IMMM 2019

The Ninth International Conference on Advances in Information Mining and Management

ISBN: 978-1-61208-731-3

July 28 – August 2, 2019

Nice, France

IMMM 2019 Editors

Dirk Labudde, University of Applied Sciences Mittweida, Germany

IMMM 2019

Forward

The Ninth International Conference on Advances in Information Mining and Management (IMMM 2019), held between July 28, 2019 and August 02, 2019 in Nice, France, continued a series of academic and industrial events focusing on advances in all aspects related to information mining, management, and use.

The amount of information and its complexity makes it difficult for our society to take advantage of the distributed knowledge value. Knowledge, text, speech, picture, data, opinion, and other forms of information representation, as well as the large spectrum of different potential sources (sensors, bio, geographic, health, etc.) led to the development of special mining techniques, mechanisms support, applications and enabling tools. However, the variety of information semantics, the dynamic of information update and the rapid change in user needs are challenging aspects when gathering and analyzing information.

The conference included academic, research, and industrial contributions. It had the following tracks:

- Information mining and management
- Mining features

We take here the opportunity to warmly thank all the members of the IMMM 2019 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to IMMM 2019. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also thank the members of the IMMM 2019 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that IMMM 2019 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the field of information mining and management. We also hope that Nice, France provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

IMMM 2019 Chairs

IMMM Steering Committee

Nitin Agarwal, University of Arkansas at Little Rock, USA
Andreas Schmidt, Karlsruhe University of Applied Sciences, Germany
Michele Melchiori, Università degli Studi di Brescia, Italy
Bernhard Bauer, University of Augsburg, Germany
Mehmed Kantardzic, University of Louisville, USA
Daniel Thalmann, Institute for Media Innovation (IMI) | Nanyang Technological University, Singapore
Duarte Trigueiros, ISCTE - University Institute of Lisbon, Portugal, Portugal

IMMM Industry/Research Advisory Committee

Dirk Labudde, Hochschule Mittweida, Germany
Adrienn Skrop, University of Pannonia, Hungary
Stefan Brüggemann, Airbus Defence and Space, Germany
Xuanwen Luo, Sandvik Mining, USA
Maria Luisa Villani, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Italy
Xiang Ji, Bloomberg LP, USA

IMMM 2019

Committee

IMMM Steering Committee

Nitin Agarwal, University of Arkansas at Little Rock, USA
Andreas Schmidt, Karlsruhe University of Applied Sciences, Germany
Michele Melchiori, Università degli Studi di Brescia, Italy
Bernhard Bauer, University of Augsburg, Germany
Mehmed Kantardzic, University of Louisville, USA
Daniel Thalmann, Institute for Media Innovation (IMI) | Nanyang Technological University, Singapore
Duarte Trigueiros, ISCTE - University Institute of Lisbon, Portugal, Portugal

IMMM Industry/Research Advisory Committee

Dirk Labudde, Hochschule Mittweida, Germany
Adrienn Skrop, University of Pannonia, Hungary
Stefan Brüggemann, Airbus Defence and Space, Germany
Xuanwen Luo, Sandvik Mining, USA
Maria Luisa Villani, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Italy
Xiang Ji, Bloomberg LP, USA

IMMM 2019 Technical Program Committee

Nitin Agarwal, University of Arkansas at Little Rock, USA
Akhlaq Ahmad, College of Engineering and Islamic Architecture - Umm Al Qura University, Saudi Arabia
Zaher Al Aghbari, University of Sharjah, UAE
Samer Al-Khateeb, Creighton University, USA
Aletéia Araújo, University of Brasília, Brazil
Kiran Kumar Bandeli, University of Arkansas at Little Rock, USA
Liliana Ibeth Barbosa-Santillan, University of Guadalajara, Mexico
Cristina Barros, University of Alicante, Spain
Bernhard Bauer, University of Augsburg, Germany
Stefan Brüggemann, Airbus Defence and Space, Germany
Erik Cambria, Nanyang Technological University, Singapore
Mirko Cesarini, University of Milan Bicocca, Italy
Nadezda Chalupova, Mendel University in Brno, Czech Republic
Chih-Chuan Chen, National Taitung University, Taitung, Taiwan
Zhiyong Cheng, School of Computing | National University of Singapore, Singapore
Hui-Chi Chuang, Cheng Kung University, Tainan, Taiwan
Pascal Cuxac, INIST-CNRS, Nancy, France
Mirela Danubianu, "Stefan cel Mare" University of Suceava, Romania
František Dařena, Mendel University Brno, Czech Republic

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

The Effect of Integrated Experiential Marketing on Brand Identification and Customer Intention

A Study on Global Fast Fashion Brands in Taiwan

Shu-Fang Luo
Department of Business Administration
Tainan University of Technology
Tainan City, Taiwan, R.O.C.
e-mail: t20011@mail.tut.edu.tw

Chih-Chuan Chen
Interdisciplinary Program of Green and Information Technology
National Taitung University
Taitung, Taiwan, R.O.C.
e-mail: ccchen@nttu.edu.tw

Yi-Chung Cheng
Department of International Business Management
Tainan University of Technology
Tainan City, Taiwan, R.O.C.
e-mail: t20042@mail.tut.edu.tw

Hui-Chi Chuang
Institute of Information Management
National Cheng Kung University
Tainan City, Taiwan, R.O.C.
e-mail: huichi613@gmail.com

Abstract—**This study aims to examine the effect of integrated in-store and online experiential marketing on customer brand identification and purchase intention. This study selects the international fast fashion brand ZARA as the main research object, mainly because ZARA is a major international clothing retailer and pioneer of fast fashion principles which is very popular among young ethnic groups in Taiwan. Different from local clothing brands, most of which either sell via online stores or brick-and-mortar stores, ZARA has successfully promoted online marketing in Taiwan, and still invests in physical stores to create an experience atmosphere. This study believes that fast fashion brands, such as ZARA, can provide high service quality by integrating virtual and real experiential marketing, which shall enable consumers to generate experiential value and brand identity, and hence to enhance customers' purchase intention of their products. The main target customers of these international fast fashion brands are young people, both online and offline. When most fast fashion brands are quite successful at online sales, why do they still open physical stores as luxury brands? Research results show that physical store management can help to improve the customer relationship and hence has positive effect on brand identification and purchase intention.**

Keywords- Experiential marketing; Internet communication; Brand identification; Customer intention; Fast fashion.

I. INTRODUCTION

There are four international apparel retailing chains popular in Taiwan among young adults, namely, ZARA, Uniqlo, H&M, and GAP. These so-called fast fashion brands have revolutionized the fashion industry by following a strategy known as fast fashion, democratizing couture and bringing trendy, affordable items to the masses, especially to the Internet generations of young adults and teenagers. Therefore, these international brands are quite successful in online sales. In Taiwan, all of these clothing brands have both physical and online stores. One might ask why would these fast fashion chains set up physical stores like high-priced clothing brands? All of the four major brands have opened physical stores in big cities in Taiwan. While young customers tend to choose the cheaper products among similar styles, they also favor higher quality. One of the differences between virtual and physical stores is that online stores can offer cheap products and special discounts to entice customers to buy. On the other hand, for the physical stores, in addition to completely delivering the style and concept of the brand and the goods, they provide customers more services with the brand's specific shopping atmosphere and experiences. Therefore, it raises the question: can integrating real and virtual sales establish a stronger and better relationship with consumers?

While online shopping is growing, it is difficult to completely present the style and the corresponding concepts of the brand and its products to the customers. The problem can be resolved in the physical stores by providing more services, store atmosphere and shopping experience to consumers. Therefore, this study aims to examine if the physical stores can strengthen the quality of online store communication, and if integrating real and virtual sales can establish a stronger relationship with consumers and better relationship quality. To this end, this study takes the international fast fashion chain ZARA as the research object, and tries to understand if clothing quality can improve the quality of relationships with customers, if the experience marketing of physical stores can improve the quality of relationships with young consumers, and if the store atmosphere of a physical store can improve the quality of relationships with young consumers.

The rest of the paper is structured as follows. In Section II, we present the literature review. In Section III, we describe our research method. The experiment and discussion are shown in Section IV. Finally, the conclusion and ideas for future work are presented in Section V.

 ISBN: 978-1-61208-731-3

II. LITERATURE REVIEW

This study focuses on issues such as service quality, store atmosphere, brand identification, experiential marketing, experiential value, purchase intention, customer satisfaction, and Internet communication quality.

A. Service Quality

Services consist of a series of processes and activities. Services possess three well documented characteristics, namely, intangibility, heterogeneity, and inseparability. Services and consumption occur and exist simultaneously. Customers feel the service in the process of interactions, and the service quality is the subjective perception of consumers [1]. Customers perceive the service through the overall perception of the overall quality of the services. When the services are provided, whether these services meet consumers' perceptions and expectations will critically influence their repurchase intention [2][3]. However, it could be difficult for a firm to understand how the customers perceive services and service quality.

Nowadays, consumers are demanding increasingly higher service quality. When they are provided the services, they have expectations. If the consumers feel that their demands are satisfied and respected, they will think about buying the products. Moreover, if they feel that the products or services are satisfactory during the purchasing process, they would have the idea of repurchase or recommending it to friends and relatives. In addition, if the service does not provide what the customers wants, one will have to let the customer feel satisfied during the waiting process [4].

Many researchers have discussed the service quality dimensions. Juran proposed dimensions for service quality, such as internal quality, hardware quality, software quality, timely response, psychological quality [5]. Zeithaml, Parasuram and Berry proposed a service quality model based on tangibility, reliability, responsiveness, assurance and empathy [6].

To improve the service quality, ZARA tries to give warm and professional assistance to the consumers and to quickly respond to customers' preference on the product. To this end, ZARA takes a different selling approach than the traditional ways by introducing the Self-Service Check Out Technology that links the online and offline experiences together and create more value for their customers.

B. Store Atmosphere

Kotler et al. [7] have shown that elaborately designed shopping environments could trigger certain emotions and further increase customers' purchase intention. According to Kotler, the sensory channels for atmosphere are sight, sound, scent and touch [8]. Taking these channels into account for the design of the stores could establish some specific atmosphere, which would create some physical as well as psychological experience to the customers, and very likely impact their buying decision [9]. To understand how consumers choose products, one must take both "personal factors" and "situational factors" into consideration, and "situational factors" tend to be more critical than "personal factors" [10][11]. Since the service is intangible, the customer will assess its quality based on some of the most common clues: the store's physical environment design, furnishings, lighting, sound effects, interaction with the service personnel, etc. Berman and Evans [12] categorize the store's atmosphere into five different kinds: (1) external environment: awnings, display windows, buildings, parking lots; (2) general indoor environment: floor plan, color, sound, lighting, smell, temperature, cashier location; (3) store layout: product combination, customer flow, traffic flow planning, waiting time, department location; (4) internal display: product display, shelf/box, poster, label, promotion tag; (5) human variables: crowdedness, customer traits, salesperson attitude and employee uniforms. ZARA conveys the brand's concept with the image of the display window. Its spacious shopping space with bright light and neutral color creates a stylish and comfortable shopping environment. In addition, there are mirrors everywhere in the accessories area for customer to try on items.

C. Brand Identification

Brand identity can convey to consumers information about the logos, concepts and products of the brand. Consumers' recognition of brands includes symbolism, emotions, associations, self-identification, etc. [13]. Consumers are more willing to purchase products that are identifying to their self-images [14], and these products make it easier to build brand identity [15]. Some consumers tend to associate their social status with brands, and think they can improve their personal image by buying brand goods [16], and therefore, to them, brands usually have a strong appeal [17].

When consumers recognize a brand, they are more willing to pick the brand's products as the first choice for purchase. Many consumers choose to use their favorite brands when they purchase goods. They also recommend relatives and friends to use these brand goods, which indirectly helps to promote the brand companies. Some consumers think that ZARA clothing is more fashionable and diversified, and it not only fits their own body shape, but also represents their personal styles.

Brand identity is highly related to perceived values which can be defined with different dimensions, such as active value and passive value [18], functional value, social value, and emotional value [19]. Kepferer defines brand identity with six facets, namely, physique, personality, culture, relationship, reflection, and self-image [20].

D. Experiential Marketing

Schmitt proposed a strategic framework for experiential marketing, in which consumers are viewed as rational and emotional human beings who take pleasurable experiences into purchase decision [21]. Therefore, marketers can create various experiences to intrigue customers and to affect their purchase intention, such as sensory experiences, affective experiences, cognitive experiences, physical experiences, and social-identity experiences. These experiences can be implemented through communications, visual and verbal identity, product presence, electronic media, etc.

 ISBN: 978-1-61208-731-3

Pine and Gilmore pointed out that an experience occurs when a company intentionally uses services as the stage, and goods as props, to engage individual customers in a way that creates a memorable event [22]. In the process of purchasing, consumers not only focus on functions, but also on personal taste, stimulating or certain feelings thus creating an unforgettable experience. The impression will make consumers connect with the brand of the product. When the consumer feels a stimulating and special experience, it will affect the consumer's willingness to purchase.

To create pleasurable and memorable shopping experiences for the consumers, ZARA offers fresh assortments of designer-style garments and accessories for relatively low prices in sophisticated stores in prime locations in order to draw masses of fashion-conscious repeat customers. For Taiwanese consumers, ZARA has been well-known before it has set up a store in Taiwan. Its unique operating model and stylish clothing brings to consumers a different experience than other stores.

E. Experential Values

Value refers to the results of consumers' assessments and comparison about the goods. Perceived value is the consumer's overall assessment of the utility of a product based on perceptions of what is received and what is given, which represents a tradeoff of the critical give and get components [23]. The value of experience comes from the consumers' interactions with the retail entity and services [24]. Experiential value varies with different people, time, and places [18]. Experiential value can be enhanced through interaction, but it may hinder the consumer's purpose [25]. After comparing the services and products, the value perceived by the consumers, and the results will influence the purchase intention.

F. Purchase Intention

Purchase intention refers to three measuring indices, namely, the possibility of consumers purchasing the product [26], whether they will consider purchasing the product, and whether they will recommend the product to friends and relatives as a measure of purchase intention [23]. The higher the willingness to buy, the greater the chance of purchase. Therefore, purchase intention can be used as a predictive consumer's subjective view of a product [27].

G. Internet Communication Quality

In the era of e-commerce, the Internet is used interactively for two-way communication and transactions, with the main method of online communication being e-mail. Some of the sites reviewed also provided online ordering and payment systems, although these varied greatly regarding geographical limitations, merchandise ranges and levels of security [28]. Zeithaml et al. pointed out that communication and control processes in the delivery of service quality. Instantly sharing the brand's information on the Internet to the customers, such as backgrounds, content, plans, activities, goals, would help to make customers understand the firm and its brand [6]. On ZARA's official website, customers can see the background information, design methods, sales models, and annual reports. Classified information, such as lists of new products and best selling products, size chart, detailed compositions and maintenance methods, are also provided for the convenience of the consumers.

III. RESEARCH METHODOLOGY

In this study, a questionnaire survey was conducted to explore the relationship between service quality, experience marketing, store atmosphere, Internet communication quality, and relationship quality, and to see how three aspects of service quality, experience marketing, and store atmosphere, impact on relationship quality with the moderation effect of Internet communication quality. Figure 1 illustrates the hypothesis framework.

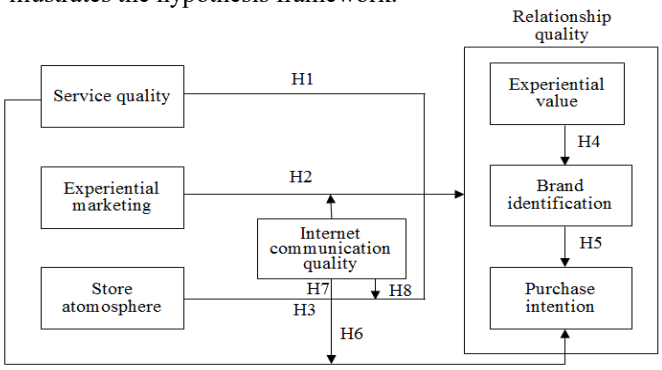

Figure 1. Research hypothesis framework.

The research hypotheses are as follows:

H1: Service quality positively affects relationship quality.

H2: Experiential marketing positively affects relationship quality.

H3: Store atmosphere positively affects relationship quality.

H4: Experiential value positively affects brand identification.

H5: Brand identification positively affects purchase intention.

H6: Internet communication quality positively enhances the relationship between service quality and purchase intention.

H7: Internet communication quality positively enhances the relation between experiential marketing and relationship quality.

H8: Internet communication quality positively enhances the relation between store atmosphere and relationship quality.

IV. RESULTS AND ANALYSIS

A. Statistical Analysis

It took three months to complete the questionnaire survey. The participants of this research were mostly young adults aged 18-40, who have ever shopped or browsed in ZARA store. A total of 592 valid questionnaires were collected. Among them, there were 248 males (47.1%) and 279 females (52.9%). The age distribution was 18-22 (29.8%), 30~40 (20.7%), 40 and above (4.2%), 18 and under (3.8%). The distribution of education was high school (19%), college (73.4%), graduate school (7.6%). The distribution of

 ISBN: 978-1-61208-731-3

monthly cost on apparel was $35 and less (12.3%), $35~$90 (31.5%), $90~$200 (23.7%), $200~$400 (27.9%), $400 and more (4.6%). About 47.6% of the participants have visited ZARA's website.

B. Reliability Analysis

The Cronbach's α is used to check the reliability of factors. According to Nunnall [10], the greater the value of α is, the greater the reliability is. Overall, when α value is greater than 0.7, it is acceptable. When α value is less than 0.5, it is unacceptable. From Table I, one can see that all subfactors have high α value, it means the survey is reliable.

TABLE I. CRONBACH'S α VALUE OF EACH

Factor	Subfactor	Cronbachs'α
Service quality		0.904
Store atmosphere		0.946
Experiential marketing	Sense	0.828
	Feel	0.855
	Think	0.901
	Action	0.751
	Relate	0.957
Internet communication quality	Feeling of presence	0.869
	Communication effectiveness	0.895
Relationship quality	Experiential value	0.944
	Brand identification	0.897
	Purchase intention	0.880

C. Tests of Research Hypothesis

For the first three hypotheses, this study uses multiple regression analysis to test the research hypotheses on service quality, experiential marketing and store atmosphere (independent variables) with respect to relationship quality (dependent variable). F test is used to measure the significance of the regression model. The standardized regression coefficient (beta) is applied to evaluating the predictive or explanative power of an independent variable. The results are shown in Table II where we can see that hypotheses H1, H2, H3 are correct. Moreover, since the Beta value of experiential marketing (0.638) is greater than that of store atmosphere (0.161) and service quality (0.111), it means that experiential marketing has more impact on relationship quality.

TABLE II. MULTIPLE REGRESSION ANALYSIS ON SERVICE QUALITY, EXPERIENTIAL MARKETING, AND STORE ATMOSPHERE VS RELATIONSHIP QUALITY

	UnStd. coef.		Std. coef.	t	P
	Beta	Std. Error	Beta		
Constant	0.630	0.134		4.703	0.000
Service quality	0.101	0.032	0.111	3.139	0.002
Store atmosphere	0.154	0.034	0.161	4.567	0.000
Experiential marketing	0.621	0.035	0.638	17.951	0.000
	R^2			0.708	
	F			422.416***	

dependent variable : relationship quality
*** indicate significance at the 1% levels

UnStd. coef.: Unstandardized coefficients ; Std. coef.: Standardized coefficients

In simple linear regression, F test and t test have the same statistical significance. The t value represents the significance and it always converted to p value to measure the hypothesis. To test hypothesis H4, simple linear regression analysis is conducted for experiential value (independent) and brand identification (dependent). The results are demonstrated in Table III where one can see that H4 is acceptable, that means, the experiential value has positive impact on brand identification.

TABLE III. SIMPLE LINEAR REGRESSION ON BRAND IDENTIFICATION AND EXPERIENTIAL VALUE

	UnStd. coef.		Std. coef.	t	P
	Beta	Std. Error	Beta		
Constant	2.192	0.183		11.961	0.000
Experiential value	0.588	0.035	0.594	16.940	0.000
	R^2			0.353	
	F			286.954***	

dependent variable: brand identification
*** indicate significance at the 1% levels

To test hypothesis H5, simple linear regression analysis is conducted for brand identification (independent) and brand purchase intention (dependent). The results are demonstrated in Table IV where one can see that H5 holds, that says, brand identification positively affects purchase intention.

TABLE IV. SIMPLE LINEAR REGRESSION ON BRAND IDENTIFICATION AND PURCHASE INTENTION

	UnStd. coef.		Std. coef.	t	P
	Beta	Std. Error	Beta		
Constant	2.306	0.175		13.208	0.000
Brand identification	0.554	0.033	0.593	16.872	0.000
	R^2			0.352	
	F			284.667***	

dependent variable: Purchase intention
*** indicate significance at the 1% levels

When adding brand identification to the above simple linear regression, Beta value is reduced from 0.641 to 0.446. It indicates that brand identification is a mediator between experiential value and purchase intention, and the impact of experiential value on purchase intention is reduced.

TABLE V. HIERACHICAL REGRESSION ANALYSIS ON EXPERIENTIAL VALUES, BRAND IDENTIFICATION AND PURCHASE INTENTION

	UnStd. coef.		Std. coef.	t	P
	Beta	Std. Error	Beta		
Constant	2.127	0.164		13.001	0.000
Experiential value	0.592	0.031	0.641	19.125	0.000
	R^2			0.411	
	F			365.781***	
Constant	1.455	0.173		8.387	0.000
Experiential value	0.412	0.036	0.446	11.383	0.000
Brand identification	0.306	0.037	0.328	8.370	0.000
	R^2			0.480	
	F			241.980***	

dependent variable: Purchase intention
*** indicate significance at the 1% levels

To test hypothesis H6, linear regression analysis is conducted for service quality (independent) and purchase intention (dependent) with moderation variable, Internet communication quality. In statistics, the Variance Inflation Factor (VIF) is usually applied in multiple regression analysis, which is the ratio of variance in a model with multiple terms, divided by the variance of a model with one term alone. The results are demonstrated in Table VI, where one can see that H6 holds, that is, good Internet communication quality will enhance the impact of service quality on purchase intention.

TABLE VI. LINEAR REGRESSION ANALYSIS OF INTERNET COMMUNICATION QUALITY SERVICE QUALITY ON PURCHASE INTENTION WITH MODERATION EFFECT

	UnStd. coef.		Std. coef.	t	VIF
	Beta	Std. Error	Beta		
Constant	5.220	0.033		159.219***	0.000
Service quality	0.172	0.040	0.244	4.290***	1.576
Communication quality	0.336	0.039	0.504	8.694***	1.645
Service X communication	0.050	0.025	0.094	2.016*	1.005
	R^2			0.777	
	F			80.585***	

dependent variable: purchase intention
*** indicate significance at the 1% levels

In Table VII, the results of regression of Internet communication as moderator on experiential marketing and relationship quality indicates that hypothesis H7 does not hold.

TABLE VII. LINEAR REGRESSION ANALYSIS OF INTERNET COMMUNICATION QUALITY AS A MODERATOR ON EXPERIENTIAL MARKETING AND RELATIONSHIP QUALITY

	UnStd. coef.		Std. coef.	t	VIF
	Beta	Std. Error	Beta		
Constant	5.243	0.020		267.174***	
Experiential marketing	0.390	0.029	0.623	13.528***	2.337
communication quality	0.190	0.028	0.312	6.794***	2.330
Experiential X communication	0.020	0.012	0.049	1.636	1.005
	R^2			0.776	
	F			285.391***	

dependent variable: relationship quality
*** indicate significance at the 1% levels

In Table VIII, the results of regression of Internet communication as moderator on store atmosphere and relationship quality indicates that hypothesis H8 does not hold.

TABLE VIII. LINEAR REGRESSION ANALYSIS OF INTERNET COMMUNICATION QUALITY AS A MODERATOR ON STORE ATMOSPHERE AND RELATIONSHIP QUALITY

	UnStd. coef.		Std. coef.	t	VIF
	Beta	Std. Error	Beta		
Constant	5.242	0.021		248.501***	
Store atmosphere	0.268	0.023	0.465	11.432***	1.611
communication quality	0.299	0.025	0.490	11.991***	1.621
atmosphere X communication	0.021	0.015	0.044	1.367	1.011
	R^2			0.746	
	F			241.822***	

dependent variable: relationship quality
*** indicate significance at the 1% levels

V. CONCLUSION

This study takes the international fast fashion chain ZARA as the research object and aims to understand the influences of service quality, experiential marketing, and store atmosphere on customer relationship quality. The results show that service quality, experience marketing and store atmosphere will affect the quality of relationship with customers. Experience marketing (ß value 0.638) has the greatest impact, store atmosphere (ß value 0.161) ranks second, and service quality (ß value 0.111) has the least impact. The value of experience affects the willingness to purchase through the mediator of the brand identity. The quality of Internet communication will positively moderate the quality of service to the willingness to purchase.

The brick-and-mortar store manages to create a customer perceptive experience to improve the quality of relationship with customers, including experience value, brand identification and purchase intention. While the store atmosphere can enhance customers' experience value and brand identification, service quality can enhance customer brand identification.

For the international fast fashion apparel chain implementing experiential marketing is useful to create experience value and further increase customers' purchase intention, while brand identification plays a role as mediator to enhance the process.

Among the participants of the questionnaire survey, 251 (47.63%) subjects have visited the ZARA webpage verification page. The Internet communication quality has shown a positive moderation effect on the degree of how the service quality, experience marketing and store atmosphere impact on the purchase intention. Therefore, for the international fast fashion apparel chain which mainly targets the young generations, to maintain and improve Web presence and communication performance can effectively increase the impact of the physical storefront experience marketing, store atmosphere and service quality on customers' purchase intention.

 ISBN: 978-1-61208-731-3

ACKNOWLEDGMENT

This study was supported in part by the Ministry of Science and Technology, ROC, under contract MOST 107-2410-H-143-005.

REFERENCES

[1] D. A. Garvin, "Quality on the line," Harvard Business Review, pp. 66-75, 1983.

[2] V. A. Zeithaml, A. Parasuraman, and L. L. Berry, "Delivering Quality Service," New York: The Tree Press, pp. 9-12, 1990.

[3] M. J. Bitner, "Servicescapes: The impact of physical surroundings on customers and employees," Journal of Marketing, vol. 56, no. 2, pp. 57-71, 1992.

[4] C. Gronroos, "Strategic Management and Marketing m the Service Sector," Boston: Marketing Science Institute, May 1983.

[5] J. M. Juran, "The Quality Trilogy: A Universal Approach to Managing Quality," Quality Progress, vol. 19, pp. 19-24, 1986.

[6] V. A. Zeithaml, L. L. Berry, and A. Parasuraman, "Communication and control processes in the delivery of service quality," Journal of marketing, vol. 52, no. 2, pp. 35-48, 1988.

[7] P. Kotler and S. J. Levy, "Demarketing, yes, demarketing," Harvard Business Review, vol. 79, pp. 74-80, 1971.

[8] P. Kotler, "Atmospherics as a Marketing Tool," Journal of Retailing, vol. 49, pp. 48-64, 1973.

[9] P. Martineau, "The Personality of the Retail Store," Harvard Business Reviews, vol. 36, no. 4, pp. 47-55, 1958.

[10] J. F. Engel, R. D. Blackwell, and D. T. Kollat, "Consumer Behavior," Dryden Press, Hinsdale, IL, 1978.

[11] S. Ward and T. S. Robertson, "Consumer behavior: theoretical sources," Prentice-Hall, 1973.

[12] B. Berman and J. R. Evans, "Retail Management: A Strategic Approach," 6th ed., 1995.

[13] D. A. Aaker, "Measuring brand equity across products and markets," California Management Review, vol. 38, no. 3, pp.102-120, 1996.

[14] D. Aaker, "Brand extensions: The good, the bad, and the ugly," MIT Sloan Management Review, vol. 31, no. 4, pp. 47-56, 1990.

[15] D. W. Rock and S. J. Levy, "Psychology themes in consumer grooming rituals," Advances in Consumer Research, vol. 10, pp. 329-333, 1983.

[16] A. B. Del Rio, R. Vazquez, and V. Iglesias, "The effects of brand associations on consumer response," Journal of Consumer Marketing, vol. 18, no. 5, pp. 410-425, 2001.

[17] C. B. Bhattacharya and S. Sen, "Consumer-company identification: A framework for understanding consumers' relationships with companoes," Journal of Marketing, vol. 67, pp. 76-88, April 2003.

[18] M. B. Holbrook, "The Nature of Customer Value: An Axiology of Services in the Consumption Experience," In Service Quality: New Directions in Theory and Practices, edited by R. Rust and R. L. Oliver, Sage, CA: Newbury Park, pp. 21-71, 1994.

[19] J. C. Sweeney and G. Soutar, "Consumer perceived value: The development of multiple item scale," Journal of Retailing, vol. 77, no. 2, pp. 203-222, 2001.

[20] J. N. Kepferer, "Strategic brand Management," New York: The Free Press, 1992.

[21] B. Schmitt, "Experiential marketing," Journal of marketing management, vol. 15, no. 1-3, pp. 53-67, 1999.

[22] B. J. Pine and J. H. Gilmore, "Welcome to the experience economy," Harvard business review, vol. 76, pp. 97-105, 1998.

[23] V. A. Zeithaml, "Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence," Journal of marketing, vol. 52, no. 3, pp. 2-22, 1988.

[24] M.B. Holbrook and K. P. Corfman, "Quality and value in the consumption experience: Phaedrus rides again," Perceived Quality, MA, Lexington, pp. 31-57, 1985.

[25] C. Mathwick, N. Malhotra, and E. Rigdon, "Experiential Value: Conceptualization, Measurement and Application in the Catalog and InternetShopping Environment," Journal of Retailing, vol. 77, pp. 39-56, 2001.

[26] B.W. Dodds, K. Monroe, and D. Grewal, "Effects of price, brand, and store information on buyers' product evaluations," Journal of Marketing Research, vol. 28, pp. 307-319, 1991.

[27] M. Fishbein and I. Ajzen, "Belief, Attitude, Intention and Behavior: an Introduction to Theory and Research," Addison-Wesley Boston, MA, 1975.

[28] C. Hart, N. Doherty, and F. Ellis-Chadwick, "Retailer adoption of the internet–implications for retail marketing," European Journal of Marketing, vol. 34, no. 8, pp. 954-974, 2000.

 ISBN: 978-1-61208-731-3

Technology Evolution and Technology Forecasting Based on Engineering Big Data

Fuhua Wang
School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
e-mail: fuhuawanghz@sjtu.edu.cn

Zuhua Jiang
School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
e-mail: zhjiang@sjtu.edu.cn

Hong Zheng
School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
e-mail: zhhong@sjtu.edu.cn

Xiaoming Sun
School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
e-mail: xmsun@sjtu.edu.cn

Haili Wang
School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
e-mail: hlwang@sjtu.edu.cn

Xinyu Li
School of Mechanical Engineering
Shanghai Jiao Tong University
Shanghai, China
e-mail: lxy2003jacky@sjtu.edu.cn

***Abstract*—Big-data-driven technology innovation management and technology forecasting in engineering design represent a new challenge that we have to cope with. Data mining algorithms, technology evolution methods and technology forecasting models desiderate theoretical breakthroughs and practical innovations. In this paper, the future development trend of technology is forecasted by analyzing engineering big data. The influence of internal and external factors on the evolution path of technology is researched. Hotspots and frontier fields of current technical development are analyzed. Based on technological innovation path and paradigm shift, we explain the mechanism of technology evolution from multiple perspectives, such as individual/group, short-term/long-term, sudden-change/gradual-change. Technology forecasting and results evaluating methods based on the multi-stage evolution mechanism are proposed. The proposal helps enterprises improve their ability to forecast technological development trends of industries, as well as decision ability of technological Research and Development (R&D) innovation.**

Keywords-data mining; technology evolution; technological paradigm shift; technology forecasting.

I. INTRODUCTION

The fierce market competition has prompted enterprises, which are engaged in product Research and Development (R&D) and technology management, to continuously study the entire process of technology evolution. A diversified and forward-looking technology development layout is proposed for R&D, design and manufacture of the next generation products [1]. However, in the face of massive, real-time, multi-source, heterogeneous engineering knowledge creation and technological paradigm shift records, the challenges are how to make technical knowledge managers and design innovators to: in-depth grasp the rules of engineering technology innovation, comprehensively have an insight into development trends of engineering fields, and systematically plan strategies of engineering product innovation [2]. These challenges have been the basic and urgent work of product development and decision-making management.

The rest of the paper is structured as follows. In Section 1, the background and significance of this topic are introduced. In Section 2, related works about technology forecasting are investigated and summarized. In Section 3, a forecasting model of technical evolution is proposed, considering both the technology supply side and the social demand side. Also, technology forecasting framework for engineering big data is proposed. In Section 4, the advantages and disadvantages of the methods mentioned in this paper are summarized, as well as the direction of future work.

II. RELATED WORKS OF TECHNOLOGY FORECASTING

Technology forecasting began with the US Department of Defense in 1950s, aiming at predicting military technology. Technology forecasting is a method that continuously observes and researches current and future development trends of technology, and evaluates the potential and development prospects of the technology in future application fields, then provides decision-makers with in-depth decision support [3]. Previous research has been based on the Technology Road Mapping Model (TRM) [4][5],

 ISBN: 978-1-61208-731-3

TABLE I. CHARACTERISTICS OF TECHNOLOGY FORECASTING METHODS

Methods	Source	Detailed Methods	Applicability	Efficiency
Technology Roadmapping	Patent texts	Interviews, observations, questionnaires, and statistical analysis	High, for any technology forecasting	Low, require the assistance of customized system
Bibliometrics	Patents and academic journals	analysis of word frequency and citations	High, high data acquisition and availability	High, mature aided software
Information Analysis	Patents and academic journals	analysis of scenario and cycle theory	Medium, affected by data bias	Low, dependent on many experts
Complex Network Analysis	Patents, academic journals and web forums	analysis of co-words, co-citation networks	Medium, affected by available knowledge set	Medium, interpret and verify by expert knowledge

Bibliometrics [6][7], Information Analysis [8], and Complex Network Analysis [9][10].

Most scholars qualitatively developed TRM-based forecasting standards and schemes [11][12], following the first proposed structural TRM model by Phaal et al. [13]. Press, subject category, author, country and keywords of academic journals are mainly analyzed by bibliometrics, especially the common analysis based on word frequency and citation. Sinha [14], Liu et al. [15], Wang et al. [16] have analyzed and explored the long-term development profiles and principles through bibliometrics studies in fields of biology, medicine and environment.

With the promotion and application of patent information analysis, many scholars have performed technology forecasting by information science. Mi [17] forecasts the technological development prospects for the laptop industry using information from the past 10 years by patent information analysis method, in order to provide reasonable decision for the strategic layout of patents.

Complex network analysis has also been used to mine hotspots and track research trends in recent years. This method mainly analyzes key co-words, co-introduction, co-authoring and their relationships in published papers by evaluation indexes of a complex network, to identify frontier topics and forecast technological development trends [18]. This method often combines with traditional methods. For example, Zhang et al. [19] establish a forecasting model of technological topics combined with the TRM and the bibliometrics, based on a complex network. Cheng et al. [20] combine traditional bibliometrics with anomaly testing methods of a network. An anomaly event detection model was proposed to explore the development trends in social computing fields.

The characteristics of the above technology forecasting methods are summarized in Table 1. However, there are some shortcomings in the existing researches because existing technology roadmapping needs to consult a large number of literature works in a specific field and the process is time consuming. Bibliometric measurement and information analysis focus on quantitative descriptions of statistics. Only the "explicit" statistical knowledge can be obtained and it is difficult to identify the potential development trend. Complex network analysis pays much attention to relationships between research variables, and thus it cannot consider particularities of variables. On the other hand, the researches worked by Fye et al. [21] have shown that the accuracy of technology forecasting for short-term (1-5 years) and medium-term (6-10 years) is about 38-39%, while that for long-term (11 years and above) is only 14%. However, there are few studies on “accuracy of technology forecasting”. There are few papers to retrospect and evaluate accuracy for their models of technology forecasting [22]. Most accuracy evaluations of forecasting are for commercial and medical fields. The evaluation system based on engineering/scientific technology is more limited.

Previous studies in this field [23]-[25] only focus on two highly structured data sources: technology patents and journals. Their methods of technology forecasting are limited to qualitative or semi-quantitative research (such as Gartner's Hype Cycle [26]), without considering the inherent mechanism of technology development, such as individual/ group, short-term/ long-term, sudden-change/ gradual-change. Also, these works seldom consider the impact of external social environment and internal user needs on technological paradigm shift. As a result, the accuracy and reliability of existing models are insufficient. Therefore, in big data environment, we propose a novel forecasting model of technological development trend, which considers both external market demand and internal technology evolution law. In addition, we propose an algorithm framework of technology forecasting based on "Technical Knowledge Evolution-Technological Paradigm Transition-Technological System Revolution". It integrates Natural Language Processing (NLP), semantic network analysis, text meta-analysis, and deep learning algorithm into the framework. The proposal improves accuracy and validity of forecasting results and provides new ideas for the research in this field.

III. FRAMEWORK OF TECHNOLOGY FORECASTING BASED ON ENGINEERING BIG-DATA

A. Data mining and semantic analysis for massive multi-source heterogeneous big-data in product development and engineering design

Driven by product development and engineering design big data, the technological paradigm shift presents new

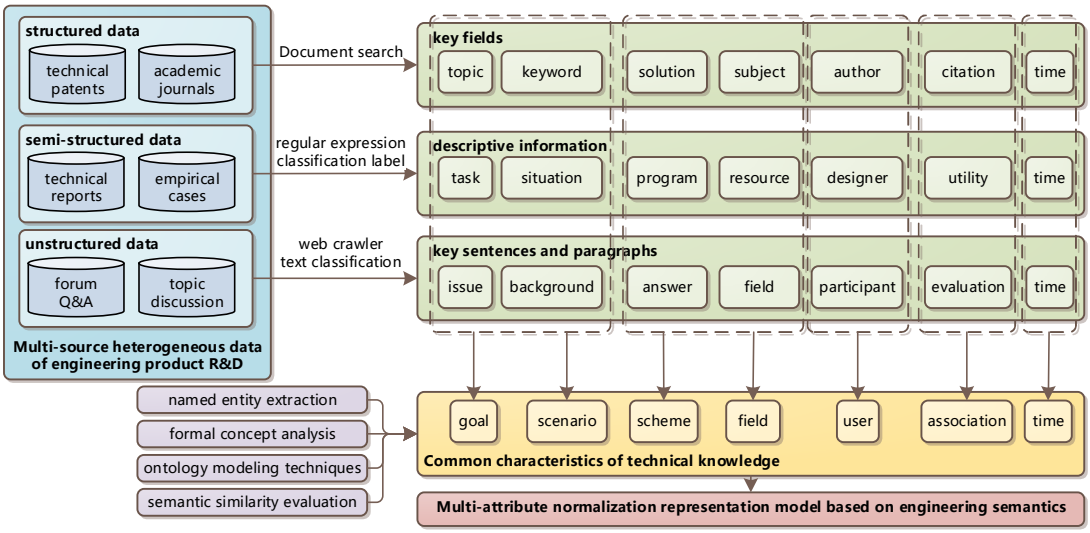

Figure 1. Data mining and semantic analysis for massive big-data in product development and engineering design

features and new forms that are different from previous single areas, single audiences, single dynamics and single models. This paper exploits and analyzes in depth massive multi-source heterogeneous engineering big data with multi-view, multi-temporal, multi-dimensional and multi-granularity, as shown in Figure 1. Our method integrates a variety of advanced text mining methods, such as natural language processing, ontology modeling, semantic network, semantic similarity evaluation, etc. Different sources and different structures engineering text, such as technical patents, journal articles, empirical case documents, professional Q&A forums are fully considered. Data mining and semantic analysis algorithms for engineering big data are developed and improved. Technical knowledge in massive multi-source heterogeneous data is automatically extracted. The engineering semantics is expressed in the form of multi-attribute. It is convenient for computer system to process knowledge automatically.

B. Technology forecasting based on the framework of "Technical Knowledge Evolution-Technological Paradigm Shift-Technological System Revolution"

Patterns, potentials and rates of technology hotspots are evaluated with focus on the future trend of technological paradigm development, based on the framework of "Technical Knowledge Evolution-Technological Paradigm Shift-Technological System Revolution". Multiple technical routes and their possibilities within the limits of technological evolution are forecasted. Considering the micro-breakthroughs, macro-continuation and multi-domain cross-impact features of engineering product technological paradigm shifts, the key characteristics at different stages are extracted, and branch sets of technological evolution are created. This paper objectively and quantitatively describes future possible scenarios of technology developments from aspects of formation conditions, classification and potential estimation of the branches. Based on historical data of technology developments, technologies in different lifecycles and prospects are tracked from the perspective of technology supply side and social demand side. The technology forecasting model and the estimating method for forecasting results are constructed based on a multi-stage technology evolution mechanism. This provides theoretical guidance and method support for technological management decision-making in the field of big data-driven engineering.

C. Forecasting model of technological paradigm shift for technology supply side and social demand side

With the rapid development of technology in current hotspots, technological paradigm shifts present characteristics of micro-breakthrough, macro-continuation, and multi-domain intersection. Enterprises and governments need to grasp the technological development context. Previous subjective, one-sided, qualitative analysis is no longer satisfied technology forecasting. As shown in Figure 2, in this paper, we use a deep learning algorithm called Long-Short Term Memory (LSTM), which can effectively take advantage of historical data. A technology forecasting model considering both the technology supply side and the social demand side is constructed, by means of regression analysis and multi-domain text meta-analysis. The training set is composed of the data in the entire time series of the technological paradigm transition and the testing set is composed of branch the data sets of the technological evolution. Our approach improves the ability of LSTM cells to process information by the improved "gate structure" and enhance the long-term memory ability of the forecasting model. Through the modified items of the assessment model of technological maturity, the forecasting algorithm adopts different optimization strategies at different stages of technological evolution. By decomposing the branch set of technological evolution, the evolution of each technological subsystem is researched and the process of model training is accelerated. Both the technology supply side (basic theoretical accumulation, bottleneck breakthrough, mature technical diffusion, etc.) and the social demand side (potential user demand, regional market layout, industrial macro-policy, etc.) are considered. Technological forecasting

 ISBN: 978-1-61208-731-3

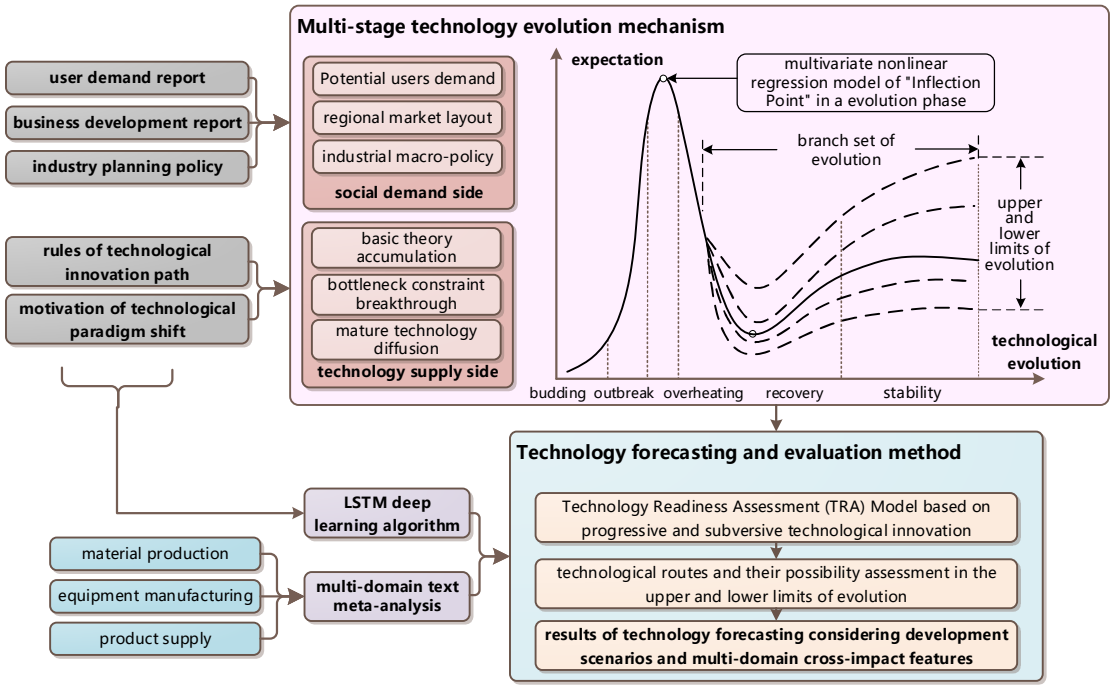

Figure 2. Multi-stage technology evolution and technology forecasting

methods and evaluation methods for hotspots and frontier fields are proposed. Various technological routes and their possibilities for the upper and lower limits of technological evolution are forecasted and analyzed. The trend of engineering technological paradigm shift is forecasted scientifically and reasonably from both perspectives of single-focus technology development and multi-domain cross-impact features.

IV. CONCLUSION AND FUTURE WORKS

Previous technology forecasting methods based on technology hype cycle just consider the external market demand. However, we propose a novel technology forecasting model, which considers both the external market demand and the internal technology evolution law. In this paper, we propose an algorithm framework of technology forecasting based on "Technical Knowledge Evolution-Technological Paradigm Transition-Technological System Revolution". It unifies natural language processing and deep learning in the algorithm framework. The proposal improves the accuracy and validity of forecasting results and provides new ideas for the research in this field. Our research group is carrying out a case study of the algorithm framework of technology forecasting. Some results have been obtained and the experimental results will be published later. In addition, we will compare the proposed model with other models, emphasizing the importance of the proposed technology forecasting methods and expanding this paper.

In big-data driven engineering product innovation, the development routes of technical knowledge and technological paradigm are affected by individual user preferences, R&D team habits, technical field consensus and industrial value concepts. This influence is sometimes abnormal or anomalous. The complexity and variability of knowledge evolution have brought great difficulties to technology forecasting. In this regard, big-data driven technology innovation management and prediction of engineering products must complete three tasks, namely: aggregating and mining the technological knowledge from multi-dimensional and multi-granularity perspectives, modeling and analyzing for technological innovations from multi-temporal and multi-modal perspectives, and observing and forecasting technological paradigms shift from multi-domains and multi-perspectives.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (No. 71671113) and China High-Tech Ship Project of the Ministry of Industry and Information Technology No. [2016] 545.

REFERENCES

[1] Y. Huang et al. "A hybrid method to trace technology evolution pathways: a case study of 3D printing," Scientometrics, vol. 111, no. 1, pp. 185-204, 2017.

[2] L. Ardito, D. D'Adda, and A. M. Petruzzelli, "Mapping innovation dynamics in the Internet of Things domain: Evidence from patent analysis," Technological Forecasting and Social Change, vol. 136, pp. 317-330, 2017.

[3] S. C. Aleina, N. Viola, R. Fusaro and G. Saccoccia, "Approach to technology prioritization in support of moon initiatives in the framework of ESA exploration technology roadmaps," Acta Astronautica, vol. 139, pp. 42-53, 2017.

[4] J. Liu, F. Ma, and Y. Zhang, "Forecasting the Chinese stock volatility across global stock markets," Physica A: Statistical Mechanics and its Applications, vol. 525, pp. 466-477, 2019.

 ISBN: 978-1-61208-731-3

[5] M. H. Cho, and J. H. Cho. "A Study on the Role of Input Stabilization for Successful Settle down of TRM in Production Process: A Case of Display Industry," Journal of the Society of Korea Industrial and Systems Engineering, vol. 39, no.1, pp. 140-152, 2016.

[6] B. Wang, Y. Liu, Y. Zhou, and Z. Wen, "Emerging nanogenerator technology in China: A review and forecast using integrating bibliometrics, patent analysis and technology roadmapping methods," Nano Energy, vol. 46, pp. 322-330, 2018.

[7] T. H. Daim, G. Rueda, H. Martin, and P. Gerdsri, "Forecasting Emerging Technologies: Use of Bibliometrics and Patent Analysisa," World Scientific Series in R&D Management, vol. 305, 2018.

[8] S. Altuntas, T. Dereli, and A. Kusiak, "Forecasting technology success based on patent data," Technological Forecasting and Social Change, vol. 96, pp. 202-214, 2015.

[9] F. Moretti, S. Pizzuti, S. Panzieri, and M. Annunziato, "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling," Neurocomputing, vol. 167, pp. 3-7, 2015.

[10] L. Saad Saoud, F. Rahmoune, V. Tourtchine, K. Baddari, "Fully complex valued wavelet network for forecasting the global solar irradiation," Neural Processing Letters, vol. 45, no. 2, pp. 475-505, 2017.

[11] D. K. R. Robinson and T. Propp, "Multi-path mapping for alignment strategies in emerging science and technologies," Technological Forecasting & Social Change, vol. 75, no. 4, pp. 517-538, 2008.

[12] T. A. Tran and T. Daim, "Taxonomic review of methods and tools applied in technology assessment," Technological Forecasting & Social Change, vol. 75, no. 9, pp. 1396-1405, 2008.

[13] R. Phaal, C. J. P. Farrukh and D. R. Probert, "Technology roadmapping—A planning framework for evolution and revolution," Technological Forecasting & Social Change, vol. 71, no. 1, pp. 5-26, 2004.

[14] B. Sinha. "Trends in Global Biopesticide Research: A Bibliometric Assessment," Social Science Electronic Publishing, vol. 82, no. 2, pp. 95-101, 2008.

[15] X. Liu, Z. Liang and H. Song, "Global biodiversity research during 1900-2009: a bibliometric analysis," Biodiversity & Conservation, vol. 20, no. 4, pp. 807-826, 2011.

[16] M. H. Wang, J. Li and Y. S. Ho, "Research articles published in water resources journals: A bibliometric analysis," Desalination and Water Treatment, vol. 28, pp. 353-365, 2011.

[17] M. Lan. "An analysis of patent information of 1997-2006 laptops and forecast of their development," Library &Information Studies, vol. 10, pp. 24-27, 2008.

[18] L. Zhao and Q. Zhang, "Mapping knowledge domains of Chinese digital library research output, 1994–2010," Scientometrics, vol. 89, no. 1, pp. 51-87, 2011.

[19] Y. Zhang, G. Zhang, H. Chen, A. L. Porter, D. Zhu and J. Lu, "Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research," Technological Forecasting & Social Change, vol. 105, pp. 179-191, 2016.

[20] Q. Cheng, X. Lu, Z. Liu and J. Huang, "Mining research trends with anomaly detection models: the case of social computing research,". Scientometrics, vol. 103, no. 2, pp. 453-469, 2015.

[21] S. R. Fye, S. M. Charbonneau, J. W. Hay and C. A. Mullins, "An examination of factors affecting accuracy in technology forecasts," Technological Forecasting & Social Change, vol. 80, no. 6, pp. 1222-1231, 2013.

[22] A. Kott and P. Perconti, "Long-term forecasts of military technologies for a 20–30 year horizon: An empirical assessment of accuracy, " Technological Forecasting and Social Change, vol. 137, pp. 272-279, 2018.

[23] D. P. D. Alcantara and M. L. Martens, "Technology Roadmapping (TRM): a systematic review of the literature focusing on models," Technological Forecasting and Social Change, vol. 138, pp. 127-138, 2019.

[24] M. J. Cobo, M. A. Martínez, M. Gutiérrez-Salcedo, H. Fujita, and E. Herrera-Viedma, "25 years at knowledge-based systems: a bibliometric analysis," Knowledge-Based Systems, vol. 80, pp. 3-13, 2015.

[25] Y. Zhang, G. Zhang, H. Chen, A. L. Porter, D. Zhu, and J. Lu, "Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research," Technological Forecasting and Social Change, vol. 105, pp. 179-191, 2016.

[26] V. Sriram, V. Shukla, and S. Biswas, "Metal Powder Based Additive Manufacturing Technologies—Business Forecast," 3D Printing and Additive Manufacturing Technologies, pp.105-118, 2019.

 ISBN: 978-1-61208-731-3

Multiclass Neural Network for Codec Classification

Seungwoo Wee

Department of Electronics and
Computer Engineering
Hanyang University
Seoul, South Korea
Email: slike0910@hanyang.ac.kr

Jechang Jeong

Department of Electronics and
Computer Engineering
Hanyang University
Seoul, South Korea
Email: jjeong@hanyang.ac.kr

***Abstract*—In this paper, we suggest to remove or modify the denoted class of codec in a bitstream for military purposes in image communication. In that case, a decoder first needs to determine the codec type to restore the original data. This paper proposes a codec classification method which has not been studied much yet. For extracting the feature of a bitstream, Recurrent Neural Network (RNN) model is used since it is suitable for time series data used for training on classification. Video codecs have their own distinctive header structures, which can be considered features in the encoded bitstreams. The proposed method extracts the feature of an encoded bitstream and classifies the bitstream into the specific codec. Three standard codecs, MPEG-2, H.263, and H.264/AVC, are used to generate the training and the test data set in the experiment. We analyze several components affecting the performance and compare to conventional algorithm. The performance degrades when two kinds of bitstreams generated by H.263 and H.264/AVC are trained together. However, when the training data includes both H.263 and H.264/AVC, performances improved with increasing training data set sizes.**

Keywords–Feature extraction; Classification; Bitstream; Multiclass neural network; Recurrent neural network.

I. INTRODUCTION

In image communication, decoders decode images by parsing the received bitstreams. Since the class of codec is denoted in the header of the bitstream, a decoder does not need additional processing to determine the codec type of received bitstreams. Therefore, codecs have been developed focused on compression rate and complexity. For this reason, codec classification methods have not been studied yet. However, if the class of codec written in the header is removed or modified for military purposes, a decoder first needs to classify the codec type to restore the original data.

Classification algorithms, usually, have been applied to image and language [2][4]. To determine the codec type using partial bitstreams, Recurrent Neural Network (RNN) is used for codec classification [13]. This method exploits the fact that standard codecs are hierarchically structured, which can be used to extract features of the specific codec.

In this paper, we propose a multiclass neural network model for codec classification considering MPEG-2, H.263, and H.264/AVC. Our proposed method classifies an unknown bitstream into a specific codec utilizing the fact that the encoded bitstream consists of its unique data form. Through the experimental results, we show the different tendencies of the performance according to the size and composition of the training data set.

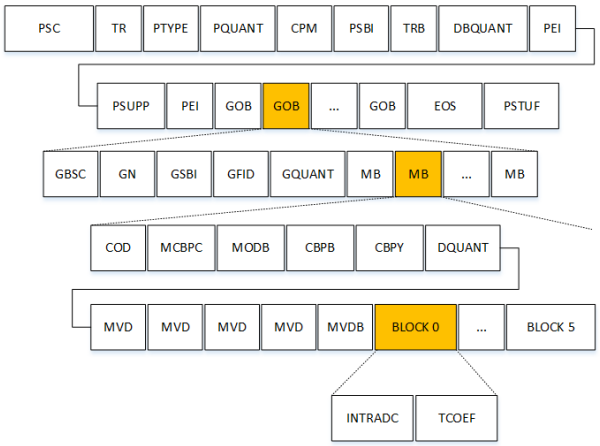

Figure 1. Hierarchical structure of H.263.

We organize the rest of the paper as follows. First, we introduce some backgrounds in Section II to help to understand the proposed algorithm about multiclass neural network. Section III introduces the proposed codec classification algorithm based on RNN model. In Section IV, experimental results are shown. Finally, this paper is concluded in Section V.

II. BACKGROUND

In this section, we give some backgrounds before introducing the proposed algorithm. After introducing the characteristics of the encoded bitstreams, we describe a neural network which extracts features of each codec for classification.

A. Video Coding

Video coding is an essential part to store or transmit video data efficiently. Two main organizations of video coding standards, Moving Picture Experts Group (MPEG) [14] and Video Coding Experts Group (VCEG) [15], try to compress data size while keeping the quality of decoded contents as high as possible. Each codec has hierarchical structure.

Figure 1 shows the hierarchical structure of H.263 to compress data efficiently. Likewise, the header structures of MPEG-2 and H.264/AVC are hierarchical, too [8] [9].

Based on each structure, codecs have their own start bit codes, which occur rarely at general data. Figure 2 represents bitstream generated by MPEG-2 encoder, and each element is expressed in hexadecimal. The bitstream begins with ’00 00

 ISBN: 978-1-61208-731-3

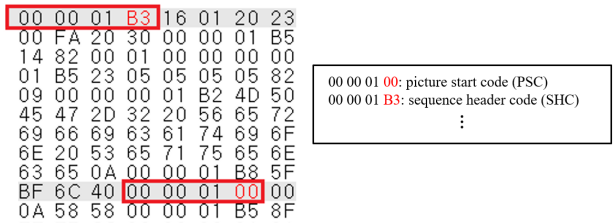

Figure 2. An example of the MPEG-2 bitstream in hexadecimal format.

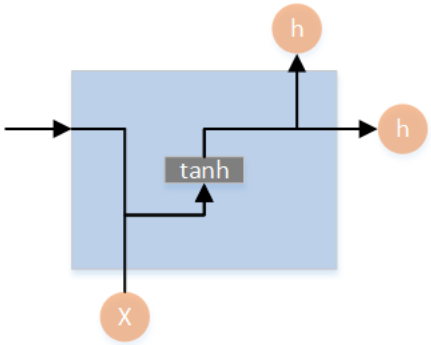

Figure 3. Basic RNN layer

01 B3', Sequence Header Code (SHC), which indicated that the following parts include sequence information.

Likewise, other standard codecs, such as H.263 and H.264, include their own start bit codes, respectively. These codes can be regarded as unique features of each codec.

B. Neural network

A Deep Neural Network (DNN) has layers between the input and output layers [1]. The inputs of DNN pass through the layers to calculate the probability of each output. The appearance of AlexNet which uses DNN improved the image classification performance substantially [2].

Convolutional Neural Network (CNN, or ConvNet) is a deep, feed-forward Artificial Neural Network (ANN) and it uses convolution for feature extraction. Therefore, it is usually utilized in image recognition and natural language processing[3][4][10][11].

Recurrent Neural Network (RNN) is an artificial neural network and has connections between nodes. The nodes can indicate time dynamic behavior for the time sequence data like handwriting or voice signals. RNN stores the internal state, which handles inputs of the network and influences the near layers. For tasks such as handwriting and speech recognition, this network is used [5]–[7]. RNN is known as a model suitable for processing data that appears sequentially, such as voice and text.

Figure 3 represents a basic RNN layer. The output h of the current state is updated by receiving previous output and current input x. The activation function of the hidden state is *tanh*, a nonlinear function. The hidden node stores a state and is connected to the next layer.

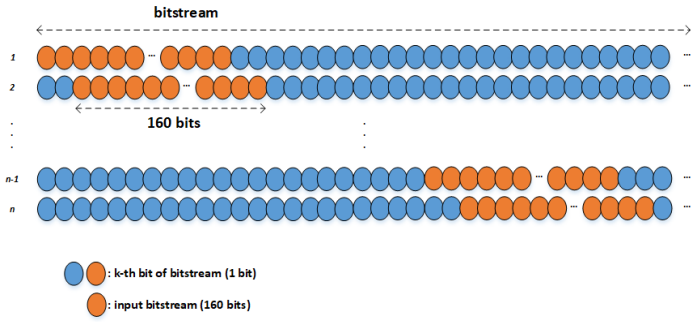

Figure 4. Composition of training and test data set.

As shown in Figure 3, RNN has an advantage that it can create various flexible structures according to needs, because it can accept inputs and outputs regardless of the sequence length. Each RNN layer stores a state and the state affects the input of the next layers. The closer the distance between the two states, the greater the impact on each other. This time dependent characteristic motivates our proposed method for applying RNN to codec classification. Since the bitstream generated by a codec is time series data and each codec has its own structure, the RNN based model is suitable to extract those features. We describe how the data set is organized and the structure of the neural network in the next section.

III. Proposed Algorithm

Our proposed algorithm depends on the following two assumptions. First, the RNN based model can be trained to extract the features of an unknown bitstream. In our previous work [13], we have already shown that an RNN based model with a simple RNN structure is trainable. Second, the longer the length of each input or the deeper the layer of the model, the higher the accuracy of codec classification. We explain the second assumption with analyze the experimental results in the next section

As mentioned in Section II, here we describe the construction of data set in detail Figure 4 presents the composition of the training and test data set. All rows represent the same bitstream of a codec. The orange circles of the n-th row are labeled as the codec class, which is a basis for training our network. n is the number of items in the data set, and the consecutive orange circles, 160 bits, are the n-th input of the proposed network.

Since the characteristics of each codec are concentrated in the header section, the data set is constructed at the beginning of the bitstream. As shown in Figure 4, the data set is created from the beginning of the bitstream by shifting because the characteristics of each codec are concentrated in the header part. We analyzed the results according to n values and the composition of the data set.

The overall network structure of the proposed multiclass neural network based on RNN is shown in Figure 5. Before the input bitstream passes through RNN layers for the training process, we apply an embedding process to improve performance [7]. Since the bitstream consists of binary data, the difference between input data that will be considered feature is too subtle to discriminate.

 ISBN: 978-1-61208-731-3

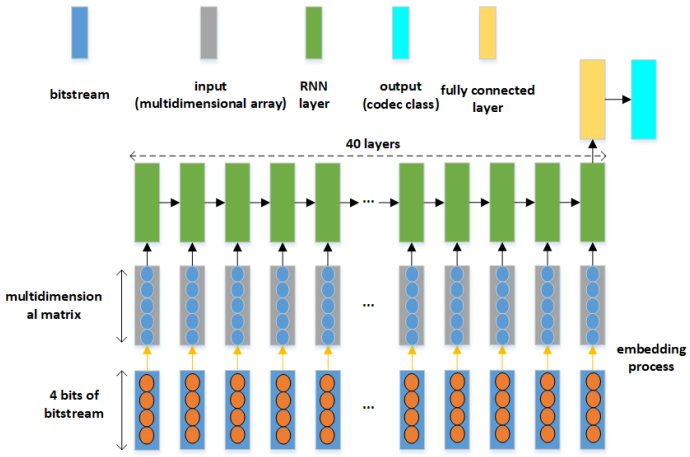

Figure 5. Network structure of the proposed algorithm.

Through the embedding process, 4 bits of the bitstream are transformed into a multidimensional matrix, which is useful to extract features and to be updated. The proposed algorithm converts each input to a 100-dimensional matrix whose elements consist of rational numbers. We compose 40 RNN layers for training and a fully connected layer estimates a specific codec class of the current input.

IV. Experimental Results

In this section, we present experimental results and the analysis of the results. Through the experimental results, we found the different tendencies of the performance according to the size and composition of the training data set.

We used RNN layer to classify the unkown bitstream into the specific codec. Three standard video codecs, MPEG-2, H.263, and H.264/AVC, were used in the experiment to generate bitstreams and test the performance. To construct the training and test set, Common Intermediate Format (CIF) video sequences were used for a total of 3,000 labeled data items. The experiments were performed using Python 3.6, PyTorch and Window 10 Pro x64 environment.

The hidden layer and the batch size were 100 and 256, respectively. We created the training data set by shifting, as shown in Figure 4. Each input was composed of 4-bit units and an input passed through each RNN layer. Before inputs passed through the RNN layer, the inputs was transformed into a 100-dimensional matrix by the embedding process. 40 matrices passed through 40 RNN layers at once.

We compared the results according to the composition and size of the training data set. The organization of the training data set was composed to calculate the accuracy of binary codec classification and multiple codec classification. The size of the test data set was the fixed length of 3,000. The sizes of training data sets were of 3,000, 6,000, 9,000, and 12,000, respectively.

Table I shows the accuracy of our proposed algorithm according to the composition and size of the training data set. Performance degradation occurred when both bitstreams generated by H.263 and H.264/AVC trained together. Based on this, we can assume that H.263 and H.264/AVC have relatively similar header structure compared to MPEG2. As shown in Table I, performances improved with larger training data set sizes except when the training data includes both H.263 and H.264/AVC.

TABLE I. Codec classification accuracy according to the composition and size of training data set.

composition of training data set	*training data set size*			
	3,000	*6,000*	*9,000*	*12,000*
MPEG-2 and H.263	0.82	0.85	0.85	0.87
MPEG-2 and H.264	0.76	0.79	0.81	0.82
H.263 and H.264	0.77	0.78	0.77	0.76
MPEG-2, H.263, and H.264	0.70	0.71	0.72	0.72

V. Conclusion

In this paper, we proposed an RNN based multiclass neural network for codec classification. Unlike the previous work, this work used three standard video codecs to test the performance of the method. Besides, further analysis is done considering the size and composition of the training data set.

We experimentally verify that the size and composition of the training data set affected performance and the proposed method is trainable for extracting the features of each codec. Experimental results show that increasing the size of the training dataset improves performance because the various structures in the training dataset help to generalize the model.

Acknowledgment

This work was supported by the Brain Korea 21 plus Project in 2014.

References

[1] J. Schmidhuber, “Deep Learning in Neural Networks: An Overview,” *Neural Networks*, vol. 61, 2015, pp. 85–117.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Proc. Conf. Adv. Neural Inform. Process. Syst.*, 2012, pp. 1097–1105.

[3] A. Van den Oord, S. Dieleman, and B. Schrauwen, “Deep Content-Based Music Recommendation,” in *Advances in neural information processing systems*, 2013, pp. 2643–2651.

[4] R. Collobert and J. Weston, “A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning,” *Proceedings of the 25th International Conference on Machine Learning, ACM*, 2008, pp. 160–167.

[5] R. Bertolami et al., “A Novel Connectionist System for Improved Unconstrained Handwriting Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, 2009, pp. 855–868.

[6] H. Sak, A. Senior, and F. Beaufays, “Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling,” in *15th annual conference of the international speech communication association*, 2014, pp. 338–342.

[7] X. Li and X. Wu, “Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition,” 2014.

[8] B. G. Haskell, A. Puri, and A. N. Netravali, Digital video: an introduction to MPEG-2, Springer Science Business Media, 1996.

[9] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H. 264/AVC video coding standard,” *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, 2003, pp. 560–576.

[10] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–13.

[11] C. Szegedy et al., “Going Deeper with Convolutions,” in *Proc. IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[12] K. Jack, Video demystified: a handbook for the digital engineer, Elsevier, 2011.

 ISBN: 978-1-61208-731-3

[13] S. Wee and J. Jeong “RNN-based bitstream feature extraction method for codec classification,”in *Proc. SPIE 11049, International Workshop on Advanced Image Technology (IWAIT)*, vol. 11049, 2019, p. 110493N.

[14] The MPEG website. [Online]. Available: https://mpeg.chiariglione.org/

[15] The VCEG document archive site. [Online]. Available: https://www.itu.int/wftp3/av-arch/video-site/

 ISBN: 978-1-61208-731-3

A Regression Model of Location Selection for Beverage Chain in Taiwan

Hui-Chi Chuang
Institute of Information Management
National Cheng Kung University
Tainan City, Taiwan, R.O.C.
e-mail: huichi613@gmail.com

Yi-Chung Cheng
Department of International Business Management
Tainan University of Technology
Tainan City, Taiwan, R.O.C.
e-mail: t20042@mail.tut.edu.tw

Chih-Chuan Chen
Interdisciplinary Program of Green and Information Technology
National Taitung University
Taitung, Taiwan, R.O.C.
e-mail: ccchen@nttu.edu.tw

Abstract—**Location selection plays a crucial role in the restaurant industry, especially for the beverage chains. A comprehensive location selection model and appropriate analytical technique can improve the quality of location decisions, attracting more customers and substantially impacting market share and profitability. Location selection is a significant part of strategic management activities. In Taiwan, thanks to its special cuisine culture, the environment of beverage industry is very competitive and versatile. Therefore, it is very important to have an efficient location selection process for a beverage chain, for both franchiser and franchisees. This study establishes a regression model for the turnover of a local beverage chain in the city of Tainan in southern Taiwan. The model is based on the factors that would affect the location selection, such as crowd flow in front of the chain stores, resident population, complementary stores, and competitive stores. For franchiser, the model can be applied to stores in different areas, and for the franchisees, the model can help them to decide the location of the stores.**

Keywords- Location selection; Regression; Beverage Chain; Customer intention.

I. INTRODUCTION

In the food and beverage industry, right location selection is one of most important strategic decisions for the success of the business. In Taiwan, thanks to its warm climate and special cuisine culture, the beverage industry is very prosperous. The most common and the easiest way to start a business is to be a franchisee of a beverage chain for its low entry barrier.

In general, the franchiser would provide the know-hows, such as business process and techniques and help with the planning, while it is the franchisee who is in charge of the location selection. Location selection results in the convenience of service and how many customers can be attracted, and, as a consequence, influences customer loyalty and operation performance [1]. Therefore, for a franchisee, location selection is crucial to the success of the business. There are more than 200 chain brands in Taiwan's beverage industry. In 2017, there were more than 21,000 beverage stores. On average, every person drinks seven cups of beverage monthly. Due to fierce market competition, the location of the beverage store is even more important.

Most studies on location selection focus on the major factors that influence the turnover. This study takes into account the "shop rent to gross profit" ratio. Although it is crucial to the franchisee's profit, most location selection methods just apply an estimated value, which could be quite different from the real value. Moreover, the better the location is, the higher the rent would be. It is an important issue that the franchisee has to take into account.

This study proposes a regression model with turnover as the dependent variable and with location selection factors as independent variables. Both franchiser and franchisee can determine the store location based on the predicted turnover and the store rent. The proposed model can help the franchiser to make the decision on whether to develop the business in a certain area, while it can help the franchiser to determine the location of the store.

The rest of this paper is organized as follows. Section 2 provides some necessary literature review while Section 3 describes the methodology. The model evaluation results and discussion are summarized in Section 4. Conclusions are drawn in Section 5.

II. LITERATURE REVIEW

Franchising is an important way for enterprises to expand their markets, especially for the business environment in Taiwan, where small and medium-sized enterprises are established, and to join the franchise of small brands is the first choice for young people to start their own businesses. Thanks to the special cuisine culture in Taiwan, the beverage industry is very well developed and various beverage chains are competing in the market. It is the most common and the easiest way for a young entrepreneur to join a beverage chain as a franchisee. In the extremely competitive beverage market, store location is crucial to the success of the beverage store.

Right location selection is one of the strategic decisions that carries importance for the success of the business. The location selection problem for all kinds of organizations could be assessed in a wide range of issues with regard to the

 ISBN: 978-1-61208-731-3

following critical aspects: plant location selection, store and warehouse location selection, shopping centre location selection, retail site location selection and healthcare centre and hospital location selection that have the similar problems [2]. Davis et al. identified a number of factors that have influence on the success of food and beverage industry, among which location of food service facility is the most important feature [3]. Location selection results in the convenience of service and how many customers can be attracted, which, as a consequence, would affect customer loyalty, operation performance, and the market share and profitability of a company [1]. A location decision usually involves a long-term commitment of resources, making this issue very important [4]. Several factors need to be taken into account, such as choice of commercial area, competitive stores, complementary stores, and traffic flow.

The choice of commercial area is a decision problem that needs to take into account a large number of criteria and to identify the best option among alternatives, such as ease in accessibility, parking facilities, and located at a street corner [5][6]. The selection of commercial area directly affects the performance of the store [7]. Therefore, it is important to have a thorough investigation on the commercial area to assess the performance of the store [8]-[10].

Jaravaza and Chitando investigated the role of store location in influencing customers' store choice [11]. The research shows that customer store choice decisions are heavily hinged on store convenience, such as shorter travelling distance, complementary services, and convenient public transport. Traffic density is another important issue. Although high traffic density means high crowd flow, it also has a downside of traffic jam, which would impair consumers' intention of coming into the commercial area [12]. For high traffic density areas, one needs to consider the parking facility or if the store is on the street in the location selection process.

In Taiwan, the density of beverage stores is extremely high, that is, it is easy to find various beverage shops of different brands on the same street, and in some areas of the same brand. They compete with each other even if they provide different kinds of drinks. The competition factor should be taken into account in the location selection process. The number of competitors and intensity of competition are included as two criteria. The number of competitors refers to the number of similar restaurants in the vicinity. The intensity of competitors refers to the scale of beverage stores in the vicinity [13]. Although competitors would compete for the customers, in some cases, high intensity of competitive stores could attract more customers.

During the last few decades, there has been a significant increase in one-stop shopping strategies. Shoppers tend to economize on the amount of time they spend on shopping by making multi-purpose shopping trips, combining purchases for different product categories and reducing the number of trips during a particular time period [14]. Multi-purpose shoppers often bypass closer stores to visit agglomerated stores that are further away in order to shop for different types of goods or indulge in different activities on the same trip [15]. Therefore, this study takes into account the complementary stores.

The most often used conventional location selection methods include checklist methods, analog approaches, regression models and location allocation models [13]. The checklist method uses a list of location factors, and systematically evaluates each of the candidate locations, and then identifies the most suitable locations. The analog approach aims to determine the boundary of the interested commercial area and predict the turnover of the new location [16]. To this end, researchers have to find out some similar stores in the earlier stage and investigate their drawing powers in the different areas and locations, then the turnover of the new location can be evaluated. The regression model has been widely used in various fields, which determines the relationship between the major factors and the business performance, by identifying a regression function with the business performance as the dependent variable and the major factors as the independent variables.

Some techniques of artificial intelligence have been applied to location selection problems, such as artificial neural networks and fuzzy set theory [13][17][18]. Jungthirapanich developed a decision support system incorporating a linear additive multi-attribute utility method and a database system which collects the location data extracted from public documents and reports [19].

III. RESEARCH METHODOLOGY

In this study, a regression model is applied to predict the turnover of a beverage store. Based on the studies on location selection in the literature [12][20], and local domain experts' advice, it takes into account several independent variables such as crowd flow, residential population, complementary stores, and competitive stores.

A. Regression Model

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression [21].

Given a data set $\{y_i, x_{I,1}, x_{I,2}, \ldots, x_{I,k}\}_{i=1}^{n}$ of n statistical units, a linear regression model assumes that the relationship between the dependent variable y and the k-vector of regressors x is linear. This relationship is modeled through a disturbance term or error variable ε — an unobserved random variable that adds "noise" to the linear relationship between the dependent variable and regressors. Thus, the model takes the form

$$y = b_0 + b_1 x_1 + b_2 x_2 + ?? b_k x_k + \varepsilon, \qquad (1)$$

where the constant b_0 is the intercept term, and $b_1, b_2, ?? b_k$ are known as effects or regression coefficients, ε is the error term. Usually, in statistics, the linear least squares method is applied for estimating the unknown parameters in a linear regression model. Let $\hat{b}_i$ denote the

squares estimator of b_i $i = 0, 1, 2, \ldots, k$. Then, the estimator of y, denoted as $\hat{y}$, can be presented as follows.

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_k x_k. \quad (2)$$

The coefficient of determination, denoted R^2, is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). It provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model [22]. The coefficient of determination normally ranges from 0 to 1. The greater the value of the coefficient, the stronger the relationship between the independent variable and the dependent variables.

The most general definition of the coefficient of determination is

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

Two kinds of hypothesis testing, *t*-tests are applied to test the significance of each coefficient, respectively, and *F*-test the model significance. The corresponding statistics area is as follows.

$$t_i = \frac{\hat{b}_i}{se(\hat{b}_i)}, \; i = 0, 1, 2, \ldots, k, \quad (4)$$

where $se(\hat{b}_i)$ denotes the standard error of $\hat{b}_i$.

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \quad (5)$$

B. Model development

Based on the studies on location selection in the literature, and local domain experts' advice, at the initial stage, this study takes into account eight major factors that are crucial to location selection of beverage chain [12][20]. They are (1) crowd flow, (2) residential population, (3) target customers in the commercial area, (4) complementary stores, (5) mutually exclusive shops (6) competitive stores, (7) advantages in establishing the store, and (8) disadvantages in establishing the store. During the initial stage, the results show that three variables, target customers, advantages in establishing the store, and disadvantages in establishing the store, are not significant in this case. Meanwhile, to the research object of this study there are no mutually exclusive shops in the commercial areas. Therefore, this study takes into account four major factors for location selection of the beverage chain, and they are (1) crowd flow in front of the store, (2) residential population in the commercial area, (3) complementary stores, and (4) competitive stores.

The dependent variable Y represents the turnover of the store, and the four independent variables are X_1: crowd flow, X_2: residential population, X_3: number of complementary stores, and X_4: number of competitive stores. The regression model can be described as follows.

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + \varepsilon \quad (6)$$

where the constant b_0 is the intercept term, and $b_1, b_2, \ldots, b_4$ are known as effects or regression coefficients, and ε is the error term. Let $\hat{b}_i$ denote the squares estimator of b_i $i = 0, 1, 2, \ldots, 4$. Then, the estimator of y, denoted as $\hat{y}$, can be presented as follows.

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \hat{b}_4 X_4. \quad (7)$$

IV. Results and Analysis

This study investigated a beverage chain including 27 stores and located in the North, Central, East, South, Anping and Yongkang districts of Tainan City. These districts are areas where the population of Tainan is more concentrated. We collected crowd flow, residential population, complementary stores, and competitive stores to predict the turnover of a beverage store. The description of the independent variables is as follows.

(1) crowd flow in front of the chain stores: slack season (July to September); peak season (October to May); non-holiday (Monday to Friday) and holidays (Saturday to Sunday or national holidays). Pick one hour from the daytime (10:00 to 17:00) and evening (17:00 to 22:00) to collect the crowd flows.

(2) residential population in the commercial area: collect the number of households within a radius of 250 meters centered on the location. If it is a general household, 4 people are calculated. If it is an office building, it is based on the number of employees registered in the business. If it is a factory, the number of people who enter the factory at 8:00 is considered.

(3) complementary store: such as restaurant, seafood shop, barbecue restaurant, Internet café.

(4) competitive store: the number of stores serving drinks.

This study utilizes crowd flow, residential population, complementary stores, and competitive stores as independent variables and turnover of a beverage store as dependent variable to develop the prediction model. The regression model is as follows:

Turnover = 0.354* *'crowd flow'* + 0.270* *'residential population'* + 0.655* *'complementary stores'* + 0.088* *'competitive stores'*.

In this paper, we use Analysis of Variance (ANOVA), which is a collection of statistical models and their associated estimation procedures, to analyze the differences among group means in a sample. From Table I, the F-value ($MS_{reg.}$ / MS_{err}) of the regression model is 26.347 and p-value is .000. The P value is the probability of finding the observed results when the null hypothesis of a study question is true. If the value less than 0.05, it means that the prediction model is statistically significant. Table II shows that the predictive power of the model is 79.6% (adjusted R^2). The p value of crowd flow, residential population and complementary stores are less than 0.05, respectively, it means these three factors are significant and positive influence to turnover, on the contrary competitive stores are not significant to turnover. The results show that the number of complementary stores has the greatest influence on the business of the beverage chain, followed by the crowd flow in front of the stores, and the resident population in the region has the least influence.

 ISBN: 978-1-61208-731-3

TABLE I. THE ANOVA TABLE OF THE PREDICTION MODEL FOR THE TURNOVER OF THE FRANCHISEES OF A BEVERAGE CHAIN IN TAIWAN

Model	df	SS	MS	F	P
regression	4	8.727E11	2.182E11	26.347	0.000
error	22	1.822E11	8.281E9		
total	26	1.055E12			

df : degree of freedom
SS : sum of square
MS : mean square
F : $MS_{reg.}$ / $MS_{err.}$
P value : significance

TABLE II. COEFFICIENTS OF THE REGRESSION MODEL FOR THE TURNOVER OF THE FRANCHISEES OF A BEVERAGE CHAIN IN TAIWAN

Model	UnStd. coef.		Std. coef β	t	P
	B	Std. Error			
(constant)	6164.964	52532.577		0.117	0.908
Crowd flow	39.441	18.519	0.354	2.130	0.045
Resident population	55.378	19.397	0.270	2.855	0.009
Complementary stores	6473.280	1336.359	0.655	2.844	0.000
Competitive stores	-2465.031	3260.591	0.088	-0.756	0.458

$R = 0.910$, $R^2 = 0.827$, adjusted $R^2 = 0.796$
UnStd. coef. : Unstandardized coefficient
Std. coef. : Standardized coefficient

V. CONCLUSION

Thanks to Taiwan's unique culture of cuisine, an extremely competitive beverage industry exists. For a beverage chain, usually the franchisor takes the responsibility for planning, counseling as well as offering services, such as training and support. On the other hand, while the franchisees take the franchisor's business systems, training and know-how and put it into practice in their location, they typically are responsible for most of the costs and risks. As the franchisor holds all the data of each franchisee, such as turnover and the location information, if it can provide services, such as location selection advice and turnover forecasting to its franchisees, it can help franchisees to avoid unnecessary losses, and hence it would be mutually beneficial to both parts. Moreover, it can also improve corporate reputation and establish a better brand image.

This study focuses on the location selection problem of a local beverage chain in the city of Tainan in southern Taiwan, where the climate is warm all the year round and it is very common to see people with a cup of tea in hand, and therefore, it forms a special beverage industry with tea shops and beverage booths everywhere. A regression model was proposed for location selection. The results show that the number of complementary stores in the area has the greatest impact on the business of the beverage chain, followed by the crowd flow in front of the stores, and the resident population in the region has the least impact.

The number of competing stores did not show significant impact on turnover, indicating that the target beverage chain has established reputation in the city such that even though there are stores of different beverage chains in the area, consumers would still choose their favorite brand.

Although this study only focuses on the chain beverage stores in Tainan, it can be applied to other cities and different cuisine industries, to identify the factors that have impact on the business.

ACKNOWLEDGMENT

This study was supported in part by the Ministry of Science and Technology, ROC, under contract MOST 107-2410-H-143-005.

REFERENCES

[1] G. H. Tzeng, M. H. Teng, J. J. Chen, and S. Opricovic, "Multicriteria selection for a restaurant location in Taipei," International Journal of Hospitality Management, vol. 21, no.2, pp. 171-187, 2002.

[2] T. Hernandez, "Enhancing retail location decision support: The development and application of geovisualization," Journal of Retailing and Consumer Services, vol. 14, no. 4, pp. 249-258, 2007.

[3] B. Davis, A. Lockwood, P. Alcott, and I. S. Pantelidis, "Food and beverage management," Routledge, 2018.

[4] T. Y. Chou, C. L. Hsu, and M. C. Chen, "A fuzzy multi-criteria decision model for international tourist hotels location selection," International journal of hospitality management, vol. 27, no. 2, pp. 293-301, 2008.

[5] M. Akalina, G. Turhanbi, and A. Sahinc, "The application of AHP approach for evaluating location selection elements for retail store: A case of clothing store," International Journal of Research in Business and Social Science, vol. 2, no. 4, pp. 1-20, 2013.

[6] C. Y. Shen and K. T. Yu, "A generalized fuzzy approach for strategic problems: The empirical study on facility location selection of authors' management consultation client as an example," Expert Systems with Applications, vol. 36, no.3, pp. 4709-4716, 2009.

[7] L. Simkin, "SLAM: Store Location Assessment Model-Theory and practice," OMEGA International journal of Management Science, vol. 17, no. 1, pp. 53-58, 1989.

[8] C. A. Ingene and R. F. Lusch, "Market selection decisions for department stores," Journal of Retailing, vol. 56, no. 3, pp. 21-40, 1980.

[9] D. Grewal, M. Levy, and V. Kumar, "Customer experience management in retailing: an organizing framework," Journal of retailing, vol. 85, no. 1, pp. 1-14, 2009.

[10] J. A. Pope, W. R. Lane, and J. Stein, "A multiple-attribute decision model for retail store location," Southern Business Review, vol. 37, no. 2, pp. 15-25, 2012.

[11] D. C. Jaravaza and P. Chitando, "The role of store location in influencing customers' store choice," Journal of Emerging Trends in Economics and Management Sciences (JETEMS), vol. 4, no. 3, pp. 302-307, 2013.

[12] H. Erbiyik, S. Özcan, and K. Karaboğa, "Retail store location selection problem with multiple analytical hierarchy process of decision making an application in Turkey," The 8th International Strategic Management Conference, Procedia – Social and Behavioral Sciences, vol. 58, pp. 1405-1414, 2012.

[13] R. J. Kuo, S. C. Chi, and S. S. Kao, "A decision support system for selecting convenience store location through integration of fuzzy AHP and artificial neural network," Computers in Industry, vol. 47, pp. 199-214, 2002.

[14] P. T. P. Leszczyc, A. Sinha, and A. Sahga, "The effect of multi-purpose shopping on pricing and location strategy for grocery stores," Journal of Retailing, vol. 80, no. 2, pp. 85-99, 2004.

 ISBN: 978-1-61208-731-3

[15] A. Ghosh and S. L. MacLafferty, “Location strategies for retail and service firms,” Lexington Books, 1987.

[16] S. B. Cohen and W. Applebaum, “Evaluating store sites and determining store rents,” Economic Geography, vol. 36, no. 1, pp. 1-35, 1960.

[17] J. Wang and B. Malakooti, “A feedforward neural network for multiple criteria decision making,” Computers & Operations Research, vol. 19, no. 2, pp. 151-167, 1992.

[18] G. S. Liang and M. J. J. Wang, “A fuzzy multi-criteria decision-making method for facility site selection,” The International Journal of Production Research, vol. 29, no. 11, pp. 2313-2330, 1991.

[19] C. Jungthirapanich, “An intelligent decision support system for facility location.” University of Missouri/Rolla Rolla, MO, USA”1992.

[20] F. F. Wang, L. F. Chen and C. T. Su, “Location selection using fuzzy-connective-based aggregation networks: a case study of the food and beverage chain industry in Taiwan.” Neural Computing and Applications, vol. 26, no. 1, pp. 161-170, 2015.

[21] D. A. Freedman, “Statistical models: theory and practice,” Cambridge University Press, 2009.

[22] R. G. D. Steel and J. H. Torrie, “Principles and Procedures of Statistics with Special Reference to the Biological Sciences,” McGraw Hill, 1960.

 ISBN: 978-1-61208-731-3

Multimodal Deep Neural Networks for Banking Document Classification

Deniz Engin, Erdem Emekligil, Mehmet Yasin Akpınar, Berke Oral, Seçil Arslan

R&D and Special Projects Department, Yapi Kredi Technology
Istanbul, Turkey
Email: {deniz.engin, erdem.emekligil, mehmetyasin.akpinar, berke.oral, secil.arslan}@ykteknoloji.com.tr

***Abstract*—In this paper, we introduce multimodal deep neural networks to classify petition based Turkish banking customer order documents. These petition based documents are commonly free-formatted texts, which are created by customers, but some of them do have a specific format. According to the structure of the banking documents, some documents containing tables and specific forms are convenient for visual representation, while some documents consisting of free-formatted text are convenient for textual features. Since the texts of these documents are obtained via Optic Character Recognition technology which does not work well on handwritten, noisy, and low-resolution image documents, text classification methods can fail on them. Therefore, our proposed deep learning architectures utilize both vision and text modalities to extract information from different types of documents. We conduct our experiments on our Turkish banking documents. Our experiments indicate that combining visual and textual modalities results in better recognition of documents compared to text or vision classification models.**

Keywords–Multimodal Deep Learning; Document Classification.

I. INTRODUCTION

Every bank puts different channels *e.g.*, fax, email, scanner into service to receive its customers' orders for banking transactions. More than 6.5 million transactions are completed in a medium-large scale bank in Turkey received from these channels yearly [1]. Customers share their orders in free-formatted petitions to declare money transfer, tax payments, salary payments which leads to more than 60 various banking process types. Those orders received in image format are mostly multi-page and low in resolution. In traditional process workflow, when the customer order is received, a back-office operator views the order and investigates all pages of the document to detect the process types in the document in order to split (if needed) and direct the order to the correct back-office data entry team. Accordingly, the classification of the customer order is one of the most time-consuming and human workforce required steps of the overall workflow management. Therefore, document classification systems play a crucial role in the banking domain. An overview of our document classification flow can be seen in Figure 1.

Document classification methods can be based on image classification, text classification on obtained text from Optical Character Recognition (OCR) of images, and multimodal classification. In recent works, several deep learning based methods have been focused on document classification by using only the image of documents [2]-[5]. In [2], Convolutional Neural Networks (CNNs) are used as a feature extractor on a specific-region in a document. Also, these region-based

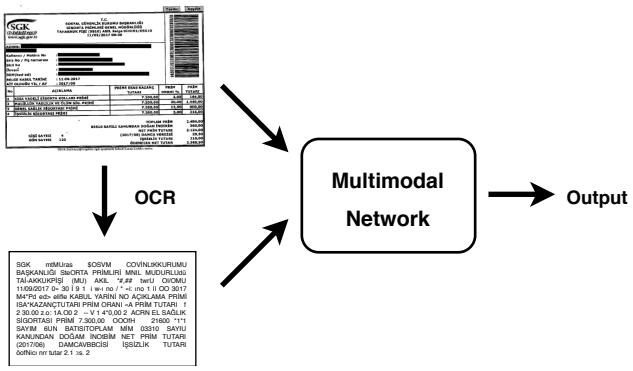

Figure 1. An overview of document classification flow.

features are concatenated before classification. Data augmentation has been applied and CNN architectures have been proposed in [3]. Several well-known CNN architectures, *e.g.*, AlexNet [6], VGG-16 [7], GoogLeNet [8], Resnet-50 [9] have been investigated for document classification by using transfer learning in [4]. Transfer learning from VGG-16 network pre-trained on ImageNet dataset [10] is utilized for region-based document classification in [5]. These proposed methods have been trained on the Ryerson Vision Lab Complex Document Information Processing (RVL-CDIP) dataset [2], which has 16 classes, such as letter, form, email, etc. While classes of this dataset are distinguishable from each other and images of inter-class are consistent, banking order documents have different structures, which can be seen in Figure 2.

Different structures can be categorized as follows:

- free-formatted texts,
- large tables,
- customer arranged forms which are unique for each customer,
- forms that are pre-defined by certain organizations.

Due to structural variation of documents in our dataset, only the vision method is not sufficient for our classification task. Similarly, documents belonging to the same class can be a form or a free-formatted text. Sample documents for this problem can be shown in Figure 3. To overcome these difficulties, we decide to utilize textual information obtained via OCR besides visual information. After the text is obtained from documents, this problem becomes a text classification task. The main idea behind the recent methods is to capture the document representations from characters, words or sentences, by using CNN or Long Short-Term Memory (LSTM), to

 ISBN: 978-1-61208-731-3

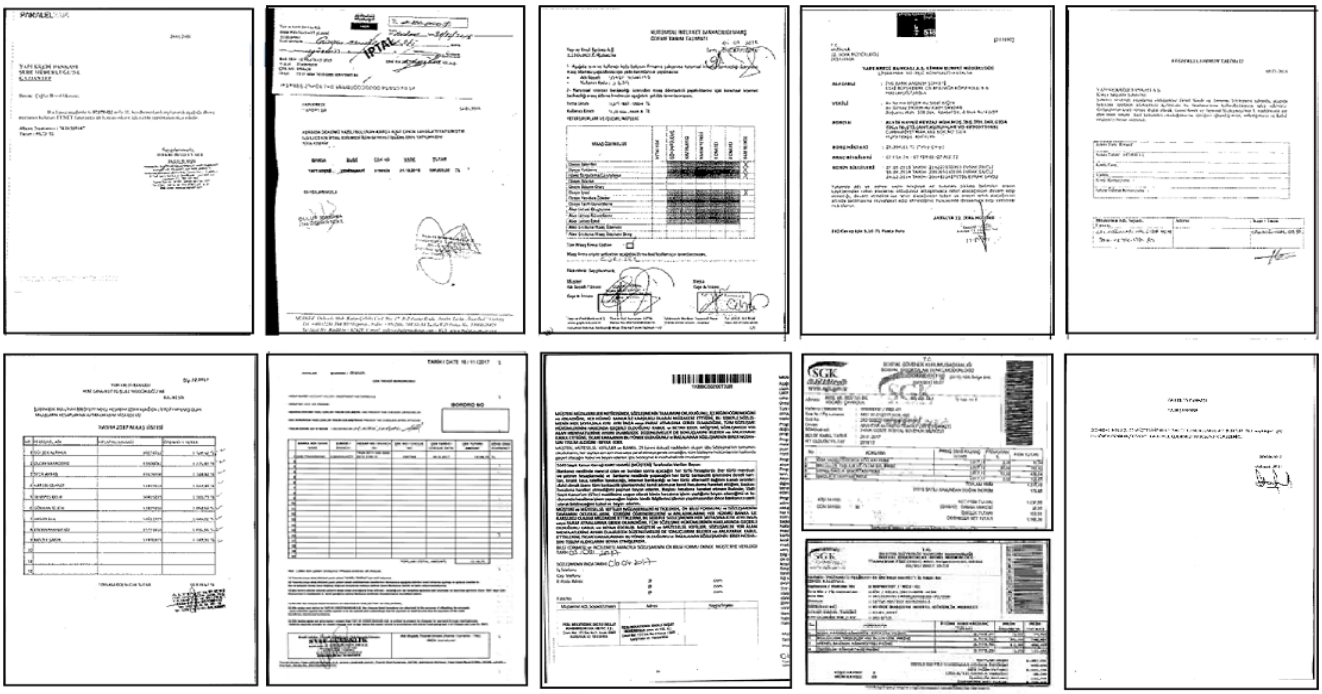

Figure 2. Sample images from the Turkish Banking Documents.

classify documents using these representations. Using CNNs to extract information from characters or word embeddings in order to classify texts have been proposed in [11][12]. In addition, [13] and [14] suggested to use embeddings with sequential models like Gated Recurrent Units (GRUs) or LSTMs.

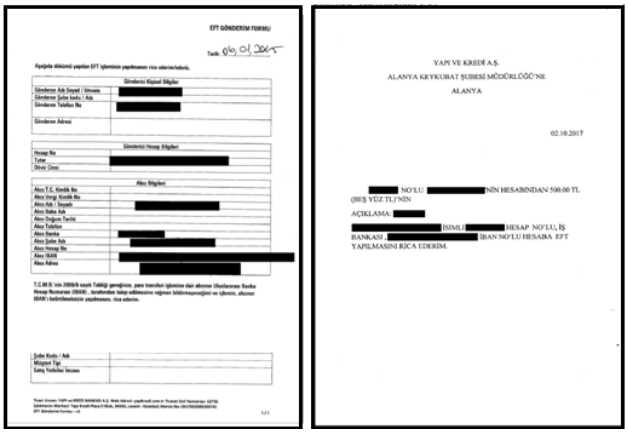

Figure 3. Sample images for the formatted document (left) and free text document (right) belonging to the money transfer class.

Although multimodal deep learning methods have been proposed for classification [15][16], visual question answering [17]-[19], image captioning [20][21], photo editing [22][23], and many other tasks, to the best of our knowledge, these methods have not been used for multimodal document classification. One of the recent works on this specific field proposed methods that use hand-crafted features and SVM for classification in early and late fusion strategies [24]. We propose multimodal classification networks by utilizing the deep neural networks on both modalities, which is different from previous works on document classification.

In this work, our main purpose is to be able to classify Turkish banking documents by utilizing the images of documents and the text which belongs to document. Textual and visual modalities are combined in two methods as early and late fusion in our proposed multimodal deep neural networks. In late fusion, we use our pre-trained LSTM model for text modality and pre-trained CNN model for vision modality to obtain probabilities. Then, we train a small decision level network, which predicts a class by using these probabilities. On the other hand, by feeding FastText embedding to LSTM and images to CNN, the proposed early fusion method jointly learns the image and text representations.

The rest of the paper is organized as follows: our methods are described in Section II, and the experimental results are discussed in Section III. Finally, the conclusion is given in Section IV.

II. METHODS

A. Word Vectors

To train word vectors, we organized an unsupervised dataset with 4.9M documents. Then, we used the Abbyy FineReader OCR tool [25] to extract texts from the data. The obtained text data consists of a total of 787M words and 55.5M vocabulary size. Since OCR generates faulty texts because of noisy images and misspellings, our vocabulary size is unusually large. To overcome problems caused by OCR noises and misspellings, we chose FastText embedding [26], since it works in agglutinative languages more effectively and is able to capture spelling errors better. In Table I, we have show that the most similar words to word ”nezdinizdeki” (a frequently used Turkish word that means *in care of*) are the

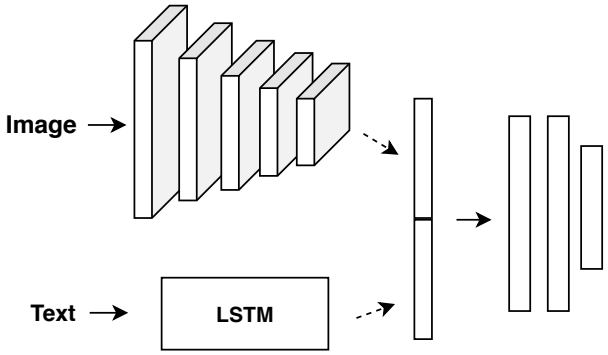

Figure 4. An overview of early fusion network.

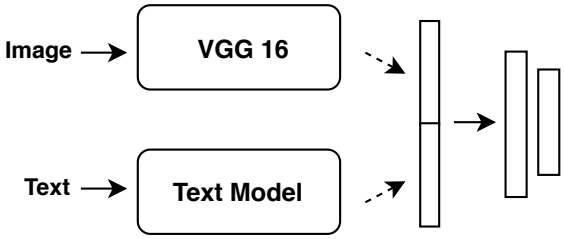

Figure 5. An overview of late fusion network.

words with OCR noises and misspellings. We trained 50, 100 and 300 dimensional FastText embedding vectors and chose to use 100 dimensional vectors in our experiments since it achieved the best accuracy.

B. Text Model

We used a rather simple neural network model with approximately 123k trainable parameters to classify documents using their texts. Our model consists of an embedding layer, a dropout layer with 0.4 rate, an LSTM layer with 128 hidden nodes and, finally, a dense layer that outputs the class predictions. Sequential capabilities of LSTMs fit well in our problem since each document in the data has different number of tokens and each of them is fed to LSTM in a single step. This rather simple model is able to perform good predictions thanks to our pre-trained word embeddings.

TABLE I. WORDS THAT ARE MOST SIMILAR TO "nezdinizdeki"

Word	Cosine Distance
nezdinizdekı	0.033
nezdinîzdeki	0.033
nezdînizdeki	0.035
nezdınizdeki	0.037
nezdinizdekî	0.038
nezrînizdeki	0.041
nezdınızdeki	0.043
nezinizdeki	0.046

C. Vision Model

We employed VGG-16 [7] network, which is a well-known architecture as the winner of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014 [27]. This network includes 5 convolutional blocks, which have 13 convolutional layers, and these layers are followed by three fully connected layers. VGG-16 model, pre-trained on ImageNet dataset which includes approximately 1 million images from 1000 classes, is utilized for transfer learning. Firstly, this pre-trained model is fine-tuned on the RVL-CDIP dataset [2] which consists of 320,000 train, 40,000 validation, and 40,000 test images from 16 classes: letter, memo, email, filefolder, form, handwritten, invoice, advertisement, budget, news article, presentation, scientific publication, questionnaire, resume, scientific report, and specification. Thus, the pre-trained model is utilized to understand document images. Finally, this model is fine-tuned on our dataset to classify 45 classes.

D. Multimodal Classification Models

Multimodal learning allows using different modalities, *e.g.*, text, vision, speech, sensor data, which are related to each other during learning [28]. Utilizing different modalities, called fusion, can take place in a different phase of learning. Correspondingly, fusion methods can be categorized into late fusion, early fusion, and hybrid fusion. Early fusion methods are based on feature learning for different modalities, while late fusion methods are defined as the decision-based.

Early Fusion. The proposed early fusion multimodal network learns embeddings jointly for vision and text modality. Our proposed network is demonstrated in Figure 4. Firstly, the output of the last convolutional layer of the VGG-16 model pre-trained on ImageNet was used to obtain $7 \times 7 \times 512$ dimensional visual features. After this convolutional layer, global average pooling layer is added to reduce the dimension of the visual features to 512. Weights of the vision model are fixed and fine-tuned during training. Also, the text model, which is explained in Section II-B, is trained from scratch by using FastText word embedding as an input. The output of the text model is 128 dimensional. Textual and visual features are concatenated and fed into three fully-connected layers to classify documents. These fully connected layers are shared during training.

Late Fusion. The late fusion method combines different model probabilities to capture two model results as a kind of decision mechanism. According to our preliminary analysis on results of text and vision models, while text model predicts the wrong label for some documents, the vision model can predict the correct label even if the vision model accuracy is lower than the text model accuracy on classification. Our late fusion network takes as an input the concatenated probabilities obtained from text and vision models. This network consists of two fully-connected layers for the training of the classification model. An overview of our late fusion network is illustrated in Figure 5. Since each model has 45 probability score, the input is 90 dimensional vector for this network.

Implementation Details. We implemented our models in Keras [29]. Training of all models was done over GTX 1080Ti GPU with the batch size 32. We performed between 40-50 epochs for all models. We used ADAM [30] optimizer with the learning rate in the range of 0.001 and 0.0001 and early

 ISBN: 978-1-61208-731-3

stopping on the validation dataset by controlling validation loss for specified consecutive epochs.

III. Experimental Results

A. Dataset

In the banking domain, customer orders have over 200 distinct process types, however, most of these processes are scarcely used. Therefore, we selected the most commonly used 45 distinct process types and limited our problem scope to these. We create a dataset with approximately 27k banking order documents labeled with 45 different classes. Sample images from several classes are shown in Figure 2. Each class has a variable number of instances that changes between 100 and 1000. We split the data by 70%, 15%, and 15% in order to create train, validation, and test sets, respectively. Since we require a different train set for training vision/text model and late fusion model, we split our main train by 75% train1 and 25% train2; similarly, we split the main validation by 75% validation1 and 25% validation2.

Challenges on the dataset.

The main challenge is that the documents were wrongly labeled by the back-office operator at the bank due to operational mistakes. The second significant challenge is that several classes have similar documents visually and textually. In addition, inter-class variation is quite common in the dataset. Moreover, some of the documents are filled with handwriting and such documents do not have textual information since the OCR tool does not support handwritten documents. In addition to the valid customer orders, irrelevant documents also reside in the data.

These irrelevant documents are:

- ID cards,
- driving licenses,
- property ownership documents (deeds),
- credit cards photocopies,
- registry newspapers,
- blank documents.

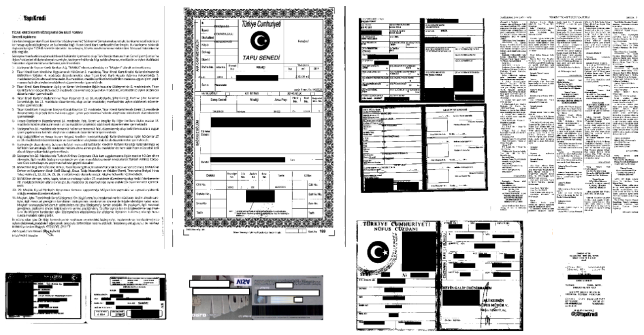

Figure 6. Sample images from the irrelevant documents.

Sample images of the irrelevant documents are illustrated in Figure 6. These irrelevant documents also have labels, similarly to the main documents. This situation causes confusion during training and testing. These dataset specific problems make our task harder than a document classification. Therefore, we considered our dataset a noisy dataset.

B. Results and Discussion

We evaluated our proposed methods on our Turkish banking dataset. We primarily investigated the effects of training the vision model on the early fusion multimodal network. To be able to do this, we employed two training strategies in the early fusion model. Firstly, ImageNet pre-trained model was utilized without updating weights, namely the fixed vision. Secondly, the last convolution block was fine-tuned in the model called fine-tuned vision. The text model was trained from scratch in both of them. As seen in Table II, fine-tuning vision model enhances the classification accuracy since the vision model adapts to the document images.

TABLE II. RESULTS ON EARLY FUSION

Learning joint embeddings	Accuracy (%)
fixed vision trained text from scratch	83.82
fine-tuned vision trained text from scratch	**85.29**

We have mainly four models to conduct experiments in order to classify banking documents: vision, text, early fusion and late fusion models, hence we have four main results. To compare all proposed models fairly, all experimental results are reported on the same test set. The results of all models are provided in Table III.

TABLE III. RESULTS ON ALL MODELS

Method	Accuracy (%)
Vision Model	70.53
Text Model	84.56
Early Fusion	**85.29**
Late Fusion	**85.42**

We also analyzed the performance of our multimodal networks by comparing with text and vision models. The highest improvement on the accuracy is observed when comparing the proposed multimodal networks with the vision model. However, we obtained a slightly better improvement when we compared the multimodal networks with the text model. This indicates that text embeddings are more beneficial than vision embedding for this task on these kinds of documents. Most of our documents in the dataset are free-formatted, thus this difference in accuracy between the vision and the text model is expected. On the other hand, multimodal networks achieve better results than the text model for class-based accuracies, especially when a class has visually rich documents, as expected. In addition, these improvements can be seen as fairly good results because of the constraints discussed in Section III-A. We observed that our fusion methods especially benefit from the text model since the text model and the multimodal model have approximately equal results. 1% improvement of the accuracy might seem unsatisfactory, but in a real-world scenario, where the prediction of each document is critical, such improvement diminishes the requirement of manpower.

 ISBN: 978-1-61208-731-3

IV. Conclusion

In this work, we proposed deep multimodal networks for document classification by using two fusion methods. The petition based customer order documents have different types of format, therefore, we focused to understand all types of documents by learning visual and textual features. We performed our experiments on our Turkish banking order dataset. The experimental results indicate that both early and late fusion multimodal models outperform text and vision models.

Acknowledgment

This work was supported by The Scientific and Technological Research Council of Turkey with the project no 3180571. We would like to thank our colleagues for their valuable discussions and support.

References

[1] EY Consulting LLC (UAE) and Microsoft, "Artificial Intelligence Maturity in Middle East Africa," Tech. Rep., 2019.

[2] A. W. Harley, A. Ufkes, and K. G. Derpanis, "Evaluation of deep convolutional nets for document image classification and retrieval," in 2015 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2015, pp. 991–995.

[3] C. Tensmeyer and T. Martinez, "Analysis of convolutional neural networks for document image classification," in 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1. IEEE, 2017, pp. 388–393.

[4] M. Z. Afzal, A. Kölsch, S. Ahmed, and M. Liwicki, "Cutting the error by half: Investigation of very deep cnn and advanced training strategies for document image classification," in 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1. IEEE, 2017, pp. 883–888.

[5] A. Das, S. Roy, U. Bhattacharya, and S. K. Parui, "Document image classification with intra-domain transfer learning and stacked generalization of deep convolutional neural networks," in 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018, pp. 3180–3185.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[8] C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[10] O. Russakovsky et al., "Imagenet large scale visual recognition challenge," International Journal of Computer Vision (IJCV), vol. 115, no. 3, 2015, pp. 211–252.

[11] Y. Kim, "Convolutional neural networks for sentence classification," Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1746–1751.

[12] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in Advances in Neural Information Processing Systems, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 649–657.

[13] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2015, pp. 1422–1432.

[14] Z. Yang et al., "Hierarchical attention networks for document classification," in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2016, pp. 1480–1489.

[15] N. Srivastava and R. R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," in Advances in Neural Information Processing Systems, 2012, pp. 2222–2230.

[16] D. Wang, K. Mao, and G.-W. Ng, "Convolutional neural networks and multimodal fusion for text aided image classification," in 2017 20th International Conference on Information Fusion (Fusion). IEEE, 2017, pp. 1–7.

[17] C. Xiong, S. Merity, and R. Socher, "Dynamic memory networks for visual and textual question answering," in International Conference on Machine Learning, 2016, pp. 2397–2406.

[18] S. Antol et al., "Vqa: Visual question answering," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2425–2433.

[19] R. Hu, J. Andreas, M. Rohrbach, T. Darrell, and K. Saenko, "Learning to reason: End-to-end module networks for visual question answering," in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 804–813.

[20] K. Xu et al., "Show, attend and tell: Neural image caption generation with visual attention," in International Conference on Machine Learning, 2015, pp. 2048–2057.

[21] P. Anderson et al., "Bottom-up and top-down attention for image captioning and visual question answering," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6077–6086.

[22] A. Brock, T. Lim, J. M. Ritchie, and N. Weston, "Neural photo editing with introspective adversarial networks," ICLR 2017, 2016.

[23] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in European Conference on Computer Vision. Springer, 2016, pp. 597–613.

[24] M. Rusiñol, V. Frinken, D. Karatzas, A. D. Bagdanov, and J. Lladós, "Multimodal page classification in administrative document image streams," International Journal on Document Analysis and Recognition (IJDAR), vol. 17, no. 4, 2014, pp. 331–341.

[25] Abbyy finereader. [Online]. Available: https://www.abbyy.com/en-us/finereader/

[26] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," Transactions of the Association for Computational Linguistics, vol. 5, 2017, pp. 135–146.

[27] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision (IJCV), vol. 115, no. 3, 2015, pp. 211–252.

[28] J. Ngiam et al., "Multimodal deep learning," in Proceedings of the 28th International Conference on Machine Learning (ICML), 2011, pp. 689–696.

[29] F. Chollet et al., "Keras," https://keras.io, 2015.

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," ICLR, 2015.

 ISBN: 978-1-61208-731-3

Fake News Detection Method Based on Text-Features

Ahlem Drif

Networks and Distributed Systems Laboratory
Faculty of Sciences
University of Sétif 1
Sétif, Algeria
Email: adrif@univ.setif.dz

Zineb Ferhat Hamida

Computer Science Department
University of Sétif 1
Sétif, Algeria
Email: zineb.ferhat@yahoo.com

Silvia Giordano

Networking Lab, SUPSI
University of Applied Sciences of Southern Switzerland
Lugano, Switzerland
Email: silvia.giordano@supsi.ch

***Abstract*—Feature extraction is a critical task in fake news detection. Embedding techniques, such as word embedding and deep neural networks, are attracting much attention for textual feature extraction, and have the potential to learn better representations. In this paper, we propose a joint Convolutional Neural Network model (CNN) and a Long Short Term Memory (LSTM) recurrent neural network architecture, taking advantage of the coarse-grained local features generated by CNN and long-distance dependencies learned via LSTM. An empirical evaluation of our model shows good prediction accuracy of fake news detection, when compared to Support Vector Machine and CNN baselines.**

Keywords–Fake news detection; social networks; deep learning; convolutional neural network; text classification; words embedding technique.

I. INTRODUCTION

Social media have pushed the ability to exchange information at a much greater pace, to a far wider audience than ever before. This information is not always truthful. Because anyone can publish anything on the Internet, the information obtained from this source can be inaccurate or even intentionally false (fake news) [1]. The term "fake news" became mainstream during the 2016 US presidential election campaign when hundreds of websites published falsified or heavily biased stories - many of them in the pursuit of capitalising on social media advertising revenue. The fake news term, popularised by the US President Donald Trump, is so prevalent now that it is hard to believe that just a few years ago the term was barely used. Besides, there are a variety of reasons for fake news and misinformation growing in levels, and rising in importance. These include how easy it is now to set up a website or even to manipulate a webpage to include the information desired, as well as how suited social media is for fake news broadcasting, combined with the rise of online social media. This work presents a comprehensive study on the features of different fake news datasets. To this extent, we implement methods based on both recent deep learning methods and machine learning methods to effectively detect fake news based on text features.

The rest of the paper is organized as follows: related literature survey is given in Section 2, Section 3 regroups fake news characteristics across different dimensions and summarizes some datasets features, the details of the proposed framework are introduced in Section 4, experimental results are presented in Section 5, and the study is concluded in Section 6.

II. RELATED WORK

The problem of fake news detection is more challenging than detecting deceptive reviews, since the political language on TV interviews, posts on Facebook and Twitters consists mostly short statements. The dissemination of fake news may cause large-scale negative effects, and sometimes can affect or even manipulate important public events. For example, within the final three months of the 2016 US presidential election, the fake news generated to favor either of the two nominees was believed by many people and was shared by more than 37 million times on Facebook [1]. There has been a large body of work surrounding features analysis of fake news. For example, Jin et al. [2] analyzed news articles' images for fake news detection based on multimedia datasets. They explored various visual and statistical image features to predict respective articles' veracity. Moreover, they proposed a fake news detection method utilizing the credibility propagation network built by exploiting conflicting viewpoints extracted from tweets. Yang et al. [3] proposed an efficient model for early detection of fake news through classifying news propagation paths using a multivariate time series. They realized a new deep learning model, which was comprised of four major components, i.e., propagation path construction and transformation, Recurrent Neural Network (RNN) based propagation path representation, CNN-based propagation path representation, and propagation path classification, which were integrated together to detect fake news at the early stage of its propagation.

 ISBN: 978-1-61208-731-3

Fake news detection based on surface-level linguistic patterns is also a popular trend in this area, such as building classifiers to detect whether tweets are factual or not. Ruchansky et al. [4] proposed an architecture of three components; the first module is a recurrent neural network to capture the temporal pattern of user activity on articles, and, the second module learns the source characteristic based on the behavior of users, and the two were integrated with the third module to classify an article as fake or not. Wang et al. [5] proposed an Event Adversarial Neural Network (EANN), which consists of three main components: the multi-modal feature extractor, the fake news detector, and the event discriminator. The multi-modal feature extractor is responsible for extracting the textual and visual features from posts. It cooperates with the fake news detector to learn the discriminable representation for the detection of fake news. Hardalov et al. [6] used a combination of linguistic, credibility and semantic features to differentiate between real and fake news. In their work, linguistic features include (weighted) n-grams and normalized number of unique words per article. Credibility features include capitalization, punctuation, pronoun usage and sentiment polarity features generated from lexicons. Text semantics were analyzed using embedding vectors method. All feature categories were tested independently and in combination based on self-created datasets. The best performance was achieved using all available features. In addition, Ma et al. [7] observed changes in linguistic properties of messages over the lifetime of a rumor using Support Vector Machine (SVM) based on time series features, then, they showed good results in the early detection of an emerging rumor. Moreover, Conroy et al. [8] illustrated that the best results for fake news detection could be achieved while combining linguistic and network features. Ciampaglia et al. [9] proposed to map the fact-checking task to the well-known task of finding the shortest path in a graph in order to utilize the information provided by knowledge networks. In that case, a shortest path indicates a higher probability of a truthful statement. Wang [10] [11] designed a hybrid Convolutional Neural Network (CNN or ConvNet) to integrate metadata with text. The best performance was achieved when incorporating different metadata features. Lendavi and Reichel [12] investigated contradictions in rumors sequences of micro-posts by analyzing posts at the text similarity level. The authors argue that vocabulary and token sequence overlap scores can be used to generate cues to veracity assessment, even for short and noisy texts. Joulin et al. [13] proposed a text classification model based on n-gram features, dimensionality reduction, and a fast approximation of the softmax classifier. This fast text classifier is built upon a product quantization method in order to minimizes the softmax loss l over N documents, therfore, it gives accurate results with less training and evaluation time [14]. For a full review of the state of the art in fake news detection in social media, see Zhou et al. [15].

In this work, we aim at building a new solution for addressing the detection of fake news based on the textual content of the news. For this reason, we realize a joint CNN-LSTM model, which can be defined by adding CNN layers in the front, followed by Long Short Term Memory (LSTM) layers with a dense layer on the output. Indeed, when analyzing fake news with such combination, the CNN acts like a trainable feature detector for the fake news content. It learns powerful convolutional features, which operate on a static spatial input. The LSTM, instead, receives a sequence of such high-level representations and generates a description of the content or maps it to some static class of outputs. We show that this combined approach works better than baselines approaches.

III. Exploring features extraction

The propagation of false information on social media is related to several factors, such as the information content and the users' behaviors. In this Section, in order to build a deep learning model that extracts discriminative characteristics of fake news, we study the most relevant attributes at the content level, user level, and social level [16]. Below, we elaborate on each level.

A. Content level

In order to capture the different aspects of fake news and real news, existing work relies on news content. Basically, the useful features that mostly are extracted from news content are linguistic-based and visual-based. Different kinds of linguistic features can be built: (i) lexical features, including character level and word-level features, such as total words, characters per word, frequency of large words, and unique words; (ii) syntactic features, including sentence level features, such as frequency of function words and phrases, i.e., n-grams and bag of words approaches [17], or Punctuation and Parts of Speech (POS) tagging. Also, visual-based characteristics have been shown to be an important manipulator for fake news propaganda [18]. As we have characterized, fake news exploits the individual vulnerabilities of people and thus often relies on sensational or even fake images (or fake videos) to provoke anger or other emotional response in the consumers.

B. User level

User based features represent the characteristics of those users who have interactions with the news on social media. These user level features are extracted to infer the credibility and reliability of each user using various aspects of user demographics, such as: registration age, number of followers and followees, number of tweets the user has authored, etc. [19]. Further, the users engagement in news dissemination process ranges from users response to a post up to spreading news pieces. Several works have observed that there are major psychological and cognitive factors that heavily increase the user engagement to fake news spreading:

- naive realism: consumers tend to believe that their perception of reality is the only accurate view, while others who disagree are regarded as uninformed, irrational, or biased [20].
- confirmation bias: consumers prefer to receive information that confirms their existing views [21].

Prospect theory describes decision making as a process by which people make choices based on the relative gains and losses as compared to their current state. This desire of maximizing the reward of a decision, to have social gains, can be modeled from an economic game theoretical perspective [22] by formulating the news generation and consumption cycle as a two-player strategy game. In fact, there are two kinds of key players in the information ecosystem: publisher and consumer. The utility for the publisher stems from two perspectives:

 ISBN: 978-1-61208-731-3

- short-term utility: the publisher's profit which is positively correlated with the number of consumers reached.
- long-term utility: the publisher's reputation in terms of news authenticity.

The utility for consumers consists of two parts:

- information utility: obtaining true and unbiased information.
- psychology utility: receiving news that satisfies their prior opinions and social needs, e.g., confirmation bias and prospect theory.

Both publisher and consumer try to maximize their overall utilities in the strategy game that is the news consumption process.

C. Social level

Social dimensions refer to the heterogeneity and weak dependency of social connections within different social communities. Users' perceptions of fake news pieces are highly affected by their like minded friends on social media, while the degree differs along different social dimensions. Thus, it is worth exploring why and how different social dimensions play a role in spreading fake news. Recent findings [23] showed that users on Facebook tend to select information that adhere to their system of beliefs and to form polarized groups, i.e., echo chambers. For example, users on Facebook always follow like-minded people and thus receive news that promote their favored existing narratives. The echo chamber effect facilitates the process by which people consume and believe fake news due to the following psychological factors:

- Social credibility, which means that people are more likely to perceive a source as credible if others perceive the source as credible, especially when there is not enough information available to access the truthfulness of the source.
- Frequency heuristic, which means that consumers may naturally favor information they hear frequently, even if it is fake news. Del Vicario et al. [24] showed that social homogeneity is the primary driver of content diffusion, and one frequent result is the formation of homogeneous, polarized clusters. Most of the time, the information is taken by a friend having the same profile (polarization), i.e., belonging to the same echo-chamber.

In Table I, we categorize the methods discussed in Section II, based on the features of the fake information analyzed. The majority of fake news detection algorithms are content feature based, in that they rely on developing efficient features of news content that individually or jointly are able to distinguish between real and fake information.

IV. Model Construction

In this work, we combined a CNN and a LSTM, which is a type of Recurrent Neural Network. Figure 1 shows the overview of our combined CNN-LSTM neural network for fake news detection. In fact, there are many interesting properties that one can get from combining convolutional neural networks and LSTM network, as we will discuss later in this work.

We build our CNN-LSTM deep neural networks model as follows: the embedding layer is the first layer in the model and it represents each statement (text) as a row of vectors. Each vector represents a token based on the word-level used. Each word in the statement, which is one token in the word level, is embedded into a vector with length of 300. This layer is a matrix of size $w \times v$, where v is the length of the vector and w is the number of tokens in the statements. The value of w is the maximum length of a statement. Any statement that contains less than the maximum number of tokens in the statement will be padded with $<$ Pad$>$ to have the same length as the maximum statement length. Each matrix in the word level has the size of 50 $\times$ 300. The vocabulary size is 10,000. We used a pre-trained 300-dimensional Google News Vectors method (GloVe) [25] of learning word embeddings from text. After that, we added a drop-out layers [26] to reduce overfitting and set the drop-out probability to 0.2 when training. The output is fed to the next layer. Then, we added a Convolutional Neural Networks layer that extract features from local input patches. We used 10 filters with size 3 to extract features of words from statement. Each filter detects multiple features in the text using ReLu [27] activation function in order to represent them in the feature map. Then, a standard max pooling operation is performed on the latent space, followed by a LSTM layer. The forward one-dimensional (1D) max pooling layer is a form of non-linear down sampling of an input tensor $X \in \mathbb{R}^{n_1 \times n_2 \times \ldots \times n_p}$. 1D max pooling partitions the input tensor data into 1D subtensors along the dimension k, selects an element with the maximal numeric value in each subtensor, and transforms the input tensor to the output tensor Y by replacing each subtensor with its maximum element. The max operation or function is the most commonly used technique for this layer and it is used in our experiments. The reason for selecting the highest value is to capture the most important feature and reduce the computation in the advanced layers. The LSTM layer has a set number of units and the input of each cell is the output from the previous max pooling layer. In fact, the output vectors of the max pooling layer become inputs to the LSTM networks to measure the long-term dependencies of feature sequences. The outputs from LSTMs are merged and then passed to a fully connected layer. We need the expressive power of two fully connected layers. The last dense layer converts the array into a single output in the range $\{0, 1\}$. Thus, the sigmoid function is used [28].

For comparison, we used two baselines: a Support Vector Machine classifier (SVM) [29], and a Convolutional Neural Network model [30]. For SVM, we used Scikit-learn library which provides very strong performances on short text classification problems. For CNN, we used TensorFlow [31] for the implementation. The CNN baseline model is obtained as follows: we performed unsupervised learning of word-level embeddings. Then, we added a drop-out layer with probability equal to 0.2. The output of the drop-out layer is fed to a convolution layer ConvNet1D with 10 filters with size 3 and the activation function ReLu. Then, a standard max pooling operation is performed followed by a 1D global average pooling layer. The average pooling layer output is passed to a fully connected layer followed by a drop-out layer with 0.6 drop-out probability. We added a fully connected layer to trade network-depth for increasing the chances to get a

 ISBN: 978-1-61208-731-3

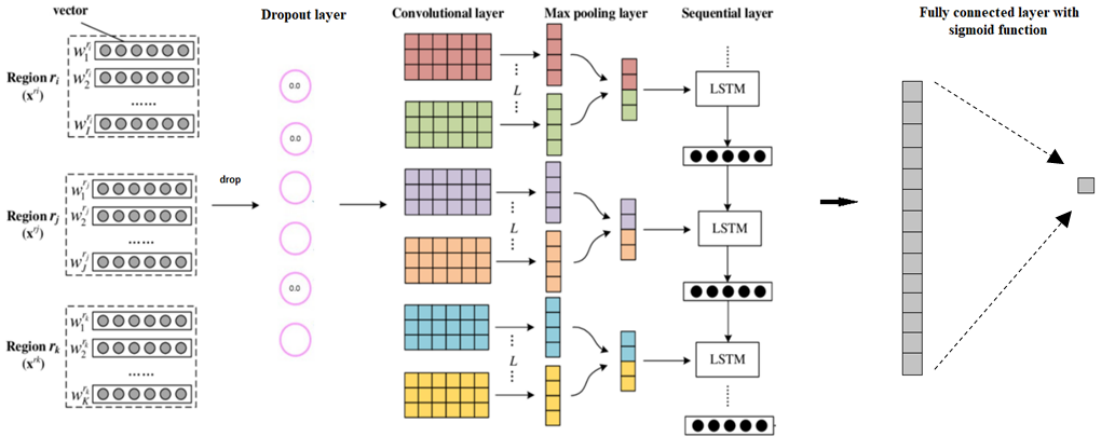

Figure 1. The architecture of the proposed fake news detection model

better learned layer. The sigmoid activation function is used to generate the final classification. For this CNN baseline, we have added two drop-out layer in order to improve the accuracy. This choice was made empirically.

V. Experimental settings and results

A. Dataset pre-processing

To fairly evaluate the performance of the proposed model, we conduct the experiments on two real social media datasets: Liar dataset [10] and News Articles dataset [32]. These two datasets contain a rich metadata that would help to descriminate text-features.

The Liar dataset [10] is collected from the fact-checking website PolitiFact through its API [33]. The website PolitiFact.COM focused on looking at specific statements made by politicians and rating them for accuracy. The Liar dataset includes a rich set of metadata for each speaker: statement, party affiliations, current job, home state, as well as historical counts of inaccurate statements. These various metadata can be granular enough to define features at the content level. The Liar dataset comprises 12,836 short statements labeled for truthfulness, subject, context/venue, speaker, state, party, and prior history, as illustrated in Figure 2. This dataset considers six fine-grained labels for the truthfulness ratings: pants-fire, false, barely-true, half-true, mostly-true, and true. In our work, we analyze the correlation between these labels. Figure 3 shows that, from the perspective of the description space, some labels might well be just correlated noise. For this reason, we merge the mostly-true and the half-true labels into the true label, and merge the barely-true and the pants-fire labels into the false label. The association between labels may give birth to a better classification standard.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	11972.json	true	Building a wall on the U.S.-Mexico border will...	immigration	rick-perry	Governor	Texas	republican	30	30	42	23	18	Radio interview
1	11685.json	false	Wisconsin is on pace to double the number of l...	jobs	katrina-shankland	State representative	Wisconsin	democrat	2	1	0	0	0	a news conference
2	11096.json	false	Says John McCain has done nothing to help the ...	military,veterans,voting-record	donald-trump	President-Elect	New York	republican	63	114	51	37	61	comments on ABC's This Week.
3	5209.json	half-true	Suzanne Bonamici supports a plan that will cut...	medicare,message-machine-2012,campaign-adverti...	rob-cornilles	consultant	Oregon	republican	1	1	3	1	1	a radio show
4	9524.json	pants-...	When asked by a reporter whether hes at ...	campaign-finance,legal-...	state-democratic-...	NaN	Wisconsin	democrat	5	7	2	2	7	a web ...

Figure 2. Liar dataset attributes

The News Articles dataset [32] comprises 20,800 stories labeled as unreliable or reliable, as shown in Figure 4. The News Articles dataset contains text, author, and title.

In the pre-processing phase, we have dropped the rows with missing values. Also, we have removed from each text the punctuations marks and the stop-words, which represent the most common words in a language, such as ”are”, ”as”, ”the”, etc. In addition, we have applied a stemming process to cut off the end or the beginning of the word, taking into account a list of common prefixes and suffixes that can be found in an inflected word. For example, for News Articles dataset , after

TABLE I. COMPARISON OF FEATURES-BASED FAKE NEWS DETECTION METHODS.

Methods	Content level		User level		Social level	
	Linguistic	Visual	User profile	Credibility features	Diffusion network	Freindship network
Ma et al. (2015) [7]	✓					
Conory et al. (2015) [8]	✓				✓	
Ciampaglia et al. (2015)[9]	✓				✓	
Lendavi et al. (2016)[12]	✓					
Hardalov et al. (2016)[6]	✓		✓			
Julian et al. (2016)[13]	✓					
Wang 2016 [5]	✓					
Jin et al. (2017)[2]		✓		✓		
Ruchansky et al. (2017)	✓		✓			
Wang et al. (2018)	✓	✓				
Yang et al. (2018)			✓	✓	✓	

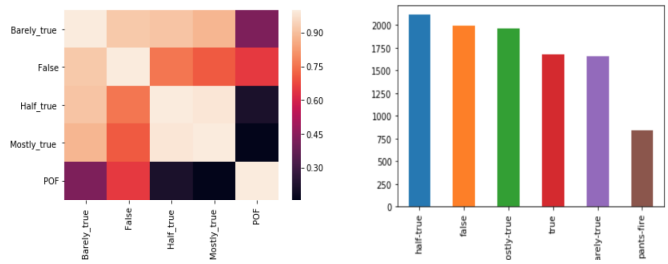

Figure 3. a) Correlation betwen labels of Liar dataset b) The Liar dataset labels

dropping the rows with missing labels or with an empty text, we have obtained 7,924 real labels and 10,361 fake labels. After that, we have applied a tokenization technique which is the process of splitting the given text into smaller pieces called tokens (words, numbers and others can be considered as tokens). Finally, we have created sequences with a vocabulary size of 10,000 for the Liar datset and 50,000 for the News Articles dataset. We have used a padding to obtain equally sized sequences.

	id	title	author	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1

Figure 4. News Articles dataset attributes

B. CNN-LSTM Implementation

For the experiment, we needed to separate training and testing sets. We have randomly split the dataset into approximately 80% training set and 20% testing set. In order to fine-tune the model hyperparameters, we needed a validation dataset; therefore, we split again the training dataset into 70% training set and 10% validation set. Table II shows the corpus statistics.

TABLE II. DATASETS STATISTICS.

Dataset Statistics	Liar	News articles
Training set size	10,240	11,703
Validation set size	1,284	2,925
Testing set size	1,267	3,657
Real label	7,134	7,924
Fake label	5,657	10,361

We implement our CNN-LSTM framework in Keras [34],following a pattern composed of 7 layers as described in Section IV. We train the network for 400 epochs with a batch size equal to 64 (the number of training examples utilized in one iteration) using Stochastic Gradient Descent (SGD) as optimization for loss function, employing the ReLu as activation function at convolution layer and the sigmoid as activation function at the output layer. We tune these hyperparameters on a validation set (10 % of the data). Table III shows a summury of the proposed CNN-LSTM model.

TABLE III. MODEL SUMMARY

Layer	Input shape	Output shape
Embedding	(None, 50)	(None, 50, 300)
drop-out	(None, 50, 300)	(None, 50, 300)
Conv1D	(None, 50, 300)	(None, 48, 10)
Max Pooling	(None, 48, 10)	(None, 24, 10)
LSTM	(None, 24, 10)	(None, 30)
Dense	(None, 30)	(None, 64)
Output layer: Dense	(None, 64)	(None, 1)

TABLE IV. THE RESULTS OF DIFFERENT METHODS ON TWO DATASETS.

Dataset	Method	Accuracy	Precision	Recall
Liar dataset	SVM	0.608	0.603	0.608
	CNN	0.614	0.611	0.614
	CNN-LSTM	**0.623**	**0.620**	**0.623**
News Articles dataset	SVM	0.683	0.680	0.683
	CNN	0.708	0.701	0.708
	CNN-LSTM	**0.725**	**0.721**	**0.725**

Table IV shows the experimental results of baselines and the proposed approaches on two datasets. We can observe that the overall performance of the proposed CNN-LSTM is much better than the baselines in terms of accuracy, precision and recall on both datasets. On the Liar dataset, the CNN-LSTM outperformed all models, resulting in an accuracy of 62.34%. On News Articles dataset, the highest value 72.50% of accuracy shows that we can well describe fake news content using such CNN-LSTM pair. Therefore, it is more efficient to apply our model on a large dataset in order to improve the fake news detection as opposed to a small datasets. Furthermore, since we have found that the CNN-LSTM model based on text-features discriminates the truthfulness of fake news , we are going to incorporate various metadata in our framework deep learning model. This could help to improve the accuracy of the fake news detection results

VI. CONCLUSION

In this work, we study the problem of fake news detection. We focus on fake news detection methods based on text-features. We propose a hybrid CNN-LSTM model as a combination of a convolution layer, used to extract unlabeled features, and a LSTM layer used to capture long-term dependencies between the sequences in order to learn a regulatory grammar to improve predictions. Experiments on two real-world datasets demonstrate the heigh accuracy of the CNN-LSTM model in classifying fake news.

The achieved results open several interesting directions for future work. First of all, we believe that fake news detection performance can be further improved. For this reason, we are studying the advantage of using all the other metadata (statement, author, title, and subject) for fake news detection. Second, to better understand the fake news detection characteristics and how to better use deep learning for that, more thorough experiments are required in the future and will be conducted on different datasets. Finally, we aim to understand the correlation between data diffusion, influence [35] and fake news, and we started designing a scenario for studying this aspect.

 ISBN: 978-1-61208-731-3

REFERENCES

[1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of Economic Perspectives, vol. 31, no. 2, 2017, pp. 211–36.

[2] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," IEEE transactions on multimedia, vol. 19, no. 3, 2016, pp. 598–608.

[3] Y. Liu and Y.-F. B. Wu, "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks," in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[4] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, 2017, pp. 797–806.

[5] W. Yaqing et al, "Eann: Event adversarial neural networks for multi-modal fake news detection," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018, pp. 849–857.

[6] M. Hardalov, I. Koychev, and P. Nakov, "In search of credible news," in International Conference on Artificial Intelligence: Methodology, Systems, and Applications. Springer, 2016, pp. 172–180.

[7] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong, "Detect rumors using time series of social context information on microblogging websites," in Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, 2015, pp. 1751–1754.

[8] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, 2015, pp. 1–4.

[9] G. L. Ciampaglia et al, "Computational fact checking from knowledge networks," PloS one, vol. 10, no. 6, 2015, p. e0128193.

[10] W. Y. Wang, LIAR, 2017, https://www.cs.ucsb.edu/william/data/liardataset.zip, Last access: June 12, 2019.

[11] W. Ferreira and A. Vlachos, "Emergent: a novel data-set for stance classification," in Proceedings of the conference of the North American chapter of the association for computational linguistics: Human language technologies, 2016, pp. 1163–1168.

[12] P. Lendvai and U. D. Reichel, "Contradiction detection for rumorous claims," arXiv preprint arXiv:1611.02588, 2016.

[13] A. Joulin et al, "Fasttext. zip: Compressing text classification models," arXiv preprint arXiv:1612.03651, 2016.

[14] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," arXiv preprint arXiv:1607.01759, 2016.

[15] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," arXiv preprint arXiv:1812.00315, 2018.

[16] K. Garg, V. Arnaboldi, and S. Giordano, "A novel approach to predict retweets and replies based on privacy and complexity-aware feature planes," in Proceedings of the 5th International Workshop on Complex Networks and their Applications, 2016., 2016, pp. 459–471.

[17] J. Fürnkranz, "A study using n-gram features for text categorization," Austrian Research Institute for Artifical Intelligence, vol. 3, 1998, pp. 1–10.

[18] A. Gupta et al, "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy," in Proceedings of the 22nd international conference on World Wide Web. ACM, 2013, pp. 729–736.

[19] C. Castillo, M. Mendoza, and B. Poblete, "Predicting information credibility in time-sensitive social media," Internet Research, vol. 23, no. 5, 2013, pp. 560–588.

[20] E. S. Reed, E. Turiel, and T. Brown, "Naive realism in everyday life: Implications for social conflict and misunderstanding," in Values and knowledge. Psychology Press, 2013, pp. 113–146.

[21] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises." Review of general psychology, vol. 2, no. 2, 1998, p. 175.

[22] M. Gentzkow, J. M. Shapiro, and D. F. Stone, "Media bias in the marketplace: Theory," in Handbook of media economics. Elsevier, 2015, vol. 1, pp. 623–645.

[23] W. Quattrociocchi, A. Scala, and C. R. Sunstein, "Echo chambers on facebook," Available at SSRN 2795110, 2016.

[24] M. Del Vicario et al, "The spreading of misinformation online," Proceedings of the National Academy of Sciences, vol. 113, no. 3, 2016, pp. 554–559.

[25] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.

[26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research, vol. 15, no. 1, 2014, pp. 1929–1958.

[27] A. Kulkarni and A. Shivananda, "Natural language processing recipes: Unlocking text data with machine learning and deep learning using python," 2019.

[28] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in International Workshop on Artificial Neural Networks. Springer, 1995, pp. 195–201.

[29] F. Pedregosa et al, "Scikit-learn: Machine learning in python," Journal of machine learning research, vol. 12, no. Oct, 2011, pp. 2825–2830.

[30] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, 2014.

[31] M. Abadi, M. Isard, and D. G. Murray, "A computational model for tensorflow: an introduction," in Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Languages. ACM, 2017, pp. 1–7.

[32] Fake-news, 2015, https://www.kaggle.com/c/fake-news/data, Last access: June 12, 2019.

[33] W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," arXiv preprint arXiv:1705.00648, 2017.

[34] F. Chollet, "Keras: Theano-based deep learning library, 2015," URL http://keras. io, Last access: June 12, 2019.

[35] L. Luceri, A. Vancheri, T. Braun, and S. Giordano, "On the social influence in human behavior: Physical, homophily, and social communities," in Proceedings of the Sixth International Conference on Complex Networks and Their Applications., 2017, pp. 856–868.

 ISBN: 978-1-61208-731-3

A Novel Feature Selection Method Based on Clustering

Jonathan A. Mata-Torres
Reynosa-RODHE Multidisciplinary Unit
Autonomous University of Tamaulipas
Reynosa, Tamaulipas, México
e-mail: a2093010058@alumnos.uat.edu.mx

Edgar Tello-Leal
Faculty of Engineer and Science
Autonomous University of Tamaulipas
Victoria, Tamaulipas, México
e-mail: etello@uat.edu.mx

Uluses M. Ramirez-Alcocer
Faculty of Engineer and Science
Autonomous University of Tamaulipas
Victoria, Tamaulipas, México
e-mail: a2093010066@alumnos.uat.edu.mx

Gerardo Romero-Galván
Reynosa-RODHE Multidisciplinary Unit
Autonomous University of Tamaulipas
Victoria, Tamaulipas, México
e-mail: gromero@docentes.uat.edu.mx

Abstract— **Nowadays, there is a great interest from academia, the industry, and the government to find potentially useful information to build a prediction model from data with high dimensionality, which has become one of the most important challenges in data mining and machine learning approaches. In this way, feature selection is the process of selecting the most useful features for building models in tasks like classification, regression or clustering, in order to reduce the dimensionality and facilitating the visualization and understanding of the data. In this paper, we propose a feature selection method based on the mean shift clustering algorithm and the Pearson correlation coefficient to contribute to solving some of the challenges in the data analytics systems, of real-time execution. Furthermore, we compare the mean shift method with the renowned Recursive Feature Elimination (RFE) method, as well as with the feature selection method designed by a human expert in the domain. Finally, the subsets of data generated with the attributes selected by the methods are evaluated by the J48 classification algorithm based on a decision tree, using a historical public safety data set. The clustering method proposed has a great advantage over the other methods in the computing time required to recommend a group of selected attributes.**

Keywords–feature selection; mean shift; clustering; data mining; J48.

I. INTRODUCTION

Nowadays, the trend in the administration of resources, infrastructure, and services in cities is increasingly based on their ability to make decisions using knowledge bases, as well as their potential to anchor external knowledge and the implementation of knowledge-based strategies, in order to provide a better quality of life to the citizens and visitors. This way, the concept of smart cities emerged, in which a smart city can understand how an urban environment is capable of offering advanced and innovative services for citizens in order to improve the quality of life in general by using widespread support of systems (system of systems) based on Information and Communication Technologies (ICT) [1]-[3]. ICT software applications and the intensive use of digital devices such as sensors, actuators, and mobiles are essential means for realizing smartness in any of smart city domains [4]. A concept closely related to smart cities is the Internet of Things (IoT) [5], representing an extension of the Internet with a large number of objects (physical or virtual things) with pervasive sensing, detection, actuation, and computational capabilities allowing these devices to generate, exchange, and consume data with minimal human intervention [6][7]. In smart cities, specific areas of application have been identified through smart systems, e.g., transportation, public safety, sustainability, healthcare, energy, transportation and mobility, environment, education, and governance [8][9].

The automation of a large number of business processes and transactions that run on inter-organizational information systems within smart cities, embedded systems, smart systems based on IoT technology, as well as the intensive use of social networks through smartphones and software systems that use cloud computing technology, have caused the generation of massive volumes of data (known as big data), of different types: structured, semi-structured or unstructured [10][11]. The knowledge extraction and the hidden correlations of big data is a growing trend in information systems to provide better services to citizens and support decision-making processes [12].

There is a great interest from academia, the industry, and government for the development and deployment of big data analysis applications, both for general use and specific use in smart cities, which face different challenges. Hence, finding potentially useful information to build a prediction model from data with high dimensionality has become one of the most important challenges of data extraction and knowledge discovery [13][14]. One of the effects of the high dimensionality in the data sets can cause prediction models with a low precision measure. In addition, these is a high computational cost associated with processing of a big

 ISBN: 978-1-61208-731-3

volume of data to predict an event [15]. To solve problems of high dimensionality in data sets (dimensionality reduction), approaches based on the feature selection and feature extraction methods have been proposed.

Feature selection methods are used in data mining and machine learning, commonly in the pre-processing stages, and include both supervised and unsupervised techniques [16]. A feature is an individual measurable property of the instance being observed, and, through a set of attributes, a data mining or machine learning algorithm can perform data classification or clustering [17]. The feature selection approaches aim to select a small subset of features that minimize redundancy and maximize relevance to the target such as the class labels in classification [18]. In other words, the feature selection consists of selecting a subset of features representative of the original data set, but that can efficiently describe the input data.

The feature selection algorithms can be classified into the following categories: filter, wrapper, and hybrid methods [15]. The filter methods select the most relevant features using variable ranking techniques as the principle criteria for attribute selection. In filter methods, the features weights are individually calculated based on some criteria (e.g., correlation coefficient), the attributes that satisfy these conditions are considered as selected features and the remaining ones are removed from the subset [19]. Furthermore, the wrapper methods use a predictor as a black box and the predictor performance as the objective function to evaluate the variable subset [17]. In wrapper methods, several search algorithms can be used to find a subset of variables which maximizes the objective function which is the classification performance. That is, in wrapper method uses the information of the classifier to find the best feature subset, usually by performing computationally expensive searches on the feature space. Additionally, the hybrid methods try to exploit the best functionalities of the filters and wrappers approach, trying to reduce the computational cost but maintaining the effectiveness in the objective task associated with using the selected functions [20].

In this paper, we propose a feature selection method based on the mean shift clustering algorithm in combination with the Pearson correlation measure, allowing to identify a subset of relevant and non-redundant attributes. In addition, we compare the mean shift method with the renowned Recursive Feature Elimination (RFE) method [21], as well as with a feature selection method designed by a human expert in the crime domain. Finally, the methods are evaluated through their implementation in a classification algorithm based on J48 decision tree. In the proposal, a data set of crime incidents from the last 17 years is used, where their records are collected by a set of software systems implemented in a smart city. The method based on the mean shift clustering algorithm has a great advantage over the other methods in the processing time required to recommend a group of attributes selected.

The rest of the paper is organized as follows. The feature selection algorithm is the subject of Section 2. The experimental study and results are covered in Section 3. The conclusions and future research are presented in Section 4.

II. CLUSTERING FEATURE SELECTION METHOD

The proposed feature selection algorithm integrates the concepts of the mean shift clustering algorithm and the Pearson correlation analysis. The former is an unsupervised learning algorithm that clusters the data based on its natural distribution. This algorithm is characterized by not requiring prior knowledge of the number and location of the centroids. In the other hand, the Pearson correlation measures the statistical relationship between two variables, specifically the dependence of one variable on another variable. Therefore, in a statistical correlation, the two variables that are correlated are dependent on each other and one may be used to predict the other. The mean shift algorithm is a statistical clustering method based on non-parametric kernel density estimation, which is expressed by (1). Given n data points x_i, $i = 1,\ldots, n$ in the d-dimensional space R^d, the multivariate kernel density estimator with kernel $K(x)$ and a symmetric positive definite $d \times d$ bandwidth matrix $\mathbf{H}$, computed in the point x is given by [22][23]:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_H(x - x_i), \qquad (1)$$

In Fig. 1 shows the operation of the method. The input data set is represented by a numerically encoded matrix. This data set turn into a data set or transposed matrix, i.e., let A be a matrix of dimension $m \times n$, we denote the element of row i and column j as $A(i, j)$, where $i < m$ and $j < n$. Then, the transposed matrix of A is defined as the matrix A^T of dimension $n \times m$ such that $A^T(j, i) = A(i, j)$, where $i < m$ and $j < n$. Next, the mean shift algorithm initializes a window on all data points; with the first data point, its distance from all data points is calculated. The data points are used to find a new mean $m(x)$ of the window according to the kernel $k(x)$ [24]. The iterations continue until the mean of a window becomes fixed. Then, the algorithm will move on to the second data point and repeat the same procedure. The iterations will continue until the system converges.

The mean shift algorithm generates a list with the set of clusters, which contains all the data set objects. The elements of the list are compared to each other to determine if they belong to the same clusters. If that is the case, the Pearson correlation coefficient between the two objects (i,j) is calculated by (2), using the original data set.

$$R(i) = \frac{cov(x_i, Y)}{\sqrt{var(x_i) * var(Y)}}, \qquad (2)$$

where x_i is the i^{th} variable, Y is the output (class labels), $cov()$ is the covariance and $var()$ the variance. Correlation

ranking can only detect linear dependencies between variable and target.

The calculation of the Pearson correlation generates a correlation matrix, to which a threshold (filter) of $\geq +0.5$ and -0.5 is applied, which allows us to remove the attributes with a value below the threshold, that is, with low linear correlation, allowing to select the representative attributes of the data set.

Algorithm 1 Feature Selection Method

Input: Dataset: Numeric encode Matrix
Output: FeatureSelected: list
MSHIFTP(Dataset) :

```
Clusters ← MeanShift(Dataset^T)
for i in range(0, len(Clusters)) do
    for j in range(0, len(Clusters)) do
        if Clusters(i) == Clusters(j) then
            PearsonL.Add(Pearson(Dataset[i], Dataset[j
        end if
    end for
end for
for x in range(0, len(PearsonL)) do
    if PearsonL[x] ≤ −0.5 or PearsonL[x] ≥ 0.5 then
        FeatureSelected.Add(PearsonL[x])
    else
        Discard()
    end if
end for
return FeatureSelected
```

Figure 1. Feature selection algorithm.

III. EXPERIMENTAL STUDY

In this section, we compare the performance of our proposed clustering feature selection method with the RFE method and with a human expert method. A crime incidents data set of a smart city is used in our experiment. In this data set, crime data has been collected through a set of information systems and IoT technologies. The incidents of crime are from the City of Chicago, in the period from 2001 to 2017, consisting of a total of 6.4 million records and 22 attributes [25].

Table I shows a description of the attributes contained in the data set. These attributes can be values of data type: string, numeric, date, location or Boolean. Further, the total number of cases or values contained by attribute are shown. The ID, Case Number, and Date attributes are not used in the execution of the attribute selection method, and the arrest attribute represents the class label of the data set, of Boolean type.

A. Feature Selection Results

In the feature selection mean shift-based method, the total number of records contained in the data set (6.4 million) was used to make the recommendation of the attributes to be selected. This method selects 9 attributes (UICR, FBI Code, Y coordinate, Latitude, Location, Beat, District, Ward, and Community Area) from a total of 17.

TABLE I. DESCRIPTION OF THE ATRIBUTES CONTAINED IN THE DATA SET

Feature	*Description*	*No cases*
ID	Unique identifier for the record	6,457,411
Case Number	Unique identifier of the incident assigned by the Chicago policy department.	6,457,411
Date	Date when the incident occurred.	2,740,512
Block	Extract of the address where the incident occurred.	60.144
IUCR	Codes used to classify criminal incidents by law enforcement agencies.	350
Primary Type	The primary type description of the IUCR code.	35
Description	The secondary description of the IUCR code	380
Location Description	Description of the location where the incident occurred.	180
Domestic	Indicates whether the incident is related to violence domestic.	2
Beat	Indicates the police district where the incident occurred.	25
Ward	The ward (City council district) where the incident occurred.	50
Community Area	Indicates the community area where the incident occurred	77
FBI Code	Indicates the crime classification based in the FBI system	26
X coordinate	The X coordinate of the location where the incident occurred in the state of Illinois.	78,528
Y coordinate	The Y coordinate of the location where the incident occurred in the state of Illinois.	129,825
Year	Year when the incident happened	17
Update On	Date and time when the record was updated.	2,593
Latitude	The latitude of the location where the incident took place.	861,599
Longitude	The longitude of the location where the incident happened	861,046
Location	This attribute is formed with the data of latitude and longitude attributes.	862,781
Arrest	A binary variable that indicates whether a criminal was arrested.	2

 ISBN: 978-1-61208-731-3

Table II shows the attributes selected by the method. The order of occurrence corresponds to the existing correlation between them, according to the approach presented in the previous section. The proposed mean shift method presents as a relevant characteristic a required computation time of 148.95 seconds (see Table II), to select the attributes. This time consists of 17.63 seconds for loading the data set, 107.29 seconds for the creation of clustering, and 24.03 seconds to execute the correlation, allowing the selected attributes to be displayed in minimum processing time.

TABLE II. COMPARISON OF THE ATTRIBUTES SELECTED BY THE METHODS

Method	*Selected Features*	*Time*
Mean Shift	1,10,12,15,17,6,7,8,9	148.95s
RFE	1,4,10,11,12,16,17	2096.49s
Human Expert	1,2,4,5,7,8,9,10	0.00s

In the RFE-based method, 80% of the data set is used for the training of the algorithm, and the remaining 20% of instances of the data set is used for the validation phase required by the method. This method requires 2069.49 seconds to recommend the attributes to be selected (see Table II). The RFE method selects 7 attributes (UICR, Location Description, FBI Code, X coordinate, Y coordinate, Longitude, and Location) from a total of 17.

Additionally, a group of experts was consulted (we call this a human expert-based method), composed of business analysts (employees of the police and criminalistics department) and software engineers who manage public safety systems. The human expert method required 25 hours to analyze the attributes and values of the data set, proposing the following 8 attributes: UICR, Primary Type, Location Description, Domestic, District, Ward, Community Area, and FBI Code.

The mean shift method coincides with the RFE method in 4 proposed attributes to be selected (UICR, FBI Code, Y coordinate, and Location), but only coincide in the position of occurrence of one attribute (UICR), in the list of attributes selected by the methods. On the other hand, the mean shift method coincides with the human expert method in 5 attributes (UICR, District, Ward, Community Area, and FBI Code), and the RFE method agrees with the human expert method only in 3 selected attributes (UICR, Location Description, and FBI Code). The three methods recommend selecting as a first attribute the UICR code, but the order of the rest of the concordant attributes among the methods does not match.

B. J48 Algorithm Results

The J48 decision tree algorithm is used to evaluate and compare the performance of the proposed feature selection algorithm with the RFE method and the human expert method, in terms of predictive accuracy.

The instances of the data subsets used in the experiment were selected and extracted by a random method, automatically, from the original data set. In our experiment a 60-20-20 approach was applied, that is, 60% of the observations were used to train our model, 20% of the instances were used for the test phase of the model and the remaining 20% of records in the data set were used for the validation of the class label prediction model.

We formed the reduced data sets (sub data set) containing those features selected by different feature selection methods applied to the full experimental data set. Then, we trained, tested, and evaluated the J48 classifier on the reduced data sets. The obtained classification accuracies are shown in Table III.

TABLE III. THE ACCURRACY OBTAINED BY J48 ALGORITHM FOR EACH ATTRIBUTE SELECTION METHOD

Algorithm	*Testing Accuracy*	*Evaluation Accuracy*
Mean Shift	0.886173	0.886952
RFE	0.887452	0.887816
Human Expert	0.886638	0.887259

It can be observed that the classifier trained on the mean shift data set tends to exhibit slightly lower classification accuracy in testing phase (0.886173). The classifier trained on the data set containing features selected by the RFE method constantly performed better than the classifier trained with mean shift and human expert data sets, both in the testing phase and evaluation phase.

The next important result that can be observed in Table III is that the mean shift method exhibits an improvement classification performance in the evaluation phase (0.886952), compared to the accuracy achieved in testing phase. Additionally, the mean shift method reduces the distance with the precision obtained by the other two methods.

IV. CONCLUSIONS

Feature selection provides an effective way to solve the dimensionality problem by removing irrelevant and redundant data, which can reduce computation time, improve learning accuracy, and facilitate a better understanding of the learning model or data.

The mean shift method proposed allows obtaining the necessary features without human intervention, because the clustering is carried out automatically without the need to define a K number a priori. The selection of the most representative features is made with the support of Pearson's linear correlation, pre-defining a threshold of +-0.5, allowing to discard irrelevant attributes.

 ISBN: 978-1-61208-731-3

On the other hand, in the RFE method, it is necessary to define initially how many attributes we want the algorithm to select. In addition, since it is a wrapper-type method, it depends entirely on the learning algorithm with which it was trained.

In our experiment, it is observed that the computation time required by the RFE method is very high compared to the processing time required by the mean shift method. This method only needs a 7.19% of the time of the RFE method to determine the attributes to be selected. Therefore, we consider that in data mining and automatic learning tools, with real-time execution, it is feasible to use the proposed mean shift method, because the computation time required to select the attributes by mean shift method is better than RFE and human expert methods.

ACKNOWLEDGMENT

This work was supported by the National Council of Science and Technology (CONACYT) and Basic Science Research Project SEP-CONACYT of Mexico under Grant 256922.

REFERENCES

[1] C. Harrison, et al., "Foundations for smarter cities," IBM Journal of Research and Development, vol. 54, no. 4, July 2010, pp. 1–16.

[2] H. Schaffers, et al., "Smart cities and the future internet: Towards cooperation frameworks for open innovation," in The Future Internet, J. Domingue, et al., Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 431–446. [Online]. https://doi.org/10.1007/978-3-642-20898-0_31 [retrieved: May, 2019].

[3] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: The internet of things is the backbone," IEEE Consumer Electronics Magazine, vol. 5, no. 3, July 2016, pp. 60–70.

[4] C. Yin, et al., "A literature survey on smart cities," Science China Information Sciences, vol. 58, no. 10, 2015, pp. 1–18.

[5] A. H. Alavi, P. Jiao, W. G. Buttlar, and N. Lajnef, "Internet of things-enabled smart cities: State-of-the-art and future trends," Measurement, vol. 129, 2018, pp. 589 – 606.

[6] C. M. Sosa-Reyna, E. Tello-Leal, and D. Lara-Alabazares, "Methodology for the model-driven development of service oriented IoT applications," Journal of Systems Architecture, vol. 90, 2018, pp. 15 – 22.

[7] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," IEEE Communications Surveys Tutorials, vol. 17, no. 4, 2015, pp. 2347–2376. https://doi.org//10.1109/COMST.2015.2444095 [retrieved: Jun, 2019].

[8] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, and J. Al-Jaroodi, "Applications of big data to smart cities," Journal of Internet Services and Applications, vol. 6, no. 1, Dec 2015, p. 25. [Online]. https://doi.org/10.1186/s13174-015-0041-5 [retrieved: Jun 2019].

[9] C. Lim, K.-J. Kim, and P. P. Maglio, "Smart cities with big data: Reference models, challenges, and considerations," Cities, vol. 82, 2018, pp. 86 – 99. [Online]. https://doi.org/10.1016/j.cities.2018.04.011 [retrieved: Apr, 2019].

[10] M. Ge, H. Bangui, and B. Buhnova, "Big data for internet of things: A survey," Future Generation Computer Systems, vol. 87, 2018, pp. 601 – 614.

[11] C.-W. Tsai, C.-F. Lai, H.-C. Chao, and A. V. Vasilakos, "Big data analytics: a survey," Journal of Big Data, vol. 2, no. 1, Oct 2015, p. 21. [Online]. https://doi.org/10.1186/s40537-015-0030-3 [retrieved: May, 2019].

[12] A. M. S. Osman, "A novel big data analytics framework for smart cities," Future Generation Computer Systems, vol. 91, 2019, pp. 620 – 633.

[13] M. Han and W. Ren, "Global mutual information-based feature selection approach using single-objective and multi-objective optimization," Neurocomputing, vol. 168, 2015, pp. 47 – 54.

[14] A. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 2, Feb 1997, pp. 153–158.

[15] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," Neurocomputing, vol. 300, 2018, pp. 70 – 79.

[16] R. Sheikhpour, M. A. Sarram, S. Gharaghani, and M. A. Z. Chahooki, "A survey on semi-supervised feature selection methods," Pattern Recognition, vol. 64, 2017, pp. 141 – 158.

[17] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Computers Electrical Engineering, vol. 40, no. 1, 2014, pp. 16 – 28, 40th-year commemorative issue.

[18] J. Tang, S. Alelyani, and H. Liu, Feature selection for classification: A review. Data Classification: Algorithms and Applications. CRC Press, 2014, pp. 37–64.

[19] M. Moradkhani, A. Amiri, M. Javaherian, and H. Safari, "A hybrid algorithm for feature subset selection in high-dimensional datasets using FICA and IWSSr algorithm," Applied Soft Computing, vol. 35, 2015, pp. 123 – 135.

[20] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," Artificial Intelligence Review, 2019, pp. 1– 42.

[21] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," Machine Learning, vol. 46, no. 1, 2002, pp. 389–422. [Online]. https://doi.org/10.1023/A:1012487302797 [retrieved: May, 2019].

[22] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, May 2002, pp. 603–619.

[23] M. A. Carreira-Perpinan, Handbook of Cluster Analysis. New York: Chapman and Hall/CRC, 2016, ch. Clustering Methods Based on Kernel Density Estimators: Mean-Shift Algorithms.

[24] A. Tehreem, S. G. Khawaja, A. M. Khan, M. U. Akram, and S. A. Khan, "Multiprocessor architecture for real-time applications using mean shift clustering," Journal of Real-Time Image Processing, 2017, pp. 1– 17.

[25] Chicago Police Department, "Reported Crime - Public Safety dataset". [Online]. https://data.cityofchicago.org/Public-Safety/Crimes-2001-topresent/ ijzp-q8t2/data, 2018, [retrieved: Apr, 2019].

 ISBN: 978-1-61208-731-3