

Online Health Community Recommendation System with LDA-FCM Approach

Shan Shan

*Department of Strategy,
Analytics and Operations,
Nottingham Business School
Nottingham Trent University
Nottingham, United Kingdom
Shan.shan@ntu.ac.uk*

Umar Farooq

*School of Strategy and Leadership
Coventry University
Coventry, United Kingdom
ad9061@coventry.ac.uk*

Jialou Wang

*School of Computer and Information Science
Northumbria University
Newcastle Upon Tyne, United Kingdom
jialou2.wang@northumbria.ac.uk*

Yulei Li

*School of Computer and Information Science
Northumbria University
Newcastle Upon Tyne, United Kingdom
Yulei3.li@northumbria.ac.uk*

Düriye Betül Gü m ü ş

*Department of Strategy, Analytics and Operations,
Nottingham Business School
Nottingham Trent University
Nottingham, United Kingdom
betulgumus@ntu.ac.uk*

Abstract—Online health communities generate vast knowledge on various topics, making it challenging to track latent themes and correlations among key variables. Identifying these hidden variables and their relationships can enhance user engagement and content curation. This paper introduces a novel approach using fuzzy cognitive maps to uncover these variables and causal relationships in online health communities, providing users with valuable insights for disease management and treatment. The study employed Latent Dirichlet Allocation (LDA) to generate topics from an online health forum, using this as evidence to construct fuzzy cognitive maps. The proposed system reflects disease development and offers significant insights for managing health conditions. The findings have theoretical and practical implications for developing online recommendation systems in the healthcare domain. This study contributes to the literature by proposing a method for identifying key variables and relationships, aiding health professionals and patients in understanding and managing health conditions. Future research will explore the system's effectiveness in real-world settings and the role of user-generated content in enhancing recommendation systems.

Index Terms—Online Health Community; Diabetes; Recommendation Systems; Dirichlet Allocation and Fuzzy Cognitive Map

I. INTRODUCTION

The use of online health communities (OHCs) has significantly increased over the past decade, driven by widespread internet access and advancements in digital technologies. OHCs provide convenient platforms for users to access medical information, interact with healthcare professionals, and share experiences with peers. In the United Kingdom, over 90% of the population uses the internet, with 53% actively seeking health-related information online. Please observe the conference page limits [1]. OHCs are particularly effective for individuals managing chronic conditions, such as diabetes, which serves as a case study for this research due to its

public health significance, high engagement levels within OHCs, and the complexity of its management. Diabetes is one of the most prevalent chronic diseases globally, ranked as the ninth leading cause of death in 2019, with 1.5 million deaths directly attributed to it [5]. The online diabetes community is notably active, engaging in discussions about their condition, sharing experiences, and seeking peer support [6], [7]. Effective management of diabetes requires ongoing medical care, patient education, and psychosocial support to prevent complications and improve quality of life [8], [9]. The complexity and global prevalence of diabetes, coupled with the vibrant online discussions surrounding it, make it an ideal focus for developing advanced recommendation systems within OHCs.

Current OHCs are enriched with various recommendation systems designed to provide personalized advice, including collaborative filtering, content-based filtering, hybrid approaches, and context-aware recommendations [13]–[21]. Collaborative filtering is particularly valued for its effectiveness and scalability, operating on the premise that users with similar preferences will benefit from similar recommendations [13], [14]. Content-based filtering complements this by focusing on a user's past interests to suggest relevant content, a method that has proven beneficial across various domains, including diabetes management [15]–[20].

However, these methods have limitations. Hybrid filtering systems, which combine multiple techniques, can potentially improve performance by addressing the weaknesses of individual methods [21], [27]. Despite this, they often rely on explicit user feedback, which can introduce biases and fail to capture the nuanced, implicit needs of users [21]–[27]. Additionally, discerning the credibility of online content remains challenging, as the abundance of information—including personal

anecdotes—can be misleading. This, combined with indirect digital interactions with healthcare professionals, can lead to risks such as self-diagnosis [10]–[12].

To address these gaps, this study introduces a hybrid recommendation system that integrates Latent Dirichlet Allocation (LDA) with Fuzzy Cognitive Maps (FCMs). FCMs provide a sophisticated framework for modelling the complexities of causality in healthcare information, similar to cognitive mapping in human problem-solving [28]. They represent causal reasoning with a degree of fuzziness, acknowledging the uncertainty inherent in medical knowledge and user health narratives. The graphical structure of FCMs supports systematic causal propagation, essential for predictive accuracy in recommendation systems, and allows for forward and backward chaining, reflecting the multifaceted nature of healthcare knowledge. This resemblance to neural networks makes FCMs a powerful tool for constructing and expanding knowledge bases, connecting disparate information in a coherent, context-sensitive manner [28].

LDA is employed to categorise user comments into topics, facilitating the exploration of connections among various nodes. By combining LDA's topic categorisation with the complex system modelling of FCMs, the proposed recommendation system aims to overcome existing limitations in OHCs. This approach is designed to authenticate and personalise information, ensuring users receive relevant recommendations rooted in clinically verified data. The integrative method promises to enhance the precision and reliability of health information, paving the way for safer and more informed use of OHCs for chronic disease management and beyond. This research contributes to the field by introducing a novel, intelligent recommendation system tailored to the intricate demands of digital health communication.

II. MATERIALS AND METHODS

OHCs use various recommendation systems to offer users personalized suggestions aligned with their preferences and interests. These systems include collaborative filtering, content-based filtering, hybrid filtering, knowledge-based recommendation, and context-aware recommendation. Collaborative filtering, a widely used technique, operates on the premise that individuals with similar tastes tend to make comparable choices, leveraging these patterns to generate recommendations [13]. This method has gained popularity in OHCs due to its effectiveness, scalability, and ability to suggest items or content to new users without requiring explicit data about the items themselves [14].

In contrast, content-based filtering focuses on extracting users' characteristics and using these features to make item recommendations [15]. This technique assumes that users maintain consistent preferences for specific attributes and is commonly applied in domains like movies, books, research papers, and music [16]–[19]. In OHCs, content-based filtering is valuable for suggesting articles similar to those a user has shown interest in, such as those related to a specific medical condition like diabetes. This approach is prevalent in OHC

recommender systems, where it can suggest relevant threads based on a user's past interactions [20].

Hybrid filtering systems combine aspects of various recommendation methods to enhance performance, often delivering better results than systems relying on a single approach [21], [27]. Despite their advantages, hybrid systems have limitations. They frequently depend on explicit user feedback, such as ratings and comments, which can introduce biases and incompleteness into the recommendation process. Additionally, these systems may struggle to account for hidden variables or complex relationships, limiting their ability to provide truly accurate and personalized recommendations.

A. The Nature of FCM in online communities

Fuzzy Cognitive Maps (FCMs) are graphical structures representing causal reasoning with a degree of fuzziness to account for imprecise causality between objects. Their graph structure enables systematic causal propagation, encompassing both forward and backward chaining, and allows for the expansion of knowledge bases by connecting different FCMs [28]. FCMs resemble neural networks and serve as powerful tools for modelling complex systems. They are characterized by a directed graph of nodes representing various concepts, such as events and policies, and edges that signify causal relationships [29].

Each concept within an FCM represents a distinct entity, such as variables, states, events, or conditions. The value associated with each concept is derived from a fuzzy value representing the concept and is typically constrained to the $[0,1]$ interval. Causal relationships between concepts in an FCM can be positive, negative, or neutral, indicating how changes in one concept affect another. FCMs simulate system behaviour by calculating the effects of interconnected concepts, updating the values iteratively until a steady state is reached [30].

FCMs are particularly suitable for modelling complex systems, such as those found in healthcare. They are used to develop advanced diagnostic and decision-support systems by capturing human knowledge and system behaviour in a structured form [33]. FCMs allow for the representation of causal relationships in a way that reflects the dynamics of the system, making them valuable for understanding and predicting system behaviour [34]–[36].

One of the key strengths of FCMs is their ability to model fuzzy causal relations (FCRs), which predict behavioural tendencies based on changes in one or more concepts. These relations help to understand and simulate the effects of various factors within a system, such as the impact of a drug on a patient's symptoms [37–39]. Causal associations are essential for understanding how phenomena are organized and for making decisions based on causal evaluations [40].

B. The construction LDA-FCM

FCMs work based on topics generated from LDA and experts' opinion modelling reality/phenomena as a conglomerate of classes/concepts and causal relations amongst them. According to the law of large numbers (LLN), in probability

theory and statistical analysis, which is a theorem describing the outcomes of performing the same experiment numerous times. The law posits that the results average, obtained from more number of trials is closer to an ideal expected value, with a tendency to become closer and closer to an expected value with respect to the number of trials performed [41]. LLN is considered important since it provides long-term stable results for averages of random concepts or events. The LLN thus makes sure, in a sense, that reliability of knowledge increases with the sample size of experts. The outcome of a causal effect is normally a variation in single or multiple

concepts. Employing Numerical Weights Knowledge (NWK) on complex system behaviour is relatively subjective and so to

develop a model with more accuracy for a complex system, the experience/opinion of experts is considered useful [42]. Experts are usually pooled together to examine relevant aspects and factors, which represent the nodes in FCM. They, therefore, determine the concepts that constitute the FCM, as well as the attributes of the system the respective concept represent. The fact that this method functions based on expert opinion is considered a major advantage, and this translates to increased quality of information. Whenever vast data appears to be unsupervised, such as that found in OHCs, FCMs can be very useful. It is the sole fuzzy technique which unveils the hidden patterns in a situation. As LLN has demonstrated that the data strength depends on the degree of expert opinion, FCMs permit incorporation of opinions from experts [43], [44].

It is therefore noteworthy that, in general, the quality of the information can be enhanced by FCMs because they represent human knowledge in a system operation. Experts create FCMs based on their knowledge and experience of a complex system. Methodologies of construction depend on exploitations of experiences of experts about a system's model along with its behaviour. The experts are also the ones who determine the types and the number of concepts which an FCM will constitute, as well as the inter-relationships amongst them. The experts have knowledge about the key factors which determines the behaviour of a complex system, with each one being represented as a concept. According to experience, the experts determine the concepts of FCM which represent events, trends, actions, values, and goals of a complex system. Lastly, experts also know the kind of influence elements in a system have on other elements [45]. FCMs denote fuzzy graph structures that represent causal reasoning. Their structure allows causal propagations, in particular patterns, allowing the development of knowledge bases by linking different FCMs [46]. FCMs are especially suitable where data is largely unsupervised. They are able to work with opinions from professionals and experts to ensure only credible information is extracted from the vast data [47].

While, the Latent Dirichlet Allocation (LDA) stands as a prominent and extensively researched topic modelling technique [48]. Its output delineates a compilation of topics associated with word groupings that appear concomitantly in documents, contingent upon configurations. To elucidate the topics pertinent to diabetes research, topic modelling was employed

for topic extraction. LDA adopts the Bayesian methodology, positing that the lexemes present in the textual content are independently sourced from an amalgamation of containers, each containing a specific word assortment, coupled with the generative procedure related to each post [48].

At its core, LDA utilises the Dirichlet distribution for both topics and documents. Formally, if θ is a K-dimensional Dirichlet random variable, then its probability density function is given by:

$$p(\theta | \alpha) = \frac{\Gamma(\sum_{i=1}^K \alpha_i)}{\prod_{i=1}^K \Gamma(\alpha_i)} \prod_{i=1}^K \theta_i^{\alpha_i - 1} \quad (1)$$

where α is the parameter of the Dirichlet distribution, and Γ denotes the gamma function.

In the realm of topic modelling, a myriad of methodologies exists alongside LDA. Notably, Non-Negative Matrix Factorisation (NMF) offers an iterative optimisation-based approach to decompose matrices, whereas Latent Semantic Analysis (LSA) employs singular value decomposition to discern patterns within textual relationships [49], [50]. Another pertinent methodology is the Correlated Topic Model (CTM), which enhances LDA by accounting for potential correlations amidst topics, thus offering a nuanced topic structure [49]. While each technique bears its distinct advantages, LDA stands out due to its inherent probabilistic framework, often yielding more discernible and coherent topics [51]. This quality of LDA, coupled with its scalability and adaptability, informed our choice of the model. Furthermore, LDA's diverse applications, spanning from document clustering to its intriguing use in crime data analysis as evinced by the paper on dispatch centre crime identification, underscore its versatility and efficacy in handling large, complex datasets [52].

According to clinical research, the data strength depends on the number of opinions by experts and in LDA-FCMs can be used with several experts' opinions [47]. FCMs as cognitive maps enable a system to accept Falsehood, Truth, and the Intermediary and such situations are common when dealing with vast unsupervised data and that which also entails personal feelings; for instance, when forming symptom-ailment model and personality treatment model [53]. A major advantage of FCM is that it is based on experts' opinion to reveal the hidden patterns of a situation [54].

Soft computing and modelling techniques like FCMs have been observed to be useful for achieving solutions in healthcare, including events/processes like diagnosis, treatment, and behaviour/trend prediction. FCMs constitute a powerful tool since it is able to combine the experience of numerous experts with the knowledge extracted from historical data to form cognitive maps. Those models have been observed to be valuable and useful, especially where inexperienced people are involved. In this case, the report has demonstrated FCMs usefulness in OHCs, which denote social forums that are open to the public for discussing and sharing healthcare-related information. One major problem that has been highlighted in OHC forums is the vast amount of unregulated and unverified

healthcare information that also come from people whose credibility cannot be ascertained. Nonetheless, considering that FCMs are dependent on experts, having diverse back-grounds, knowledge, and experiences, renders them subjective. Therefore, to obtain models that are more reliable and correspond to the real situation, a parameter for ‘credibility’ ought to be included. Additionally, particularly for medical issues, time can be a significant factor as it may cause different influences on the concepts and subsequently the output. Thus, the constructing framework of LDA-FCM is demonstrated as following figure 1.

Besides, the landscape of online communities, especially health-related forums, is intricate. Users, content, interactions, and underlying sentiments shape these platforms. Some hidden variables represent underlying patterns, trends, or factors that aren’t immediately obvious but influence user behaviour and content dynamics. By identifying and understanding these variables, it can enhance personalization. By understanding the hidden variables that drive user preferences, behaviours, and needs, online platforms can tailor content delivery, ensuring that users receive information most relevant to them [55]. Also, it can improve user engagement. Recognizing the hidden factors that keep users engaged can help online communities retain their members and encourage more active participation. For instance, understanding the variables that drive users to respond to certain posts can lead to better content curation and presentation [56].

To verify the validity of this model, this research built the corpus which are based on users’ comments in the “diabetes.co.uk” forum and further established the Diabetes LDA-FCM study. To extract effective nodes and demonstrates the relations, this research proposed the LDA -FCM methods.

The workflow of this study consists of four steps: (1) collect data in the diabetes forum; (2) pre-process the data by removing unnecessary noise; (3) extract nodes using FCM to find relations between nodes (4) explore the nodes related to diabetes to make recommendations for future search.

C. Data Collection

The ScrapeStorm software (version 3.5.1) was used to crawl the user comments under the topic posts about “Diabetes” in the diabetes Forum. To filter out non-relevant content, such as advertising posts, copyright disclaimers, and irrelevant image links, the research focused on crawling the most recent 154 topics. These topics encompassed a total of 4540 user

comments and were collected until November 23, 2020, to form the dataset. The topic posts are generally those made by users on the 1st floor of the comments, and the replies from users on the other floors are replies to the 1st floor users. The crawled data was used as a corpus for the diabetes domain, on which the diabetes FCM was constructed.

D. Data Pre-processing and Topics Selection

In this research, a series of data preprocessing steps were undertaken to clean and prepare the user comments related to “Diabetes” for analysis:

- 1) Non-English Comment Filtering: Non-English comments were filtered out to ensure that the analysis focused exclusively on English-language content.
- 2) Text Cleaning: The text was cleaned by removing numbers and punctuation marks, and all text was converted to lowercase to standardize the text for analysis.
- 3) Tokenization: The text was tokenized, which involved breaking it into individual words to facilitate further analysis.
- 4) Stopword Removal: Common stop words (e.g., “the,” “and,” “in”) were removed from the text, as they typically do not contribute to the analysis of meaningful content.
- 5) Stemming: Words were reduced to their root form (stem) to reduce dimensionality and group similar words together.
- 6) Removal of Quote Parts: To prevent double counting and maintain data integrity, parts of quotes were removed from the text.
- 7) Standardization of Terms: Some terms were standardized to their root forms for consistency. For example, “eye” was used instead of “eyesight,” “feet” instead of “circulation feet,” and “t1” and “t2” were standardized as “type 1” and “type 2,” respectively.
- 8) Removal of Irrelevant Comments: Irrelevant comments, such as those containing emoticons or generic phrases like “thanks for sharing,” “top,” and “like,” as well as empty comments, were removed from the dataset.

The cleaned data was then analysed using the Natural Language Toolkit (NLTK) in Python, which facilitated tasks such as word splitting, annotation, frequency counting, and keyword extraction related to diabetes.

Keywords for constructing the FCM in the OHC were selected through a combination of co-occurrence analysis of user comments and expert input. These keywords serve as conceptual nodes within the FCM, representing user concerns, disease characteristics, and treatment-related topics. Co-word analysis was employed to identify key terms and their relationships, revealing developmental trends and structural evolution within the field [57]. This analysis involved standardising the co-occurrence matrix using Pearson co-efficients and performing clustering analysis with SPSS software.

The FCM serves as a fuzzy graphical structure designed to depict cause-and-effect reasoning. Within the context of the

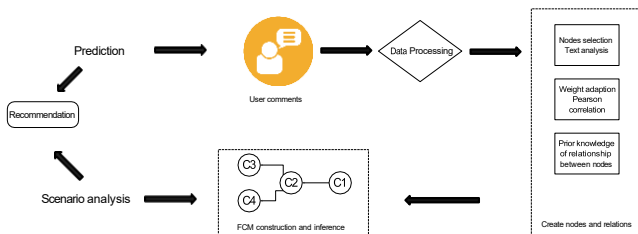


Fig. 1. Recommendation Framework of LDA-FCM.

FCM, the semantic meaning of keywords primarily pertains to the medical interpretation of terms related to the disease, along with associated medical terminology such as complications, therapies, drugs, and more. The primary objective is to elucidate whether there exists a causal relationship between these medical terms, as determined by medical knowledge.

III. RESULTS

A. Establishment of a diabetic FCM

1) *Determining the number of topics*: After approximating the LDA posterior distribution, the K topics are represented as multinomial distributions over V. Each topic distribution contains every word but assigns a different probability to each of the words; the words within topics with high probability are words that tend to co-occur more frequently. Typically, the top 10 or top 15 words with the highest probabilities are used to interpret and semantically label the topics. However, the number of topics, denoted as K, plays a critical role in shaping the outcomes of the topic modeling process. A low value of K results in fewer topics or very broad topics, while

a high value of K can lead to uninterpretable topics or topics that ideally should have been merged. Therefore, selecting the appropriate value of K is a crucial task in the context of topic modeling algorithms [58].

Automated coherence score measures were used for determining topic number K in this research, where $\hat{V}((t)) = v_{-1}((t)), \dots, v_M((t))$ comprises a list of the M most probable words in topic t. A smoothing count of 1 is included

to avoid adopting the logarithm of zero. Topic Coherence measures score a single topic by measuring the degree of semantic similarity between high scoring words in the topic [59].

We calculated the Coherence Value for different numbers of topics. When there are 2 topics, the Coherence Value is 0.519. With 8 topics, the Coherence Value increases to 0.534. However, when the number of topics reaches 14, the Coherence Value decreases to 0.518. Therefore, the highest Coherence Value is attained when there are 8 topics. The result is illustrated in figure 2

After establishing the optimal number of topics, experts proceeded to allocate relevant nodes within these topics. We extracted comment IDs and their respective comments using the method outlined earlier, as illustrated in Table I below. Notably, comments 3, 8, 9, and 13 did not yield extractable keywords. These extracted keywords serve as nodes in the diabetes FCM. It's worth mentioning that applying a threshold to the adjacency matrix can prove advantageous. This thresholding process aids in noise reduction and enhances the performance of subsequent analyses that rely on the adjacency matrix [60].

The similarity values between the keywords in these comments were obtained from the previous chapter's calculated keyword similarity. These similarity values are used as the weights between nodes. The semantic relationship dictionary is established to represent each node's semantic relations in the FCM (positive and negative values of arrows and weights),

TABLE I
SAMPLE OF TOPICS EXTRACTED FROM THE COMMENTS ON OHC.

Topic number	Topics	Weight of tokens	The keywords extracted from user comment (C _i)
1	Type	18.8	complication(C4), type2(C6), metformin(C7)
2	Blood test	18.5	liver(C2), glucose(C9)
3	Liver	15.9	hypos (C14), glucose(C9), insulin (C3)
4	Insulin	13.1	hba1c (C10), LCHF (11)
5	Eye	10.6	type2(C6), metformin(C7), eye(C8)
6	Diet	10.5	type1(C5), blood test(C1), liver(C2), insulin (C3), glucose(C9)
7	Glucose	7	complication (C4), hypos (C14)
8	Medicine	5.6	gliclazide (C12)

TABLE II
RECOMMENDATION FRAMEWORK OF LDA-FCM¹

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
C1	0	-0.15	-0.29	-0.18	-0.15	-0.15			-0.27	0.23				
C2		0	0.28	0.38	0.24		-0.29		0.23	-0.31		-0.21		0.24
C3			0.28	0	0.22	0.30	-0.26	-0.18		-0.39	-0.22		-0.18	0.19
C4			0.38	0.22	0	0.41	0.19	-0.34		0.20	-0.30		-0.18	0.37
C5			0.24	0.30	0.41	0	0.26	-0.38	-0.19	0.07	-0.33	-0.21	-0.17	0.30
C6				0.19	0.26	0	-0.20	-0.12		-0.19				
C7				-0.29	-0.26	-0.34	-0.38	-0.20	0	-0.02	-0.23		0.24	-0.21
C8				-0.16	-0.18	-0.19		0		-0.15			0.22	-0.17
C9		-0.27	0.23		0.20		-0.23	-0.15	0	-0.2		-0.16		
C10		0.23	-0.31	-0.39	-0.30	-0.33	-0.19		-0.23	0				-0.25
C11			-0.23	-0.16	-0.21						0			-0.17
C12			-0.21	-0.09	-0.18	-0.17	-0.15	0.24		-0.16		0		
C13				-0.18				0.22					0	
C14		-0.15	0.24	0.19	0.37	0.30		-0.21	-0.17	-0.25	-0.17			0

thus constructing fuzzy cognitive mapping of diabetes in online health communities. The adjacency matrix table for the diabetes FCM in the online health community is shown in table II.

¹C1:Blood Test,C2: Liver, C3: Insulin, C4: Complication, C5: Type 1, C6: Type 2, C7: Metformin, C8: Eye, C9: Glucose, C10: Hba1c, C11: LCHF, C12: Gliclazide, C13: Feet, C14:Hypoglycaemia.

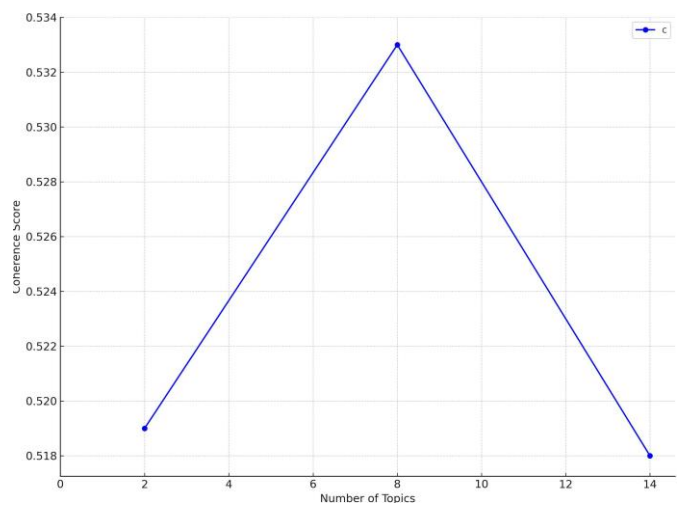


Fig. 2. Coherence value of topics of OHC comments.

TABLE III
TYPE 1 DIABETES NODE'S VALUE FOR EVERY ITERATION ²

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
A0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A1	0.5	0.67	0.71	0.77	0.5	0.68	0.23	0.35	0.55	0.26	0.34	0.37	0.5	0.71
A2	0.35	0.83	0.55	0.90	0.86	0.53	0.02	0.22	0.48	0.02	0.23	0.23	0.46	0.81
A3	0.31	0.93	0.83	0.97	0.94	0.71	0.01	0.20	0.69	0.01	0.20	0.15	0.46	0.92
A4	0.26	0.96	0.88	0.98	0.97	0.74	0.00	0.14	0.73	0.00	0.15	0.12	0.42	0.94
A5	0.25	0.97	0.90	0.98	0.97	0.76	0.00	0.13	0.76	0.00	0.15	0.11	0.40	0.95
A6	0.25	0.97	0.90	0.99	0.97	0.76	0.00	0.12	0.76	0.00	0.14	0.11	0.39	0.95
A7	0.25	0.97	0.90	0.99	0.97	0.76	0.00	0.12	0.76	0.00	0.14	0.11	0.39	0.95
A8	0.25	0.97	0.90	0.99	0.98	0.76	0.00	0.12	0.76	0.00	0.14	0.11	0.39	0.95
W	0.25	0.30	0.19	0.22	0.48	0.08	0.23	0.23	0.21	0.26	0.20	0.26	0.11	0.24

TABLE IV
TYPE 2 DIABETES NODE'S VALUE FOR EVERY ITERATION ²

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
A0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A1	0.5	0.5	0.5	0.64	0.68	0.5	0.35	0.40	0.5	0.35	0.5	0.5	0.5	0.5
A2	0.40	0.72	0.41	0.81	0.65	0.51	0.05	0.25	0.39	0.04	0.25	0.26	0.5	0.73
A3	0.34	0.88	0.74	0.93	0.90	0.65	0.03	0.26	0.62	0.03	0.25	0.19	0.48	0.83
A4	0.28	0.95	0.84	0.98	0.96	0.72	0.09	0.16	0.70	0.08	0.17	0.13	0.44	0.93
A5	0.26	0.96	0.89	0.98	0.97	0.75	0.00	0.14	0.75	0.05	0.15	0.11	0.41	0.95
A6	0.25	0.97	0.90	0.99	0.97	0.76	0.00	0.12	0.76	0.00	0.14	0.11	0.41	0.95
A7	0.25	0.97	0.91	0.99	0.97	0.76	0.00	0.12	0.76	0.00	0.14	0.11	0.39	0.95
A8	0.25	0.97	0.90	0.99	0.97	0.76	0.00	0.12	0.76	0.00	0.14	0.11	0.39	0.95
W	0.25	0.47	0.4	0.35	0.29	0.26	0.35	0.28	0.26	0.35	0.36	0.39	0.11	0.45

B. Process of Diabetes FCM

FCM was proposed by Kosko to improve decision-makers ability to understand the dynamic behavior of cause-and-effect diagrams [63], [64]. The online health community FCM process can reflect the development of diseases. Based on the FCM constructed above, different state values at the moment of A(0) are activated to represent the condition scenario of a patient with diabetes. After three iterations, $A(8) \approx A(9)$ indicates that the whole diabetic FCM system reaches equilibrium. The state values of each node at each time are shown in Tables III and IV.

C. Analysis of results for diabetes FCM

The Latent Dirichlet Allocation - Fuzzy Cognitive Map (LDA-FCM) model utilizes an iterative process to study the development and correlations of diabetes conditions within online health communities. This process demonstrates the dynamic relationships and changes in concept node values as they relate to the progression of diabetes. When we initialize our model with $A(0)=(0,0,0,0,1,0,0,0,0,0,0,0,0,0)$ or $A(0)=(0,0,0,0,0,1,0,0,0,0,0,0,0,0)$ – which correspond to "type 1 diabetes" and "type 2 diabetes" respectively – the system iterates to $A(7) \approx A(8)$. This equating of values is indicative of the system achieving a steady state. Such equilibrium provides a stable representation of the key concepts related to the respective diabetes types, revealing significant medical correlations based on word co-occurrence.

Interestingly, as we journey from A(1) to A(7), the varying values of each concept node emphasize the dynamic nature of the FCM model. These variations validate the model's ability to encapsulate the real-world intricacies of the disease and the multitudes of discussions around it within the online community.

²C1:Blood Test,C2: Liver, C3: Insulin, C4: Complication, C5: Type 1, C6: Type 2, C7: Metformin, C8: Eye, C9: Glucose, C10: HbA1c, C11: LCHF, C12: Glucalide, C13: Feet, C14:Hypoglycaemia, W: Range of change

A salient feature of the iterative process is its reflection of the intensity and proximity of the correlations. Specifically, a more modest change in the value of a concept node implies its stronger tie to the initial state, A(0). This observation proves crucial when comprehending the evolution of the disease, as these minor value fluctuations mirror the closely-knit aspects of the disease's trajectory. Thus, identifying nodes with the smallest changes in value could potentially highlight the most significant factors or areas of concern for those initialising with A(0).

Taking a closer inspection, a comparative analysis between table 3 and table 4 yields fascinating insights. For instance, while the value of the "Insulin" node stands at 0.7112 in A(1) for table 3, it is 0.5000 in table 4. Beyond mere numbers, this differential signifies insulin's robust relationship with diabetes. Grounded in established medical literature, insulin is a cornerstone in diabetes management. Moreover, the more pronounced change in value in the context of type 1 diabetes accentuates its stronger association with this particular diabetes variant as compared to type 2.

In summation, the iterative process in the LDA-FCM model offers a nuanced lens to discern the intricate dynamics of diabetes. By analyzing the changes and steadiness in node values, medical professionals, researchers, and even patients can gain invaluable insights into the disease's trajectory and the medical relationships interwoven within online health community discussions. As technology and data analytics continue to merge with medical research, tools like the LDA-FCM model will undoubtedly spearhead innovative approaches to understanding and managing diseases.

IV. DISCUSSION

Integrating the robust foundation of existing recommendation system literature with the advances presented in this study, it's evident that our approach to enhancing OHCs is both innovative and necessary. The existing body of work recognizes the importance and utility of collaborative filtering, which operates on the premise that users with similar interests will have similar needs [13], [14]. Content-based filtering complements this by focusing on the user's own characteristics or features to recommend items [15]. Hybrid systems attempt to combine the strengths of both to improve recommendation quality [21], [27].

However, while these systems have laid the groundwork, they fall short when addressing the complexity and dynamic nature of user interactions in OHCs. This study's integration of LDA with FCMs pushes beyond these limitations. Existing methods have been criticized for depending heavily on explicit user feedback, which can be biased and incomplete. Furthermore, these methods may not effectively capture the nuanced and complex relationships between different health-related variables [21], [27].

The proposed framework harnesses the predictive power of LDA for probabilistic topic modeling and the nuanced causal reasoning of FCMs, which allows for fuzziness in representing complex inter-variable relationships [28]. This combination

not only fills the gap left by the reliance on explicit feedback but also affords a more sophisticated understanding of user behavior and preferences. The use of FCMs particularly stands out as they offer the potential to systematically propagate causality, which is critical for accurately predicting user needs in a healthcare context [28].

The discussion of the study's findings resonates with the literature's identified need for more personalised and precise recommendations in OHCs. By incorporating expert knowledge in the construction of FCMs, the study confronts the issue of credibility, ensuring the information processed is both reliable and relevant. This strategy directly addresses the challenges identified in the literature, where recommendation systems may not fully consider the credibility and accuracy of the health information suggested to users.

V. CONCLUSION

Our research has demonstrated that the fuzzy cognitive map approach can effectively mine the key variables and build casual relationships among these variables, based on which the online recommendation system could be built to recommend the relevant knowledge to users. This approach has given highlights to the existing online community research where the user comments are unstructured and have no internal coherence. By implementing this approach, online community practitioners can gain valuable insights into enhancing the overall user content recommendation process. Furthermore, it offers users the opportunity to receive more precise recommendations tailored to their content preferences.

The use of LDA-FCM method ensured that the resulting FCM was based on high-quality data and effectively captured the key concepts related to diabetes in OHCs. This approach is consistent with previous research on the application of FCMs in healthcare, which have shown to improve decision-making, clinical outcomes, and patient satisfaction. The study's findings suggest that FCMs could also serve a valuable role within OHCs by providing recommendations for users seeking information about diabetes.

A. Limitations of Study:

One limitation of this study is that the data was collected from a single diabetes forum, which may limit the generalizability of the results. Future research could expand the sample size and diversity of online health communities to obtain more representative and diverse data on diabetes-related discussions and recommendations. Additionally, future studies could explore the effectiveness and efficiency of the proposed LDA-FCM methodology in identifying key concepts and establishing semantic relationships in other disease domains or health-related topics.

The second salient limitation of the present study lies in its omission of discussions on user privacy rights, especially when dealing with patient data. As we increasingly navigate the digital age, the ethical collection, storage, and analysis of personal health information emerge as paramount concerns [65]. The aggregation of patient data, even when anonymised,

carries inherent risks, including potential breaches of confidentiality and unauthorised access [66]. To address these concerns, a transparent informed consent process, where patients are fully apprised of the data's intended use, can bolster trust and ethical integrity [67]. Additionally, regular audits and adherence to international data protection standards, such as the General Data Protection Regulation (GDPR) in Europe, ensure compliance with the evolving landscape of privacy regulations [68]. Any comprehensive exploration of online health community recommendation systems ought to encompass a thorough examination of these pivotal privacy considerations, coupled with proactive strategies to safeguard user rights.

B. Implications of Study:

The proposed LDA-FCM methodology can be applied in various disease domains or health-related topics, facilitating the extraction of meaningful nodes and the construction FCMs for decision-making and recommendation purposes. The FCM constructed in this study offers a comprehensive and visual representation of the semantic relationships among diabetes-related concepts and help identify the key factors that may influence the disease's development and management.

Furthermore, the FCM can be used as a tool for patient education and engagement, helping patients understand the interconnections among different aspects of the disease and make informed decisions about their health. In future studies, there is an opportunity to integrate clinical and demographic data of the forum contributors to examine the influence of these factors on the discussions and recommendations. This opens avenues for investigating the use of FCMs as a tool for patient-centred decision-making and personalized medicine in diabetes and other disease domains.

REFERENCES

- [1] R. E. Costello, A. Anand, M. J. Evans, and W. G. Dixon, "Associations between engagement with an online health community and changes in patient activation and health care utilization: longitudinal web-based survey," *Journal of medical Internet research*, vol. 21, no. 8, p. e13477, 2019.
- [2] P. Chamakiotis, D. Petrakaki, and N. Panteli, "Social value creation through digital activism in an online health community," *Information Systems Journal*, vol. 31, no. 1, pp. 94-119, 2021.
- [3] Y. Li and X. Yan, "How could peers in online health community help improve health behavior," *International Journal of Environmental Research and Public Health*, vol. 17, no. 9, p. 2995, 2020.
- [4] S. Joglekar et al., "How online communities of people with long-term conditions function and evolve: network analysis of the structure and dynamics of the asthma UK and British lung foundation online communities," *Journal of Medical Internet Research*, vol. 20, no. 7, p. e238, 2018.
- [5] World Health Organization. "Diabetes." <https://www.who.int/news-room/fact-sheets/detail/diabetes> (accessed 10/10, 2023).
- [6] M. L. Litchman, L. S. Edelman, and G. W. Donaldson, "Effect of diabetes online community engagement on health indicators: cross-sectional study," *JMIR diabetes*, vol. 3, no. 2, p. e8603, 2018.
- [7] J. Guo, S. Shan, and Y. Ali Khan, "What are the impetuses Behind E-health applications' self-management services' ongoing adoption by health community participants?," *Health Informatics Journal*, vol. 29, no. 1, p. 14604582231152801, 2023.
- [8] J. Min, Y. Chen, L. Wang, T. He, and S. Tang, "Diabetes self-management in online health communities: an information exchange perspective," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1-12, 2021.

- [9] A. Hughes, N. Heydarian, D. Gerardo, I. Solis, and O. Morera, "Seeking Health Information and Social Support in the Diabetes Online Community," *Frontiers in Clinical Diabetes and Healthcare*, vol. 2, p. 708405, 2021.
- [10] A. Kamis, Y. Yao, and S. Kim, "An Empirical Validation of the Patient-centered e-Health Framework in Patient-focused web-sites," *Communications of the Association for Information Systems*, vol. 34, no. 1, p. 25, 2014.
- [11] J. Otono, M. Olaosebikan, O. Shaer, O. Nov, and M. P. Ball, "Understanding Users Information Needs and Collaborative Sensemaking of Microbiome Data," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1-21, 2019.
- [12] Q. B. Liu, X. Liu, and X. Guo, "The Effects of Participating in a Physician-Driven Online Health Community in Managing Chronic Disease: Evidence from Two Natural Experiments," *MIS Quarterly*, vol. 44, no. 1, 2020.
- [13] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56-58, 1997.
- [14] A. K. Sahoo, C. Pradhan, R. K. Barik, and H. Dubey, "DeepReco: deep learning based health recommender system using collaborative filtering," *Computation*, vol. 7, no. 2, p. 25, 2019.
- [15] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in *Proceedings of the fifth ACM conference on Digital libraries*, 2000, pp. 195-204.
- [16] S. Philip, P. Shola, and A. Ovy, "Application of content-based approach in research paper recommendation system for a digital library," *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 10, 2014.
- [17] T. Di Noia, R. Mirizzi, V. C. Ostuni, D. Romito, and M. Zanker, "Linked open data to support content-based recommender systems," in *Proceedings of the 8th international conference on semantic systems*, 2012, pp. 1-8.
- [18] S. Sahu, R. Kumar, M. S. Pathan, J. Shafi, Y. Kumar, and M. F. Ijaz, "Movie popularity and target audience prediction using the content-based recommender system," *IEEE Access*, vol. 10, pp. 42044-42060, 2022.
- [19] S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, "Content-based movie recommendation system using genre correlation," in *Smart Intelligent Computing and Applications: Proceedings of the Second International Conference on SCI 2018, Volume 2*, 2019: Springer, pp. 391-397.
- [20] C. Young, "Community management that works: how to build and sustain a thriving online health community," *Journal of medical Internet research*, vol. 15, no. 6, p. e119, 2013.
- [21] R. Burke, "Knowledge-based recommender systems," *Encyclopedia of library and information systems*, vol. 69, no. Supplement 32, pp. 175-186, 2000.
- [22] X. Li and T. Murata, "Customizing knowledge-based recommender system by tracking analysis of user behavior," in *2010 IEEE 17th International Conference on Industrial Engineering and Engineering Management*, 2010: IEEE, pp. 65-69.
- [23] A. Felfernig, G. Friedrich, D. Jannach, and M. Zanker, "An integrated environment for the development of knowledge-based recommender applications," *International Journal of Electronic Commerce*, vol. 11, no. 2, pp. 11-34, 2006.
- [24] C. C. Aggarwal and C. C. Aggarwal, "Knowledge-based recommender systems," *Recommender Systems: The Textbook*, pp. 167-197, 2016.
- [25] D. Chen, R. Zhang, K. Liu, and L. Hou, "Knowledge discovery from posts in online health communities using unified medical language system," *International journal of environmental research and public health*, vol. 15, no. 6, p. 1291, 2018.
- [26] X. Wang, K. Zhao, and N. Street, "Social support and user engagement in online health communities," in *International Conference on Smart Health*, 2014: Springer, pp. 97-110.
- [27] M. Lee, P. Choi, and Y. Woo, "A hybrid recommender system combining collaborative filtering with neural network," in *Adaptive Hypermedia and Adaptive Web-Based Systems: Second International Conference, AH 2002 Malaga, Spain, May 29-31, 2002 Proceedings 2*, 2002: Springer, pp. 531-534.
- [28] B. Kosko, "Fuzzy cognitive maps," *International journal of man-machine studies*, vol. 24, no. 1, pp. 65-75, 1986.
- [29] E. I. Papageorgiou and J. L. Salmeron, "A review of fuzzy cognitive maps research during the last decade," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 1, pp. 66-79, 2012.
- [30] C. Stylios and V. C. Georgopoulos, "Develop Fuzzy Cognitive Maps based on recorded data and information," in *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2015: IEEE, pp. 1-6.
- [31] G. Xirogiannis, J. Stefanou, and M. Glykas, "A fuzzy cognitive map approach to support urban design," *Expert Systems with Applications*, vol. 26, no. 2, pp. 257-268, 2004.
- [32] S. D'Onofrio, S. M. Müller, E. I. Papageorgiou, and E. Portmann, "Fuzzy reasoning in cognitive cities: an exploratory work on fuzzy analogical reasoning using fuzzy cognitive maps," in *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2018: IEEE, pp. 1-8.
- [33] A. Amirkhani, E. I. Papageorgiou, A. Mohseni, and M. R. Mosavi, "A review of fuzzy cognitive maps in medicine: Taxonomy, methods, and applications," *Computer methods and programs in biomedicine*, vol. 142, pp. 129-145, 2017.
- [34] A. J. Jetter and K. Kok, "Fuzzy Cognitive Maps for futures studies—A methodological assessment of concepts and methods," *Futures*, vol. 61, pp. 45-57, 2014.
- [35] O. Osoba and B. Kosko, "Causal modeling with feedback fuzzy cognitive maps," *Social-Behavioral Modeling for Complex Systems*, pp. 587-615, 2019.
- [36] P. P. Groumpos, "Fuzzy cognitive maps: basic theories and their applications in medical problems," in *2011 19th Mediterranean Conference on Control & Automation (MED)*, 2011: IEEE, pp. 1490-1497.
- [37] J. P. Carvalho, "On the Implementation of Evolving Dynamic Cognitive Maps," in *International Fuzzy Systems Association World Congress*, 2019: Springer, pp. 201-213.
- [38] P. Zdanowicz and D. Petrovic, "New mechanisms for reasoning and impacts accumulation for rule-based fuzzy cognitive maps," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 543-555, 2017.
- [39] J. L. Salmeron, "Fuzzy cognitive maps for artificial emotions forecasting," *Applied Soft Computing*, vol. 12, no. 12, pp. 3704-3710, 2012.
- [40] O. Osoba and B. Kosko, "Beyond DAGs: modeling causal feedback with fuzzy cognitive maps," *arXiv preprint arXiv:1906.11247*, 2019.
- [41] E. Bolthausen and M. V. Wu'thrich, "Bernoulli's law of large numbers," *ASTIN Bulletin: The Journal of the IAA*, vol. 43, no. 2, pp. 73-79, 2013.
- [42] M. Alemi-Ardakani, A. S. Milani, S. Yannacopoulos, and G. Shokouhi, "On the effect of subjective, objective and combinative weighting in multiple criteria decision making: A case study on impact optimization of composites," *Expert Systems with Applications*, vol. 46, pp. 426-438, 2016.
- [43] E. Bourgani, C. Stylios, V. Georgopoulos, and G. Manis, "A study on fuzzy cognitive map structures for medical decision support systems," in *8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13)*, 2013: Atlantis Press, pp. 784-791.
- [44] Y. Li, X. Ma, J. Song, Y. Yang, and X. Ju, "Exploring the effects of online rating and the activeness of physicians on the number of patients in an online health community," *Telemedicine and e-Health*, vol. 25, no. 11, pp. 1090-1098, 2019.
- [45] M. F. Dodurka, E. Yesil, and L. Urbas, "Causal effect analysis for fuzzy cognitive maps designed with non-singleton fuzzy numbers," *Neurocomputing*, vol. 232, pp. 122-132, 2017.
- [46] J. P. Carvalho, "On the semantics and the use of fuzzy cognitive maps and dynamic cognitive maps in social sciences," *Fuzzy Sets and Systems*, vol. 214, pp. 6-19, 2013.
- [47] W. Pedrycz, A. Jastrzebska, and W. Homenda, "Design of fuzzy cognitive maps for modeling time series," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 1, pp. 120-130, 2016.
- [48] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993-1022, 2003.
- [49] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788-791, 1999.
- [50] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American society for information science*, vol. 41, no. 6, pp. 391-407, 1990.
- [51] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National academy of Sciences*, vol. 101, no. suppl_1, pp. 5228-5235, 2004.
- [52] M. P. Basilio, V. Pereira, and G. Brum, "Identification of operational demand in law enforcement agencies," *Data Technologies and Applications*, vol. 53, no. 3, pp. 333-372, 2019, doi: 10.1108/DTA-12-2018-0109.

- [53] M. S. Islam, M. M. Hasan, X. Wang, and H. D. Germack, "A systematic review on healthcare analytics: application and theoretical perspective of data mining," in *Healthcare*, 2018, vol. 6, no. 2: Multidisciplinary Digital Publishing Institute, p. 54.
- [54] G. Na'poles, E. Papageorgiou, R. Bello, and K. Vanhoof, "Learning and convergence of fuzzy cognitive maps used in pattern recognition," *Neural Processing Letters*, vol. 45, no. 2, pp. 431-444, 2017.
- [55] J. A. Konstan and J. Riedl, "Recommender systems: from algorithms to user experience," *User modeling and user-adapted interaction*, vol. 22, pp. 101-123, 2012.
- [56] R. Burke, "Hybrid recommender systems: Survey and experiments," *User modeling and user-adapted interaction*, vol. 12, pp. 331-370, 2002.
- [57] K. Mls, R. Cimler, J. Vas'c'a'k, and M. Puheim, "Interactive evolutionary optimization of fuzzy cognitive maps," *Neurocomputing*, vol. 232, pp. 58-68, 2017.
- [58] S. Syed and M. Spruit, "Full-text or abstract? examining topic coherence scores using latent dirichlet allocation," in *2017 IEEE International conference on data science and advanced analytics (DSAA)*, 2017: IEEE, pp. 165-174.
- [59] D. Mimno, H. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," in *Proceedings of the 2011 conference on empirical methods in natural language processing*, 2011, pp. 262-272.
- [60] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059-1069, 2010.
- [61] M. J. Cobo, W. Wang, S. Laengle, J. M. Merigo', D. Yu, and E. Herrera-Viedma, "Co-words analysis of the last ten years of the International journal of uncertainty, fuzziness and knowledge-based systems," in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 2018: Springer, pp. 667-677.
- [62] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059-1069, 2010.
- [63] B. Kosko, "Fuzzy Cognitive Maps" *international journal of man-machine studies*, January, pp. 62-75, 1986.
- [64] B. Kosko, "Hidden patterns in combined and adaptive knowledge networks," *International Journal of Approximate Reasoning*, vol. 2, no. 4, pp. 377-393, 1988.
- [65] D. J. Solove, *Understanding privacy*. Harvard University Press, 2008.
- [66] P. Ohm, "Broken promises of privacy: Responding to the surprising failure of anonymization," *UCLA L. Rev.*, vol. 57, p. 1701, 2009.
- [67] N. Manson, "The medium and the message: tissue samples, genetic information and data protection," 2009.
- [68] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (gdpr)," *A Practical Guide*, 1st Ed., Cham: Springer International Publishing, vol. 10, no. 3152676, pp. 10-5555, 2017.