchasing of organic foods.

Dias et al. (2016) evaluated the scale of consumer loyalty for organic foods by collecting data through a web-based survey and by subsequently solving structural equations for data analysis. The researchers came up with a one-dimensional validated scale consisting of eight questions and a composite reliability level to measure the loyalty of consumers for organic foods. The stability of value basis for consuming organic foods in China from 2009 to 2014 was studied (Thøgersen et al., 2016). Data from ordinary Chinese consumers have also been collected in 2009 and 2012 outside supermarkets in Guangzhou, China. Results showed that people in this city purchased more organic foods in 2012 compared to that in 2009. In Norway, a study based on two surveys was carried out in 2000 and 2013, indicating that people in 2013, compared to 2000, had a more positive tendency to purchase organic foods and to appreciate the importance of these foods in health (Vittersø et al., 2015). Although there are no continuous data for all the years, it is evident that people's willingness to buy organic foods from available markets has been increasing.

Researchers in Taiwan administered questionnaire surveys in the context of decisional factors driving consumption of organic foods and consumer purchase intentions (Teng & Wang, 2015). The surveys were conducted of customers in four main urban supermarkets and major health food stores in three different cities. It was concluded that trust attitudes about organic foods and the subjective norm exert a significant influence on consumer purchasing behavior. Consumer attitudes towards the mandatory EU logo for organic foods and consumer willingness to pay for a variety of governmental organic labels and farmer association ones were analyzed in a study conducted in Germany (Janssen et al., 2014). These researchers also provided recommendations for governmental and private owners of organic certification labels. Findings revealed that the awareness of the new EU logo, the trust for the logo, and the understanding of the logo were supposed to be key points for the EU to strengthen the organic sector. During the 1990s in Sweden, the largest share of organic food-related products was in dairy and about 3% of milk sales in 1998 were labeled as organic (Cederberg & Mattsson, 1999). The large dairy producers doubled their organic milk production. Cederberg et al., (2000) concentrated their work on the most important categories of organic foods based on sales in local areas. There are still such problems to research in the next decade, which inevitably depend on geographical locations and people's living habits.

In recent years, many studies on organic food interests rely on commercial data since purchase behavior can well reflect people's interest in the market. Hansen et al. (2018) shows the interplay between consumer motivations and values influences organic food identity. Nguyen et al. (2019) focus on studying the relationship between consumers' personal factors and organic food purchases in food stores.

To sum up, the positive attitudes towards healthy lifestyle and environment friendliness were the main factors for consumers to purchase and consume organic foods. The food choice is more influenced by the psychological interpretation of product properties than their physical attributes (Rozin et al., 1986). Consumers purchase organic foods for different reasons, including concerns about the effects of conventional farming practices on the environment, human health, animal welfare, and perceptions that organic foods are tastier than their conventional alternatives. To study the psychology and interest of consumers, scientific experiments and extensive data support will be required to arrive at global scientific conclusions. The conclusions in the existing literature are often limited to the countries and products at a certain time (Thøgersen et al., 2015).

B. GOOGLE SEARCH ENGINE DATA APPLIED IN RESEARCH AND BUSINESS

As mentioned earlier, data from Google Trends can be collected more easily than using a traditional survey questionnaire. It can be a valuable source of data or reference for research after appropriate preprocessing. In addition, the greater transparency and accessibility will improve data reliability as an effective research tool (Sudhakar et al., 2014).

With high volume and accessibility, these data are used by more and more researchers for their studies and analyses. For example, the data have been extensively used in medical studies (Ramos-Casals et al., 2015; Yao et al., 2015). A large number of users search online for information on diseases when they or their family members feel ill or are diagnosed with a disease. The records on search engines allow us to collect worldwide disease outbreak information, which cannot be easily completed by other ways. Researchers used the correlation data from Google Trends and autoregressive integrated moving average (ARIMA) models to improve the quality of predictions on epidemics and found that the result was consistent with the actual data during the Zika epidemic period (Teng et al., 2017). This study provides an evidence for the potential use of stream editor in epidemic prediction. Another study on the Zika epidemic collected Zika-related data from Google searches, Twitter microblogs, and the HealthMap digital surveillance system (McGough et al., 2017). Based on these historical data of suspected cases, scholars found that the time span with the best prediction performance for Google and Twitter is 2-3 weeks (McGough et al., 2017). This finding helps other researchers to utilize these data in a more efficient way.

After a disaster, a traditional statistical method is ineffective for timely data collection and analysis. Information from social media, such as Facebook and Twitter, can be a powerful tool in disaster management (Kim & Hastak, 2018). Researchers combined the social media data and SED to explore the critical role of social media in emergency information propagation, helping the emergency agencies develop appropriate strategies for emergency responses. Liu et al. (2017) studied the periodicity of consumer behavior on an online search site for the US hotels and built a structure model to describe the SED-based behavior. Tang et al. (2018) compared the periodic behavior of consumer online searches for restaurants in the US and China based on SED and found that the online search behavior in the US is much more strongly governed by weekly cycles than that in China. Predicting the number of visitors to a place is important for tourism management. Yang et al. (2015) used web search query volume to predict the number of visitors for a popular destination in China and compared with the actual number. Yelowitz and Wilson (2015) used Google Trends data to study determinants of interest in Bitcoin. Based on anecdotal evidence about Bitcoin users, proxies for four possible clienteles were built. Maringer et al. (2016) conducted research on using the Google search engine for capturing the most popular and widely used systems capable of integrating with other data collection systems, in which, SED played an import role in data collection and data analysis.

When looking for help or recommendations, more and more people tend to use search engines for information and advice, thus making the search history data increasingly valuable for studying a particular topic. Google search engine data were widely used in business analysis field. However, in organic food area, previous studies were mainly based on manually collected datasets, and Google search engine data were not applied for analysis.

C. RECURRENT NEURAL NETWORKS (RNN) AND LONG SHORT-TERM MEMORY (LSTM) MODELS FOR DATA ANALYSIS AND PREDICTION

Because the Google search engine provides search interests over time, it is possible to build a deep learning model to represent the search records and to predict the trend of people's interest in organic food, which cannot be easily achieved by collecting data through traditional survey. Deep learning can be applied in various fields such as medical research (Litjens et al., 2017), geography studies (Zhao et al., 2015), agricultural forecasts (Kamilaris et al., 2018), and market analysis (Leng et al., 2018). Recurrent neural networks (RNN) are one of the most popular deep learning algorithms (Zhu et al., 2018), which can predict the trends in a time series. In recent years, many researchers have shown that the RNN is able to have good performance in modeling time series data. The long short-term memory (LSTM) model composed of a cell, an input gate, an output gate, and a forget gate is well suited for RNN-based classification, processing, and predictions. Several recent studies on RNN and LSTM models for predictions are summarized as follows.

A cell is a unit that takes an input and stores it for some arbitrary intervals, and the three gates regulate information flowing into or out of, or remaining in the cell, which correspond to an input gate, an output gate, and a forget gate. The number of cells and the weights for each cell are initialized randomly (usually with values in a normal distribution) as done in the literature (Thimm, et al., 1995). When an LSTM neural network is trained, the differences between network predictions and real values are measured by a loss function. The training algorithm back-propagates the errors to the cells for the three types of gates to control and adjust the weights to optimize the output (Greff et al., 2017). In this work, the commonly-used cross validation technique was used and the data were divided into training and validation datasets. We used historical data segments to train the LSTM network and future data segments to validate the model as done in Karpathy et al. (2015).

Fig. 1 illustrates the basic structure of an LSTM network, which includes inputs, outputs, memory blocks, and activation functions (Gers et al., 1999; Shi et al., 2015).

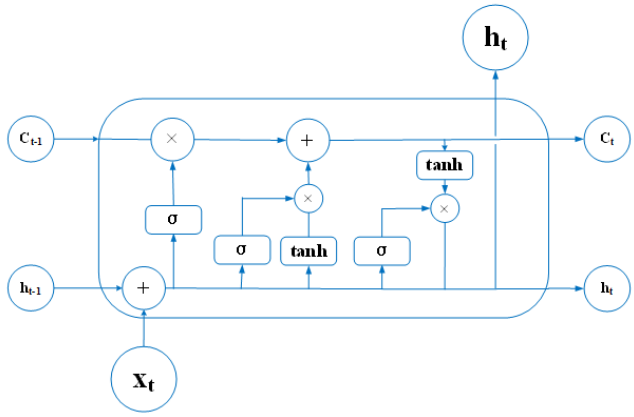


FIGURE 1. An illustration of LSTM network. X_t: Input Vector. h_t: Output of current block. C_{t-1}: Memory from previous block. h_{t-1}: Output of previous block. C_t: Memory from current block. σ/ tanh: Activation function.

There are RNN/LSTM models successfully applied in different areas. Tax et al. (2017) showed that data extracted from historical event logs and used in an RNN/LSTM model could outperform existing techniques in monitoring and predicting the business process. For business data analytics, scholars proposed a new model based on an LSTM network which performed well on real-world business datasets (Gan et al., 2017). Stock price prediction has been a hot topic for many years. Scholars from Japan proposed a sequential learning model based on RNN or LSTM-RNN for prediction of single-stock prices and it performed better than conventional models based on corporate action event information and macro-economic indices (Minami & Shotaro, 2018).

An LSTM model can also be combined with other methods. A hybrid method of wavelet transform and LSTM has been proposed to predict nonlinear systems. The data were first decomposed into constitutive series through wavelet transform, and then an LSTM model was used to find the bias vectors and weighting coefficients for prediction. This model could obtain a higher accuracy than the wavelet transform alone and had a shorter training time compared with an LSTM-only model (Sugiartawan et al., 2017).

Based on the prior work, it is clear that RNN and LSTM models are quite suitable for modeling of time series, and thus this method was applied for predicting trends in organic foods based on Google searches conducted in the present work.

III. MATERIALS AND METHODS

Google is the most popular search engine in the world. Google Trends is a public web facility of Google Inc. Google Trends based on Google searches shows how often a particular search term is used relative to the total search volume across various regions of the world. In this work, a list of organic foods-related search terms (all the data are available by request) including organic milk, edible oil, meat, grains, fruits, and vegetables was used as keywords for searches in Google Trends for the past 10 years (1/1/2008 – 12/31/2017, https://trends.google.com/trends/). Among these 6 major categories, there are 3 keywords in the milk category (milk, baby formula, and soybean milk), 8 keywords in the grains category (rice, corn, wheat etc.), 6 keywords in the edible oil category (oil, coconut oil, olive oil etc.), 10 keywords in the meat category (chicken, beef, meat, etc.), 14 keywords in the fruits category (apple, fruit, orange, etc.), and 17 keywords in the vegetables category (vegetable, garlic, pumpkin, etc.). All the keywords in each category are listed in the pie charts in Fig. 3. By visiting the webpage of Google Trends, we may search by different keywords, compare search results from different locations, define the search timeframe, and download data in ".csv" format. Google Trends provides a very friendly interface and clear instructions.

Google Trends provides search interest over time and search interest over a region. Search interest with monthly time resolution is provided if the search duration is set as 10 years. If daily search interest data are wanted, the longest search duration should not be more than half a year. Since the maximal search interest value for each half a year is always scaled to 100 by Google automatically, other values in this half a year is a relative number to 100. If we want to compare the daily search interest in a time span longer than half a year, we must normalize the daily search interest data. In this work, we first retrieved the data for the 10 years and summed up the data every 6 months. Then we used the 6-month sums to normalize the daily search interests retrieved with search duration set at 6 months. For the search interest over a region, Google Trends provides search interest values for all the subregions during the specified timeframe. Values are scaled from 0 to 100, where 100 is the location with the highest search frequency. For search interest over time, daily search data in the past 10 years were collected. For search interest in a region, thirty-two countries in 6 major categories of organic food were collected. These data are all relative search interests, which are adequate for comparative analysis as done in this work.

To analyze the raw data and to find correlations between them and other potential factors, data on GDP, life expectancy, and educational attainment were collected from the official website of the World Bank (https://datacatalog.worldbank.org), whereas data on cultural dimensions were obtained from the website at https://www.hofstede-insights.com/models/national-culture. SPSS software (SPSS Inc., Chicago, IL, USA) was used for the correlation analysis. Deep learning was conducted in TensorFlow (Abadi, Martín, et al. 2016). RNN and LSTM models were developed to forecast trends in people's interest in organic foods a few weeks ahead. A suitable network architecture was obtained by comparing the prediction results from varied structures based on the relative fitting errors (RTF) as the loss function.

IV. RESULTS

A. ORGANIC FOODS OF GREATEST INTEREST

Fig. 2(a) shows the relative search interest among nine major categories of organic foods, in which keywords "organic oil" and "organic egg" receive the highest and lowest search interest, respectively. Fig. 2(b) shows the variations in search interest in some organic foods over the years (For clarity, only six keywords are shown). Visually, the relative search interests have remained relatively steady in the past ten years, except for some fluctuations in fruits and vegetables, which appear to be associated with seasonality.

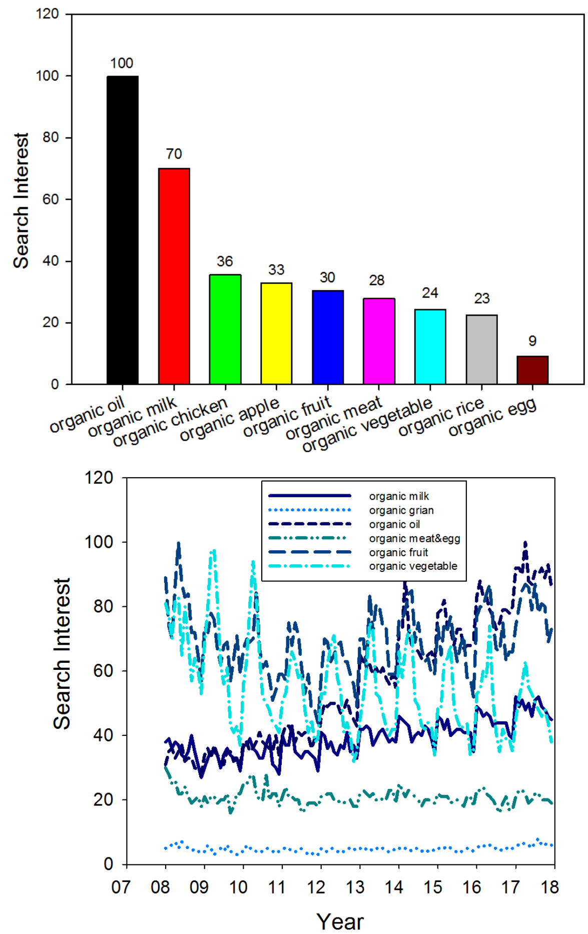
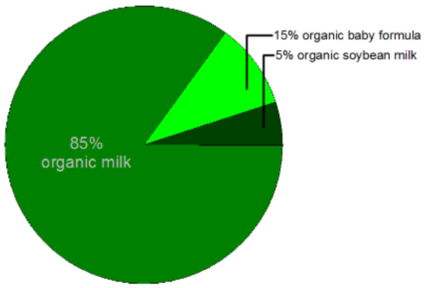
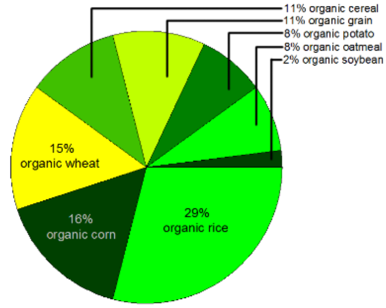


FIGURE 2. The organic foods people are most interested in (a) comparison of the search interest for nine keywords and (b) search interest variations over the past 10 years for six keywords.

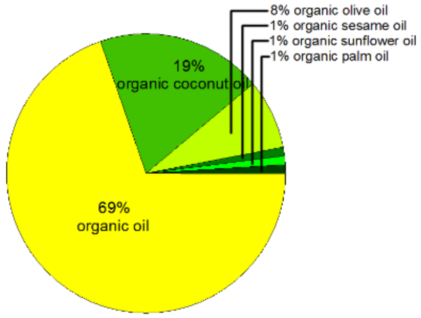
The pie charts in Fig. 3 show the interest level of each keyword within its category. In categories of milk, grains, edible oil, meat, fruits and vegetables, the keywords



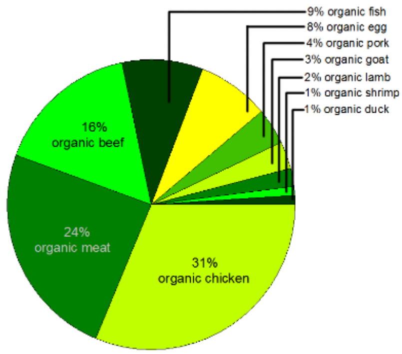
(a) Search interest in organic milk category and its relative search interest levels.



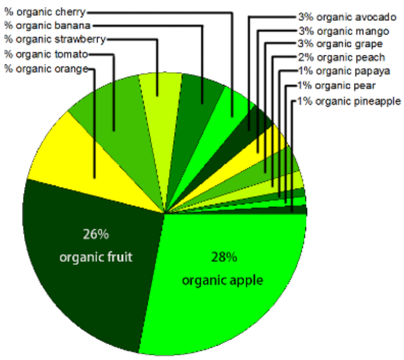
(b) Search interest in organic grain category and its relative search interest levels.



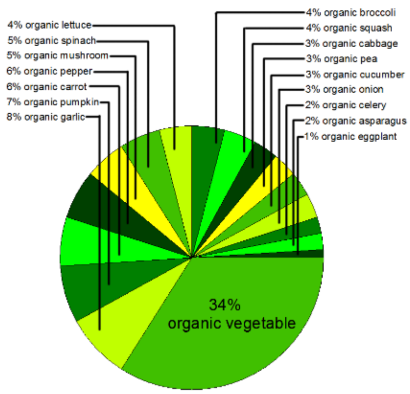
(c) Search interest in organic oil category and its relative search interest levels.



(d) Search interest in organic meat category and its relative search interest levels.



(e) Search interest in organic fruits category and its relative search interest levels.



(f) Search interest in organic vegetables category and its relative search interest levels.

FIGURE 3. Search interest in organic foods for each category and their relative search interest levels in their category: (a) milk, (b) grains, (c) edible oils, (d) meat, (e) fruits, and (f) vegetables.

receiving the highest search interest are organic milk (85%), organic rice (29%), organic oils (69%), organic chicken (31%), organic apple (28%), and organic vegetable (34%), respectively. Among them, the keywords for some specific foods receive even higher search interest than the keyword for the category. For example, the keywords of organic rice, organic chicken, and organic apple respectively receive higher search interests than the category words of organic grain, organic meat, and organic vegetables.

B. WHO ARE INTERESTED

Fig. 4 shows the countries in which people are most interested in organic foods in each category. From the results, it can be observed that people from Singapore, the U.S., New Zealand,

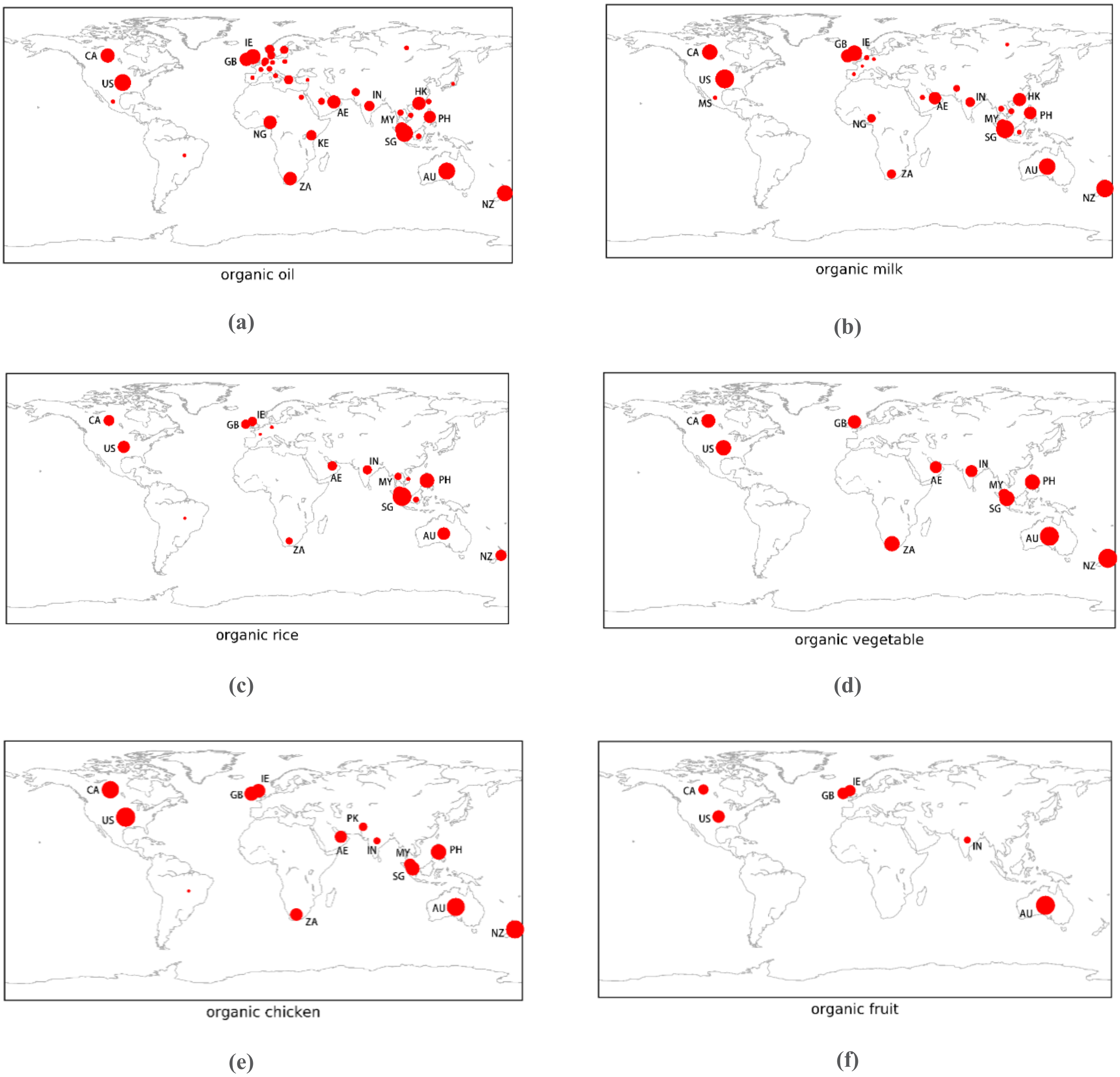


FIGURE 4. **Countries in which people are most interested in organic foods for each category, (a) edible oils, (b) milk, (c) rice, (d) vegetables, (e) chicken, and (f) fruits. The countries are shown by the standard internet domain names and the size of the dots indicates the level of interest.**

Australia, the United Kingdom, and Canada pay more attention to organic foods compared to other countries.

C. CORRELATION OF SEARCH INTEREST IN ORGANIC FOODS WITH OTHER FACTORS

Table 2 shows a correlation analysis of people's interest in different kinds of organic foods with factors including geography (latitude and longitude), GDP, life expectancy, education, and culture (e.g., power distance, individualism, uncertainty avoidance, and long-term orientation). By calculating the *p*-value of people's search times of organic foods and some objective factors stated above, Table 2 reveals that the search interest in organic milk, organic edible oils, organic chicken, and organic vegetables have a significant correlation with the cultural dimensions of individualism, uncertainty avoidance, and long-term orientation. A significant correlation between search interests in organic milk and organic edible oils with education was found. However, people's search interest in organic foods had no significant correlation with GDP or life expectancy. People at low latitudes often search for organic vegetables, but the longitude has no significant correlation with interest in organic foods. People with a higher education degree care more about organic milk and edible oils. People who live in low-power-distance countries also care more about organic fruits. Besides, in general, people with higher individualism tend to be more interested in organic products such as milk, edible oil, fruits, chicken, and vegetables.

TABLE 2. Correlations between people's interest in organic foods and some factors[a,b].

	Latitude	Longitude	GDP	LE	Education	PD	Individualism	UA	LO
Milk	-.126	.091	.323	.266	.325*	-.298	.385*	-.519**	-.438*
Edible oils	-.148	.140	.216	.276	.323*	-.272	.367*	-.570**	-.477**
Fruits	-.040	-.033	.256	.197	.261	-.418*	.546**	-.200	-.369*
Chicken	-.149	.026	.317	.233	.276	-.351	.491**	-.383*	-.551*
Rice	-.187	.199	.104	.203	.279	.049	.072	-.539**	-.303
Vegetables	-.351*	.194	.164	.174	.310	-.228	.407*	-.383*	-.465**

a LE- life expectancy; PD-power distance; UA- uncertainty avoidance; LO- long term orientation
b Significant levels:* $p< 0.05$; ** $p<0.01$

People who are in low-uncertainty-avoidance countries care more about organic food materials of milk, edible oils, chicken, rice, and vegetables. Finally, people with low levels of long-term orientation generally care more about most organic foods.

D. SEARCH INTEREST PREDICTION WITH LSTM MODEL

LSTM networks were trained for search interest prediction. In order to reduce computation time, only parts of the data (5-year data) from the original dataset (10-year data) were used. The six categories of search interest data from 1 Jan 2013 to 1 May 2017 were used to train models to predict future search interest levels from 2 May 2017 to 31 Dec 2017. Optimized neural network structures were obtained experimentally based on prediction performance measured with a relative fitting error. The output of each model was the weekly interest level for each category in the future period of prediction. The relative fitting errors are given in Table 3 and the prediction performance is displayed graphically in Fig. 5.

TABLE 3. Performance of LSTM networks for predicting search interest.

Organic foods	RFT[a,b]	
	Training	Validation
Milk	0.03	0.02
Edible oils	0.03	0.01
Fruits	0.06	0.05
Grains	0.13	0.14
Meats	0.10	0.12
Vegetables	0.04	0.03

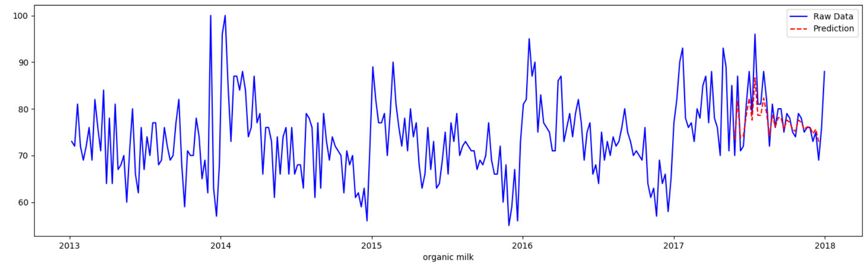
a RFT: relative fitting error
b Training period: 1/1/2013-5/1/2017, Validation period: 5/2/2017-12/31/2017

Predications for all six categories were close to the actual values, indicating the usefulness of the models. The prediction errors for grains and meats were relatively larger than the errors for milk and oil, but they are still small. From the plots, it is evident that, as staple foods, grains and meats do not show a clear periodic seasonality pattern as the other foods do and thus lack the regularity makes prediction more difficult.

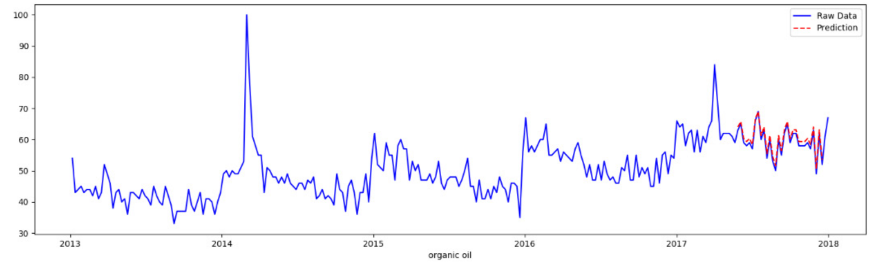
V. DISCUSSION

As stated earlier, the global sales of organic foods have steadily increased in the last 18 years according to the FiBL-IFOAM-SOEL surveys (Willer et al., 2018). Although organic products make up a minor share of the world food market, the proliferation of certified commodities and their increasing availability in mainstream supermarkets have made organic products one of the fastest growing segments in the food industry (Raynolds, 2004). The customer analysis in organic foods and the influencing factors and prediction of consumer interests in organic foods can reveal business opportunities. Given the relatively recent and unexpected growth in organic foods, there are currently few sources of comparable international data for analysis. Therefore, data analysis based on Google search engine provides a useful alternative research tool for studying interest in organic foods.

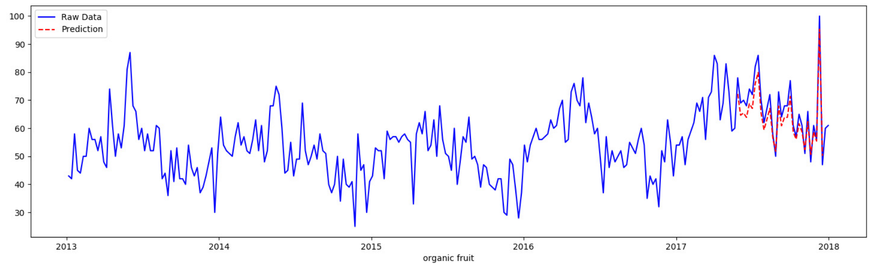
Vegetables, as an important food in daily diets, are expected to receive very high search interests; nonetheless, Fig. 1 shows that people's interest in organic vegetables does not appear in the top four. It was surprising that people's interest in organic foods does not have a significant correlation with GDP. Although the prices of organic foods are generally much higher than those of non-organic ones, this has not hampered a globally growing interest in organic foods. There is no significant correlation between people's interest in finding organic foods and life expectancy. However, the results indicated that the people's search interest in organic foods is significantly correlated with education and several cultural dimensions, such as individualism, uncertainty avoidance, and long-term orientation. This is in excellent agreement with results from traditional questionnaire-based research in the literature (Ramesh & Divya, 2015; Schifferstein, 1998), proving the reliability of using Google–SED to analyze people's interest in organic foods. Moreover, people's interest in organic foods can be practically forecast by a structured RNN model. Planning food production, transportation, and storage in a timely manner is important for the food industry. An ability to predict people's interest in organic foods is useful for the food industry to make decisions on production and inventory. Such information may be also useful for organic food companies to determine when and where to advertise their products. By using the method provided in this work,



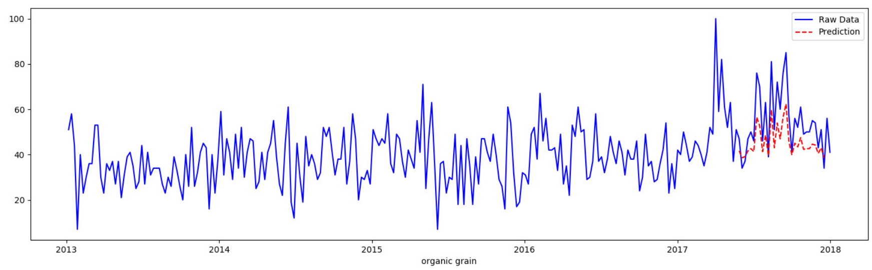
(a) Weekly interest level (blue line) and its predictions (red line) for organic milk in the US



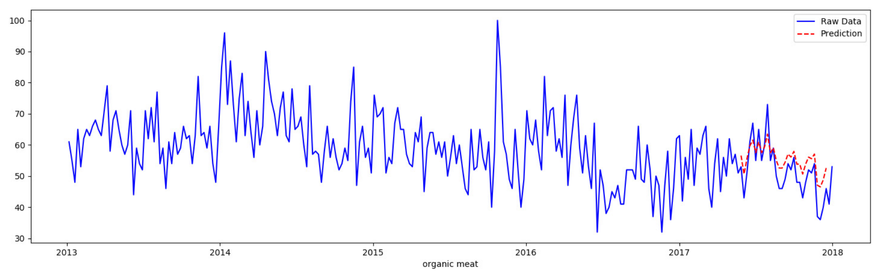
(b) Weekly interest level (blue line) and its predictions (red line) for organic oil in the US



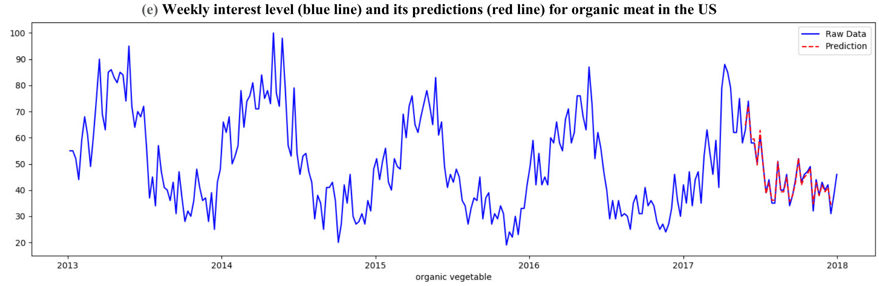
(c) Weekly interest level (blue line) and its predictions (red line) for organic fruit in the US



(d) Weekly interest level (blue line) and its predictions (red line) for organic grain in the US



(e) Weekly interest level (blue line) and its predictions (red line) for organic meat in the US



(f) Weekly interest level (blue line) and its predictions (red line) for organic vegetable in the US

FIGURE 5. **Actual (blue line) and predicted (red line) interest levels for each category of organic food: (a) milk, (b) edible oils, (c) fruits, (d) grains, (e) meat, and (f) vegetables.**

consumers' interest in more varieties of organic food for a specific country can be investigated in details. In future work, the correlation between people's interest in organic food and sales or sustainability issues can also be studied.

VI. CONCLUSION

A movement in consumption and popularity of organic foods is occurring with the improvement of living standards. Results of the current study on global interest in organic foods based on query data from the Google search engine showed that (i) Organic foods of the greatest interest include organic edible oils, milk, chicken, and apples; (ii) People from Singapore, the US, New Zealand, Australia, the United Kingdom and Canada are most interested in organic foods; (iii) People's interest in organic foods has no significant correlation with GDP and life expectancy, but has a significant relationship with education level and several cultural dimensions; (iv) People with higher educational degrees pay more attention to organic foods; (v) The cultural dimensions of power distance, uncertainty avoidance, and long-term orientation have negative correlations with people's interest in organic foods, while individualism shows a positive correlation. These findings are in close agreement with results obtained from conventional questionnaire-based research in the literature. An RNN model structure can be used to predict people's interest in organic foods with small relative errors.

REFERENCES

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, and M. Kudlur, "TensorFlow: A system for large-scale machine learning," in *Proc. OSDI*, vol. 16, Nov. 2016, pp. 265–283.

[2] B. Beharrell and J. H. MacFie, "Consumer attitudes to organic foods," *Brit. Food J.*, vol. 93, no. 2, pp. 25–30, 1991.

[3] T. Briz and R. W. Ward, "Consumer awareness of organic products in Spain: An application of multinominal logit models," *Food Policy*, vol. 34, no. 3, pp. 295–304, 2009.

[4] C. Cederberg and B. Mattsson, "Life cycle assessment of milk production—A comparison of conventional and organic farming," *J. Cleaner Prod.*, vol. 8, no. 1, pp. 49–60, 2000.

[5] H. Choi and H. Varian, "Predicting the present with Google trends," *Econ. Rec.*, vol. 88, no. 1, pp. 2–9, 2012.

[6] A. Davies, A. J. Titterington, and C. Cochrane, "Who buys organic food?" *Brit. Food J.*, vol. 97, no. 10, pp. 17–23, 1995.

[7] V. da Veiga Dias, M. da Silva Schuster, E. Talamini, and J. P. Révillion, "Scale of consumer loyalty for organic food," *Brit. Food J.*, vol. 118, no. 3, pp. 697–713, 2016.

[8] L. Eriksson, T. Byrne, E. Johansson, J. Trygg, and C. Vikström, *Multi- and Megavariate Data Analysis Basic Principles and Applications*, vol. 1. Umeå, Sweden: Umetrics Academy, 2013.

[9] Z. Gan, P. D. Singh, A. Joshi, X. He, J. Chen, J. Gao, and L. Deng, "Character-level deep conflation for business data analytics," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2222–2226.

[10] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. 9th Int. Conf. Artif. Neural Netw. (ICANN)*, Sep. 1999, pp. 850–855.

[11] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, no. 7232, pp. 1012–1014, 2009.

[12] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[13] P. Honkanen, B. Verplanken, and S. O. Olsen, "Ethical values and motives driving organic food choice," *J. Consum. Behav., Int. Res. Rev.*, vol. 5, no. 5, pp. 420–430, 2006.

[14] T. Hansen, M. I. Sørensen, and M.-L. R. Eriksen, "How the interplay between consumer motivations and values influences organic food identity and behavior," *Food Policy*, vol. 74, pp. 39–52, Jan. 2018.

[15] M. Janssen and U. Hamm, "Governmental and private certification labels for organic food: Consumer attitudes and preferences in Germany," *Food Policy*, vol. 49, pp. 437–448, Dec. 2014.

[16] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.

[17] A. Karpathy, J. Johnson, and L. Fei-Fei, "Visualizing and understanding recurrent networks," 2015, *arXiv:1506.02078*. [Online]. Available: https://arxiv.org/abs/1506.02078

[18] J. Kim and M. Hastak, "Social network analysis: Characteristics of online social networks after a disaster," *Int. J. Inf. Manage.*, vol. 38, no. 1, pp. 86–96, 2018.

[19] A. Krystallis and G. Chryssohoidis, "Consumers' willingness to pay for organic food," *Brit. Food J.*, vol. 107, no. 5, pp. 320–343, 2005.

[20] E. Lea and T. Worsley, "Australians' organic food beliefs, demographics and values," *Brit. Food J.*, vol. 107, no. 11, pp. 855–869, 2005.

[21] J. Leng, Q. Chen, N. Mao, and P. Jiang, "Combining granular computing technique with deep learning for service planning under social manufacturing contexts," *Knowl.-Based Syst.*, vol. 143, pp. 295–306, Mar. 2018.

[22] J. Liu, X. Li, and Y. Guo, "Periodicity analysis and a model structure for consumer behavior on hotel online search interest in the US," *Int. J. Contemp. Hospitality Manage.*, vol. 29, no. 5, pp. 1486–1500, 2018.

[23] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[24] M. Maringer, S. Ekman, A. Normann, N. Klepacz, M. Verain, M. Raats, and A. Geelen, *Inventory of Types of Consumption Data and Data Collection Methodologies for Consumer Generated Food Purchase Data*, document INFRADEV-1-2014, 2017.

[25] S. F. McGough, J. S. Brownstein, J. B. Hawkins, and M. Santillana, "Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data," *PLoS Neglected Tropical Diseases*, vol. 11, no. 1, 2017, Art. no. e0005295.

[26] S. Minami, "Predicting equity price with corporate action events using LSTM-RNN," *J. Math. Finance*, vol. 8, no. 1, pp. 58–63, 2018.

[27] H. V. Nguyen, N. Nguyen, B. K. Nguyen, A. Lobo, and P. A. Vu, "Organic food purchases in an emerging market: The influence of consumers' personal factors and green marketing practices of food stores," *Int. J. Environ. Res. Public Health*, vol. 16, no. 6, p. 1037, 2019.

[28] S. V. Nuti, B. Wayda, I. Ranasinghe, R. P. Dreyer, S. I. Chen, and K. Murugiah, "The use of Google Trends in health care research: A systematic review," *PLoS ONE*, vol. 9, no. 10, 2014, Art. no. e109583.

[29] S. Padel and C. Foster, "Exploring the gap between attitudes and behaviour," *Brit. Food J.*, vol. 107, no. 8, pp. 606–625, 2005.

[30] C. Pelat, C. Turbelin, A. Bar-Hen, A. Flahault, and A. J. Valleron, "More diseases tracked by using Google trends," *Emerg. Infectious Diseases*, vol. 15, no. 8, pp. 1327–1328, 2009.

[31] S. V. Ramesh and M. Divya, "A study on consumers' awareness attitude and satisfaction towards select organic food products with reference to Coimbatore," *Int. J. Interdiscipl. Multidisciplinary Stud.*, vol. 2, no. 4, pp. 81–84, 2015.

[32] M. Ramos-Casals, P. Brito-Zerón, B. Kostov, A. Sisó-Almirall, X. Bosch, D. Buss, A. Trilla, J. H. Stone, M. A. Khamashta, and Y. Shoenfeld, "Google-driven search for big data in autoimmune geoepidemiology: Analysis of 394,827 patients with systemic autoimmune diseases," *Autoimmunity Rev.*, vol. 14, no. 8, pp. 670–679, 2015.

[33] L. T. Raynolds, "The globalization of organic agro-food networks," *World Develop.*, vol. 32, no. 5, pp. 725–743, 2004.

[34] H. N. J. Schifferstein and P. A. M. O. Ophuis, "Health-related determinants of organic food consumption in The Netherlands," *Food Qual. Preference*, vol. 9, no. 3, pp. 119–133, 1998.

[35] C. Smith-Spangler, M. L. Brandeau, G. E. Hunter, J. C. Bavinger, M. Pearson, P. J. Eschbach, V. Sundaram, H. Liu, P. Schirmer, C. Stave, I. Olkin, and D. M. Bravata, "Are organic foods safer or healthier than conventional alternatives?: A systematic review," *Ann. Internal Med.*, vol. 157, no. 5, pp. 348–366, 2012.

[36] P. Sugiartawan, R. Pulungan, and A. K. Sari, "Prediction by a hybrid of wavelet transform and long-short-term-memory neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 2, pp. 326–332, 2017.

[37] H. Tang, Y. Qiu, Y. Guo, and J. Liu, "Comparison of periodic behavior of consumer online searches for restaurants in the U.S. and China based on search engine data," *IEEE Access*, vol. 6, pp. 34109–34117, 2018.
[38] A. Tarkiainen and S. Sundqvist, "Subjective norms, attitudes and intentions of Finnish consumers in buying organic food," *Brit. Food J.*, vol. 107, no. 11, pp. 808–822, 2005.
[39] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, "Predictive business process monitoring with LSTM neural networks," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.* Cham, Switzerland: Springer, 2017, pp. 477–492.
[40] C.-C. Teng and Y.-M. Wang, "Decisional factors driving organic food consumption," *Brit. Food J.*, vol. 117, no. 3, pp. 1066–1081, 2015.
[41] Y. Teng, D. Bi, G. Xie, Y. Jin, Y. Huang, B. Lin, X. An, D. Feng, and Y. Tong, "Dynamic forecasting of Zika epidemics using Google trends," *PLoS ONE*, vol. 12, no. 1, 2017, Art. no. e0165085.
[42] G. Thimm and E. Fiesler, "Neural network initialization," in *Proc. Int. Workshop Artif. Neural Netw.* Berlin, Germany: Springer, Jun. 1995, pp. 535–542.
[43] J. Thøgersen, Y. Zhou, and G. Huang, "How stable is the value basis for organic food consumption in China?" *J. Cleaner Prod.*, vol. 134, pp. 214–224, Oct. 2016.
[44] J. Thøgersen, M. D. de Barcellos, M. G. Perin, and Y. Zhou, "Consumer buying motives and attitudes towards organic food in two emerging markets," *Int. Marketing Rev.*, vol. 32, nos. 3–4, pp. 389–413, 2015.
[45] L. Vaughan and E. Romero-Frías, "Web search volume as a predictor of academic fame: An exploration of Google trends," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 4, pp. 707–720, 2014.
[46] G. Vittersø and T. Tangeland, "The role of consumers in transitions towards sustainable food consumption. The case of organic food in Norway," *J. Cleaner Prod.*, vol. 92, pp. 91–99, Apr. 2015.
[47] J. Wang and P. Liu, "The research comment on organic food consumer behavior," *Asian Agricult. Res.*, vol. 6, no. 6, pp. 8–15, 2014.
[48] H. Willer, J. Lernoud, B. Huber, and A. Sahota, "The world of organic agriculture at BIOFACH 2018," Res. Inst. Organic Agricult. FiBL, Nuremberg, Germany, Tech. Rep., 2018. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=The+world+of+organic+agriculture+at+BIOFACH+2018&btnG= and http://orgprints.org/32677/1/arbenz-2018-woa.pdf
[49] S. H. I. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
[50] X. Yang, B. Pan, J. A. Evans, and B. Lv, "Forecasting Chinese tourist volume with search engine data," *Tourism Manage.*, vol. 46, pp. 386–397, Feb. 2015.
[51] Q. Yao, Y. Tian, P.-F. Li, L.-L. Tian, Y.-M. Qian, and J.-S. Li, "Design and development of a medical big data processing system based on Hadoop," *J. Med. Syst.*, vol. 39, no. 3, p. 23, 2015.
[52] A. Yelowitz and M. Wilson, "Characteristics characteristics of bitcoin users: An analysis of Google search data," *Appl. Econ. Lett.*, vol. 22, no. 13, pp. 1030–1036, 2015.
[53] R. M. W. Yeung and J. Morris, "Food safety risk," *Brit. Food J.*, vol. 103, no. 3, pp. 170–187, 2001.
[54] E. K. Yiridoe, S. Bonti-Ankomah, and R. C. Martin, "Comparison of consumer perceptions and preference toward organic versus conventionally produced foods: A review and update of the literature," *Renew. Agricult. Food Syst.*, vol. 20, no. 4, pp. 193–205, 2005.
[55] S. D. Young, E. A. Torrone, J. Urata, and S. O. Aral, "Using search engine data as a tool to predict syphilis," *Epidemiology*, vol. 29, no. 4, pp. 574–578, 2018.
[56] W. Zhao, Z. Guo, J. Yue, X. Zhang, and L. Luo, "On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery," *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3368–3379, 2015.
[57] N. Zhu, X. Liu, Z. Liu, K. Hu, Y. Wang, J. Tan, M. Huang, Q. Zhu, X. Ji, Y. Jiang, and Y. Guo, "Deep learning for smart agriculture: Concepts, tools, applications, and opportunities," *Int. J. Agricult. Biol. Eng.*, vol. 11, no. 4, pp. 32–44, 2018.

XU LIU is a Research Assistant with Dr. Y. Guo Lab., Jiangnan University.

MEIYING ZHONG is a Research Assistant with Dr. Y. Guo Lab., Jiangnan University.

BIN LI is an Associate Research Scientist with the Beijing Research Center for Information Technology in Agriculture.

YE SU is a Faculty Member of the University of Nebraska Kearney.

JINGLU TAN is a Faculty Member of the University of Missouri.

SEYED MOHAMMAD TAGHI GHARIBZAHEDI is a Young Research Leader of Islamic Azad University at Lahijan Branch, Lahijan.

YA GUO received the Ph.D. degree from the University of Missouri. He has rich interdisciplinary research experience and published extensively in peer-reviewed journals. He currently serves as an associate editor/section editor-in-chief/editor for several peer-reviewed international journals.

JUAN LIU received the Ph.D. degree in hospitality management from the University of Missouri. Her research papers have been accepted by prestigious peer-reviewed international journals. Her research interests include hospitality management and big data analytics.