

Original Paper

# Identifying Influences in Patient Decision-making Processes in Online Health Communities: Data Science Approach

Mingda Li<sup>1</sup>, PhD; Jinhe Shi<sup>1</sup>, PhD; Yi Chen<sup>2</sup>, PhD

<sup>1</sup>Ying Wu College of Computing, New Jersey Institute of Technology, Newark, NJ, United States

<sup>2</sup>Martin Tuchman School of Management, New Jersey Institute of Technology, Newark, NJ, United States

**Corresponding Author:**

Yi Chen, PhD

Martin Tuchman School of Management

New Jersey Institute of Technology

184-198 Central Ave

Newark, NJ, 07102

United States

Phone: 1 973 596 6302

Email: [yi.chen@njit.edu](mailto:yi.chen@njit.edu)

## Abstract

**Background:** In recent years, an increasing number of users have joined online health communities (OHCs) to obtain information and seek support. Patients often look for information and suggestions to support their health care decision-making. It is important to understand patient decision-making processes and identify the influences that patients receive from OHCs.

**Objective:** We aimed to identify the posts in discussion threads that have influence on users who seek help in their decision-making.

**Methods:** We proposed a definition of influence relationship of posts in discussion threads. We then developed a framework and a deep learning model for identifying influence relationships. We leveraged the state-of-the-art text relevance measurement methods to generate sparse feature vectors to present text relevance. We modeled the probability of question and action presence in a post as dense features. We then used deep learning techniques to combine the sparse and dense features to learn the influence relationships.

**Results:** We evaluated the proposed techniques on discussion threads from a popular cancer survivor OHC. The empirical evaluation demonstrated the effectiveness of our approach.

**Conclusions:** It is feasible to identify influence relationships in OHCs. Using the proposed techniques, a significant number of discussions on an OHC were identified to have had influence. Such discussions are more likely to affect user decision-making processes and engage users' participation in OHCs. Studies on those discussions can help improve information quality, user engagement, and user experience.

(*J Med Internet Res* 2022;24(8):e30634) doi: [10.2196/30634](https://doi.org/10.2196/30634)

**KEYWORDS**

influence relationship; decision-making threads; online health communities; patient engagement; deep learning; text relevance measurement

## Introduction

**Background**

In recent years, online health communities (OHCs) such as the Cancer Survivors Network (CSN), MedHelp, DoctorLounge, WebMD, and Health-boards message boards have become one of the most important resources that patients leverage [1]. An OHC is defined as an asynchronous web-based message board system for patients that contains multiple message boards, each

of which typically focuses on 1 disease. OHCs provide a web-based channel that enables information exchange, facilitates communication, and provides support to patients and caregivers [2-4]. They are especially valuable for patients with chronic diseases to learn about their conditions and seek social support [5,6].

Empowering and supporting patients to make informed health care decisions is a key component of patient-centered health care and is a social, economic, and technical necessity [7,8]. A

lot of patients seek information and advice on OHCs. Existing work has found that nearly half of the threads in a breast cancer forum [9] are related to patient decision-making [1]. Studies have also shown that patients are often influenced by web-based sources and social media in their health care decision-making [10,11].

**Objectives**

The goal of this study was to identify the influence relationship of posts in discussion threads related to health care decision-making. Specifically, we defined the influence relationships and identified post replies that influenced the initial author, who had questions posted on OHCs.

The outcomes of this study are important for health care professionals to help patients make informed decisions for several reasons. First, analyzing the writing style and pattern of posts that have influence may help explain why they have influence and provide insights to health care professionals on effective communication with patients. Second, if the information provided by posts that have an influence is not accurate, it will mislead patients. It is important to check the information quality in such posts to improve the quality of influence. Furthermore, a patient who has questions but does not receive any replies that have an influence may need further help.

**Literature Review**

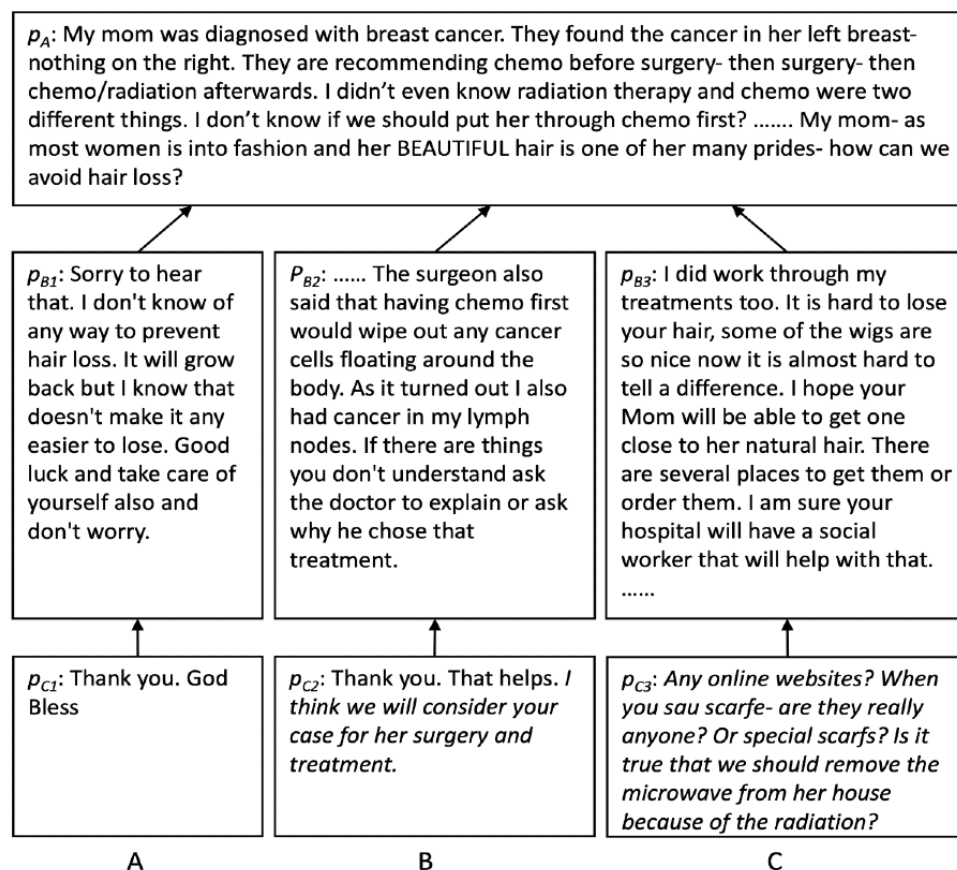
There is a lot of research conducted on OHC analysis, although with limited study on identifying influence relationships of

posts. Several studies have been conducted on analyzing the reciprocal patterns between users’ replies in discussion forums [12-14]. There is also work on analyzing the patterns between post views and post replies [15]. Many studies have been conducted on identifying influential users in a community [16-20]. In those applications, a post, blog, or tweet typically expresses an opinion of the author, and the replies are considered as an indication of being influenced by the opinion of the original post. That is, the reply relationship is considered as an influence relationship. The focus is on judging the influential power of an author based on activeness of post writing [21] and social network features [17,18] such as PageRank-like algorithms or clustering algorithms.

Finding influence relationships among posts in discussion forums is different from finding influential users and requires different techniques. In an OHC, the initial author of a thread typically expresses a question, not an opinion. The influence happens when a reply to the question affects the initial author. There are only 2 existing studies that consider the influence of the replier on the initial author [21,22]. This influence is identified when the sentiment of the initial author is changed to be similar to that of the replier. However, this definition may not be accurate.

Let us look at an example of a discussion thread related to patient decision-making, shown in Figure 1. An OHC user initialized a thread asking for advice on whether to have chemotherapy before surgery for her mother’s treatment plan in post  $p_A$ .

**Figure 1.** Example of a discussion thread.



In Figure 1A, a user replied by comforting her in post  $p_{B1}$ . The reply was not informative. Even though the initial author expressed gratefulness to the author of post  $p_{B1}$ , with sentiment changing to be positive in post  $p_{C1}$ , she was not influenced by post  $p_{B1}$ . Indeed, studies show that 75% to 85% of CSN forum participants change their sentiment in a positive direction through web-based interactions with other community members [23]. A change in sentiment is not necessarily an indicator of being influenced.

In contrast, in Figure 1B, a user shared her experience in a similar situation suggesting to have chemotherapy before a surgery in post  $p_{B2}$ . The initial author expressed her gratitude and indicated that she would consider this suggestion in determining her mother’s treatment plan (the sentences in italics) in  $p_{C2}$ , showing her being influenced.

**Contribution**

Instead of considering sentiment changes, we propose using questions or future actions on relevant replies as an indicator of being influenced, as illustrated in the aforementioned example. There are 2 major challenges in identifying influence relationships. First, we need to define influence relationships of posts. We examined the semantics of post content to define influence relationships. Unlike influential users, who are defined by network features in the existing work [16-20], text content is the key to determine whether posts have influence. Second, it is hard to identify influence relationships. Unlike typical text classification problems, influence relationships involve multiple posts with reply relationships rather than a single paragraph of text. In addition, influence is an abstract concept. It is challenging to extract relevant features to capture the influence patterns considering both content and the reply relationship.

This study makes novel contributions to identifying influence relationships in discussion threads in OHCs related to patient decision-making. Specifically, (1) we defined the influence relationship between the posts based on the semantics of the post content, (2) an extensible deep learning model that extracts and combines both sparse and dense features was proposed to identify the influence relationships in OHC decision-making threads, and (3) the proposed model achieved good performance in identifying influence relationships in empirical evaluation.

**Methods**

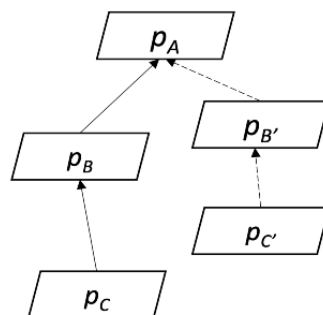
In this section, we first model the OHC data and define the influence relationship in discussion threads. We then propose a deep learning-based model to identify the influence relationships.

**Problem Definition**

**Definition of Discussion Threads**

Figure 2 presents an overview of the OHC data structure. We modeled an OHC as a set of discussion threads  $T = \{t_1, t_2, \dots, t_n\}$ . Each thread  $t_i$  is composed of a set of posts and a function  $R$  that represents the reply relationship. For example, Figure 2 illustrates a thread that contains a set of 5 posts  $\{p_A, p_B, p_C, p_{B'}, p_{C'}\}$ . One of the reply relationships,  $R(p_B) = p_A$ , represents that post  $p_B$  replies to post  $p_A$ . Each post  $p_i$  consists of a sequence of sentences  $p_i = \{s_1, s_2, \dots, s_l\}$ . Each post has an author. We denoted the author relationship using a function  $U$ .  $U(p_i)$  represents the author of post  $p_i$ . Note that a post only has a single author; however, an author may write  $\geq 0$  posts in a thread. We used  $p_A$  to present the first post of a thread and named it the *initial post*. The author of the *initial post*,  $U(p_A)$ , is referred to as the *initial author* of the thread.

Figure 2. Data structure of an online health community.



Existing work [1] has studied the thread discussions in OHCs and identified that a subset of threads is related to *patient decision-making*. Such a thread is characterized by questions in the initial post and replies with suggestions of options. Techniques have been developed to identify decision-making threads in OHCs.

In this paper, we study how to identify the cases where the initial author of a decision-making thread is influenced by a reply post. Note that our study is general to any thread discussions related to decision-making. The definition and identification of decision-making threads can be handled using the approach developed in existing work [1] or other approaches. In the rest

of this paper, we use *threads* to refer to decision-making threads for simplicity. The defined influence relationship may not be applicable to discussion threads that are not related to decision-making, such as discussion threads for casual communication or experience-sharing threads providing social support.

**Definition of Relationships**

**Overview**

Before introducing the definition of *influence relationships*, we first introduce relationships. A relationship is defined on a triple

of posts in a thread with reply relationships: an initial post, a reply to the initial post, and the initial author's subsequent reply.

### Definition 1 (Relationship)

We define the relationship among three posts  $p_A$ ,  $p_B$ , and  $p_C$ , in a thread as  $r_i = (p_A, p_B, p_C)$ , where post  $p_A$  is the initial post of the thread, post  $p_B$  replies to  $p_A$ , post  $p_C$  replies to  $p_B$ , and the authors of  $p_A$  and  $p_C$  are the same person. That is,  $R(p_B) = p_A$ ,  $R(p_C) = p_B$ , and  $U(p_A) = U(p_C)$ .

We used  $r_i = (p_A, p_B, p_C)$  to denote the relationship among  $p_A$ ,  $p_B$ , and  $p_C$ . Note that there are many such relationships in a thread, and we considered all such triples. For instance, Figure 2 shows a thread with 2 relationships,  $r_1 = (p_A, p_B, p_C)$  and  $r_2 = (p_A, p_B', p_C')$ .

Also, note that existing work on identifying influential users [16-20] does not consider the relationships among post triples but only considers the reply relationship between 2 posts.

### Definition of Influence Relationships

#### Intuition

Now, we discuss how to define *influence relationships* on relationship  $(p_A, p_B, p_C)$ , where post  $p_B$  has an influence on the initial author  $U(p_A)$ .

First, intuitively, if post  $p_B$  influences the initial author  $U(p_A)$ , then the content of these 3 posts must be relevant.

Second, we referred to the definition of *influence* in Merriam-Webster [24]—"to affect or alter by indirect or intangible means"—and the reaction of *being influenced* is to *sway* rather than being convinced. If the initial author considers the suggestion given in post  $p_B$ , even if she eventually does not take the suggestion, she is considered to have been influenced by post  $p_B$ . On the basis of this definition, we observed 2 indications that the initial author,  $U(p_A)$ , was influenced by  $p_B$ .

An observation of being influenced is that the initial author may ask questions in  $p_C$  based on the suggestions in  $p_B$ . Curiosity is a motivator for learning and influential in decision-making [25]. An existing study [26] used a statistically large sample of learning forum posts to investigate whether student participation in the forum could be influenced. They observed that students who were influenced by others' interesting answers were more likely to ask follow-up questions. This indicates that asking further questions is a sign of being influenced. The same pattern also exists in OHCs. Let us look at the example in Figure 1C. The initial author expressed concerns about hair loss in  $p_A$ . Another user replied in post  $p_{B3}$  suggesting the use of wigs. The initial author then replied in post  $p_{C3}$  with questions (the sentences in italics) for more details about the suggestion given in post  $p_{B3}$ . These questions indicate that the initial author was

thinking about the suggestion given in post  $p_B$ ; that is, being influenced.

The second indication that the initial author was influenced by a post  $p_B$  is that she expressed her intention to take action in post  $p_C$ . Adjei et al [27] found that member-to-member communication in web-based brand communities greatly influenced the members' future purchase behavior. Similarly, the communication through discussion threads in OHCs may also affect the initial author's future actions. Let us look at the example in Figure 1B again. For the treatment question asked in  $p_A$ , a forum user shared her experience and discussed the treatment in post  $p_{B2}$ . The initial author then replied with a planned action (the sentence in italics) in  $p_{C2}$ . The intention of future action based on the communications in the thread is an indicator of the influence relationship.

On the basis of these observations, we define influence relationships in decision-making threads in the following section.

### Definition 2 (Influence Relationship)

A relationship  $r_i = (p_A, p_B, p_C)$  is considered as an influence relationship—that is,  $U(p_A)$  is influenced by  $p_B$ —if and only if the following conditions are met: (1) the content of  $p_B$  is relevant to post  $p_A$ , (2) the content of  $p_C$  is relevant to post  $p_B$ , and (3)  $p_C$  contains questions or indicates future actions.

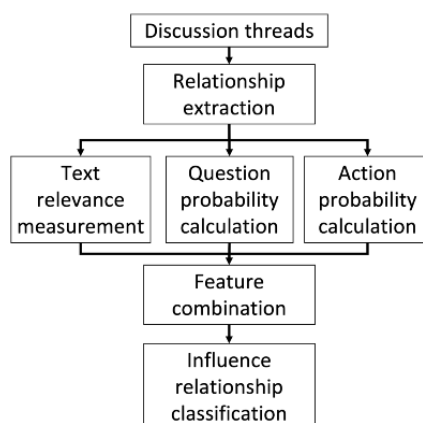
To identify influence relationships, we modeled it as a classification task. Given a set of relationships  $R = \{r_1, r_2, \dots, r_n\}$ , for each relationship  $r_i$ , we predicted its label to be either 1 or -1, where label 1 indicated that  $r_i$  was an influence relationship and label -1 indicated that  $r_i$  was not an influence relationship. The goal was to learn a model from the labels of known relationships and predict the labels for unlabeled relationships.

### Model Design

#### Overview

In this section, we present the method to identify the influence relationships in decision-making threads in OHCs. Figure 3 presents the framework of the proposed method.

Given a set of discussion threads as the input, we first extracted the triple relationships using the relationship extraction module. Text relevance features, question probability features, and action probability features were then calculated using the text relevance measurement module, the question probability calculation module, and the action probability calculation module, respectively. Finally, all the features were combined using a deep learning model in the feature combination module to generate the probability of a relationship being an influence relationship.

**Figure 3.** Workflow of influence relationship identification.

### Relationship Extraction Module

In this section, we introduce the relationship extraction module, which extracted all relationships defined in definition 1.

In the first step of relationship extraction, we built the reply tree structure based on the indented format in html files. For each adjacent post pair, the post that was posted earlier was treated as the parent of the latter post. The ancestor-descent distance between a post and the initial post was represented by the number of tab characters. The reply structure of a thread is illustrated in Figure 2. Each post is a node in the thread tree, and each edge represents a reply relationship. The root of the thread tree is the initial post (ie,  $p_A$ ) in definition 1.

Existing work observes that, in some forums, the reply structure in a discussion thread may not be fully available and proposes techniques to construct full reply structures [28]. The OHCs used in our experiments had a full reply structure. Existing techniques can be leveraged if needed for other forums.

We then navigated the thread tree to extract all relationship triples, as defined in definition 1. Each triple started with the initial post followed by a reply to the initial post written by another author and then a subsequent reply by the initial author, all of which were on the same path in the thread tree. For example,  $r_1 = (p_A, p_B, p_C)$  and  $r_2 = (p_A, p_B', p_C')$  are 2 relationships in the thread tree in Figure 2.

### Text Relevance Measurement Module

The text relevance measurement module measures the content relevance, or text semantic similarity, of 2 posts using a relevance score between 0 and 1.

There are mainly 2 types of deep learning-based methods in the literature that measure text relevance. The first type of method extracts content feature vectors of 2 input texts and then combines them to make a prediction, such as the Deep Structured Semantic Models (DSSM) [29], the Convolutional DSSM [30], and Architecture-I (ARC-I) [31]. The intuition of this method is to highlight the important information of the original texts so that irrelevant content can be removed before the feature combination phase. However, the drawback of this type of method is that it runs the risk of losing detail [32].

The second type generates the word-level relevance first and then uses neural networks to learn the hierarchical interaction patterns for content-level relevance, such as DeepMatch [33], Architecture-II (ARC-II) [31], and MatchPyramid [34]. The motivation is that making a good relevance judgment requires considering the interactions in the text relevance measurement process, starting from the interactions between words to patterns in phrases and those in whole sentences [34]. However, the training process for the second type is much more expensive than for the first one.

We evaluated both approaches to measure text relevance in experiments. We chose 2 state-of-the-art representative methods for the text relevance measurement module in the evaluation. For the first type, we chose ARC-I [31], which uses a multilayer perceptron to combine relevance feature vectors. It shows better performance than the DSSM [29] and Convolutional DSSM [30], both of which use cosine similarity [34]. We chose MatchPyramid [34] to represent the second type of method as it exhibits better performance than the other 2 methods (DeepMatch [33] and ARC-II [31]) in experiments on multiple data sets [34].

We further proposed the adaptation of Bidirectional Encoder Representations from Transformers (BERT) [35] as the embedding layer in the ARC-I and MatchPyramid models. BERT is a state-of-the-art embedding method for word representation in many natural language understanding tasks, trained on BookCorpus and English Wikipedia. We considered both BERT (trained on Wikipedia) and word2vec (trained on the training data set) as the embedding methods for both ARC-I and MatchPyramid. Different variations of the text relevance measurement module are evaluated in the *Text Relevance Evaluation* section.

### Question Probability Calculation Module

We now discuss how to calculate the probability of a post containing a question using the question probability calculation module.

There are 2 types of methods to identify question sentences in forums: a rule-based approach and a learning-based approach. In a rule-based approach, question marks and 5W1H words (what, who, when, where, why, and how) are used to identify question sentences [36]. A learning-based approach uses sequential question patterns to train a binary classifier on labeled

data [37-40]. Liu and Jansen [37] used the question mark to extract question posts from Sina Weibo. In the studies by Ranganath et al [38,39], frameworks were proposed to identify rhetorical questions by modeling the motivation of the user for posting them. In the study by Ojokoh et al [40], questions from ResearchGate were identified based on the maximum probability value of a naïve Bayes classification with part-of-speech tag features.

Both rule-based and learning-based approaches can achieve excellent performances. A study shows that a rule-based approach can outperform complicated learning-based approaches [36]. Thus, we followed a rule-based method [36] to identify question presence in the posts. In total, 2 types of rules were considered: question marks and 5W1H words. We made adaptations of this approach for OHCs. As a question mark is the most significant sign of a question, we gave a higher confidence score to a sentence with a question mark. We also set some constraints on 5W1H words to simulate the pattern of question sentences. First, 5W1H must appear at the beginning of a sentence. Second, auxiliary words were added to the original words for more specific patterns. For example, we considered *what is*, *what are*, *what does*, and *what do* instead of *what*.

After the question probability of each sentence in a post  $p_i$  was calculated, the maximum probability was used as the likelihood of post  $p_i$  containing at least one question, denoted as  $Q(p_i)$ .

**Action Probability Calculation Module**

This section presents the action probability calculation module, which generated the probability of action presence in a post.

The indication of a future action can be captured by the presence of verbs and appropriate sentence tense. The Natural Language Toolkit (NLTK) [41] tagger module defines a standard interface for augmenting each token of a text with supplementary information, such as its part of speech or its WordNet synset tag, and provides several different implementations for this

interface. We leveraged the NLTK tagger module to assess the likelihood of a post containing future actions by checking the existence of words with a future tense verb tag (eg, *will consider* in Figure 1B) or a modal auxiliaries tag (eg, *can*, *could*, *may*, and *must*). To count on the cases where future tenses may not be identified because of forum users’ typos or informal writing, we set the probability of future action to be 0.5 when the rules failed to identify future actions. Equation 1 shows the calculation formula to generate the action probability of a post  $p_i$ .

$$A(p_i) = \begin{cases} 1 & \text{if } VBF \text{ or } MD \in POS(p_i) \\ 0.5 & \text{otherwise} \end{cases} \quad (1)$$

Note that we did not consider negation in the action probability calculation module. For example, in post  $p_C$ , the initial author disagrees with the suggestions proposed in  $p_B$  and decides to do something different. For those cases, the overall meaning of  $p_B$  and  $p_C$  would be the opposite and, therefore, would be captured by the relevance vectors generated in the text relevance measurement module. Thus, we did not consider negations in this phase to avoid double counting.

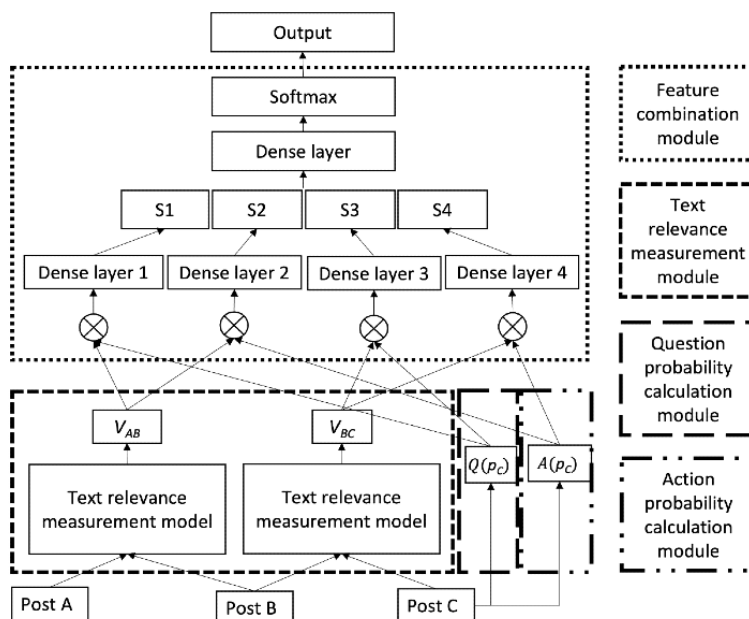
**Feature Combination Module**

**Overview**

Referring to Figure 4, the text relevance measurement module calculated  $P_{AB}$ —the relevance score between  $p_A$  and  $p_B$ —and  $P_{BC}$ —the relevance score between  $p_B$  and  $p_C$ . The question probability calculation module calculated the question probability  $Q(p_C)$ —or  $Q$  in short—and action probability  $A(p_C)$ —or  $A$  in short—based on the text of  $p_C$ .

We now discuss the feature combination module that measures the influence score based on these features. We discuss 2 alternative methods: *a baseline approach* and *a deep learning model*.

Figure 4. Architecture of the feature combination module.



### Baseline Approach

Recall that, according to definition 2, the presence of an influence relationship requires the relevance between post  $p_A$  and post  $p_B$ , the relevance between post  $p_B$  and post  $p_C$ , and the presence of a question or action in post  $p_C$ . We started with an intuitive method to detect influence relationships based on the definition using Equation 2.

$$P_{baseline} = P_{AB} \times P_{BC} \times \max [Q(p_C), A(p_C)] \quad (2)$$

We set the thresholds to 0.5, 0.5, and 0.9 for each component.

### Deep Learning Approach

We further proposed a deep learning model that combines the text relevance, the likelihood of question presence, and the likelihood of future action presence to identify influence relationships. The architecture of this model is shown in Figure 4.

Compared with the *baseline* approach, there are 3 major benefits of using a deep learning model. First, it is labor-intensive, time-consuming, and difficult to determine appropriate thresholds for cutting off the probabilities using a rule-based approach such as the *baseline* approach. A threshold that works well for one data set may not be optimal for another. Both a rule-based approach and a deep learning model require different thresholds for different data sets. A rule-based approach requires manual parameter tuning for each data set. In contrast, a deep learning approach learns thresholds from the ground truth and, thus, can easily adapt to a new data set with minimal human intervention [42]. Second, the question and action features may have different interactions with the relevance features. We observed that questions are often relevant, but actions are not necessarily. People typically express appreciation in post  $p_C$  or sometimes even mention actions totally irrelevant to post  $p_B$ , such as the plan to travel or shop. Being relevant is more important to consider in the presence of actions compared with in the presence of questions. However, in the *baseline* approach, the question and action features are merged before being combined with the relevance features, resulting in the loss of important information. Furthermore, we used relevance vectors as inputs to the deep learning model to calculate the influence score. Compared with the *baseline* approach, which uses the relevance scores as input to measure the influence score, relevance vectors provide much richer information. This can be especially helpful when there are several topics involved in the discussion. The relevance information is also leveraged during the phase of combining the relevance features with the question or action features.

Let  $V_{AB}$  denote the relevance vector between  $p_A$  and  $p_B$  and  $V_{BC}$  denote the relevance vector between  $p_B$  and  $p_C$ . We generated  $V_{AB}, V_{BC}$  from  $p_A, p_B$ , and  $p_C$  and calculated  $Q$  and  $A$  from  $p_C$ .

These features were then connected. The question or future action in  $p_C$  must be related to the content of  $p_A$  and  $p_B$ . Thus, we combined  $V_{AB}$  and  $V_{BC}$  with  $Q$  and  $A$  using one of the following two operators: (1) *cat* (concatenating each relevance vector with question or action probability) and (2) *dot*

(multiplying each relevance vector with question or action probability).

There are 2 major differences between these 2 operators for connecting the features: *cat* and *dot*. First, *dot* makes sure that  $Q$  and  $A$  affect each dimension in the relevance vectors, whereas *cat* cannot guarantee this as some neurons or nodes are dropped out. Some interactions between questions or actions and text relevance may be ignored by the *cat* operator. Second, the training process of the *cat* is more expensive than that of the *dot* because, for each dense layer 1 to 4, there is an additional dimension for the *cat* compared with for the *dot*.

In Figure 4, we use  $\otimes$  to present the combination operator, which can be either *cat* or *dot*. The combination step produces 4 feature vectors:  $V_{AB} \otimes Q$ ,  $V_{AB} \otimes A$ ,  $V_{BC} \otimes Q$ , and  $V_{BC} \otimes A$ . To extract the key information from these combined feature vectors, 4 dense (fully connected) layers were used to populate the summarized feature vectors ( $S_1, S_2, S_3, S_4$ ). The concatenation of these 4 summarized feature vectors was passed through 2 dense layers. The first one was used to further combine the summarized feature vectors. The second one aimed to generate the probability distribution over the labels. To avoid gradient vanishing and exploding [43], we chose the *Relu* function as the activation function for all the dense layers except the output layer, which uses the *softmax* function to populate the probabilities.

We trained the model using the binary cross-entropy loss function defined in Equation 3, which minimizes the distance between the probability distributions of the ground truth and those of the predicted score.

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n y_i \cdot \log(s_i) + (1 - y_i) \cdot \log(1 - s_i) \quad (3)$$

Where  $y_i$  is the ground truth label of the  $i$ th training sample and  $s_i$  is the score predicted by the model. The Adam optimizer [43] was leveraged for optimization because of its advantage of processing sparse features and obtaining faster convergence compared with the normal stochastic gradient descent with momentum.

### Ethics Approval

All materials were obtained from anonymous open-source data. Thus, ethics approval was not required.

## Results

### Experiment Setting and Evaluation Metrics

We implemented a prototype system for influence relationship identification on discussion threads. The prototype system and data sets used in the evaluation are publicly available at GitHub [44].

For empirical evaluation, we collected 25,208 threads that were publicly available in the CSN breast cancer forum [9]. The webpages were collected and processed by a web crawler we developed leveraging the Spider Crawler library [45]. There were 321,000 posts with 1.9 million sentences in total. We applied the classifier proposed by Li et al [1] on all 25,208

threads to identify the ones that were related to patient decision-making and obtained 11,815 (46.87%) such threads. Note that other models for classifying decision-making threads can also be plugged in.

We then extracted relationships from the decision-making threads using the relationship extraction module and obtained 9053 relationships. We randomly picked 853 (9.42%) of them to label. A total of 4 PhD students worked on the manual labeling. All the relationship triples and post pairs were first independently labeled. In case of disagreement, a consensus was reached after discussion. A total of 261 relationships were labeled as influence relationships. Recall that, per definition 1, each relationship is presented as a triple  $(p_A, p_B, p_C)$ . We also labeled whether posts  $p_A$  and  $p_B$  were relevant (ie,  $P_{AB}$ ) and whether posts  $p_B$  and  $p_C$  were relevant (ie,  $P_{BC}$ ). We observed some reply posts with content expressing only comfort or wishes. Although they express care about the initial author's conditions and seem relevant, they are generic. After discussion, we reached an agreement that, when the initial post and reply post shared similar medical terms (such as *chemotherapy* and *chemo*), we would label them as relevant. All 1706 post pairs  $(p_A, p_B)$  and  $(p_B, p_C)$  of the 853 relationships were labeled. Of the 1706 pairs, 1210 (70.93%) were relevant pairs, and the remaining 496 (29.07%) were irrelevant. We split the set of relationships into a training set (90%) and a testing set (10%). The post pairs in the aforementioned training and test sets were used for text relevance training and testing, respectively.

The metrics used for evaluation included precision, recall,  $F_1$  score, accuracy, area under the receiver operating characteristic curve (ROC AUC), and area under the precision-recall curve

(PR AUC). They evaluated the effectiveness of a system using different aspects: (1) *precision*, also known as positive predictive value, is the fraction of relevant instances among the retrieved instances; (2) *recall*, also known as sensitivity, is the fraction of relevant instances that are retrieved among all relevant instances; (3)  $F_1$  score measures a model's performance by calculating the harmonic mean of the precision and recall, as shown in the following equation:  $F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$  (4); (4) *accuracy* is a common evaluation metric for binary classification problems and is defined as the fraction of corrected predictions among the total number of predictions; (5) *ROC AUC* is a common evaluation metric for binary classification problems and is created by plotting the true positive rate against the false positive rate at various threshold settings; and (6) *PR AUC* is commonly used to evaluate the performance of a model on data sets with imbalanced labels.

### Text Relevance Evaluation

Table 1 presents the classification results of the text relevance measurement module. In total, 2 observations were made. The first observation was that the models using BERT achieved high recall but low precision, whereas the models with word-embedding vectors trained on OHC data obtained balanced precision and recall values. There are 2 reasons for these results. First, OHC data are domain-sensitive and can benefit from domain-specific word representation. Second, the BERT transformer tends to link words in adjacent sentences by mistake. In the text relevance measurement module, precision was more important than recall as the accuracy of influence relationship identification depended on the precision of relevance classification. Thus, we used the word vectors trained on OHC data instead of BERT in the following experiments.

**Table 1.** Text relevance measurement module results.

	Precision	Recall	$F_1$	Accuracy	ROC AUC <sup>a</sup>	PR AUC <sup>b</sup>
MatchPyramid with BERT <sup>c</sup> (trained on Wikipedia)	0.578	0.992 <sup>d</sup>	0.730	0.512	0.502	0.583
MatchPyramid with word2vec (trained on the training data set)	0.781	0.820 <sup>d</sup>	0.806	0.692	0.763	0.854
ARC-I <sup>e</sup> with BERT (trained on Wikipedia)	0.523	0.890 <sup>d</sup>	0.659	0.503	0.493	0.554
ARC-I with word2vec (trained on the training data set)	0.832	0.747 <sup>d</sup>	0.785	0.784	0.848	0.903

<sup>a</sup>ROC AUC: area under the receiver operating characteristic curve.

<sup>b</sup>PR AUC: area under the precision-recall curve.

<sup>c</sup>BERT: Bidirectional Encoder Representations from Transformers.

<sup>d</sup>The  $P$  value is statistically significant at  $P=.05$ .

<sup>e</sup>ARC-I: Architecture-I.

The second observation was that, with word vector embedding, ARC-I achieved a better performance than MatchPyramid in most of the evaluation metrics. In the ARC-I model, each input text goes through an embedding layer, a convolution layer, and a max pooling layer, and the extracted feature vectors are then concatenated together as the input to a fully connected layer that calculates the predicted relevance scores. MatchPyramid populates the local word relevance matrix first. Each cell of the matrix presents the dot product of the word-embedding vectors of the words in the text input. The patterns of these interactions

are then extracted using a convolutional neural network [46]. Thus, ARC-I focuses on checking relevance based on the meaning of the whole text, whereas MatchPyramid focuses on summarizing the important relevance features based on local word similarity. For OHC data sets, posts were relatively long and often contained noisy information; thus, considering the meaning of the entire post text was more important than focusing on adjacent words. This is why the performance of ARC-I was better than that of MatchPyramid in our evaluation. We also observed that ARC-I with word2vec outperformed



*MatchPyramidwith word2vec* in both *ROC AUC* and *PR AUC* but had an inferior  $F_1$  score. Note that  $F_1$  averages the performance of all the samples by combining the precision and recall, whereas the *ROC AUC* and *PR AUC* cumulate the precisions among all samples with different recall thresholds. This indicates that the average performance of *MatchPyramidwith word2vec* was better, but the overall performance of *ARC-Iwith word2vec* was better.

### Question and Action Probability Evaluation

Now, we present the evaluation of the question probability calculation module and the action probability calculation module. The performance is shown in [Table 2](#). Good performance was achieved for question identification. For future action identification, a high score was achieved on recall but not on precision. The following are a few examples of posts

that are classified as containing future actions but actually do not have action intent: *I will tell you though I hated my silicone* or *I would worry about it*. These sentences have verbs in the future tense, but those verbs only convey opinions or feelings rather than taking action on health care. We plan to improve action detection by training action sentence models as future work.

Recall that in the *deep learning* approach, question and action probabilities are considered as input features instead of imposing a strict requirement on their presence. We conducted an analysis on the test data in terms of their presence. All positive cases either had a probability of action presence of 1.0 or had a high probability of question presence, with an average probability of 0.986 (SD 0.033). This indicates that the *deep learning* approach captures definition 2 well, ensuring the high likelihood that either a question or a future action is present.

**Table 2.** Question and action calculation module results.

	Precision	Recall	$F_1$	Accuracy	ROC AUC <sup>a</sup>	PR AUC <sup>b</sup>
Question probability calculation module	1.000	1.000 <sup>c</sup>	1.000	1.000	1.000	1.000
Action probability calculation module	0.771	1.000 <sup>c</sup>	0.871	0.810	0.733	0.771

<sup>a</sup>ROC AUC: area under the receiver operating characteristic curve.

<sup>b</sup>PR AUC: area under the precision-recall curve.

<sup>c</sup>The *P* value is statistically significant at  $P=.05$ .

### Influence Relationship Classification Evaluation

[Table 3](#) shows the performance of the *baseline* and *deep learning* approaches with alternative ways to combine text relevance vectors, question features, and action features. Recall that, for the feature combination module, *baseline* combines the text relevance score, the likelihood of question presence, and the likelihood of future action presence to identify influence

relationships. *MatchPyramid+cat Q/A* represents the model using *MatchPyramid* to calculate the text relevance score and *cat* as the combination operator  $\otimes$ , whereas *MatchPyramid+dot Q/A* uses *dot* as the combination operator  $\otimes$ . *ARC-I+cat Q/A* represents the model using *ARC-I* to calculate the relevance score and *cat* as the combination operator  $\otimes$ , whereas *ARC-I+dot Q/A* uses *dot* as the combination operator  $\otimes$ .

**Table 3.** Influence relationship classification results.

	Precision	Recall	$F_1$	Accuracy	ROC AUC <sup>a</sup>	PR AUC <sup>b</sup>
Baseline	0.300	0.231 <sup>c</sup>	0.261	0.595	0.495	0.307
MatchPyramid+cat Q/A <sup>d</sup>	0.667	0.154 <sup>c</sup>	0.25	0.714	0.560	0.442
MatchPyramid+dot Q/A <sup>e</sup>	0.633	0.577 <sup>c</sup>	0.603	0.667	0.634	0.481
ARC-I+cat Q/A <sup>f</sup>	0.667	0.154 <sup>c</sup>	0.25	0.714	0.637	0.515
ARC-I+dot Q/A <sup>g</sup>	0.750	0.462 <sup>c</sup>	0.571	0.786	0.724	0.631

<sup>a</sup>ROC AUC: area under the receiver operating characteristic curve.

<sup>b</sup>PR AUC: area under the precision-recall curve.

<sup>c</sup>The *P* value is statistically significant at  $P=.05$ .

<sup>d</sup>MatchPyramid+cat Q/A: model using *MatchPyramid* to calculate the text relevance score and *cat* as the combination operator  $\otimes$ .

<sup>e</sup>MatchPyramid+dot Q/A: model using *MatchPyramid* to calculate the text relevance score and *dot* as the combination operator  $\otimes$ .

<sup>f</sup>ARC-I+cat Q/A: model using *Architecture-I* to calculate the relevance score and *cat* as the combination operator  $\otimes$ .

<sup>g</sup>ARC-I+dot Q/A: model using *Architecture-I* to calculate the relevance score and *dot* as the combination operator  $\otimes$ .

We also visualized the operating characteristic curves of all methods, as shown in [Figure 5](#). From [Table 3](#) and [Figure 5](#), we have the following observations.

First, all proposed *deep learning* methods, which use relevance features and consider the interaction between relevance and the presence of questions or actions, significantly outperformed the

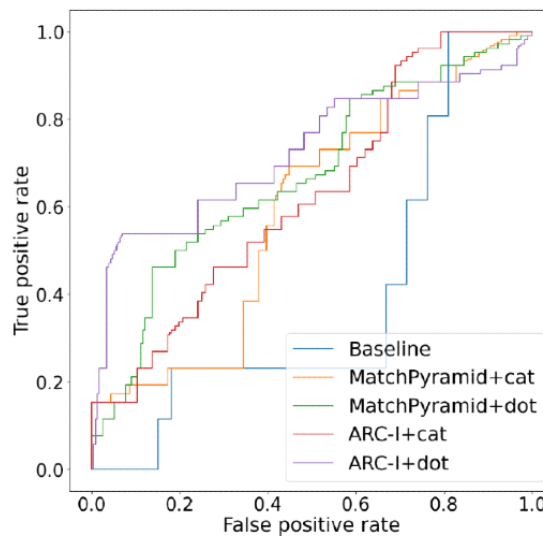
*baseline* approach. This indicates that the relevance feature vectors generated by the text relevance measurement module were effective in capturing relevant content. Combining these feature vectors with the features of question presence and action presence helped capture their interactions and achieved good performance in influence relationship classification. In contrast, the *baseline* approach, which directly follows definition 2, did not perform well. This was due to the inability to capture the interactions between text relevance and question or action presence and the challenge of manually setting an appropriate cutoff threshold for each module.

Second, the models using the *dot* operator performed better than those using the *cat* operator. There are mainly 2 reasons for this. First, question probability and action probability may interact with  $V_{AB}$  and  $V_{BC}$  relevance vectors, which can be captured well by the *dot* operator. Figure 1B shows an example in which the action in  $p_C$  is related to the discussion in  $p_A$  and  $p_B$ . The action in  $p_{C2}$  is related to *chemo*, which is the common content of  $p_A$  and  $p_{B2}$ . In this case, the action probability needs to be combined with  $V_{AB}$ . Although, in another case, the action refers to an option mentioned in  $p_B$ , the interaction between  $p_B$  and  $p_C$  is

more likely to be the context of the action and, thus, the action probability needs to be combined with  $V_{BC}$ . In contrast, the *cat* operator ignores some interactions between questions (actions) and the context because of the dropout of some neutrals. Therefore, the *cat*-based methods had a much lower recall than the *dot*-based methods. The results show that interactions between action and context are important for influence identification.

Furthermore, the *ARC-I+dot Q/A* had a much better precision, accuracy, ROC AUC, and PR AUC than *MatchPyramid+dot Q/A* but had lower recall and slightly lower  $F_1$ . This is because *ARC-I* achieved a better performance than *MatchPyramid* in the text relevance measurement module. *ARC-I+dot Q/A* was stricter than *MatchPyramid+dot Q/A* when fitting the model to the relevance factor. For applications that want to analyze the writing style and patterns of posts that have influence, precision is critical. *ARC-I+dot Q/A* is effective for locating such discussions. In contrast, for applications that want to check the information quality of the posts that have influence to prevent and mitigate the spread of misleading information, *MatchPyramid+dot Q/A* is more suitable because of its higher recall.

Figure 5. Influence relationship classification.



### A Case Study

Figure 1 shows an example of 3 relationships,  $(p_A, p_{B1}, p_{C1})$ ,  $(p_A, p_{B2}, p_{C2})$ , and  $(p_A, p_{B3}, p_{C3})$ , where  $(p_A$  is the initial post of the thread. The scores of these 3 relationships calculated using our system were 0.282, 0.793, and 0.622, respectively. Our system identified  $(p_A, p_{B2}, p_{C2})$  and  $(p_A, p_{B3}, p_{C3})$  as each containing an influence relationship, and  $(p_A, p_{B1}, p_{C1})$  does not. As we can see from the post content,  $p_{B2}$  provides suggestions to the initial author regarding the treatment decision. In post  $p_{C2}$ , the initial author expresses actions to take based on the suggestions in  $p_{B2}$ . In post  $p_{B3}$ , the replier recommends that the author use wigs. The initial author then asks further questions about the wig information. Both relationships indicate that the initial author was influenced. In contrast,  $p_{B1}$  discusses general information and comforts the initial author, and the initial author

expresses thanks in  $p_{C1}$ , but there is no indication of being influenced.

## Discussion

### Principal Findings

To the best of our knowledge, this is the first study that defines the influence relationships of discussion posts related to decision-making in OHCs. We proposed a deep learning-based natural language processing prototype to identify influence relationships. We then applied the developed techniques to identify the influence relationships in an OHC, the CSN breast cancer forum. There were 2 major observations.

First, we found that there is a significant amount of influence relationships in the OHC. Of the 9052 relationships in decision-making threads identified by Li et al [1], 3069 (33.9%)

were identified as influence relationships. That is, approximately one-third of the communications influence the initial authors on their decision-making. Furthermore, of the 5143 decision-making threads, which have at least one relationship, 2417 (47%) contain at least one influence relationship. Owing to the prevalence, it is important to study posts that have influence.

Second, we also observed that posts that have influence may contribute to engaging users in discussions. The average number of posts in threads containing at least one influence relationship was 15.5, whereas the average number of posts in threads containing no influence relationship was 12.6. Our conjecture is that posts that have an influence likely provide helpful information or good reasoning, which are thought-provoking and help engage users in discussions.

On the basis of these observations, there are several applications that can benefit from the identification and analysis of influence relationships.

First, analyzing the quality of posts that have influence helps improve the quality of the influence. As discussed in the first observation, influence relationships are common. Quality checking of those posts is more critical than that of other posts in terms of improving the effect of influences and mitigating the spread of misleading information.

On the basis of the identification of influence relationships, we can further identify influential users in OHCs. We can use existing techniques that analyze the network features to identify influential users [16-20], where this work calculates the edge weights (ie, the influence of a post). Identifying and checking influential users contributes to high-quality information dissemination.

Second, based on the second observation, analyzing the writing style of posts that have influence provides insights to health care professionals about effective communication for patient engagement.

Furthermore, identifying influence relationships contributes to effective information recommendations for addressing the information overload problem. When a user searches for information in OHCs, it is important to rank discussion threads and posts and recommend to users the most relevant and helpful discussions. On the basis of the analysis of influence relationships and the second observation, discussions that contain influence relationships are more likely to provide helpful information and encourage patient engagement. Thus, the presence of influence relationships is a positive factor in ranking.

### Limitations

Our results are not without limitations. First, our definition of relationship was based on 3 posts, including the initial post in the thread. Therefore, we only identified the posts that had an influence on the initial author. However, any 3 posts that have a sequential reply relationship with the first and third posts from the same author can represent a relationship. We conjecture that the proposed techniques can be used to identify influence relationships among the generalized relationships and plan to study that problem in the future. Second, in this study, we considered text relevance between the posts in the relationship. Sometimes, even though 2 posts,  $p_B$  and  $p_C$ , are relevant overall, the specific sentence that has a question or future action indication in  $p_C$  may not be relevant to the suggestions in  $p_B$ . In addition, the current technique for future action detection sometimes generates false positives. To address these issues, we will investigate how to leverage part-of-speech and reference resolution techniques [47] to improve natural language understanding.

### Conclusions and Future Work

We studied the problem of identifying influence relationships of web-based discussions and developed techniques and a prototype system for identifying influence relationships in OHCs. The proposed deep learning model demonstrates the performance advantage of the compared methods. As future work, we will address the aforementioned limitations to improve the generality and accuracy of the proposed techniques.

---

### Acknowledgments

This material is based on work partially supported by the Leir Foundation and a grant from the National Institutes of Health (UL1TR003017).

---

### Conflicts of Interest

None declared.

---

### References

1. Li M, Shi J, Chen Y. Analyzing patient decision making in online health communities. In: Proceedings of the 2019 IEEE International Conference on Healthcare Informatics. 2019 Presented at: ICHI '19; June 10-13, 2019; Xi'an, China p. 1-8. [doi: [10.1109/ICHI.2019.8904879](https://doi.org/10.1109/ICHI.2019.8904879)]
2. Fan H, Smith SP, Lederman R, Chang S. Why people trust in online health communities: an integrated approach. In: Proceedings of the 21st Australasian Conference on Information Systems. 2010 Presented at: ACIS '10; August 18-20, 2010; Yamagata, Japan.
3. Kim HS, Mrotek A. A functional and structural diagnosis of online health communities sustainability: a focus on resource richness and site design features. *Comput Human Behav* 2016 Oct;63:362-372. [doi: [10.1016/j.chb.2016.05.004](https://doi.org/10.1016/j.chb.2016.05.004)]
4. Demiris G. The diffusion of virtual communities in health care: concepts and challenges. *Patient Educ Couns* 2006 Aug;62(2):178-188. [doi: [10.1016/j.pec.2005.10.003](https://doi.org/10.1016/j.pec.2005.10.003)] [Medline: [16406472](https://pubmed.ncbi.nlm.nih.gov/16406472/)]

5. Nasralah T, Noteboom C, Wahbeh A, Al-Ramahi MA. Online health recommendation system: a social support perspective. Dakota State University. 2017. URL: <https://scholar.dsu.edu/cgi/viewcontent.cgi?article=1008&context=bispapers> [accessed 2022-06-10]
6. Ji X, Geller J, Chun S. Social InfoButtons for patient-oriented healthcare knowledge support. In: Proceedings of the 2014 AMIA Annual Symposium. 2014 Presented at: AMIA '14; November 15-19, 2014; Washington, DC, USA.
7. Waterworth S, Luker KA. Reluctant collaborators: do patients want to be involved in decisions concerning care? *J Adv Nurs* 1990 Aug;15(8):971-976. [doi: [10.1111/j.1365-2648.1990.tb01953.x](https://doi.org/10.1111/j.1365-2648.1990.tb01953.x)] [Medline: [2229694](https://pubmed.ncbi.nlm.nih.gov/2229694/)]
8. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. *Health Aff (Millwood)* 2013 Feb;32(2):207-214. [doi: [10.1377/hlthaff.2012.1061](https://doi.org/10.1377/hlthaff.2012.1061)] [Medline: [23381511](https://pubmed.ncbi.nlm.nih.gov/23381511/)]
9. Breast Cancer. Cancer Survivors Network. URL: <https://csn.cancer.org/forum/127> [accessed 2022-06-10]
10. Yaraghi N, Wang W, Gao GG, Agarwal R. How online quality ratings influence patients' choice of medical providers: controlled experimental survey study. *J Med Internet Res* 2018 Mar 26;20(3):e99 [FREE Full text] [doi: [10.2196/jmir.8986](https://doi.org/10.2196/jmir.8986)] [Medline: [29581091](https://pubmed.ncbi.nlm.nih.gov/29581091/)]
11. Zhang D, Yue WT. Social media use in decision making: special issue of decision support systems for the 10th workshop on e-business. *Decis Support Syst* 2014 Jul;63:65-66. [doi: [10.1016/j.dss.2013.08.007](https://doi.org/10.1016/j.dss.2013.08.007)]
12. Wu B, Jiang S, Chen H. Effects of individuals' motivations on communications in online health forums. *Soc Behav Pers* 2016 Mar 23;44(2):299-312. [doi: [10.2224/sbp.2016.44.2.299](https://doi.org/10.2224/sbp.2016.44.2.299)]
13. Pan W, Shen C, Feng B. You get what you give: understanding reply reciprocity and social capital in online health support forums. *J Health Commun* 2017 Jan;22(1):45-52. [doi: [10.1080/10810730.2016.1250845](https://doi.org/10.1080/10810730.2016.1250845)] [Medline: [28027009](https://pubmed.ncbi.nlm.nih.gov/28027009/)]
14. Li S, Feng B, Wingate VS. Give thanks for a little and you will find a lot: the role of a support seeker's reply in online support provision. *Commun Monogr* 2019;86(2):251-270. [doi: [10.1080/03637751.2018.1539237](https://doi.org/10.1080/03637751.2018.1539237)]
15. Yu J, Hu Y, Yu M, Di Z. Analyzing netizens' view and reply behaviors on the forum. *Phys A Stat Mech Appl* 2010 Aug;389(16):3267-3273. [doi: [10.1016/j.physa.2010.03.043](https://doi.org/10.1016/j.physa.2010.03.043)]
16. Tang X, Yang CC. Identifying influential users in an online healthcare social network. In: Proceedings of the 2010 IEEE International Conference on Intelligence and Security Informatics. 2010 Presented at: ISI '10; May 23-26, 2010; Vancouver, Canada p. 43-48. [doi: [10.1109/ISI.2010.5484779](https://doi.org/10.1109/ISI.2010.5484779)]
17. Cha M, Haddadi H, Benevenuto F, Gummadi K. Measuring user influence in Twitter: the million follower fallacy. *Proc Conf AAAI Artif Intell* 2010;4(1):10-17.
18. Ishfaq U, Khan HU, Iqbal K. Identifying the influential bloggers: a modular approach based on sentiment analysis. *J Web Eng* 2017;16(5-6):505-523.
19. Zhao K, Qiu B, Caragea C, Wu D, Mitra P, Yen J, et al. Identifying leaders in an online cancer survivor community. In: Proceedings of the 21st Annual Workshop on Information Technologies and Systems. 2011 Presented at: WITS '11; December 3-4, 2011; Shanghai, China p. 115-120.
20. Hou X, Lei CU, Kwok YK. OP-DCI: a riskless K-means clustering for influential user identification in MOOC forum. In: Proceedings of the 16th IEEE International Conference on Machine Learning and Applications. 2017 Presented at: ICMLA '17; December 18-21, 2017; Cancun, Mexico p. 936-939. [doi: [10.1109/icmla.2017.00-34](https://doi.org/10.1109/icmla.2017.00-34)]
21. Zhao K, Yen J, Greer G, Qiu B, Mitra P, Portier K. Finding influential users of online health communities: a new metric based on sentiment influence. *J Am Med Inform Assoc* 2014 Oct;21(e2):e212-e218 [FREE Full text] [doi: [10.1136/amiajnl-2013-002282](https://doi.org/10.1136/amiajnl-2013-002282)] [Medline: [24449805](https://pubmed.ncbi.nlm.nih.gov/24449805/)]
22. Cercel DC, Trausan-Matu S. Opinion influence analysis in online forum threads. In: Proceedings of the 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing. 2014 Presented at: SYNASC '14; September 22-25, 2014; Timisoara, Romania p. 228-235. [doi: [10.1109/synasc.2014.38](https://doi.org/10.1109/synasc.2014.38)]
23. Qiu B, Zhao K, Mitra P, Wu D, Caragea C, Yen J, et al. Get online support, feel better--sentiment analysis and dynamics in an online cancer survivor community. In: Proceedings of the IEEE 3rd International Conference on Privacy, Security, Risk and Trust and IEEE 3rd International Conference on Social Computing. 2011 Presented at: SocialCom '11; October 9-11, 2011; Boston, MA, USA p. 274-281. [doi: [10.1109/PASSAT/SocialCom.2011.127](https://doi.org/10.1109/PASSAT/SocialCom.2011.127)]
24. Definition of influence. Merriam-Webster. URL: <https://www.merriam-webster.com/dictionary/influence> [accessed 2022-06-10]
25. Kidd C, Hayden BY. The psychology and neuroscience of curiosity. *Neuron* 2015 Nov 04;88(3):449-460 [FREE Full text] [doi: [10.1016/j.neuron.2015.09.010](https://doi.org/10.1016/j.neuron.2015.09.010)] [Medline: [26539887](https://pubmed.ncbi.nlm.nih.gov/26539887/)]
26. Palloff RM, Pratt K. Lessons from the Virtual Classroom: The Realities of Online Teaching. 2nd edition. Hoboken, NJ, USA: John Wiley & Sons; 2013.
27. Adjei MT, Noble SM, Noble CH. The influence of C2C communications in online brand communities on customer purchase behavior. *J Acad Mark Sci* 2009 Nov 13;38(5):634-653. [doi: [10.1007/s11747-009-0178-5](https://doi.org/10.1007/s11747-009-0178-5)]
28. Liu Y, Chen F, Chen Y. Learning thread reply structure on patient forums. In: Proceedings of the 2013 International Workshop on Data Management & Analytics for Healthcare. 2013 Presented at: DARE '13; November 1, 2013; San Francisco, CA, USA p. 1-4. [doi: [10.1145/2512410.2512426](https://doi.org/10.1145/2512410.2512426)]
29. Huang PS, He X, Gao J, Deng L, Acero A, Heck L. Learning deep structured semantic models for web search using clickthrough data. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management.

- 2013 Presented at: CIKM '13; October 27-November 1, 2013; San Francisco, CA, USA p. 2333-2338. [doi: [10.1145/2505515.2505665](https://doi.org/10.1145/2505515.2505665)]
30. Shen Y, He X, Gao J, Deng L, Mesnil G. Learning semantic representations using convolutional neural networks for Web search. In: Proceedings of the 23rd International Conference on World Wide Web. 2014 Presented at: WWW '14 Companion; April 7-11, 2014; Seoul, South Korea p. 373-374. [doi: [10.1145/2567948.2577348](https://doi.org/10.1145/2567948.2577348)]
  31. Hu B, Lu Z, Li H, Chen Q. Convolutional neural network architectures for matching natural language sentences. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. 2014 Presented at: NIPS '14; December 8-13, 2014; Montreal, Canada p. 2042-2050.
  32. Wang YC, Kraut R, Levine JM. To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12; February 11-15, 2012; Seattle, WA, USA p. 833-842. [doi: [10.1145/2145204.2145329](https://doi.org/10.1145/2145204.2145329)]
  33. Lu Z, Li H. A deep architecture for matching short texts. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1. 2013 Presented at: NIPS '13; December 5-10, 2013; Lake Tahoe, NV, USA p. 1367-1375. [doi: [10.7551/mitpress/11474.003.0014](https://doi.org/10.7551/mitpress/11474.003.0014)]
  34. Pang L, Lan Y, Guo J, Xu J, Wan S, Cheng X. Text matching as image recognition. Proc Conf AAAI Artif Intell 2016 Mar 5;30(1):10341. [doi: [10.1609/aaai.v30i1.10341](https://doi.org/10.1609/aaai.v30i1.10341)]
  35. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv 2018 Oct 11. [doi: [10.48550/arXiv.1810.04805](https://doi.org/10.48550/arXiv.1810.04805)]
  36. Efron M, Winget M. Questions are content: a taxonomy of questions in a microblogging environment. Proc Am Soc Info Sci Tech 2011 Feb 03;47(1):1-10. [doi: [10.1002/meet.14504701208](https://doi.org/10.1002/meet.14504701208)]
  37. Liu Z, Jansen BJ. Questioner or question: predicting the response rate in social question and answering on Sina Weibo. Inf Process Manag 2018 Mar;54(2):159-174. [doi: [10.1016/j.ipm.2017.10.004](https://doi.org/10.1016/j.ipm.2017.10.004)]
  38. Ranganath S, Hu X, Tang J, Wang S, Liu H. Identifying rhetorical questions in social media. Proc Conf AAAI Artif Intell 2016;10(1):667-670.
  39. Ranganath S, Hu X, Tang J, Wang S, Liu H. Understanding and identifying rhetorical questions in social media. ACM Trans Intell Syst Technol 2018 Mar 31;9(2):1-22. [doi: [10.1145/3108364](https://doi.org/10.1145/3108364)]
  40. Ojokoh B, Igbe T, Araoye A, Ameh F. Question identification and classification on an academic question answering site. In: Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries. 2016 Presented at: JCDL '16; June 19-23, 2016; Newark, NJ, USA p. 223-224. [doi: [10.1145/2910896.2925442](https://doi.org/10.1145/2910896.2925442)]
  41. Bird S, Loper E. NLTK: the natural language toolkit. In: Proceedings of the ACL Interactive Poster and Demonstration Sessions. 2004 Presented at: ACL '04; July 21-26, 2004; Barcelona, Spain p. 214-217.
  42. Goodfellow I, Yoshua B, Courville A. Deep Learning. Cambridge, MA, USA: MIT Press; Nov 2016:4.
  43. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. 2015 Presented at: ICLR '15; May 7-9, 2015; San Diego, CA, USA.
  44. Mingda Li, Jinhe Shi, Yi Chen. NJIT-AI-in-Healthcare / Influence-Relationship-Classification. GitHub. 2022 Feb 27. URL: <https://github.com/NJIT-AI-in-Healthcare/Influence-Relationship-Classification> [accessed 2022-06-10]
  45. Scrapy. URL: <https://scrapy.org/> [accessed 2022-06-10]
  46. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017 Jun;60(6):84-90. [doi: [10.1145/3065386](https://doi.org/10.1145/3065386)]
  47. Liu Y, Shi J, Chen Y. Patient-centered and experience-aware mining for effective adverse drug reaction discovery in online health forums. J Assoc Inf Sci Technol 2017 Sep 22;69(2):215-228. [doi: [10.1002/asi.23929](https://doi.org/10.1002/asi.23929)]

## Abbreviations

- ARC-I:** Architecture-I  
**ARC-II:** Architecture-II  
**BERT:** Bidirectional Encoder Representations from Transformers  
**CSN:** Cancer Survivors Network  
**DSSM:** Deep Structured Semantic Models  
**OHC:** online health community  
**PR AUC:** area under the precision-recall curve  
**ROC AUC:** area under the receiver operating characteristic curve

*Edited by R Kukafka; submitted 03.06.21; peer-reviewed by S Liu, G Lim; comments to author 02.08.21; revised version received 05.10.21; accepted 21.03.22; published 31.08.22*

*Please cite as:*

*Li M, Shi J, Chen Y*

*Identifying Influences in Patient Decision-making Processes in Online Health Communities: Data Science Approach*

*J Med Internet Res 2022;24(8):e30634*

*URL: <https://www.jmir.org/2022/8/e30634>*

*doi: [10.2196/30634](https://doi.org/10.2196/30634)*

*PMID:*

©Mingda Li, Jinhe Shi, Yi Chen. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 31.08.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.