Thesis for the Degree of Doctor of Philosophy

EXTRACTING USER EXPERIENCE DIMENSIONS FROM QUALITATIVE DATA USING UX QUALIFIERS AND TOPIC MODELING

Jamil Hussain

Department of Computer Science and Engineering
Graduate School
Kyung Hee University
South Korea

August 2019

EXTRACTING USER EXPERIENCE DIMENSIONS FROM QUALITATIVE DATA USING UX QUALIFIERS AND TOPIC MODELING

Jamil Hussain

Department of Computer Science and Engineering
Graduate School
Kyung Hee University
South Korea

August 2019

EXTRACTING USER EXPERIENCE DIMENSIONS FROM QUALITATIVE DATA USING UX QUALIFIERS AND TOPIC MODELING

by

Jamil Hussain

Supervised by

Prof. Sungyoung Lee

Submitted to the Department of Computer Science and Engineering and the Faculty of Graduate School of Kyung Hee University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Dissertation Committee:

Prof. Tae-Seong Kim ______________________

Prof. Jee-In Kim ______________________

Prof. Sung-Ho Bae ______________________

Prof. LokWon Kim ______________________

Prof. Sungyoung Lee ______________________

Achievement of each challenging goal is possible due to self-efforts and the guidance of elders, especially those who were close to our heart.

My humble efforts, I dedicated to my beloved

Father & Mother,

Brothers & Sisters,

My beloved wife & Kid

whose endless love, encouragement, support, and prays make me able to achieve such success and honor.

Along with my all respected, hardworking, and supportive

Teachers.

Abstract

To design a good product, it is a crucial issue to identify the opportunities from the user feedback data. These days' customers want an interactive product that brings positive and initiative experiences. Most product designers and companies come to the point that positive user experience (UX) has a more significant impact on product design and success. Therefore, it is important to understand the UX of users and customers. Prior studies mentioned different methods for the UX assessment and evaluation, but still, there is no universal method exists because UX is context dependent, subjective in nature, and dynamic.

Conventional methods use qualitative data collected through traditional methods such as post-task questionnaires or surveys which have predefined question items related to users experience. These methods are essential to obtain the user data, but such methods considered limited dimensions of UX for data collection, that might have a more significant impact on how the user is feeling about the product, system, or service. Additionally, the experimental cost is too high, and the designer may not be free from biasness. To overcome these shortcomings, new methods are designed to extract the usability and UX information from online products distributed platforms in the form of user-generated content (UGC). UGC become a vital source of information for UX aspects extraction for both researcher and companies. A UX researcher tries to extract the information related to product sentiment requested features, bug reports, suggestion, complaints, and others. With these platforms, customers can share their product experience, provide useful suggestions, and comments in the form of online user reviews. These reviews are in the form of spontaneous and insightful user feedback that is usually accessible free of cost anywhere and anytime. User review is not just a summary assessment but also self-reported data of user experience in a real context. The existing literature has shown that user review used by the researchers to ex-

tract information can help to understand the user preference that the user felt during the usage of a product. Despite the availability of the vast amount of online reviews, existing literature primarily focuses on the online ratings, which is in the numerical form and ignore the actual textual context in online reviews. Compared to online ratings, the textual part often contains valuable information that can be used for improving the existing product or making the new innovative product.

However, this vast amount of user reviews are in the unstructured format written in natural language. How these online user reviews help the designer for designing a new product or improve the existing product? How to make a solution that extracts UX information from user reviews that can help to build an innovative product based on the requested features? It is still very challenging for both product designers and researchers to apply text mining techniques to derive the UX insights from the vast amount of UGC data. The sentiment analysis and opinion mining are often used on user reviews to find the opinion toward the product, but the extraction of UX information from user reviews is still limited. Most of the existing studies use topic modeling approach especially LDA for the extraction of latent dimensions along with the regression analysis on the rating data for the verification and validation of the extracted dimensions in the domain of UX. LDA uses an unsupervised generative statistical model to identify latent aspects from the collection of a textual document without any supposition about the text distribution and their syntactic information, which gives the global context of the corpus documents. Combing the online rating and user reviews content analysis by LDA enables the researchers to identify the causal relationship between the extracted dimensions and user satisfaction. However, without prior knowledge, the unsupervised models frequently generate semantically incoherent topics which are hard to understand. To resolve the shortcoming of the unsupervised models, some previous works add domain knowledge in the topic modeling using various approaches, but most of the these models cannot learn knowledge automatically. The main goal of this thesis is to resolve these challenges; we design a comprehensive framework for modeling UX from online reviews. In the proposed method, first, we filter out those reviews which are unrelated to UX domain using UX multi-criteria qualifiers (UXMCQ). Then, we extract the UXDs from the filtered reviews using enhanced topic extraction methodology called UXWE-LDA. UXWE-LDA improves the existing knowledge-based topic models by extracting more domain dependent dimensions in the UX area through UGC. UXWE-LDA combines the topics modeling, especially LDA with word embedding that automatically learns the domain knowledge from a large amount of textual data. The proposed method automatically gains the domain knowledge from the vast amount of documents using co-occurrence and word-embedding word vectors correlation of related data, which gives a more coherent topic. Then, we apply the sentiment analysis on the reviews concerning the extracted UXDs. To measures, the casual relationship of customer sentiment toward each UXDs on user satisfaction, an ensemble neural network based model (ENNM) Method is used [1]. Finally, we map each dimension on the Kano model of satisfaction. The proposed methodology presented in this thesis is evaluated at different levels by performing multiple experiments on various evaluation criteria. Firstly, we evaluate UXMCQ model for domain aspect classification by comparing with two LDA-based approaches in terms of precision, recall, and F-measure. We compare UXMCQ against the results with two LDA-based methods, UXMCQ achieves marginally higher performance. The results generated by these models required supervision for labeling the extracted topics. Our UXMCQ model gives a labeled data based on the domain aspects configuration; therefore, no of the need for manual inspection and labeling. Secondly, the UXWE-LDA model performance is evaluated based on the topic coherence by comparing with baseline topic models. UXWE-LDA regularly gives higher topic coherence scores as compared to the baseline models. Thirdly, the sentiment analyzer model employed an ensemble learning method with feature selection approaches efficiently increase the classification performance as compared to baseline classifier. Finally, the results obtained from these evaluations showed significant improvements in terms of accuracy and topic coherence. The presented study has potential implication in product design. It can extract those UX aspects from online reviews that customers are most concerned about. Additionally, they can further know the strengths and weaknesses of the product. This method allows the product designer to understand the different categories of UDXs in terms of the Kano model, which is essential for product enhancement. According to the classification results of UXDs, the priority order of UXDs for developing product enhancement plans can be determined.

Acknowledgement

First and foremost, I render my humble and sincere thanks to the ***Almighty Allah*** for showering HIS blessings upon me. The Almighty gave me the strength, courage, and patience during my doctoral study.

Special thanks to my advisor Prof. Sungyoung Lee who provided me guidance, strength, support, and courage in overcoming the difficult challenges throughout my time as his student. I have learned a lot from him in becoming a productive person in diverse situations. He inspired me with his dynamic personality and his unreserved help and guidance lead me to finish my thesis. He has great role in polishing my skills such as thinking, creativity and technical soundness which are key ingredients for high quality research.

I am grateful to my dissertation evaluation committee for their insight comments and valuable suggestions during the dissertation defense. Their valuable comments improved the presentation and contents of this dissertation.

I am extremely grateful to some of my colleagues who have always provided me time, expertise, and encouragement in my course of research. They were always present to guide me in difficult situations of my PhD duration. I would like to thank Dr Asad Masood Khattak, Dr. Muhammad Fahim, Dr. Wajahat Ali Khan, Dr. Bilal Amin, Dr. Maqbool Hussain, Dr. Muhammad Afzal, Dr. Rahman Ali, Dr. Siddiqi Muhammad Hameed, Dr. Mahmood Ahmad, Dr. Taqdir Ali, Dr. Maqbool Ali, and Shujaat Hussain. They have contributed enormously in successfully performing various academic and personal tasks that confronted me during my stay at South Korea.

I am very thankful to all of my current and former Ubiquitous computing lab fellows and colleagues for their kind support to my personal and academic life at Kyung Hee University. I am highly obliged to brilliant researchers Bilal Ali, Asif Razzaq, Usman Akhtar, Imran Ali, Musarrat Hussain, Fahad Ahmed Satti, Muhammad Sadiq, Asim Abbas, Muhammad Zaki Ansaar, Tae Ho Hur, Jae Hun Bang, Mrs. Seoungae Kim, Dr. Kifayat Ullah, Omar Farooq, Dr. Dildar Hussain, Saeed Ullah, Dr. Ahsan Raza Kazmi, and Zain Abbas. This journey would not have been possible without their support. They contributed a lot in to my personal and academic life to polish myself. Also, I appreciate all my Korean and international friends who worked as a team with me and developed my team work skills and provided me wonderful memories during my stay in South Korea.

Last but not the least, I would like to express my sincere gratitude to my mother, sisters, and brothers for their endless love, support, prayers, and encouragement. Their support and encouragement has made this dissertation possible. I would like to thank my wife Mrs. Zainab Jamil for their love, support and guidance in every difficult situation I faced during all this time. Finally, I would like to thanks of my father and brothers for his encouragement in all of my pursuits and inspiration to follow my dreams.

Jamil Hussain

August, 2019

Table of Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

1.1 Background

To design a good product; It is necessary to identify the opportunities from the user feedback data analysis is still a crucial issue. These days' customers want an interactive product that brings positive and initiative experiences. Most product designers and companies come to the point that positive user experience (UX) has a greater impact on product design and success [2, 3]. So it is essential to understand the UX of users and customers. In academia and industry, different methods used to examine and evaluate UX of product, system, or services, but still, there is no universal method exists because UX is context dependent, subjective in nature, and dynamic. In general, UX is all about user feeling about a product, system, or services [4, 5]. According to ISO 9241-11:2018(E) [6], "person's perceptions and responses resulting from the use and anticipated use of a product, system or service." UX influences from many factors such as user mental and physical state, product, and context of use that occur before, during and after use [6]. Many studies have proposed that positive UX play a vital role in motivating user loyalty such as a recommendation to other peoples, writing positive reviews or continuous the product usage. Many factors contribute to making a positive UX (e.g., user satisfaction, quality, fun, enjoyment, ease of use, and others). Most of the prior studies usually used traditional methods such as questionnaires, survey, report grand techniques (RGT) in field and lab studies to evaluate UX by crafting various scenarios [7–9]. In the scenario, they define different tasks and context-of-use while participants interact with the product [4]. They often require many efforts such as task arrangement, participants selection, selection of UX evaluation methods and training, and cost involving in the collection of sample data. These methods are essential to collect the user/customer data, but such approaches considered limited aspects UX for data collection, that might have a more significant impact on how the user is feeling about the product, system, or services. Moreover, in prior studies measurement items used in the survey developed based on the researchers knowledge which might inconsistent and most ignore the perspective of end user. In literature, we found studies [10, 11] that have designed new successful products based on the actual user experience by gathering a rich user experience data (e.g., user feeling, preferences, thought, and beliefs). However, the experimental cost is too high, and the designer may not be free from biased. To overcome these shortcomings, new methods [11–15] in the literature to extracts the usability and UX information from online products distributed platforms in the form of user-generated content (UGC). UGC become a vital source of information for UX aspects extraction for both researcher and companies. The researcher tries to extract the information related to product sentiment [16], requested features, bug reports, suggestion, complaints and others [17]. With these platforms, customers can share their product experience, provide useful suggestions, and comments in the form of online user reviews [12]. These reviews are in the form of spontaneous and insightful user feedback that usually accessible free of cost anywhere and anytime. User review is not just a summary assessment but also self-reported data of user experience in a real context. The existing literature has shown that user review used by the researchers to extract information can help to understand the user preference that the user felt during the usage of a product.

Additionally, user reviews obtained from diverse users, which contains different user opinions at the different context of use. For example, different users have different performance and experience for the same products. Such a diverse user experience gives a complete picture to design a new product by considering the influencing factors. Despite the availability of the vast amount of online reviews, existing literature primarily focuses on the online ratings, which is in the numerical form and ignore the actual textual context in online reviews. Despite the availability of the vast amount of online reviews, existing literature primarily focuses on the online ratings, which is in the numerical form and ignore the actual textual context in online reviews. Compared to online ratings, the textual part often contains valuable information related to features requested , bug reports, and much more, which can help to improve the product. However, this vast amount of user reviews in the unstructured form written in natural language. How to use these user reviews for designing new design product or improve the existing product? How to make a solution that extracts UX information from user reviews that can help to build an innovative product based on the requested features? It is still very challenging for both product designers and researchers to apply text mining techniques to derive the UX insights from large UGC data. The sentiment analysis and opinion mining are often used on user reviews to find the opinion toward the product [18], but the extraction of UX information from user reviews is still limited [12]. An evolving research in the UX area has tried to know about user satisfaction from the online reviews. These studies can be classified into two categories, (1) mining the user experience dimensions (UXDs) from online reviews [14], and (2) modelling UX from online reviews [19]. In the first category, the researcher uses the numerous text mining techniques for the extractions of different UXDs using probabilistic topic models: Latent Dirichlet Allocation (LDA) and Probabilistic Latent Semantic Analysis (PLSA) [1, 14, 15]. and analyzing the relative importance of each UXDs. For example, Tirunillai and Tellis [20] proposed a framework for extracting the UXDs from online reviews by an improved latent Dirichlet allocation (LDA). Yue Guo (2017) [14] used data having 266,544 online reviews using topic modeling and content analysis to find out user satisfaction. Most of the existing studies [1, 10, 14, 15] use topic modeling approach especially LDA for the extraction of latent dimensions along with the regression analysis on the rating data for the verification and validation of the extracted dimensions in the domain of UX. LDA uses an unsupervised generative statistical model to identify latent aspects from the collection of a textual document without any supposition about the text distribution and their syntactic information, which gives the global context of the corpus documents. Combing the online rating and user reviews content analysis by LDA enables the researchers to identify the causal relationship between the extracted dimensions and user satisfaction. However, without prior knowledge, the unsupervised models frequently generate semantically incoherent topics which are hard to understand [21, 22]. To resolve the shortcoming of the unsupervised models, some previous works add domain knowledge in the topic modeling using various approaches, but most of the models cannot learn knowledge automatically [23]. These days researchers try to use other word representation scheme as word embedding into topic modeling that reduces the dimensionality of word vector based on the co-occurrence information, considering the local context of words. By combining the global and local context gives more coherence topics. To resolve these challenges, we designed a comprehensive framework for modeling UX from online reviews. In the method, first, we filter those reviews unrelated to UX domain using UX multi-criteria qualifiers. Then, we extract the UXDs from the filtered reviews using enhanced topic extraction methodology called UXWE-LDA. UXWE-LDA improve the existing knowledge-based topic models by for the extractions of more domain dependent dimensions in the UX area through UGC. UXWE-LDA combines the topics modeling especially LDA with word embedding that automatically learns the domain knowledge from a large amount of textual data. The proposed method automatically gains the domain knowledge from the vast amount of documents using co-occurrence and word-embedding word vectors correlation of related data, which gives a more coherent topic. Then, we apply the sentiments analysis on the reviews concerning the extracted UXDs. To measures, the casual relationship of customer sentiment toward each UXDs on user satisfaction, an ensemble neural network based model (ENNM) Method is used [1]. Finally, we map each dimension on the Kano model of satisfaction. More Specially, the contributions are made in three parts:

- UX Multi-Criteria Qualifiers (UXMCQ) identify those reviews which contain useful information related to UX. This step is essential to remove trivial reviews before applying the topic modeling. The UXMCQ classify the online review based on the predefined UX aspects (user facets, situation facets, and product facets). It contains two important steps i) aspects configuration as seed word for auto labeling for model training ii) Novel methodology for feature construction and selection for enhancing the modeling accuracy.

- UXDs extraction from online reviews using propose user experience word-embedding LDA (UXWE-LDA) methodology. UXWE-LDA automatically learn the domain knowledge from the given text corpus for the generation of a more coherent topic. It contains mainly two steps i) UXWE-LDA is an improved version of LDA, that automatically learn the domain knowledge from the given text corpus and extract more coherent topics and assign labels as UXD to each extracted topic using dictionary based approach ii) Identifying the sentiment orientations of the reviews concerning each UXDs based on ensemble methodology. It classifies each review into positive or negative sentiment and associates the sentiment orientation with the extracted UXDs.

- The causal relationship of sentiments toward each UXDs on user satisfaction obtains from using existing ensemble neural network based model (ENNM) proposed by [1], which overcome the problem of existing models used to the modeling of user satisfaction from online reviews.

Our aim to extract essential aspects inducing positive UX using UGC data. By extracting UXDs from UGC data, allows to understand the customer preferences and needs more efficiently, so that the owner of product and investors can improve their product, system, or service.

1.2 Motivation

In literature, we found studies that have designed new successful products based on the actual user experience by gathering a rich user experience data (e.g., user feeling, preferences, thought, and beliefs). However, the experimental cost is too high, and the designer may not be free from biased. To overcome these shortcomings, new methods [12] in the literature to extracts the usability and UX information from a social media platform, user-generated content (UGC). UGC become a significant source of information for UX aspects extraction for both researcher and companies. The researcher tries to extract the information related to product sentiment, requested features, bug reports, suggestion, complaints and others. With these platforms, customers can share their product experience, provide useful suggestions, and comments in the form of user reviews. These reviews are in the form of spontaneous and insightful user feedback that usually accessible free of cost anywhere and anytime. User review is not just a summary assessment but also self-reported data of user experience in their own word in a real context. The existing literature has shown that user review used by the researcher to extract information can help to understand the user preference that the user felt during the usage of a product.

In this thesis, we focus on the dimensions extractions from online user reviews for modeling the user experience in the form of user satisfaction. The proposed methodology utilizes online user reviews to drive the useful insights related to UXDs and identify the causal relationships with extracted dimension on user satisfaction.

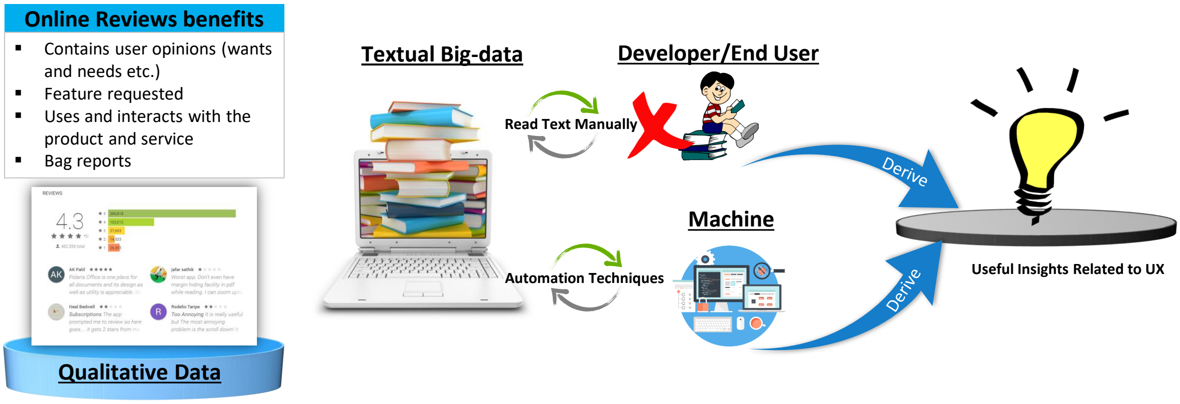


Figure 1.1: Motivations for extracting useful insights from user online reviews

1.3 Problem Statement

The user online reviews most part in the form of unstructured text. How to extract meaningful and essential information from these unstructured user online reviews? How to make a system/approach that automatically mines the UX information from reviews to help the product designer to build a new product based on these insights. It is still challenging for both designer and companies to analyze the vast amount of user online reviews to get informed decision about the product design. Most of existing methods use the unsupervised topic models for the extraction of latent dimensions. However, these unsupervised models without prior knowledge often generated semantically incoherent topics which are hard to understand. To resolve the shortcoming of LDA, prior works integrate domain knowledge in the topic modeling using various approaches, but most of the models cannot learn knowledge automatically [23]. To resolve these challenges, we designed a comprehensive framework for modeling UX from online reviews by attempting to mine the essential aspects of influencing user satisfaction. By mining information from online user reviews, enables product designer or developer can effectively understand the user requeirements to develop more innovate product. . Specially, the main aim of this thesis is to clarify the following questions:

- How do extract the expected dimensions using topic modeling approach by incorporating the domain knowledge?

- What are the essential dimensions of UX expressed in online reviews?
- What are the essential UX dimensions influencing user satisfaction based casual effect analysis?

We proposed a flexible and robust methodology to extract the more related UX dimensions from user online reviews for user satisfaction modeling. It provides the following three steps solutions to achieve the goal as mentioned earlier and to address the above challenges.

- UX Multi-Criteria Qualifiers (UXMCQ) identify those reviews which contain useful information related to UX. This step is essential to remove trivial reviews before applying the topic modeling. The UXMCQ classify the online review based on the predefined UX aspects (user facets, situation facets, and product facets). It contains two important steps i) aspects configuration as seed word for auto labeling for model training ii) Novel methodology for feature construction and selection for enhancing the modeling accuracy.
- UXDs extraction from online reviews using propose user experience word-embedding LDA (UXWE-LDA) methodology. UXWE-LDA automatically learn the domain knowledge from the given text corpus for the generation of a more coherent topic. It contains mainly two steps i) UXWE-LDA is an improved version of LDA, that automatically learn the domain knowledge from the given text corpus and extract more coherent topics and assign labels as UXD to each extracted topic using dictionary based approach ii) Identifying the sentiment orientations of the reviews concerning each UXDs based on ensemble methodology. It classifies each review into positive or negative sentiment and associates the sentiment orientation with the extracted UXDs.
- The causal relationship of sentiments toward each UXDs on user satisfaction obtains from using existing ensemble neural network based model (ENNM) proposed by [1], which overcome the problem of existing models used to the modeling of user satisfaction from online reviews.

1.4 Thesis Organization

This dissertation is organized into chapters as following.

- **Chapter 1: Introduction**. Chapter 1 provides brief introduction of the research work on UX experience model for UX assessment particularly UX structural model.In UX structural model, focus on the UX dimensions in existing literature, and their extraction methodologies.

- **Chapter 2: Related Work**. A background detail is provided in this chapter about the user experience and usability. User real expereience data in the form of user reviews,which contains useful information related to UX. This chapter also provides the state-of-the-art literature for the knowledge extracion from those reviews.

- **Chapter 3: Proposed Methodology**. A proposed solution in the form of a framework for achieving more coherent topic extraction methodology for UX dimension extraction from user reviews is presented in this chapter to overcome the limitations of current approaches.

- **Chapter 4: Usefulness of reviews**. The UX multi-criteria Qualifiers UXMCQ filter out those reviews which contain useful information related to UX. This step is very important to remove the un important text before apply the topic modeling.

- **Chapter 5: UXWE-LDA: Topic Extractor Model**. This chapter will explain about *Topic Extractor Model* to extract the expect UX dimension by automatically incorporating UX domain knowledge.

- **Chapter 6: Results and Evaluation**. The results and evaluation of different techniques used in the proposed framework are highlighted in this chapter. It explains two types of results and evaluation.

- **Chapter 7: Conclusion and Future Directions**. This chapter concludes the thesis and also provides future directions in this research area. The main contribution of the thesis is also highlighted in this chapter.

Chapter 2

Related Work

2.1 User Experience

The user experience (UX) is a multi-faceted research area that includes diverse aspects of the experiential and effective use of a product, system or service [24, 25]. A UX assessment helps uncover the important aspects of designing high-quality interactive products and providing an overall positive UX [26]. The UX involves user beliefs, preferences, thoughts, feelings, and behaviors when interacting with the product, system, or service [24]. It is thus subjective by nature, highly dependent on the use context [27], and linked to the potential benefits obtained from the product, system, or service [28]. The UX is measured using different constructs related to the usability (perspicuity, efficiency, etc.), user perception (stimulation, dependability, novelty, etc.), and human emotional reaction [29] using various methods. For example, a user's feelings can be captured if the user "thinks aloud" while performing tasks. Similarly, the UX can also be interpreted by means of a daily diary over a certain period, such as a long-term diary study [30], day reconstruction method [31], repertory grid technique (RGT) [32], and experience sampling method (ESM) [33].

Furthermore, UX plays a vital role in system design and interaction value, which is measured using diverse methods and tools. The appropriate method selection is one of the significant challenges for UX expert based on the context of use. Based on UX assessment, UX experts and practitioners benchmark the products with a competitive product and enhance their product based on the UX recommendation assessed through various measurement methods or tools. The main concern is how to validate the UX constructs or dimensions and how to measures those constructs or dimensions. Therefore, modeling UX is the baseline for creating great design and products [1, 14]. There are two types of UX modeling: measurement models and structural models. Measurement models deal with measuring the constructs using various methods or tools in a specific domain, where the structural models deal to identify the causal relations among the constructs [34]. Comprehensive measurement methods need to establish proper validation for measuring the UX. Then, descriptive or predictive structural models are required for identifying the consequences in between UX constructs for better understanding the user experience for inform decisions to make a better system, product, or service.

UX modeling is one of the vital research areas for efficient product development and quality enhancement [1, 14]. It can also be combined with other models, including quality function deployment (QFD) for improved results. UX can be measured through various techniques. One of the most used methods is customer surveys [19, 35]. There are multiple advantages to customer surveys. However, it required enough time, as well as money. Also, the survey result is heavily dependent on the willingness of the participant, the length of the survey, and the complexity of the questions asked in the survey [36]. The collected data is applicable only for a limited time and in a constraint targeted environment [37]. Therefore, it is essential to contemplate multiple sources for understanding actual user experience and satisfaction. The ease of access to internet increase number of posts, products, and its reviews tremendously [38]. The review is a rich source of customers' opinion sentiment and concerns about a product. The research can utilize these reviews to understand customers' needs, requirements, and satisfaction about a product [10]. The products reviews are publically available online. It is easy to collect, analyze, and understand real user experience with minimum cost, time, and efforts [1, 14]. Hundreds of thousands of customers contribute their opinions about a product, regarded as "wisdom of crowds" [12]. Therefore, online reviews can help in understanding user experience and customer satisfaction.

2.2 Usability and user experience dimensions

Usability and UX have commonly used terms in human-computer interaction (HCI). However, the accurate definitions, commonalities, and the differences between these terms are still under discussions. Also, it is challenging to sub-divide it into UX dimensions [39], in the case of UX hotly so [24, 40, 41]. There is no universally agreed decision about whether usability is an aspect of UX or UX is an aspect of usability. The key focus of this research is to understand the existing research work that maps dimensions to aspects, phenomena, and viewpoints in UX. The brief description of those research works is as follows.

2.2.1 Dimensions of usability

The usability defined by the ISO 9241 standard [6] using three dimensions, such as efficiency, effectiveness, and satisfaction. They define the usability as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." A detailed description of usability is mapped to five dimensions [42,43]. However, there exist some deviations in the aspects' name. The five dimensions includes: (i) Effectiveness/Errors, (ii) Efficiency, (iii) Satisfaction, (iv) Learnability, (v) Memorability. There exist variations in the definition of these five dimensions in literature, and some of the researchers use subsets only [39]. Also, the meaning of these terms varies in different studies some research focuses on its limited scope while others consider it on its broader range.

2.2.2 Dimensions of UX

Compare to usability; there exists a minimal consensus on UX definition and its mapping to aspects. According to the ISO 9241-210 [6] UX is defined as: "A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service".

This can be interpreted the same as veiled by the Satisfaction dimension of usability defined by ISO 9241-11. However, the literature interpreted it in much more nuanced. Bevan, et al. [44] considered it into four dimensions named satisfaction measures. The study performed at Nokia by Roto [41] considered as useful, however, Bevan subsequently grouped it into dimensions [40]. Hassenzahl [24] split UX analysis to three methods beyond the instrumental, emotion and affect, and the experimental with partial overlapping. However, the author later on [5] focus on the subjectivity of the product use and considered it to Self Determination Theory and Flow.

2.3 Usability and UX in online reviews

Usability measures the overall ability of a product, service, or system to achieve targeted goals effectively and proficiently. While UX evaluations provide a perception of the users' satisfaction towards achieving these goals. Both usability and UX are closely related to the specific product, defined task, user cognitive, and distinct circumstances. They play an essential role in critical product analysis and are the target of academic evaluations. Product reviews are the rich sources of identifying usability and UX of a targeted product. It helps in understanding user opinion about a product and assists in product improvements. Users typically check the reviews given by other users to take a final decision of purchasing a product. Also, the reviews reveal the real UX of a user about the product as it is given after consuming the services and using the product. The user provides product feedback in the form of reviews due to motivation, tangible, and intangible rewards. Despite benefits, there are some limitations for considering online product reviews for usability and UX evaluation. The reviews strong describe user opinion towards a product. However, in user online reviews some important information are missing such as age, gender, and preferences, which are required for usability studies. Moreover, all reviews are not credible for usability study; some reviews may contain false information or even provided by the owner of the product for promoting their products.

2.4 Mining the UX dimensions from online reviews

An evolving stream of UX research has focused to directly or indirectly assess the UX from online reviews. Online reviews are real reservoirs of the UX. These are unstructured textual documents containing a large amount of information. The quantitative analysis of these reviews generates insight by applying text mining and analytics techniques. Additionally, these techniques extract substantial information from unstructured text data and then analyzing such information. Currently, text mining is intensifying the major research areas of sentiment analysis, topic modeling, document classification, and natural language processing. Generally, the studies in these domains can be categorized into extracting UX dimensions (UXDs) from online reviews and modeling UX from online reviews [1]. The following subsection described these two categories in detail.

2.4.1 Topic modeling for UX dimension extraction

The user experience dimensions mining extracts the UXDs from online user reviews and evaluates the equal importance of each UXD. Tirunillai et al. [20] proposed a "unified framework for extracting the UXD from online reviews" using an enhanced LDA algorithm. Initially, reviews were preprocessed by applying preprocessing techniques(tokenization, stop-words filtering, stemming , and others) for further analysis. Next, they employed the improved version of LDA for latent dimension extraction along with sentiment orientation. Finally, they assigned a label to each extracted topic as dimension. Guo et al. [14] empirically determined the tourist satisfaction dimensions from online reviews. They used 266,554 online reviews for user satisfaction modeling. They used the LDA model for extraction of 19 dimensions related to tourist satisfaction. Moreover, perceptual mapping was used to determine the critical aspects of tourist satisfaction according to hotel star-ratings. However, without prior knowledge, the unsupervised models frequently generate semantically incoherent topics which are hard to understand [21, 22].

To resolve the shortcoming of the unsupervised models, some previous works add domain knowledge in the topic modeling using various approaches, but most of the models cannot learn knowledge automatically [23]. These days researchers try to use other word representation scheme as word embedding into topic modeling that reduces the dimensionality of word vector based on the co-occurrence information, considering the local context of words. By combining the global and local context gives more coherence topics. Andrzejewski et al. [45] developed topic-in-set knowledge model which controlled the assignment of words to topic based on a subset of topics. The Author extended [46] that model by integrating the general knowledge using first-order logic.

Likewise, Chemudugunta et al. [47] developed a concept model using the "Open Directory Project (ODP)." The Dirichlet Forest LDA (DF-LDA) model incorporate knowledge by a user in Andrzejewski et al. [45] model in the form of must-links and cannot-links. If two words in the same topic refer to must-link otherwise, cannot-link. Newman et al. [48] developed two models based on Bayesian regularization interpretations along with word co-occurrence to enrich topic coherence.

Chen et al. [49] proposed the Multi-Domain Knowledge (MDK-LDA) LDA model, having abilities to incorporate the multiple domains prior knowledge. The author extended the MDK-

LDA model called a knowledge-based topic model (MC-LDA) using must-link and cannot-link set. General Knowledge based LDA (GK-LDA) [50] model uses word ratio probabilities in each topic to diminish the wrong knowledge induction. Recently, Probase-LDA [51] model combined a probabilistic knowledge base with model uses the Wikipedia knowledge for a better topic generation. Yang et al. [52] incorporate the existing prior knowledge into topic modeling with large scale datasets.

While the above stated knowledge-based topic models unable to learn the domain knowledge automatically, the prior knowledge only feeds by a human expert. Automated Knowledge LDA (AKL) [53], Lifelong Topic model (LTM) [51], and topic modeling with Automatically generated Must-links and Cannot-links (AMC) [52] model resolve the issue of automatically incorporation of domain knowledge into topic modeling. These models automatically learn the domain knowledge from the given corpus. Although, these models were useful for extracting the more precise topics. These models uses frequent itemset mining technique for mining the domain knowledge without considering the word order in the corpus. The model coherence is measured using automatic approach proposed by [22] by word co-occurrence. Chuang et al. [51] measured the correspondence between a set of latent topics and a set of reference concepts. These days' word embedding has been employed to assess the topic coherence from Twitter data by Fang et al. [54].

2.4.2 Sentiment Analysis of online user reviews

Sentiment analysis (SA), also known as opinion mining. In existing studies, SA has been done at three levels [55]: document level, sentence level, and word/aspect level [56–60]. In document level, the document is classified into either positive, negative, or neutral. In sentence level, each sentence in document is classified into either positive, negative, or neutral. While, in word/aspect level, each word polarity is checked either positive, negative, or neutral. In our work, we considered the document level SA classification. In our case, each review is considered a single document. The effective SA classification of online reviews is relying on important feature construction and selection. In literature, different feature construction and selection has been discussed. Such as Bag-of-word (BOW), part-of-speech (POS), n-gram for feature construction, while, filters and wrapper base approaches for feature selection. Feature construction task converts the collection of documents into word-vector by extracting the important features that express the user opinion. In the feature selection task, the unwanted and unrelated features are filtered out by selecting the most dominant features for improving the classification.

In prior work, the feature selection process has been done in three methods: filter, wrapper, and embedded methods [61]. In the filtering method, the subset of essential features/relevant features are selected by ranking them according to specific scoring schemes based on the intrinsic properties of the features. The low scoring features are removed while the highest scoring features are selected. The filter uses a fast evaluation function and is independent of the classifier. In the wrapper method, various subsets of features are generated and evaluated through the classifier. Different methods have been used in wrapper feature selection methods such as forward selection, backward selection, and optimize selection. For example, a forward selection method starts with an empty selection of features/attributes and, in each iteration, it adds a new attribute of the given recordset. The embedded method selects features during the model training. Among the three methods, wrapper feature selection methods perform better but computationally expensive and overfitting problem. Pang et al. [56] used the supervised machine learning techniques and performed classification using three classifier (Support Vector Machine, Naive Bayes, and Maximum Entropy) for SA. They classify the movie reviews into positive and negative by employing the different combination of feature set of unigram and bigram along with POS tags. Their results show the best classification performance on unigram features. They claim that using bigram for feature construction is not appropriate for SA either using bigram separately or used with unigram. On another hand, Dave et al. [62] gains higher model performance using bigram features for SA classification. Tripathy et al. [63] performed different experiments using different classification algorithms (Maximum Entropy, Nave Bayes, Stochastic Gradient Descent, and SVM) with a different combination of n-grams (unigram, bigram, and trigram) on movie review dataset. Their experimental results show higher accuracy on a different combination of n-grams using SVM classifier.

Turney [58] classify review documents using unsupervised algorithm. The author used pointwise mutual information(PMI) and part of speech patterns to identify each phrase average sentiment orientation by comparing the similarity of a phrase with already known sentiment terms like poor and excellent. M. rushid et al. [64] perform various experiments for sentiment analysis by applying SVM algorithm with various features, weighting schemes including TFIDF, BO and TO, n-grams like unigrams, bigrams, and trigrams, data sets, and applied in different domains. The trigram model performed better compare to unigram and bigram. Ng et al. [65] also applied SVM classifier for review sentiment identification and its polarity classification. They also applied various n-grams along with dependency relation. The primary problem faced in the experiment was the review identification and polarity classification. The review identification focuses on separating reviews document from other documents while in polarity classification, reviews are classified as positive or negative against the target product or service. Most of the sentiment analysis techniques are applicable in the English language. However, Zhang et al. [66] presented an approach for Chinese reviews sentiment classification. The authors used word2vec and SVM algorithm for the review classification. The authors clustered related and similar features by using word2vec and produce training data using lexicon and part of speech based approaches. Tan and Zhang [67] used the filter based feature selection methods using four filters algorithms (Information Gain, chi-square statistics, Mutual Information, and document frequency) for SA classification using five different machine learning classifier on Chinese review dataset. These classifiers were K-nearest neighbor, centroid, winnow, Nave Bayes, and SVM. The Information Gain feature selection method gives higher performance among the other filter based algorithms using SVM for SA classification. Alireza et al. [68] proposed a hybrid approach by ensemble the filter and wrapper methods for feature selection to extract dominant features for SA classification. As ensemble learning methods increases the SA classification performance for different domains [69–76]. Xia et al. [76] studied the efficiency of ensemble learning method for feature selection for SA classification. They first extracted features set based on part-of-speech tags and word-relation. Then they used three well-known base classifiers SVM, NB and ME. Finally they combined these classification algorithms and considered three ensemble methods, fixed combination, meta-classifier combination and weighted combination. Catal [72] proposed multiple classifier systems (MCS) for sentiment classification on Turkish reviews. According to their experimental results MCS achieved better performance on sentiment classification for Turkish reviews. In addition, the ensemble of multiple filters feature selection methods [77–80] has been widely proposed for appropriate feature selection in different domains. In this study the existing ensemble based multiple feature selection techniques in different domains are considered [78, 81, 82]. Onan [77] presented genetic rank aggregation based feature selection model for sentiment classification. Their model increased the classification accuracy of sentiment classification.

According to state-of-the-art, feature extraction, selection, and ensemble techniques have been enhanced the performance of sentiment classification. In our approach, the appropriate features extraction and selection for document sentiment classification is based on the different feature sets combinations with ensemble learning methods.

2.4.3 Modeling UX from online reviews

The modeling UX from online reviews primarily concern on examining the effects of user sentiments towards product features on UX, particular on customer satisfaction. In the literature, various studies have been proposed to model user satisfaction from online reviews. Farhad et al. [19] proposed a Bayesian approach by using semi-structured data for aspect-level sentiment analysis for customer satisfaction modeling. They associated the sentiment with product aspect in each review using a probabilistic approach to produce a single rating for each attribute and their relative importance of the product or service. Decker et al. [83] used regression models (Poisson, negative binomial, and latent class Poisson) to assess the effects of user sentiments toward product features on satisfaction. Their findings show that a negative binomial regression model outperforms for identification of the causal impact of user sentiments towards product features on user satisfaction.

Xiao et al. [84] developed a model for assessing customer preferences from online reviews. They used semi-structured reviews for identification of sentiment orientations toward product attributes along with user rating to build the customer trust network.

These stated studies made substantial contributions to modeling customer satisfaction from online reviews. However, there are some issues these methods to measure the effect of UXDs for mining UXDs from online reviews [14, 20]. These methods typically rely on the supposition that the online rating (customer satisfaction) follows a Gaussian distribution [19, 83, 85]. But, this supposition is not always correct due to the user rating. In most cases the user rating is in the form of J-shaped asymmetric distribution [19, 83, 85, 86]. Customer discusses a different aspects of a product, so there might be multifaceted associations among different UXDs and customer satisfaction. Therefore, we need a model that can deal with the complicated relationship between different UXDs and user satisfaction.

Also, the Kano model developed by Kano et al. [87] were used in existing studies for modeling customer satisfaction. This model categorizes the product features different classes such as must-be, performance, excitement, indifferent, and reverse. These features values associate with user satisfaction [1]. Must-be features are essential customers requirements and expectation and are taken for granted. These features must be fulfilled; otherwise, the product customer becomes dissatisfied. One-dimensional (Performance) features related to product quality promised by the product, service provider. These features have a direct impact on customer satisfaction when fulfilled. Attractive (Excitement) feature gives satisfaction, when filled, but do not affect customer dissatisfaction. Indifferent features neither influence on user satisfaction nor dissatisfaction. The reverse features state to a more significant degree of achievement, causing more customer dissatisfaction.

Chapter 3

Proposed Methodology

For the extraction of UXDs, we proposed the three steps methodology for the modeling UX from online user reviews is shown in Figure 3.1

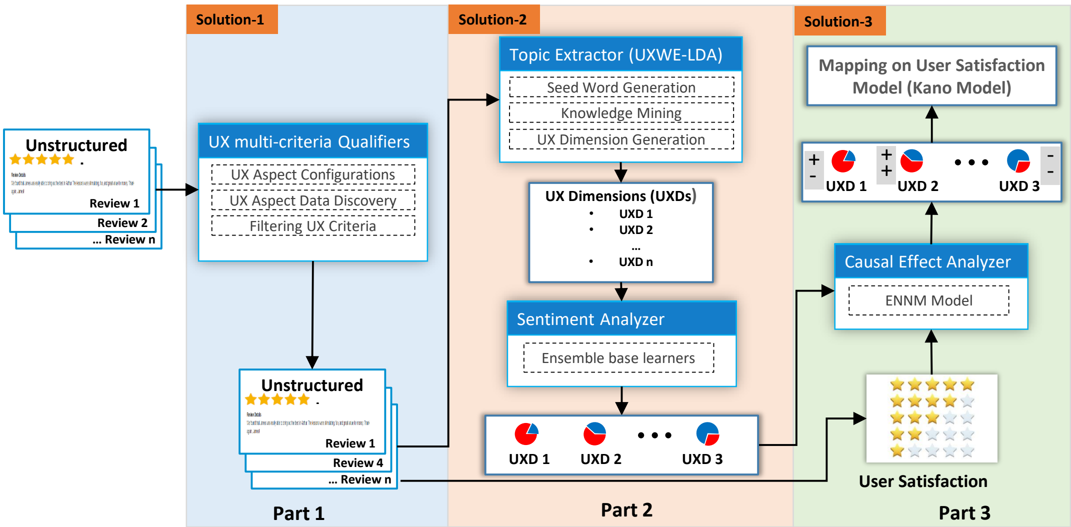


Figure 3.1: Abstract view of proposed methodology

The part-1 is related to find the usefulness of user reviews from the collection of the corpus, which contains information related to UX and usability. We propose a methodology called UX multi-criteria qualifier (UXMCQ). The UXMCQ uses an almost unsupervised method that required minimal configuration of domain seed words for auto labeling the data based on the context window. UXMCQ classify the UX aspects (product, user, and sentiment) for the given input text. The part-2 is related to the UXDs extraction from the filtered useful reviews in part-1 using an improved knowledge-based topic modeling methodology called UXWE-LDA. Additionally, identifying the user positive and negative sentiment association towards each UXDs using ensemble learning methodology. Part-3 is about measuring the casual relations toward each UXD on user satisfaction using ensemble neural network based model (ENNM) proposed by Bi, Jian-Wu, et al. [1].

In the following section, we briefly described each methodology internal function details as shown in Figure 3.2.

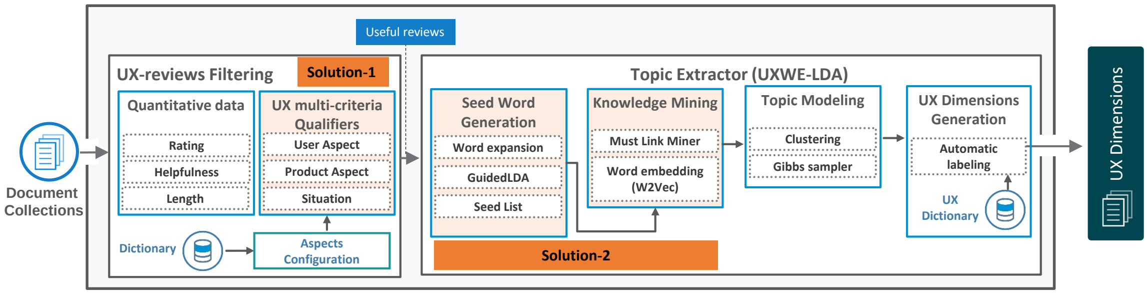


Figure 3.2: The overall proposed methodology workflow

3.1 UX Multi-Criteria Qualifier (UXMCQ)

UX Multi-Criteria Qualifiers (UXMCQ) identify those reviews which contain useful information related to UX. This step is essential to remove trivial reviews before applying the topic modeling. The UXMCQ classify pieces of text into a predefined set of UX domain aspects.

3.1.1 Aspects and sentiment configuration

UXMCQ need minimal domain aspect as seed words. The domain aspect terms used as gold terms for auto labeling of the model training. The bootstrapped mechanism is used based on the word occurrences on the unlabeled domain corpus. For the occurrence, we use the context window size [+3, -3] for generating the training instances. Finally, the ensemble learning methodology for model training based on these labeled instances.

3.1.2 Features construction and selection in conjunction with ensemble learning methods for model training

For model training, we propose features selection methodology, that will select the appropriate features from the training instances created through bootstrap method using filters base features selection and majority voting technique. The construction of an adequate feature space from the raw and unstructured text for better learning performance is necessary for text classification. It is essential to include only relevant/appropriate features for text representation. In this study, we used BOW, POS tags, semantic features (lexicons and dictionaries). For feature construction, we have applied a preprocessing step to make the initial feature vectors which are suitable for further feature extraction and selection process. The preprocessing step contains tokenization, stop-word removal, and stemming (Porter algorithm).

Feature selection is the way to extract and select the most important and relevant features. It reduces the dimensionality feature space without losing too much information for an accurate prediction. The selected features are used to train the predictive model. In the filtering method, the subset of essential features/relevant features is selected by ranking them according to specific scoring schemes based on the intrinsic properties of the features. The low scoring features are removed while the highest scoring features are selected.

3.1.3 Ensemble learner

We have employed the ensemble learning method for aspects and sentiments classification. Ensemble learning combines the predictions of multiple base learners to improve performance over a single learner. In this work, we have employed a majority voting technique in conjunction with three base learners namely, Support Vector Machine (SVM), Nave Bayes (NB) and Decision Tree. Based on the majority voting of base learners, the user reviews is classified into three UX facets. After classification, filter is applied to select the useful reviews for UXDs extraction.

3.2 UXDs Extraction from online reviews using UXWE-LDA

In this subsection, a brief introduction of UXWE-LDA then the process of UXDs from the online reviews based on the UXWE-LDA is explained.

3.2.1 User Experience word-embedding LDA (UXWE-LDA)

UXWE-LDA is an improved version of LDA, that automatically learn the domain knowledge from the given text corpus. UXWE-LDA resolve the problems of existing LDA, which knowledge often generated semantically incoherent topics. UXWE-LDA improve the existing knowledge-based topic models by for the extractions of more domain dependent dimensions in the UX area through UGC. UXWE-LDA combines topic modeling especially LDA with word embedding that automatically learns the domain knowledge from a large amount of textual data. This model automatically learns the domain knowledge from the given text corpus and extracts more coherent topics in order to assign labels as UXD to each extracted topic using dictionary based approach. UXWE-LDA mainly consists of three steps. First, it has run the guidedLDA with guided seeds words and selects topical words as seed words from the online reviews. The word vector of the seeds words is used to generated must-link knowledge-based using word embedding and other similarity computation. For the similarity, we use cosine similarity and Point-wise Mutual Information (PMI) [88]. Finally, we cluster similar must-links words and apply Gibbs sampling to find more semantically coherent topics. The goal of the first and second step is to generate the prior domain knowledge from local and global context for topic modeling, and the third step is to generate more coherent topics from the domain knowledge.

3.2.2 Identifying the sentiment orientations of the reviews concerning each CSD based on ensemble methodology

To identify the sentiment orientation, we trained the sentiment analysis model based on the ensemble methodology [18]. The sentiment analysis process consists of three steps. The first step is related to features construction based on bag-of-words (BOW), part-of-speech (POS) tagging, semantic features (lexicons and dictionaries). For feature construction, we have applied a preprocessing step to make the initial feature vectors which are suitable for further feature extraction and selection process. The preprocessing step contains tokenization, stop-word removal, and stemming (Porter algorithm). We used PENN Treebank scheme [89] for POS tagging pattern. For example, the feature excellent interface filtered by the POS tag pattern JJ NN and was disappointed feature is filtered out by the pattern VBD VBN. TF-IDF term weight scheme is used for word vector creation. The second step is about the feature selection. We have employed a filter method and the wrapper method for the most dominant features selection. The third step is about the training the ensemble learning model using voting techniques having three base classifiers base learners namely, Support Vector Machine (SVM), Nave Bayes (NB) and Decision Tree. The sentiment orientation of each review is associated with the UXDs for finding the positive and negative sentiment toward each UXDs for user satisfaction modeling.

3.3 The causal relationship of sentiments toward each UXDs on user satisfaction

For the causal relationship, we used the existing ensemble neural network based model (ENNM) proposed by [1] which overcome the problem of existing models used to the modeling of user satisfaction from online reviews. This model combines the user rating and extracted dimensions for measuring the causal relationship of user sentiment on user satisfaction. We employed the Kano model, developed by Kano et al. [87], which is a two-dimensional model. Kano model is a well-known model of user satisfaction. This model categorizes the product features into different classes such as must-be, performance, excitement, indifferent, and reverse. These features values are associated with user satisfaction [1].

Chapter 4

Usefulness of Reviews : UX Multi-Criteria Qualifiers

In this chapter, we discussed the part 1 of the proposed methodology as shown in Figure 4.1. Part 1 is responsible to filter out the unrelated online reviews, which shadow the overall semantics of topic extraction in part 2 of the proposed methodology.

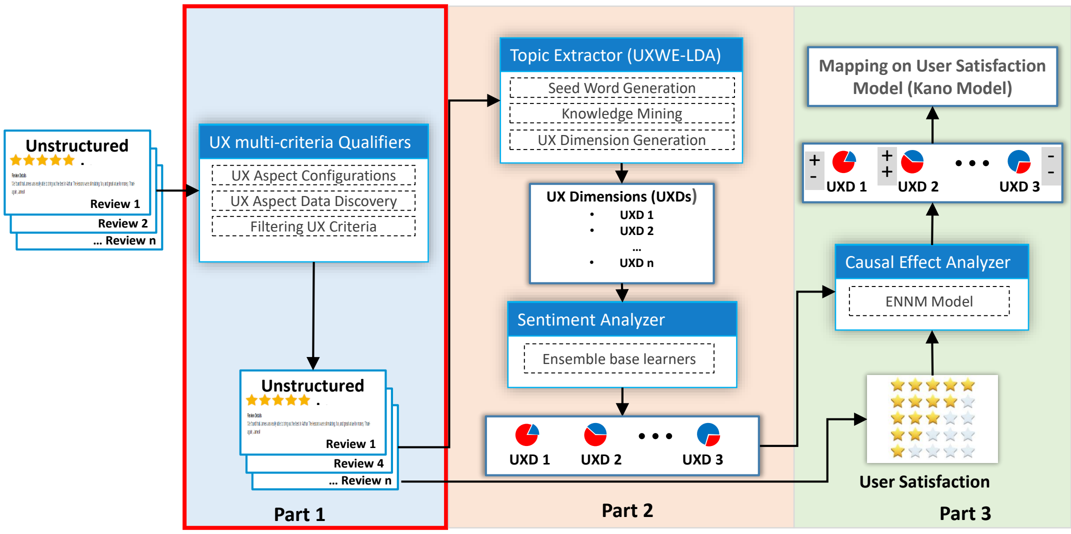


Figure 4.1: Abstract view of proposed methodology

Before apply topic modeling, it is essential to filter out those reviews which contain unrelated data to a specific domain. This type of filter can boost the topic coherence in the topic extraction methodology. The overview of the proposed UX Multi-Criteria Qualifiers (UXMCQ) filter is shown in Figure 4.2 and described in Algorithm 1. The UXMCQ select those reviews for topic modeling, which contain useful information related to UX.

The UXMCQ model creation consists of mainly three steps 1) UX aspects dictionary creation and aspects configuration 2) word occurrence mapping and context window creation for auto

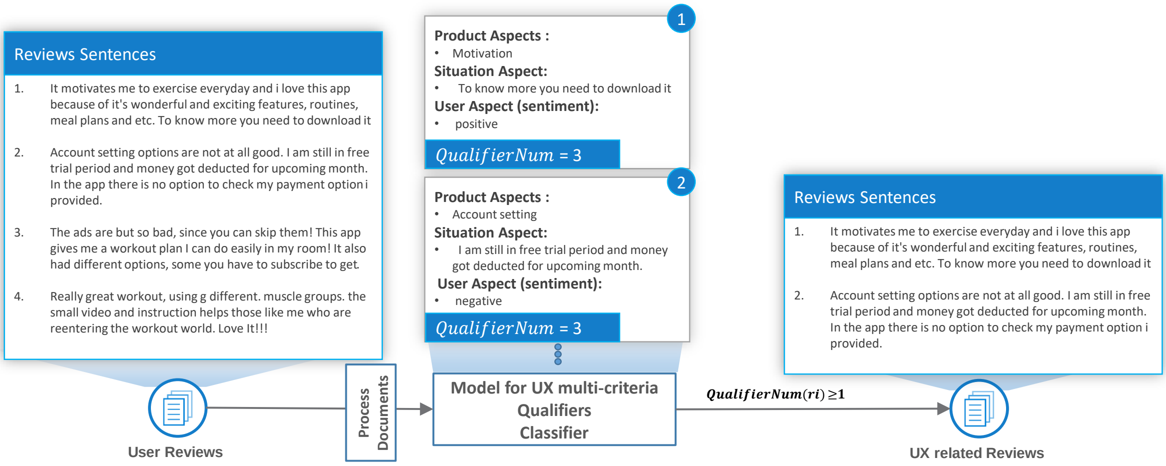


Figure 4.2: A Processing to identify the usefulness of review

labeling 3) model creation based on the influence factors selected through filter based selection process. The overall process model is shown in Figure 4.3. The details of each step are explained in the following subsections.

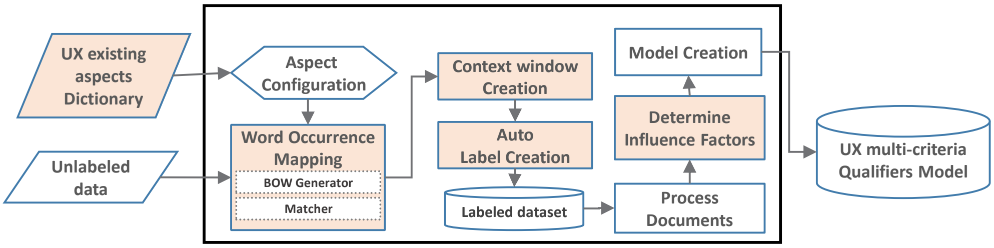


Figure 4.3: A UX multi-criteria qualifiers model overview.

4.1 UX aspects dictionary creation and aspects configuration

UX aspects configuration is the primary step for UXCQ module. Based on the selected aspects, the model automatically labeled the unlabeled data using bootstrap method based on the occurrence of

Algorithm 1: UX multi-criteria Qualifier algorithm

Input : $A_d = \{D_1, D_2, D_3, .., D_i, ..., D_n\}$ `// Aspects definition`
$B_W = \{w_1, w_2, w_3, .., w_i, ..., w_n\}$ `// Bag of words of size n`
$F_s = \{f_1, f_2, f_3, .., f_j ..., C_n\}$
λ = Threshold value

Result: UX multi-criteria Qualifiers(UXMCQ) Model

Initialization ;
$C_{size} \leftarrow [+n, -n]$;
$Y_i \leftarrow 0$;
$_f \leftarrow newSet()$;
foreach D_i *in* A_d **do**
  if *matched* D_i *in* B_W *do* **then**
    $C_W \leftarrow createContextWindow(B_W, D_i, C_{size})$;
    $Y_i \leftarrow toLabeledData(C_w)$
  end
end
foreach *f in F* **do**
  $X_F \leftarrow rankFeatures(y_i, F)$
  $X_F \leftarrow sortDES()$
  $T_F \leftarrow selectTopKFeatures(X_F, k)$
  $R_F.add(T)$
end
$R_F \leftarrow majorityVoting(R_F, >t)$;
$UXMCQModel \leftarrow trainModel(R_F, classifier)$;

a word using the context window size. It is essential to make the domain depend aspects seed for filtering the important reviews for latent dimensions extraction. In order to make the UX domain aspects, we made UX aspects dictionary using a systematic review process. The detail of UX existing aspects identification for aspect configuration is discussed in the following subsections.

4.1.1 Identifying UX existing aspects for aspect configuration

As we mentioned in the introduction section, that UX is context dependent, subject in nature, and dynamic. Due to these factors, UX researchers considered different aspects for measuring the UX. So we need to scan the prior research using systematic review process for identifying the UX dimensions or aspect in the UX domain, which help to build UX aspects dictionary for aspect configuration. Additionally, this UX aspect dictionary can help to build a more comprehensive

UX model for UX evaluation. We used a two-phase approach for the extraction of UX aspects as shown in Figure 4.4. In the First phase, we use the systematic review process to identify the UX related literature mentioned the UX aspects, dimensions, and measurements. In the later phase, we analyzed the selected papers for UX aspects selection. Finally, we construct the UX aspects dictionary.

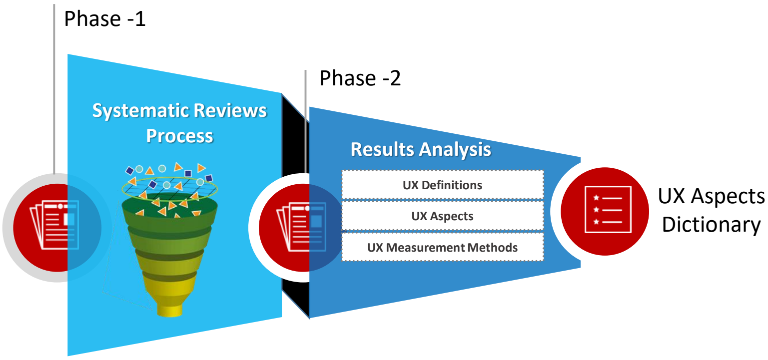


Figure 4.4: Identifying UX existing aspects for aspect configuration process.

4.1.1.1 Phase-1: Systematic review process

We used Systematic review process for articles selection in UX research. The publications is selected using four steps, borrowed from the [90]. The Systematic review process steps are shown in Figure 4.5.

In **Step-I**, we selected the sources such as Google Scholar, ACM, IEEE explore, Springer, Elsevier, and web of knowledge for searching the UX related articles by using the search terms "User Experience, UX, UX aspects, UX dimensions, and UX measurement methods". In **Step-II**, we exclude those papers having citation lower than 10, non-English, and duplicates. In **Step-III**, we narrow down the selection criteria by including the high impact journal and premium conference papers. In **Step-IV**, we filtered out those papers that discussed UX dimensions. Finally, we selected 57 research articles for the extraction of the essential UX dimensions and aspects for making a comprehensive UX dictionary.

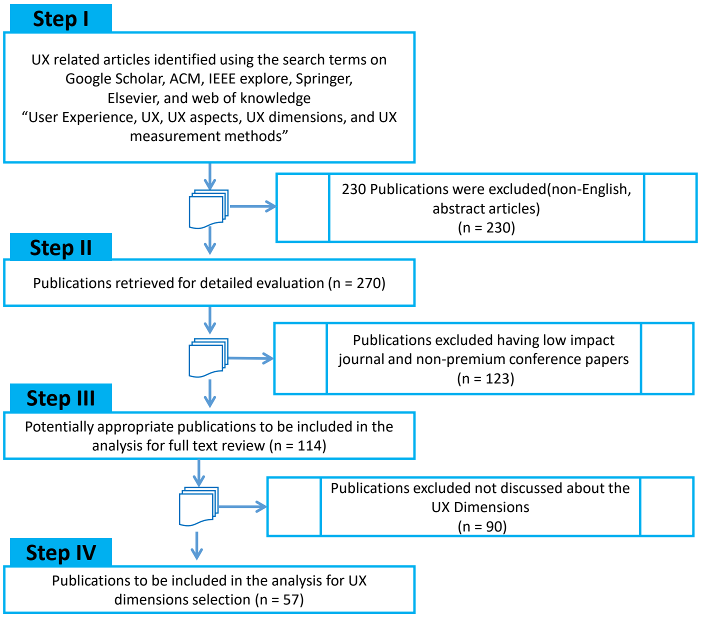


Figure 4.5: A systematic review process for UX related articles selection.

4.1.1.2 Phase-2: Results analysis

We grouped the UX aspects based on the existing conceptual UX Facet model [12] . UX facet model divided all essential factors into three main facets: user facet, product facet, and situation facet. The user facet is related to user sentiment and cognition such as background information, user preferences, intentions, and user opinions (negative, positive, or neutral). Product facet is related to product attributes such as UI, aesthetic, quality, and others. Situation facet is related to the environmental factors of the context of use such as time and place. Table 4.1 shows the selected UX aspects for two UX Facets: user and product facets are selected from articles by a focus on the UX aspects, constructs, and dimensions.

Table 4.1: Selected UX aspects from UX literature as UX dictionary.

UX Aspect	UX Facet	UX Aspect	UX Facet
Accessibility	Product Facet	Informativeness	Product Facet

Aesthetics	Product Facet	Learnability	User Facet
Affect	User Facet	Likeability	User Facet
Anticipation	User Facet	Maintainability	Product Facet
Appeal	Product Facet	Memorability	User Facet
Appraisal	User Facet	Motivation	User Facet
Attachment	User Facet	Emotions	User Facet
Attractiveness	User Facet	Novelty	Product Facet
Beauty	User Facet	Perspicuity	Product Facet
Comfort	User Facet	Physicality	User Facet
Competence	User Facet	Pleasure	User Facet
Complexity	Product Facet	Popularity	Product Facet
Context	Product Facet	Portability	Product Facet
Cost	Product Facet	Positive emotions	User Facet
Delicacy	User Facet	Pragmatic	Product Facet
Dependability	User Facet	Preciousness	Product Facet
Directness	User Facet	Predictability	Product Facet
disorientation	User Facet	Presence	Product Facet
Ease of use	User Facet	Problematic	Product Facet
Effectiveness	User Facet	Psycho-pleasure	User Facet
Efficiency	User Facet	Refresh	User Facet
Enchantment	User Facet	Relatedness	Product Facet
Engagement	User Facet	Reliability	Product Facet
Enjoyment	User Facet	Safety & Security	Product Facet
Entertaining	User Facet	Satisfaction	User Facet
Flexibility	Product Facet	Simplicity	Product Facet
Flow	Product Facet	Skill	User Facet
Fragility	Product Facet	Sociability	User Facet
Frustration	User Facet	Socio-pleasure	User Facet

Fun	User Facet	Stimulation	User Facet
Goodness	User Facet	Support	User Facet
Hedonic	User Facet	Training	User Facet
identification	User Facet	Trust	User Facet
Ideo-pleasure	User Facet	Usability	Product Facet
Immersion	User Facet	Usefulness of content	Product Facet
Impact	User Facet	User differences	User Facet

For the third UX facet (situation facet), we used the Linguistic Inquiry and Word Count (LIWC)[1] tool categories such as "Time," "Space," and "work." LIWC tool reveals common thoughts, emotions, feelings, moods, personal and social concerns, and motivation. LIWC was used to analyze the given text based on the dictionary. The percentage was calculated based on how well the words of the given text matched to the dictionary categories.

4.1.2 Aspect configuration

UXMCQ only need a small amount of domain aspect as seed words. As aforementioned, we created the UX aspects dictionary, which is used for the aspect configuration. The aspects seed words used as gold-standards to the auto-annotation of the unlabeled data based on the occurrences of these seed words according to the context window.

4.2 Word occurrence mapping and context window creation for auto labeling

We used the bootstrap method for auto labeling based on the gold aspect terms related to three UX facets. The auto labeling is based on the occurrence of the term by exact matching with the aspect terms in the unlabeled data. We used the context window having size [+3, -3], and generated the label as UX facets based on aspect terms matching. The overall bootstrapping process is described in Algorithm 2.

[1]http://liwc.wpengine.com/

Algorithm 2: Aspect based Auto Labeling

Input : $D = \{d_1, d_2, d_3, ...d_n\}$ // Collection of user Reviews
$A_T = \{T_1, T_2, T_3, ...T_k\}$ // Aspect Terms
Result: Labled Data Based On Aspects
foreach d_i *in* D **do**
  foreach $ConceptC$ *in* $operands$ **do**
    $S = \{s_1, s_2, ...s_n\}$ // Split all reviews into sentences
    for s_i *in* S **do**
      $Matched \leftarrow$ Match(s_i, T) // Match the aspect Term in each Sentence S
      if $Matched$ = *True* **then**
        $C_w \leftarrow$ create context window $[+n, -n]$
        $Label \leftarrow$ Assign label to that context window as T
      end
    end
  end
end

The Figure 4.6 depicts, how the bootstrap method assigns labeled based on the terms occurrence and context window. First, it loads all the unlabeled reviews data, split each review into sentences, then the matcher check the occurrence of the aspect terms in each reviews. If match found that created context window and assigned label as aspect terms.

4.3 UXMCQ Model creation based on the influencing factors

We apply the feature selection process on the auto labeled data generated by the bootstrap method. The word vectors are generated by applying the document process methodology. The following subsections described the overall process of selecting the influencing features from the given corpus data.

4.3.1 Process Document

The process document converts the auto labeled reviews into word vector by applying the NLP steps such as tokenization, stemming, filter stop words, POS tagging, and others. Word vector creation with adequate features is an essential task to boost classifier accuracy. It removes unwanted features by keeping the important features, while converting unstructured textual data into

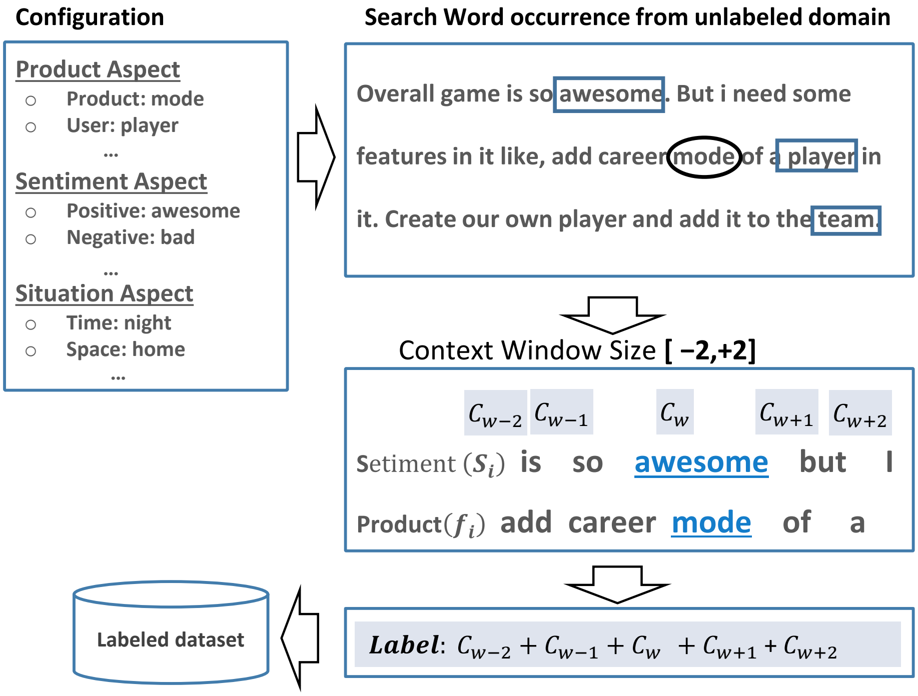


Figure 4.6: Auto labeling process based on the context window.

structured data. In recent literature, different methods and techniques have been used to create word-vector. These methods include bag-of-word(BOW), POS tagging, n-gram, semantic-based feature creation, and others. For word vector creation, we employed the processing steps (tokenization, case conversion, filter stop words, wordnet base synonym), POS tagging, and n-gram. We used the term frequency-inverse document frequency (TF-IDF) scheme for the vector creation along with term pruning. For POS tagging, the PENN [91] tree scheme is used for POS tagging pattern. The details description of each step is described in the following sub-sections.

4.3.1.1 Pre-processing of user online reviews (textual data)

To convert the user reviews textual data into a feature vector (word vector), we employed the preprocessing step. This step consists of text processing operations that convert the textual data into the numerical form using TF-IDF scheme. The pre-processing step contains the following sub-steps.

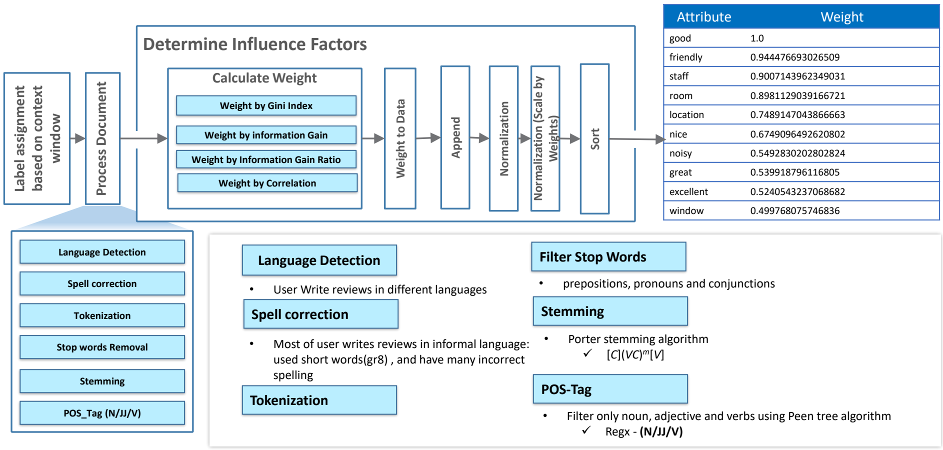


Figure 4.7: The process of vector creation and ensemble of multiple filters feature selection.

- **Language detection:** We apply the language detection process using the spacy-langdetect module of spacy [2]. We only considered the English language written reviews filtered through the spacy-langdetect module. This step is essential because of the online users reviews written by a different user across the globe in different languages.

- **Spelling correction:** Most of the user reviews written in informal language, uses abbreviated words and misspelled. We employed the spell correction process to correct the misspell and short words such as gr8 to great, goooood to good.

- **Tokenization:** After selecting reviews written in English and spelling correction, the tokenization process is applied to split the text of the review into sequences of words/tokens.

- **Case Conversion (lower case)**: To reduce the word vector space, case conversion process is applied, it converts the all tokens into either lower case or upper case. We employed the lower case conversion.

- **Filtering stops words:** As the reviews contains unwanted words such as the, an, for, it, a, be, and others. These words have no meaning related to UX domain and less weight as compared to the other features. We removed those stop-words based on English stop-

[2]https://spacy.io/

POS-pattern 1: $(J|N)^{+}\ N$
POS-pattern 2: $((J|N|C)^{+}\ |\ (J|N|C)^{*}\ (N\ I)?\ (J|N|C)^{*})\ N$

words filters operator along with dictionary-based stop-words to filter stop-words other than English stop-words.

- **Removal of URLs, numeric, quotations, and special characters:** We also filtered out the special character, numerical value, and URL, which have no relation with the UX domain facets.

- **Stemming:** This process reduces the token to its root word using Porter stemming algorithm [92], which use a rule-based word replacement technique to reduce the length of the words until word suffixes reach a minimum length.

- **Part-of-Speech (POS) Tagging:** Part-of-speech (POS) tagging is an essential step to filter out the feature based on the specified types of POS tags. This process first assigns the POS tag to each token based on PENN [91] system for English tagging and are defined by a regular expression of types as shown in Table 4.2. We used the following patterns to filter out the features, which contribute to recognized the UX facets.

Table 4.2: Penn Tree bank Part-of-Speech tags

Tag	Description
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative

LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP$	Possessive wh-pronoun
WRB	Wh-adverb

- **Term Weighting**: In text classification, different weighting schemes have been used for word-vector creation such as term frequency (TF), term occurrence, binary term occurrence, and term frequency-inverse document frequency (TF-IDF). The most widely used weighting scheme is TF-TDF. We also employed the TF-IDF scheme for word-vector creation. Additionally, we reduce the vector space by applying the pruning method having an absolute threshold to prune below absolute is 3 and above is 999.

4.3.2 Feature Selection

We used the feature selection process to select the dominant features to enhance the UX facet classification accuracy. The feature selection is the way to extract and select the most important and relevant features. It reduces the dimensionality feature space without losing too much information for an accurate prediction. The feature selection methods are classified into three main categories: filter, wrapper, and embedded methods. The selected features are used to train the predictive model. We have employed a filter method and the wrapper method for useful features selection. In the filtering method, the subset of important features/relevant features is selected by ranking them according to specific scoring schemes based on the intrinsic properties of the features. The low scoring features are removed while the highest scoring features are selected. The filter uses a fast evaluation function and is independent of the classifier. In the filter based method, we have used the filters like chi-square, Gini index (GI), gain ratio (GR), and information gain (IG). The chi-square is proposed in [93] to measure the absence of independence between the feature and the class label, using a chi-square statistic. If there is some correlation between features and class label exist, then the feature is considered more relevant otherwise feature is discarded. GI is a measure of the uncleanness of feature for classification by Huanjing et al. [94]. IG method measures the weight of features concerning class label [93]. IG calculate the value of each feature and assign the rank. The higher weight features are considering more relevant. IG is very good to measure for selecting the most relevant feature set based on features ranking. It reduces the feature space by selecting the top rank features and discard those features whose ranks less than the predefined threshold. GR resolves the bias of information gain. GR improves the IG by taking the essential information of a split into account [95].

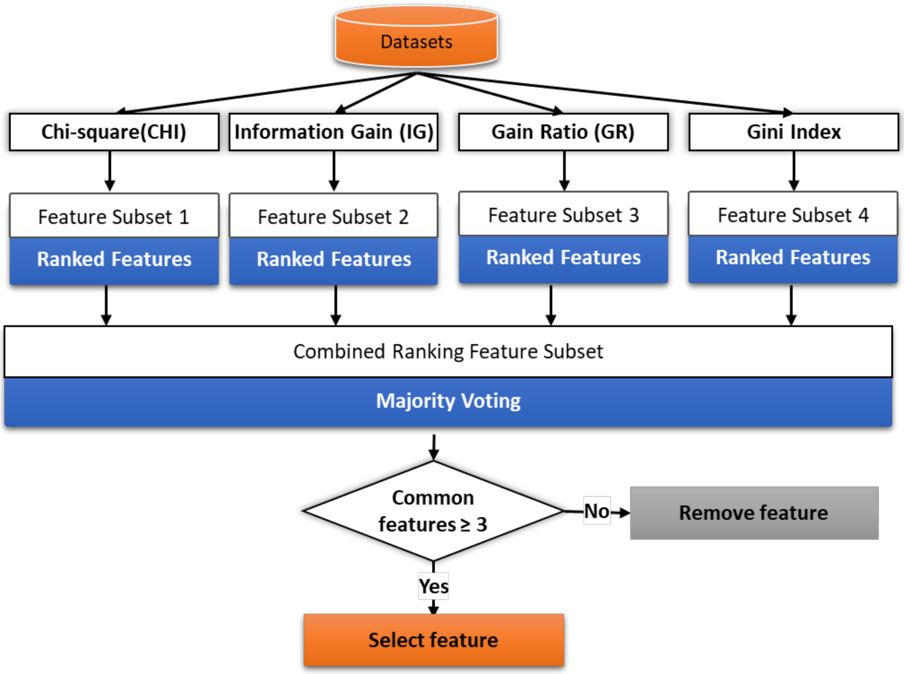


Figure 4.8: Filter base feature selection process

4.3.2.1 The ensemble method of multiple filters feature selection

In this section, we discuss the proposed ensemble method of multiple filters features selection method based on majority voting. The details of that method are described in the Algorithm 1. We used five filters, specifically IG, MRMR, CHI, GR, and GI, to select the best collective features among them for better classification, as shown in Figure 4.8. The individual filter assigns a weight to each feature using their internal logic and selects the initial subset features. We set the threshold value to 3 that checks for common features selected by at least three filters.

After the filter base feature selection, we have applied the wrapper method (forward selection process), in subset feature selection. In the wrapper method, various subsets of features are generated and evaluated. The forward selection starts with an empty selection of features/attributes and, in each iteration, it adds a new attribute of the given recordset. We have applied 10-fold cross-validation using SVM learner to estimate the performance if the added attribute gives the higher performance then is added to the selection. Then a new round is started with the modified selection. We have added the stopping behavior to stop the iteration if no significant increase in performance.

4.3.3 The ensemble learning method

We have employed the ensemble learning method for UX Facets classification. Ensemble learning combines the predictions of multiple base learners to improve performance over a single learner. In this work, we have employed a majority voting technique in conjunction with three base learners, namely, Support Vector Machine (SVM), Nave Bayes (NB) and Decision Tree. The input is the feature vector from review dataset. Based on the majority voting of base learners, the user reviews are classified into either UX Facet qualifiers or none class.

4.4 Summary

We propose a methodology called UX multi-criteria qualifier (UXMCQ). The UXMCQ uses an almost unsupervised method that required minimal configuration of domain seed words for auto labeling the data based on the context window. UXMCQ classify the UX aspects (product, user, and sentiment) for the given input text. A disadvantage of this method is that each term in the gold terms will be classified as aspect-terms, the words that not related to any of these categories will require manual supervision.

Chapter 5

UX Dimensions(UXDs) Extraction

Online user reviews are generally in an unstructured form. Therefor, the process to convert an unstructured data into structured data is an essential task for mining the UX dimensions for modeling user satisfaction. In this chapter, we discussed the part 2 and part 3 of the proposed methodology as shown in Figure 5.1. Part 2 is related to the UX dimensions extraction using enhanced topic modeling methodology, while part 3 is related to measuring the influence of user sentiments along with UXDs on user satisfaction.

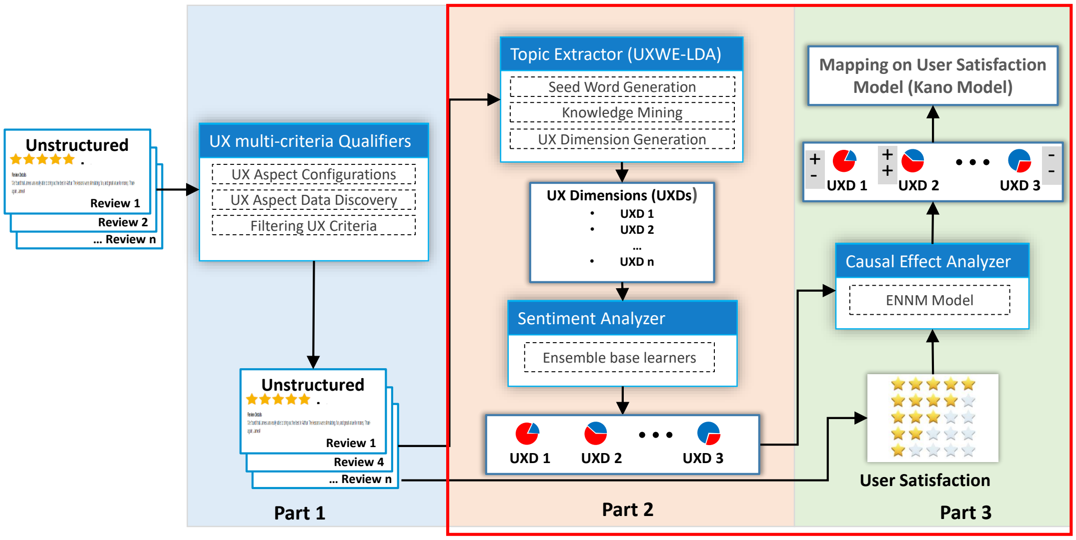


Figure 5.1: Abstract view of proposed methodology

5.1 Part 2: UX dimensions extraction and sentiment analysis

In the first part, we will discuss the process of (1) UX dimensions (UXDs) extraction using the proposed user experience word-embedding LDA (UXWE-LDA) topic modeling (2) sentiment analysis and its orientation for each extracted UXD from online user reviews.

5.1.1 User experience word-embedding LDA (UXWE-LDA)

User Experience Word-Embedding LDA (UXWE-LDA) is an improved version of LDA, that automatically learns the domain knowledge from a given text corpus. UXWE-LDA improve the existing knowledge-based topic models by extracting more domain dependent dimensions in the UX area through UGC. UXWE-LDA combines topic modeling especially LDA with word embedding that automatically learns the domain knowledge from a large amount of textual data. This model automatically learns the domain knowledge from the given text corpus and extracts more coherent topics in order to assign labels as UXD to each extracted topic using dictionary based approach. UXWE-LDA mainly consists of four steps as shown in Figure 5.2. The details description of this model is given in the following sections.

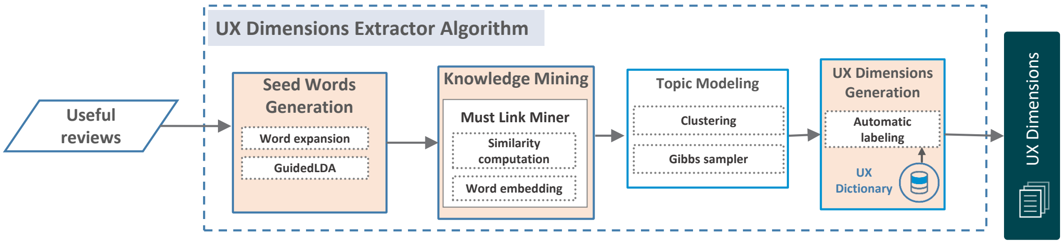


Figure 5.2: Abstract workflow for UX Dimensions Extractor

5.1.1.1 Seed Words Generation

This step generated the global context from collections of reviews filtered in part 1 of the overall methodology. First, all reviews are processed to convert the unstructured text into a structured form. For preprocessing, we applied tokenization, stemming, filter stop words and others as aforementioned in sections 4.3.1. For seed words generation, we used two steps processes. First, we run the guidedLDA with guided seeds words and selects topical words as seed words. We used the same methodology as [23] for seed words generation, but internally, our method different considered the syntactic and semantic relationships. We used the guidedLDA instead of simple LDA to generate the topic of interest seed words. Second, we expended the produced seed words using pre-trained word embedding models to make a more comprehensive global context. Algorithm 3 explains the overall process.

5.1.1.2 Guided LDA

These days, topic modeling is vital for the analysis of extensive collections of the document. LDA is one of the most popular and widely used for topic extraction. LDA is an unsupervised generative probabilistic topic model that extracts latent dimensions from a collection of documents [96]. Many extensions have been proposed for the improvement of word coherence in each topic in the LDA model. Due to unsupervised nature, it is generally hard to extract the intended topics using simple LDA model. To guide the LDA model, Jagarlamudi [97] proposed the guidedLDA, which incorporates prior lexical information in the topic model especially in the LDA model. The guidedLDA model starts from a seed list and groups similar words into the same topic. Thus it creates an array of topics based on the user intention.

5.1.1.3 Word Expansion using pre-trained word embedding model

We enhanced the global seeds generated by guidedLDA which considers the syntactical variation of the words (w) along with taking into account the semantic similarity of a given corpus. In the existing literature, semantic similarity is computed using manually built dictionary [98]. There are a number of issues with dictionary approaches such as extensive human involvement, effort and time required to hand-craft the dictionary, and also challenging to scale a dictionary to incorporate the new contexts. These days some researchers attempt to explore the automatic way for computation the semantic similarity using the word distances, but they ignore the context of the words in word embedding space [99]. Most of the prior works only focus on the implicit relationship in word context window within the document [100], but do not consider the similarity of the word with pre-trained word embedding models. We used a similar approach called CluWord [101] to

Algorithm 3: Seed words generation algorithm

Input : Useful reviews corpus C,
Seed topic words S_d
External corpur C
Vector dimension k
Vector dimension k

Result: The global context for user reviews text W_t

foreach *(doucment* $d \in C$ *)* **do**
  Sampling a topic form a topic's multiple distribution.
  $Z_d \sim Mul(\theta)$
  foreach *word* $\in$ *document* d *where* $W \in (wd_1, wd_2, ...wd_n)$ **do**
    Generate a variable weight probability from the Bernulli distribution the prbability of under t estimated by guided LDA
    $W_t = n - argmar_w \varnothing(w, s_d)$
  end
end
$W2VTrain(C, k)$
$VocabSize \leftarrow getVocabSize(C)$
$V \leftarrow initVector(vocabSize, k)$
$\theta \leftarrow initVector(vocabSize, k)$
for *(*$W_i \in C$*)* **do**
  $e \leftarrow 0$
  $Xw \leftarrow \Sigma u \in context(W_i)V(u)$
  for $(u = w_i UNEG(w_i))$ **do**
    $e \leftarrow e + g\theta^u$
  end
  for $(u \in Context(w_i))$ **do**
    $V(u) \leftarrow -V(u) + e$
  end
end
$t' = V(W_t, k)$
$expenedWord \leftarrow 0$
for $(t' \in W_t)$ **do**
  $\delta(t, t') = \frac{\sum_i^l u_i . v_i}{\sqrt{\sum_i^l u_i^2} . \sqrt{\sum_i^l v_i^2}}$
  if $(\delta(t, t') > \alpha)$ **then**
    $expendedWord \leftarrow t'$
  end
end
$W_t = W_t + expendedWord$
return W_t

exploit the word similarity based on pre-trained word embedding model to create a more general global context in terms of semantic and syntactic. We used the Word2Vec [102] for pre-trained word representation using googleNews data. Let G_V represent the global vocabulary generated by guideLDA for all documents topics D_T. Let W_E be the word embedding vector representation for each term in G_V based on pre-trained word embedding model. We compute the word expansion based on the following equation 5.1

$$W_{t,t'} = \begin{cases} \delta(t,t') & if \delta(t,t') \geq \alpha \\ 0 & otherwise \end{cases} \tag{5.1}$$

Where $\delta(t,t')$ is computed using cosine similarity matching define in equation 5.2 and α is the threshold value for filtering the most similar words to t.

$$\delta(t,t') = \frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} u_i^2}\sqrt{\sum_{i=1}^{n} v_i^2}} \tag{5.2}$$

The δ_t for term t, the expansion is limited based on the α value to remove the unrelated words that have no significant relationship to term t. If the similarity between t and t' is less than threshold value, then we discard the t'.

Table 5.1: Word expansion example based on the pre-trained word embedding model

Similarity	Term t: chat
Semantically	chatroom, conversation , conversing, talk, conversed, Live_Chat, message, interview, speak,
Syntactically	Chats , chatting , Chat, chatted

Finally, we created the global context document (G_d) for each expended topics.

5.1.1.4 Knowledge Mining

For knowledge mining, we incorporate word-embedding and other two similarity computation such as concise similarity and PMI. The overall process of knowledge mining is shown in Figure

5.3.

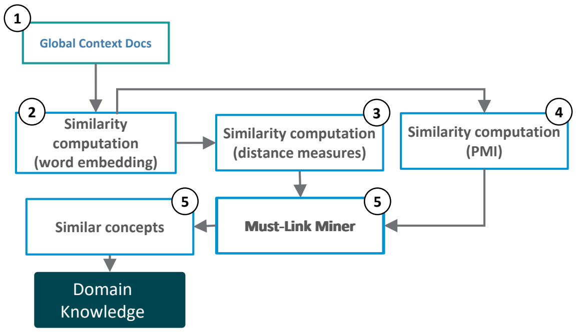


Figure 5.3: Work flow of must-link mining using similarity computation

This process consumed the global context documents G_d generated in the section above to generate the Word2Vec model. For word-embedding generation, we used the Word2Vec algorithm, that computes semantic relationship in two words based on the context window. Word2Vec is based on the algorithm called Efficient Estimation of Word Representations in Vector Space as known as skip-gram-model [103]. Algorithm 4 described the word embedding vector generation process. We use the following parameters settings for Word2Vec model generation as shown in Table 5.2.

Algorithm 4: Word embedding vector generation algorithm

Input : Global Context Corpus $C_d = \{d_1, d_2, d_3, ..., d_n\}$,
Context Window c
Result: Word Embedding Vector V

Word embedding Vector $\vec{V}$
foreach *(document $D \in C_d$)* **do**
    foreach *(word $w \in D$)* **do**
        Compute $P(W_{x+c}|W_x) = \frac{exp(vwx.vwx)}{\sum_{w=1}^{V} exp(vw.vwx)}$
        $\vec{V} \leftarrow P(W_{x+c}|W_x)$
    end
end
return $\vec{V}$

A Word2Vec model can be thought of a dictionary or hash map. This dictionary contains a word vector for every word in the training corpus as shown in the below Table 5.3.

Table 5.2: Word2Vec model generation parameters settings

Parameters	Values	Description
Minimal Vocab Frequency	5	A minimum number of occurrences each word needs to have to be considered for model generation.
Layer Size	10	Size of the vector which is generated. Typical values are between 50-500.
Window Size	5	During model generation, each text is split into windows. This specifies how large the window is. Typical values are 3-7.
Iterations	50	Number of iterations during training

Table 5.3: Word2Vec model extract vocabulary example.

Word	dimension_0	dimension_1	dimension_2	dimension_3	dimension_4	dimension_5	dimension_6	dimension_7	dimension_8	dimension_9
content	-0.617	-0.056	-0.408	0.586	0.016	0.058	-0.124	-0.192	-0.201	-0.103
hotel	-0.482	-0.109	0.258	0.326	-0.113	-0.064	0.525	0.534	-0.054	0.045
room	-0.225	0.549	-0.362	0.296	-0.466	0.065	0.124	0.085	-0.419	0.098
stay	-0.255	0.271	0.502	0.152	-0.183	0.036	0.505	-0.038	-0.147	0.523
stayed	-0.152	0.302	-0.014	0.149	0.128	-0.296	0.104	0.170	0.090	0.843
rooms	0.025	0.180	0.391	0.012	-0.243	-0.625	0.328	0.314	-0.348	0.192
staff	-0.115	-0.138	0.066	0.473	-0.212	-0.800	0.084	-0.049	0.029	-0.212
clean	-0.034	0.159	0.365	0.517	-0.145	0.463	-0.045	-0.185	0.333	0.437
night	-0.236	0.227	-0.471	0.414	-0.036	0.378	0.211	-0.267	-0.319	0.371
area	0.302	-0.085	0.263	0.363	-0.421	-0.215	-0.129	-0.289	-0.561	0.250
day	0.217	-0.172	0.208	0.630	-0.507	-0.451	-0.075	-0.100	-0.016	0.080
great	-0.259	0.364	0.194	0.671	0.128	-0.296	-0.301	0.189	-0.276	-0.073
time	-0.654	0.279	-0.203	0.550	-0.058	-0.173	0.056	-0.337	-0.022	0.013
use	0.004	0.746	-0.353	0.498	0.021	0.183	0.115	-0.138	-0.027	0.058
good	-0.385	-0.013	0.203	0.714	-0.305	-0.125	-0.186	0.244	0.156	-0.269
service	-0.200	0.276	-0.148	0.393	-0.548	0.170	-0.231	0.189	0.323	0.430
get	-0.375	0.268	0.296	0.245	-0.300	-0.151	-0.171	-0.250	0.634	-0.181

We computed the similarity between each word using cosine similarity based on the generated vectors trained by the Word2Vec. The cosine similarity of word vectors $\vec{V}$ and $\vec{U}$ is computed for $w1$ and $w2$ using equation 5.3.

$$sim(w1, w2) = \frac{\vec{U_{w1}}\vec{V_{w1}}}{|\vec{U_{w1}}||\vec{V_{w1}}|} \tag{5.3}$$

We also computed the similarity using Point-wise Mutual Information (PMI) for a must-link generation. PMI value is computed as in equation 5.4.

$$PMI(w_1, w_2) = log\frac{P(w_1, w_2)}{P(w_1)P(w_2)} \tag{5.4}$$

Finally, we combined the concise similarity with the PMI for checking the word relatedness .

We computed the coherence between $w1$ and $w2$ using the following equation 5.5.

$$Coherence(w_1, w_2) = sim(w_1, w_2)PMI(w_1, w_2) \tag{5.5}$$

Algorithm 5: Must link mining algorithm

```
Input  : Word embedding vector V⃗
Result: must link M_link
Vocabulary voc ← getVocabular(V⃗)
foreach (document w_i ∈ voc) do
    sim(w_i, w_{i+1}) = (V⃗_wi V⃗_wi+1) / (|V⃗_wi||V⃗_wi|)
    PMI(w_i, w_{i+1}) = log( P(w_i, w_{i+1}) / (P(w_i)P(w_{i+1})) )
    r(w_i, w_{i+1}) = PMI(w_i, w_{i+1}).sim(w_i, w_{i+1})
    if (r(w_i, w_{i+1}) >= 1) then
        M_link + − = w_i
    end
end
return M_link
```

Based on the following conditions we filter out all those words which meet the following equation 5.6 criteria. If the relatedness of the two words is higher than 1, then it is must-link. Otherwise, it removes those words from knowledge mining.

$$\gamma(w_1, w_2) = \begin{cases} Coherence(w_1, w_2) \geqslant 1 & is\ must - link \\ 0 & otherwise \end{cases} \tag{5.6}$$

The algorithm combines all must-link into similar groups using K-mean clustering algorithm, and create the feature vector of must-link. The overall process flow is shown in the Figure 5.4.

5.1.1.5 Topic Modeling using Gibbs sampler

Topic Modeling using Gibbs sampler is used in order to extracts topic based on automatically incorporating the domain knowledge enriched by global and local contexts using the Gibbs sampler

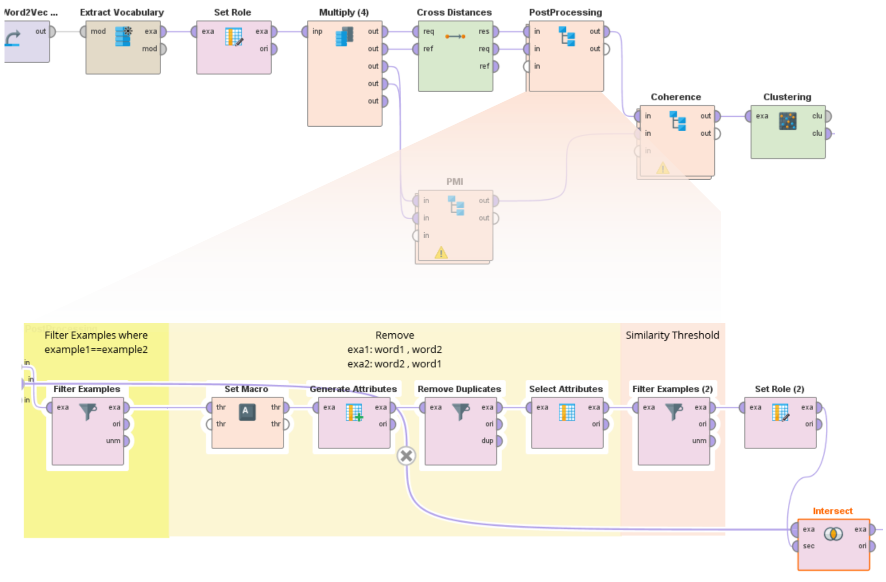
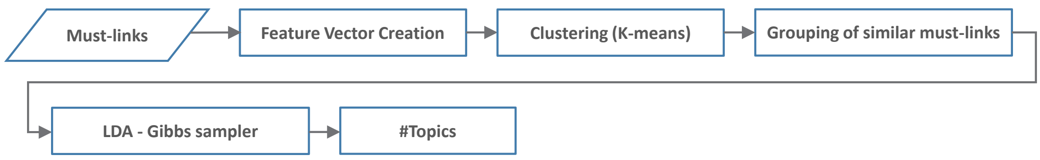


Figure 5.4: Work flow of must-link mining using similarity computation in rapidminer.

algorithm. The overall process flow is depicted in Figure 5.5.



$$P(z_i = k \mid z_{-i}, \alpha, \beta, \lambda) \propto \frac{n_{dk} + \alpha}{n_d + T\alpha} \times \frac{\sum_{w'=1}^{V} \lambda_{k,w_i,w'} \times n_{kw'} + \beta}{\sum_{v=1}^{V} \left(\sum_{w'=1}^{V} \lambda_{k,v,w'} \times n_{kw'} + \beta \right)}$$

Figure 5.5: Work flow of integrating must-link into the Gibbs sampler.

where z_i represent topic of word w_i , n_{dk} represent the number of times topic k is allocated to a word in the document d. donate document d length. is the number of times w' is allotted to topic k. α and β are Dirichlet hyper-parameters.

Table 5.4 shows an example of extracted topics with its weights using UXWE-LDA algorithm.

Table 5.4: Example of generated topics using UXWE-LDA.

Topic-1		Topic-2		Topic-3		Topic-n	
Topic-Word	**Relative-Weight**	**Topic-Word**	**Relative-Weight**	**Topic-Word**	**Relative-Weight**	**Topic-Word**	**Relative-Weight**
fun	89	accessible	50	visual	35	annoy	69
annoy	85	effective	48	effect	35	awful	64
creative	79	efficient	43	cute	33	awkward	59
enjoy	76	interface	43	trendy	33	confuse	44
exciting	71	reliable	41	technological	25	cheer	36
frustrate	67	usable	38	shape	23	rigid	35
addict	61	elegant	35	pleasurable	22	okay	33
impressive	46	error	33	color	21	trust	26
cool	45	inconsist	33	smooth	18	value	24
addict	37	delay	27	beautiful	13	dislike	24
regret	34	load	27	unusual	12	petty	23
cute	32	trouble	12	futuristic	11	help	20

5.1.1.6 UX Dimensions Generation

This section explains the process of UX dimensions by auto labeling the each extracted topic in the preceding section. We used the dictionary-based approach for classifying each topic based on top "n" words. The overall flow is depicted in Figure 5.6.

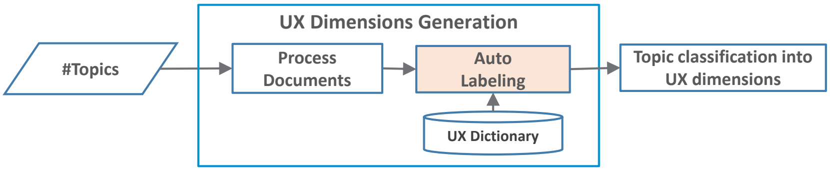


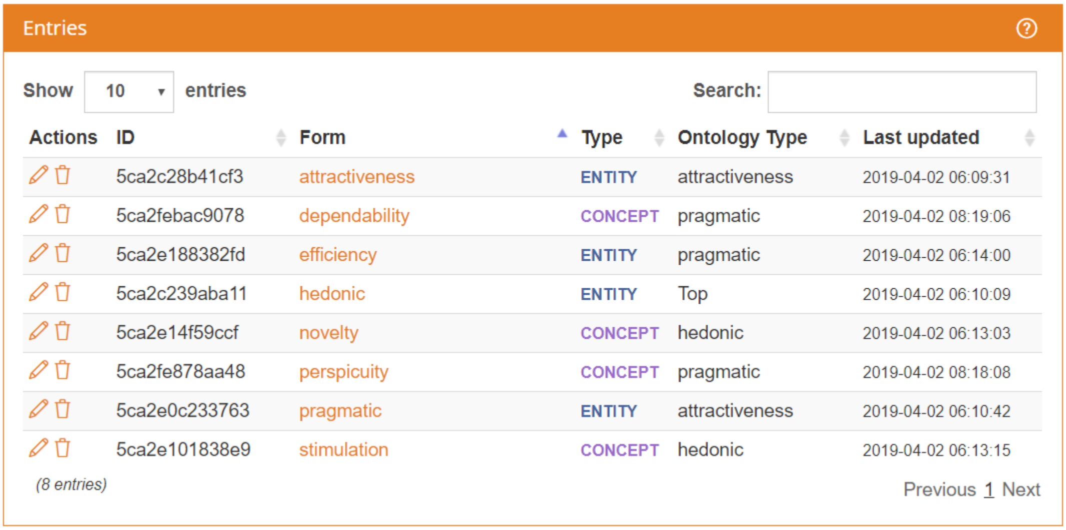
Figure 5.6: Workflow of topic labeling into UX dimensions.

We build the lexicon dictionary based on terms already used in previously validated scales [29, 104, 105] for measuring different aspects of UX using systematic review process. We selected the 223 terms, then applied the WordNet for word expression. Final thesaurus contains 500 terms by adding the synonyms to UX dictionary. We also included the aspects from previous UX dictionary mentioned in section 4.1. Finally, we validated the UX dictionary using Cohen's kappa coefficient [106] from three domain experts.

For topics classification based on dictionary, we used the MeaningCloud text mining API. MeaningCloud allow to define the custom dictionary in form of ontology. We created the UX dimensions dictionary with terms used in the existing scales. The Figure 5.7 show the UX dictionary created at MeaningCloud.

Table 5.5: UX existing Scales for UX Dictionary Creation.

Scales	Dimensions	Individual Terms
Hassenzahl Attrakdiff [112]	Pragmatic Quality	technical, human, complicated, simple, impractical, practical, cumbersome, direct, unpredictable, predictable, Confusing, clear, Unruly, manageable
	HQ-Stimulation	typical, original, standard, creative, cautious, courageous, conservative, innovative, lame, exciting, Easy, challenging, Commonplace, new
	HQ-Identification	isolating, integrating, amateurish, professional, gaudy, classy, cheap, valuable, non-inclusive, inclusive, unpresentable, presentable
User experience Questionnaire (UEQ) [25]	Attractiveness	annoying , enjoy, unattractive, friendly, unfriendly
	Efficiency	fast ,slow, inefficient, efficient, impractical , practical, organized, cluttered
	Stimulation	valuable, inferior, boring, exiting, not interesting , interesting, motivating, demotivating
	Perspicuity	not understandable, understandable, easy to learn, difficult to learn, complicated, easy, clear, confusing
	Dependability	unpredictable, predictable, obstructive, supportive, secure, not secure, meets expectations , does not meet expectations
	Novelty	creative , dull, inventive , conventional, usual ,leading edge, conservative , innovative
psychometric scales [113]	Aesthetics	aesthetic, pleasant, clear, clean, symmetric, artistic, creative, fascinating, special effects, original, sophisticated



Entries

Show 10 entries Search:

Actions	ID	Form	Type	Ontology Type	Last updated
	5ca2c28b41cf3	attractiveness	ENTITY	attractiveness	2019-04-02 06:09:31
	5ca2febac9078	dependability	CONCEPT	pragmatic	2019-04-02 08:19:06
	5ca2e188382fd	efficiency	ENTITY	pragmatic	2019-04-02 06:14:00
	5ca2c239aba11	hedonic	ENTITY	Top	2019-04-02 06:10:09
	5ca2e14f59ccf	novelty	CONCEPT	hedonic	2019-04-02 06:13:03
	5ca2fe878aa48	perspicuity	CONCEPT	pragmatic	2019-04-02 08:18:08
	5ca2e0c233763	pragmatic	ENTITY	attractiveness	2019-04-02 06:10:42
	5ca2e101838e9	stimulation	CONCEPT	hedonic	2019-04-02 06:13:15

(8 entries) Previous 1 Next

Figure 5.7: UX aspects dictionary created at MeaningCloud platform.

We used the MeaningCloud plugin in Rapidminer for the classification of each topic based on the created dictionary. The overall process workflow is shown in Figure 5.8.

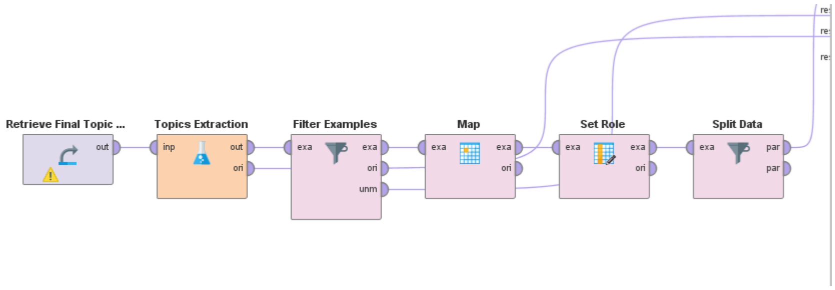


Figure 5.8: Topics classification based on UX aspect dictionary Rapidminer workflow.

5.1.2 Sentiment orientation identification of reviews for each extracted UXDs using ensemble methodology

Let $R_i = \{r_1, r_2, ., r_n\}$ is the set of useful online user reviews filtered by UXMCQ relates to the UXDs in R, where the sentiment of each review is identified using sentiment analyzer module. We trained, the sentiment analyzer using trained dataset. The workflow of sentiment analyzer is shown in Figure 5.9. The workflow consists of three main steps (1) Feature construction (2) Feature Extraction and Selection, and (3) Learning of prediction model. Details of the proposed workflow is described in the subsequent sections.

5.1.2.1 Feature construction

In text classification, conversion of text into feature vector is an essential task. The construction of an adequate feature space from the raw and unstructured text for better learning performance is necessary for text classification. It is essential to include only relevant/appropriate features for text representation. In the recent literature, different features representation methods have been used to represent text, for textual classification. These are bag-of-words (BOW), linguistic patterns using part-of-speech (POS) tags, high order n-gram features (character n-grams and word n-grams), dependency parsing tree, semantic features (lexicons and dictionaries) and structural features [89,90]. In this study, we used BOW, POS tags, semantic features (lexicons and dictionaries). For feature construction, we have applied preprocessing step to make the initial feature vectors which are suitable for further feature extraction and selection process. The preprocessing step contains tokenization, stop-word removal, and stemming (Porter algorithm). We used PENN

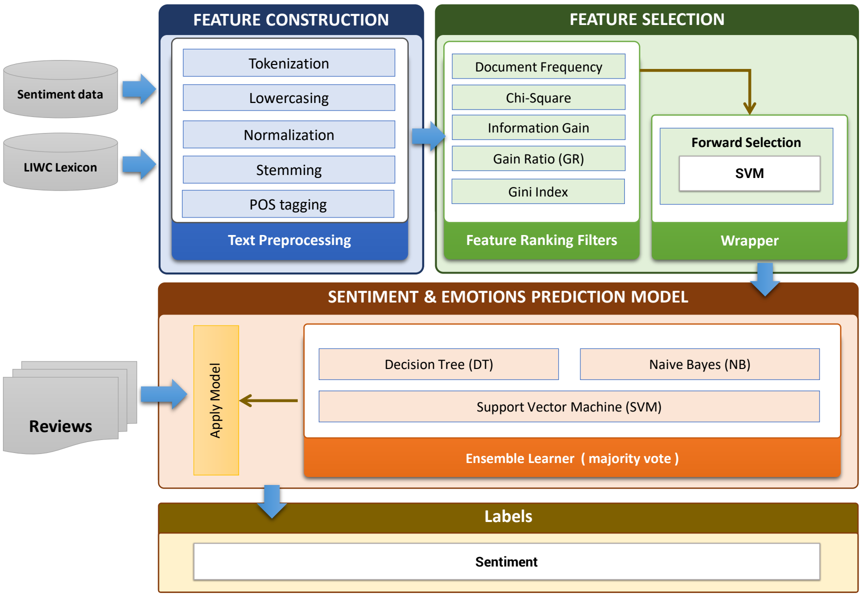


Figure 5.9: The workflow of sentiment analyzer.

Treebank scheme [91] for POS tagging pattern. For example, the feature ”excellent interface” filtered by the POS tag pattern ”JJ NN” and ”was disappointed” feature is filtered out by the pattern ”VBD VBN”. TF-IDF term weight scheme have been applied for word vector creation.

5.1.2.2 Feature Selection

Feature selection is the way to extract and select the most important and relevant features. It reduces the feature space dimensionality without losing too much information for an accurate prediction. The selected features are used to train the predictive model. We have employed both filter method and wrapper method for effective features selection. In the filtering method, the subset of important features/relevant features is selected by ranking them according to specific scoring schemes based on the intrinsic properties of the features. The low scoring features are removed while high scoring features are selected. The filter uses a fast evaluation function and is independent of the classifier. In the filter based method, we have used the filters like chi-square, Gini index, gain ratio, and information gain.

The word-vector is input in feature selection module. The individual filter assigns weight to each feature using their internal logic and select the initial subset features. We apply the majority voting method for the final feature selection. We set the threshold value to 3 that checks for common features selected by at least three filters. Then we have applied the wrapper method (forward selection process), in subset feature selection. In wrapper method, various subsets of features are generated and evaluated. The forward selection starts with an empty selection of features/attributes and, in each iteration, it adds new attribute of the given recordset. We have applied 10-fold cross-validation using SVM learner to estimate the performance, if the added attribute gives the higher performance then is added to the selection. Then a new round is started with the modified selection. We have added the stopping condition to stop the iteration if no significant increase in performance is observed.

5.1.2.3 Learning prediction model (Ensemble Learner)

We have employed the ensemble learning method for sentiment and emotion classification. Ensemble learning combines the predictions of multiple base learners to improve performance over a single learner. In this work, we have employed majority voting technique in conjunction with three base learners namely, Support Vector Machine (SVM), Nave Bayes (NB) and Decision Tree. Based on the majority voting of base learners, the user textual feedback is classified into either positive or negative class.

Based on the sentiment alignment of each review in R_i with extracted UXDs (the ith UXDs), we generated the structured data as shown in Table 5.6. We used the following equation x for sentiment orientation of D_i in online reviews R_i.

Table 5.6: Sentiment orientation toward each dimension

Online reviews	**UX Dimensions (UXDs)**					
	D1		D2		Dn	
	Pos	Neg	Pos	Neg	Pos	Neg
r1	1	0	0	1	0	0
r2					0	1
rn	0	1	1	0	1	0

$$S^{*}_{mi} = \begin{cases} 1 & if\ the\ sentiment\ orientation\ is\ * \\ 0 & otherwise \end{cases} \tag{5.7}$$

$*$ represent the sentiment orientation, where $* \in \{pos, neg\}$ as shown in Table 5.6. the sentiment values are encoded to nominal as, if the sentiment of in review is positive associated with dimension then $s_m^{pos} = 1$ and $s_m^{neg} = 0$; if the sentiment in review is negative then $s_m^{pos} = 0$ and $s_m^{neg} = 1$; if the sentiment in review is neutral then $s_m^{pos} = 0$ and $s_m^{neg} = 0$;

5.2 Part 3: Casual Effect Analyzer

In the latter part, we discussed measuring the effect of user positive or negative sentiments toward each UXDs on user satisfaction. We used the existing model called an ensemble neural network based model (ENNM) [1] for measuring that effect. The details of each part are described in the subsequent sections.

5.2.1 Ensemble neural network based model (ENNM) [1]

Jian-Wu B et al. [1] proposed the ENNM model to overcome the problems of the existing model used to user satisfaction models such as Gaussian distribution [19] and regression analysis. Most of the existing models for user satisfaction assume that online rating given by a user is a linear amalgamation of the sentiment regarding all the dimensions discussed in the online reviews. However, this assumption is not valid; there are many issues, such as there may be a complex combination of sentiments towards most of the dimensions in user online review. In order to resolve this issue, ENNM outperforms as compared to other models for modeling user satisfaction. Based on this reason, we also adopted the same model for measuring the sentiment effects towards each UXDs.

They used backpropagation neural networks (BPNNs) to handle the sparseness problem as shown in Figure 5.10. The ENNM consists of “T BPNNs” for assessing causal effect of user satisfaction on each UXDs.

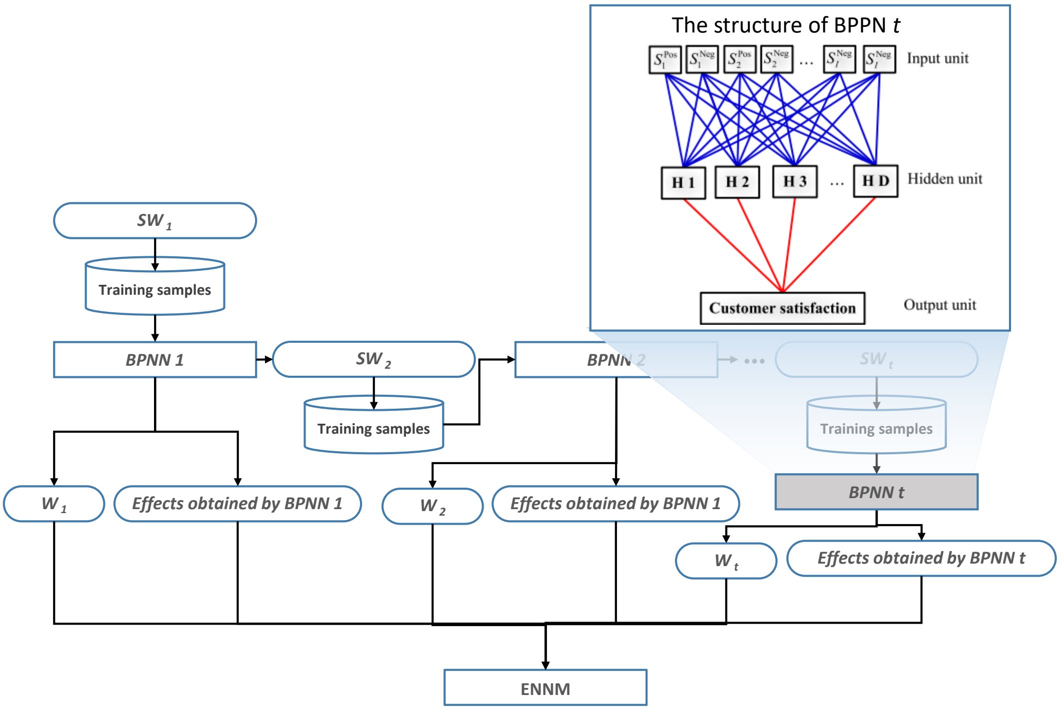


Figure 5.10: The training process of the ENNM model [1].

5.2.2 Kano Model

We employed the Kano model, developed by Kano et al. [87], which is a two-dimensional model. Kano model is a well-known model of user satisfaction. This model categorizes the product features into different classes such as must-be, performance, excitement, indifferent, and reverse. These features values are associated with user satisfaction [1]. Details of each feature are described as follows:

1. Must-be: These features are essential customers requirements and expectation and are taken for granted. These features must be fulfilled, otherwise, the product customer becomes dissatisfied.

2. One-dimensional (Performance): These features related to product quality promised by the product, service provider. These features have a direct impact on customer satisfaction when fulfilled.

3. Attractive (Excitement): These features give satisfaction, when filled, but have no effect on customer dissatisfaction.

4. Indifferent: These product features neither influence on user satisfaction nor dissatisfaction.

5. Reverse: These features state to a more significant degree of achievement, causing more customer dissatisfaction.

Based on the rules defined by [1], we also mapped the UXDs on the Kano model. We mapped the $\vec{w}_i^{pos}$ and $\vec{w}_i^{neg}$ on the Kano model 5 categories to for modeling the user satisfaction. The details of these rules are the following:

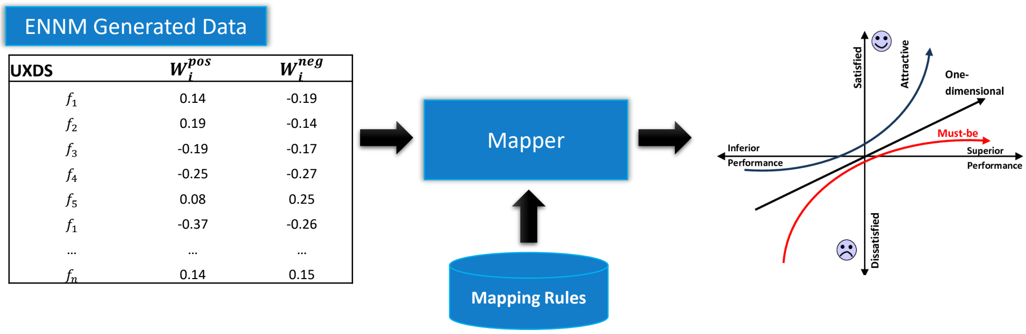


Figure 5.11: Mapping of UDXs on Kano Model.

1. If $\vec{w}_i^{pos} \leq$q 0 and $\vec{w}_i^{neg} < 0$ then UXD_i is a must-be.

2. If $\vec{w}_i^{pos} \leq 0$ and $\vec{w}_i^{neg} \geq 0$ then UXD_i is a reverse.

3. If $\vec{w}_i^{pos} > 0$ and $\vec{w}_i^{neg} < 0$ then UXD_i is a performance.

4. If $\vec{w}_i^{pos} > 0$ and $\vec{w}_i^{neg} \geq 0$ then UXD_i is a excitement.

Figure 5.11 depict the mapping on the Kano model based on the rules above.

5.3 Summary

User online reviews are generally in an unstructured form, the process to convert the unstructured data into structured data is an essential task for mining the UX dimensions for modeling user satisfaction. In this chapter, we discussed the process of (1) UX dimensions (UXDs) extraction using the proposed user experience word-embedding LDA (UXWE-LDA) topic modeling (2) sentiment analysis and its orientation for each extracted UXD from online user reviews. User Experience word-embedding LDA (UXWE-LDA) is an improved version of LDA, that automatically learn the domain knowledge from the given text corpus. UXWE-LDA improves the existing knowledge-based topic models by for the extraction of more domain dependent dimensions in the UX area through UGC. UXWE-LDA combines the topic modeling especially LDA with word embedding that automatically learns the domain knowledge from a large amount of textual data. The proposed method automatically gains the domain knowledge from the vast amount of documents using co-occurrence and word-embedding word vectors correlation of related data, which gives a more coherent topic. Then, we apply the sentiment analysis on the reviews concerning the extracted UXDs. To measure, the casual relationship of customer sentiment toward each UXDs on user satisfaction, an ensemble neural network based model (ENNM) method is used [1]. Finally, we map each dimension on the Kano model of satisfaction.

Chapter 6

Experimental Results and Evaluations

In order to evaluate the efficiency of the proposed solutions, different experiments are performed at different levels. We evaluated the different part of the proposed solution such as (i) the experimental results and evaluations of part-1 UXMCQ model (ii) the experimental results and evaluations of part-2 UXWE-LDA model and sentiment analyzer. We used different datasets for each part. The detail explanation and results are discussed in the following section.

6.1 Part-1: UX multi-criteria Qualifier (UXMCQ) Model

We evaluate UXMCQ for domain aspect classification by comparing with two LDA-based approaches.

6.1.0.1 Datasets

We used the restaurant reviews dataset [107] which contains different aspects related to food, staff, and others written in English.

6.1.0.2 Results

We compare UXMCQ against the results with two LDA-based approaches, LocLDA [108] and ME-LDA [109]; these models are based on unsupervised approach. However, results generated by these models required supervision for labeling the extracted topics. They manually labeled the extracted topic. UXMCQ assign a label to extracted topics based on the domain aspects configuration, so no manual topic labeling are needed. Table 6.1 and Figure 6.1 shows the comparison result, UXMCQ achieves slightly better overall performance over the other LDA base models.

Table 6.1: Comparison of UXMCQ model classification against other two LDA based model.

Method	Aspects											
	Staff			Food			Ambiance			Overall		
	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1	Pre.	Rec.	F1
LocLDA	**0.80**	0.59	0.68	0.90	0.65	0.75	0.60	0.68	0.64	0.77	0.64	0.69
ME-LDA	0.61	0.54	0.64	0.87	**0.79**	**0.83**	**0.77**	0.56	**0.65**	**0.81**	0.63	0.70
UXMCQ	0.78	**0.86**	**0.71**	**0.96**	0.69	0.81	0.55	**0.75**	0.63	0.70	**0.77**	**0.72**

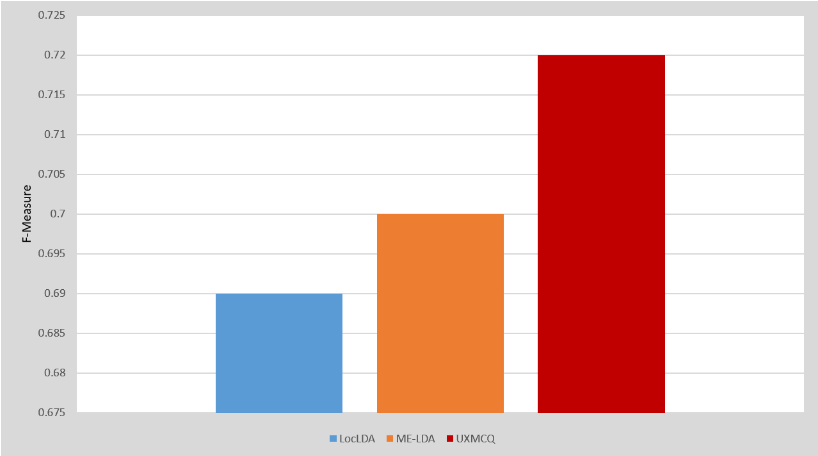


Figure 6.1: Comparison of UXMCQ with other LDA base model.

6.2 Part-2: Topic extractor and sentiment analyzer

6.2.1 Topic extractor

6.2.1.1 Datasets

We used the dataset from [110] which contains the reviews data of both electronics and non-electronics products. Each domain category consists of 50 different of products with total 1,000 reviews.

Topic coherence: For topic modeling evaluation, we use the UMass topic coherence [22] metics. The topic coherence (TC) metrics calculates the words relatedness within the topics, higher coherence values means a good topic. TC is computed as:

$$C(t; V^{(t)}) = \sum_{m=2}^{M} \sum_{l=1}^{m-1} \log \frac{D(v_m^{(t)}, v_l^{(t)}) + 1}{D(v_l^{(t)})}$$

6.2.2 Results

This section, we show an example of topics generated by UXWE-LDA, WE-LDA, and LDA to show an improvement by our proposed topic extractor. The red color in each topic in the given Table 6.2 show errors, as UXWE-LDA extracted more coherent and meaningful topic as compared to other baseline models.

Table 6.2: Example topics generated by UXWE-LDA, WE-LDA and LDA. Red marked shows an error.

Electronic products dataset (TV -Screen)			Non- electronic products (Food)		
UXWE-LDA	WE-LDA	LDA	UXWE-LDA	WE-LDA	LDA
screen	image	screen	vegetable	flavor	popcorn
headset	video	side	cook	sweet	functions
microphone	end	big	meat	salt	bag
speaker	microsoft	line	pizza	market	potato
voice	logitech	flat	soup	fruit	healthy
sound	resolution	image	baked	natural	chicken
mic	top	top	delicious	spice	meat
audio	cd	lcd	Chicken	strong	sweet
conversation	pc	resolution	bean	ingredient	number
mike	skype	sound	yummy	delicious	machine
loud	chat	strap	winter	bean	basic
bottom	picture	substitute	simple	open	spice

We did the parameters tuning of UXWE-LDA such as number of top seed words (n), and words similarity (m) and trust score(u).

we examine the sensitivity of the three parameters of UXWE-LDA such as top seed words n, most similar words m and the trust score U. The number of top 15 words gives us more coherent topic with higher TC value with other parameters setting for the given dataset. Figures 6.2 depict the average TC on the both datasets using number of top seed n = 5, 10, 15, 20, 25, 30 respectively. The results reveals that top 15 seed word gives higher TC value for electronic data, and for non-electronic dataset the top 25 seeds words gives us higher TC value. So that can conclude that either few seed not too many seeds words generate the coherent topic.

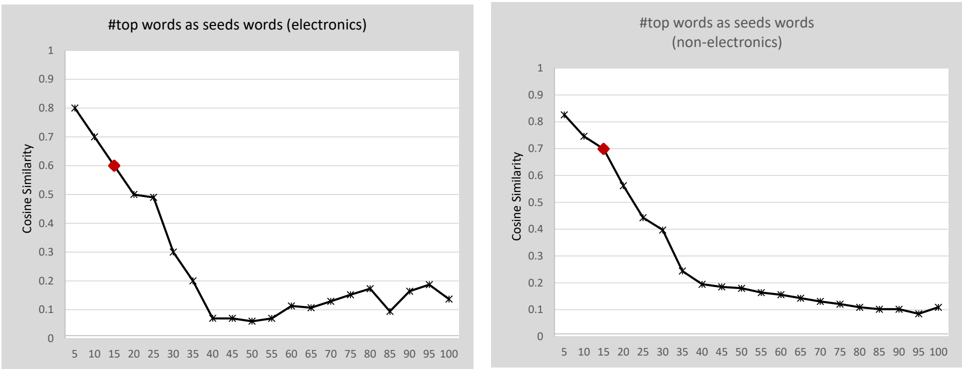


Figure 6.2: Average cosine similarity per the number of topics on both datasets.

Figure.6.3 demonstrations the average TC with m from 5 to 100. From results we reveals that TC increase with more similar words at initial stage, and gives us higher TC value at 15 on electronic dataset, for non-electronic dataset the model gives higher value at 15.

We examined for electronic products dataset, TC increases with more similar words at the beginning, then becomes unchanging and gives higher value at m = 15. For non-electronic product dataset, the UXWE-LDA model gives almost similar TC value and higher at m =25. This shows that high quality of knowledge is generated by must-links which are produced by the best seed words and word similarity. The similarity computation using TC ensure the quality of a must-link that proper knowledge is incorporated into UXWE-LDA.

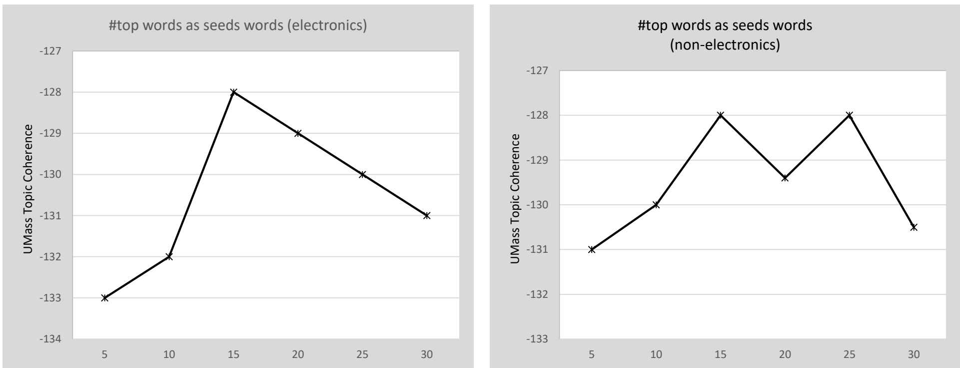


Figure 6.3: Average TC of top n words with different number of seed words on both datasets.

Figure 6.4 shows the average TC of each model using different number of topics on two datasets.

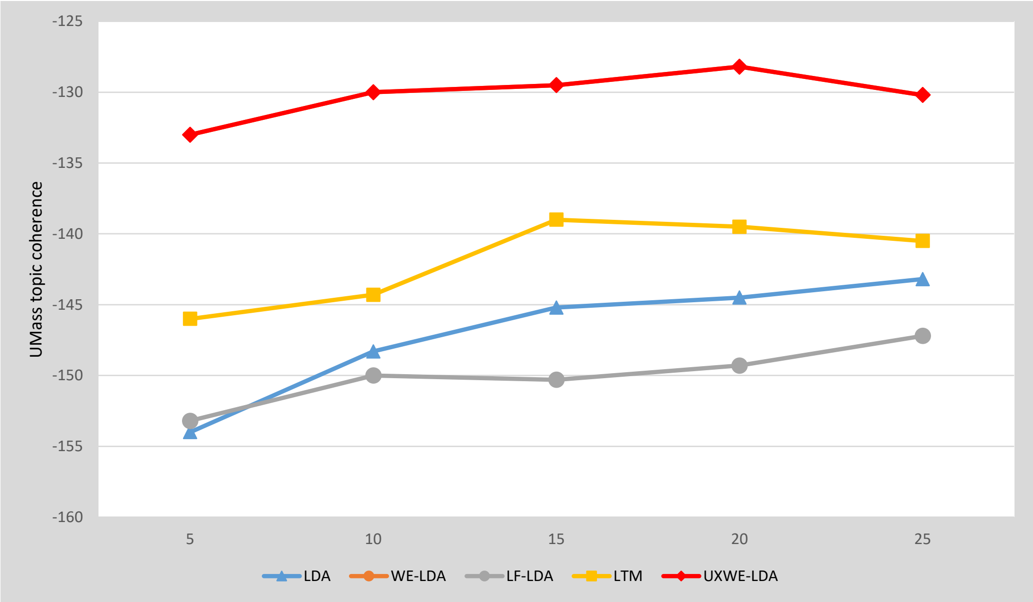


Figure 6.4: Average TC of top words with different number of topics on on both datasets.

The results show that with the different number of topics and setting, UXWE-LDA always gives higher TC value than other models, which shows that the UXWE-LDA is vigorous with a different combination of must-link clusters. Enhancements of UXWE-LDA over other models with p-value (p¡ 0.007) significant in 2-tailed paired t-test.

We also evaluated the topic consistency using a PMI score generated by the topic extractor with base-lines topic models. Figure 6.5 , Figure 6.6, and Figure 6.7 depict the comparison with PMI scores for the generated topics by Top-n topic words. Overall, Overall, this UXWE-LDA model and the WE-LDA model gives comparatively better results, both with chance wins; the LTM model and LF-LDA followed, and the LDA model is the weakest. LDA model is fragile because of the sparsity of little texts.

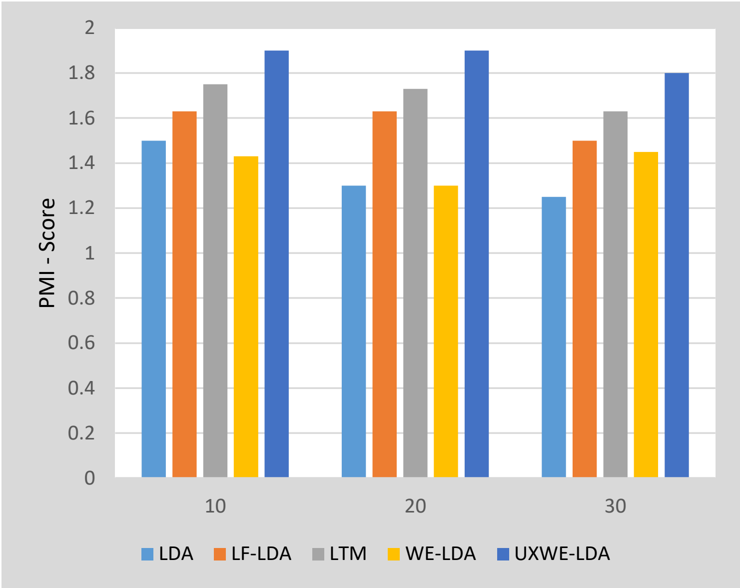


Figure 6.5: PMI-Score- TOP-5 topic words.

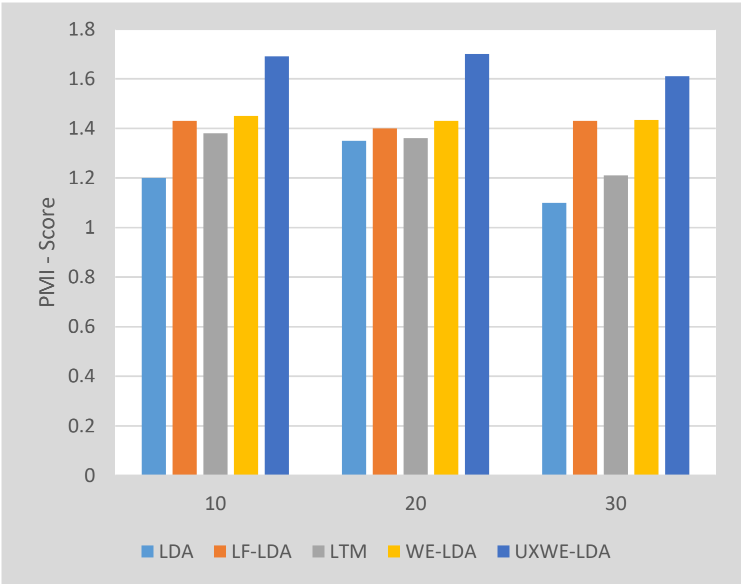


Figure 6.6: PMI-Score - TOP-10 topic words.

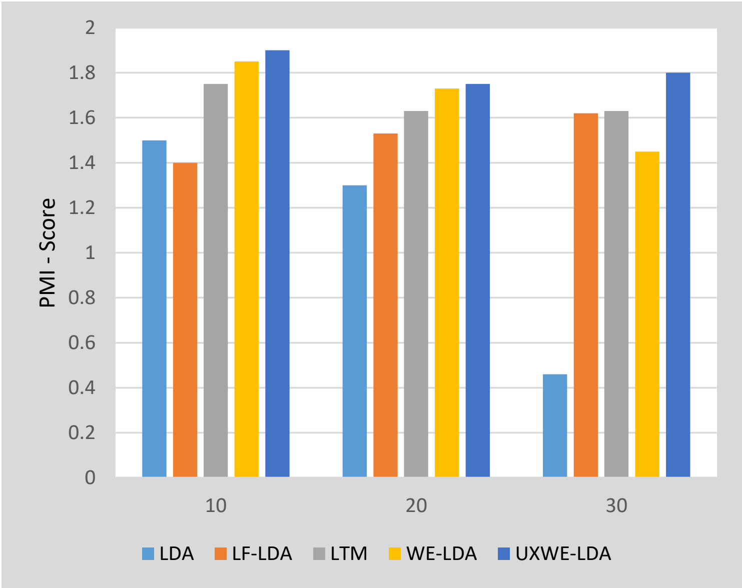


Figure 6.7: PMI-Score - TOP-20 topic words.

6.2.3 Sentiment Analyzer

6.2.3.1 Datasets

For SA, we used publically available dataset know as Cornell movie reviews dataset, 2). Amazon product reviews datasets (a) and, 3). Amazon product reviews datasets (b). The Cornell movie reviews dataset consists of total 2000 reviews with an equal number of positive and negative reviews. The Amazon product review dataset “a” consist of four types of products reviews(Book, DVD, Electronic, and Kitchen). Each product reviews set having 2000 reviews with equal distribution of positive and negative reviews. The third dataset “b” of product reviews contains 10000 reviews with 5000 are positive reviews, and 5000 are negative reviews. The dataset details are given in Table 6.3.

For SA model evaluation, we used an accuracy metric to determine the overall SA classification performance.

Table 6.3: Detailed description of the datasets.

Dataset A: Cornell Movie & Amazon product reviews				Dataset B: Amazon product reviews			
Domain	Positive Reviews	Negative Reviews	Total reviews	Domain	Positive Reviews	Negative Reviews	Total reviews
Movie	1000	1000	2000	Movie	5000	5000	10000
Book	1000	1000	2000	Book	5000	5000	10000
DVD	1000	1000	2000	DVD	5000	5000	10000
Electronics	1000	1000	2000	Electronics	5000	5000	10000
Kitchen	1000	1000	2000	Music	5000	5000	10000

6.2.3.2 Results and comparison

We performed different experiments on the different filters based feature selection algorithms with different numbers of features on different datasets. The Table 6.4 shows the classification accuracies.

Table 6.4: Sentiment analysis classification accuracies with a different number of feature combination for each dataset.

Methods	Number of features	Movie dataset				Book dataset				DVD dataset			
		SVM	NB	GLM	CE	SVM	NB	GLM	CE	SVM	NB	GLM	CE
IG	300	87.69	88	84.44	87.75	81.81	83.87	82.31	84.5	80.69	81.23	79.3	
	400	88.25	88.07	86.06	88.56	84.44	85.62	83.75	85.8	82.06	83.6	80.05	83.9
	500	88.69	89.06	87.31	88.84	85.56	87.19	83.44	85.6	85.38	84.2	82.5	84.5
	600	89.5	88.98	87.62	89.8	88.12	88.63	86	89.2	85.8	84.7	83.3	86.9
	700	90.25	89.56	88.19	90.75	89.19	89.88	86.9	89.8	86.9	85.9	84.1	87.5
	800	90.75	89.81	88.25	90.93	89.5	89.94	87.56	90.6	87.87	86.5	85.6	88.7
	900	90.8	89.85	88.28	90.95	89.53	89.96	87.59	90.62	87.89	86.53	85.62	88.74
	1000	90.84	89.88	88.33	90.97	89.55	89.97	87.61	90.64	87.94	86.57	85.65	88.77
Average		**89.6**	**89.15**	**87.31**	**89.82**	**87.21**	**88.13**	**85.65**	**88.35**	**85.57**	**84.9**	**83.27**	**86.16**
MRMR	300	89.5	88	85.87	89.81	82.25	84.38	83.31	84.7	81	81.9	80	82.36
	400	89.8	89.19	87.87	89.31	85.5	86.37	83.75	86.2	83.5	83.9	80.2	83.7
	500	91.69	90.06	88.06	91.88	86.69	87.56	84.44	88.1	85.4	85.1	83.7	85.6
	600	92.5	90.13	88.75	92.12	87.19	88.8	85	87.95	86.3	85.6	84.1	87.52
	700	90.81	90.69	88.81	91.88	88.19	88.9	86.19	89.9	87	87.5	85.9	87.77
	800	92.94	91.94	90.44	93.3	89	89.98	87.56	90.3	88.15	87.8	86.11	88.4

	900	92.95	91.96	90.47	93.35	89.05	90	87.59	90.35	88.18	87.98	86.17	88.45
	1000	92.97	91.98	90.49	93.4	89.1	90.06	88.63	90.37	88.2	88.05	86.25	88.5
Avereage		**91.46**	**90.28**	**88.85**	**91.88**	**87.12**	**88.26**	**85.81**	**88.48**	**85.97**	**85.98**	**84.05**	**86.54**
CHI	300	85.38	84.87	83.69	85.7	81.1	82.8	81.2	83.3	79.5	80	78.7	80.74
	400	86.5	85.56	84.1	87.1	82.45	83.6	82	84.1	80.05	81.15	79.4	81.43
	500	85.88	85.75	84	86.33	83.25	83.9	82.5	84.5	82.3	81.95	80	82.5
	600	86.69	86.12	84.4	86.91	84.9	85.6	82.8	85.1	83.5	82.2	80.16	82.1
	700	87	85.75	84.6	86.13	87	86.23	83.5	86.5	84.7	83.9	82	84.9
	800	87.69	86.06	85.7	88.52	88.74	87.44	85.9	88.9	85.8	84.5	82.6	84.95
	900	87.73	86.1	85.73	88.54	88.75	87.6	85.94	88.93	85.84	84.53	82.65	84.97
	1000	87.75	86.13	85.77	88.58	88.78	87.85	85.96	88.96	86.9	84.6	82.68	84.99
Average		**86.83**	**85.79**	**84.75**	**87.23**	**85.62**	**85.63**	**83.73**	**86.29**	**83.57**	**82.85**	**81.02**	**83.32**
GR	300	86.94	85.88	84	87.15	81.5	83.8	82.3	84.1	80.5	81	79.1	81.8
	400	88.44	87.87	86.1	88.92	84	85.3	83.25	85.93	81.5	82.6	80.05	82.85
	500	89.05	89	87	89.3	85.5	86.9	84.2	87.3	82.35	83.4	81.12	83.7
	600	89.88	89.56	87.3	90.5	87.81	88.5	85.9	88.2	83.9	84.5	81.98	84.6
	700	91.13	90.69	87.5	91.82	89.06	89.5	86.8	89.2	85.5	85.9	83.78	85.8
	800	91.44	91.4	88.7	91.2	89.6	89.9	87.6	90.2	86.5	87.2	84.25	86.7
	900	91.47	91.45	88.73	91.5	89.63	89.92	87.64	90.24	86.55	87.23	84.29	87.3
	1000	91.49	91.48	88.75	91.59	89.65	89.95	87.67	90.27	86.85	87.27	84.33	87.4

Average		**89.98**	**89.67**	**87.26**	**90.25**	**87.09**	**87.97**	**85.67**	**88.18**	**84.21**	**84.89**	**82.36**	**85.02**
GI	300	86.56	85.9	83.8	86.1	81	82.9	81.5	83.6	80	80.66	79.2	80.89
	400	87.19	86.8	84.2	87.92	82.9	83.7	82.5	84.7	81.4	81.86	80.13	82.16
	500	88.44	87.1	86.5	89.4	83.6	84	83.56	85.1	82.59	83.55	80.5	82.7
	600	89.38	88.13	87.56	90.5	85	85.96	83.95	85.8	83.75	84.47	81.96	83.4
	700	89.56	89.1	86.7	89.7	87.9	86.4	85.9	88.9	84.91	85.3	83.1	85.78
	800	90.62	89.6	86.9	89.9	88.85	87.6	85.8	87.98	85.17	85.93	84.6	86.29
	900	90.65	89.62	87	90.7	88.88	87.95	85.9	88.9	85.22	85.95	84.66	86.4
	1000	90.68	89.7	87.2	90.75	88.9	88.6	85.96	88.95	85.26	85.97	84.7	86.47
Average		**89.14**	**88.24**	**86.23**	**89.47**	**85.88**	**85.89**	**84.38**	**86.74**	**83.54**	**84.21**	**82.36**	**84.26**

The results of experiments shown in Table 6.5 and Figure 6.8 reveal that the ensemble method with minimal feature selection strategies can effectively increase the accuracy of classification compared with the baseline classifier.

Table 6.5: Sentiment analyzer model accuracy using different datasets.

Dataset	Classifier	# of Features	Accuracy
Movie	SVM	3625 1209	93
	NB	2400 1375	92
	DT	3816 1254	88
	Ensemble	3779 1314	94
	Average	3405	91.75
Book	SVM	2199 1066	87
	NB	2612 1074	86
	DT	2031 1048	83
	Ensemble	2956 1021	89
	Average	**2449**	**86.25**
Electronic	SVM	1323 474	85
	NB	1002 1090	89
	DT	1938 625	87
	Ensemble	1760 855	86
	Average	1505	86.75
Kitchen	SVM	1843 770	89
	NB	1566 470	86
	DT	1600 787	89
	Ensemble	1969 877	90
	Average	**1744**	**88.5**
DVD	SVM	642 296	89
	NB	819 276	87
	DT	855 267	86
	Ensemble	362 155	88
	Average	**669**	**87.5**

6.3 Overall Comparison

6.3.1 Extrinsic UXDs extraction evaluation

We performed an extrinsic evaluation by comparing the UXWE-LDA inferred topic with the gold-label topic assigned by the three human experts in the field of NLP and text mining. The human

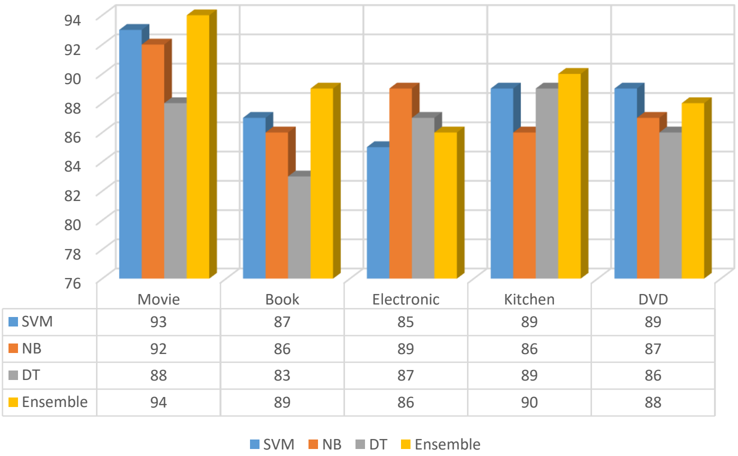


Figure 6.8: Average classification performance on top k high ranked score feature utilizing wrapper and filters feature selection, ensemble learner

experts annotated by a total of 300 online reviews, where each sentence is label based on the provided UX dimension list. Mutually agreed sentence all three annotators were considered as gold-label for the performance evaluation. We employed the topic-wise performance metrics (recall, precision, and F1 score) for comparison with LDA baseline algorithms. Precision means the percentage of correct classifications of that topic among all gold-label reviews sets, where the UXWE-LDA model predict that topic. Where recall for a topic is the portion of correct classifications of that topic out of all the cases of that topic in the gold-label reviews. The F1 score of a topic is the harmonic mean of recall and precision of that topic and is given in Equation 6.1.

$$F_1 Score = 2 \times \frac{precision_k \times recall_k}{precision_k + recall_k} \tag{6.1}$$

Where higher F1 score indicates, the model performs well for classifying the test data.

Figure 6.9 has shown that UXWE-LDA achieves 3 % improvement on average as compared with LDA due to the incorporation of semantic domain knowledge.

Table 6.6: Topic-wise performance measures.

Topics	LDA			UXWE-LDA		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
attractiveness	0.71	0.46	0.55	0.83	0.72	0.77
dependability	0.78	0.49	0.60	0.80	0.91	0.85
efficiency	0.73	0.60	0.66	0.76	0.77	0.76
perspicuity	0.80	0.47	0.59	0.80	0.72	0.76
novelty	0.76	0.51	0.61	0.81	0.76	0.78
stimulation	0.75	0.47	0.58	0.87	0.81	0.84

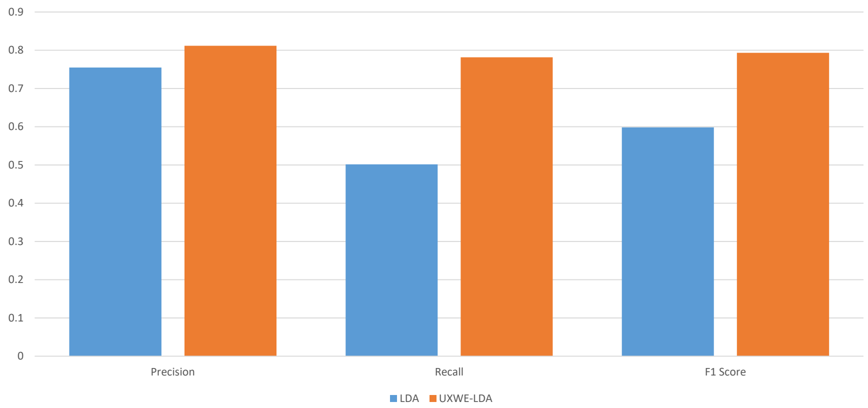


Figure 6.9: Average F-measure, Precision and Recall of LDA and UXWE-LDA.

6.3.2 Expert Base Evaluation

We compared the extracted dimensions using UXWE-LDA analysis with manually extracted by Human expert for validation. We used the Jaccard coefficient similarity [111] to check the degree of dimensions overlapping between automatic extraction using UXWE-LDA and human experts. The Jaccard coefficient is calculated as Equation 6.2

$$JC = \frac{|D_{UXWE-LDA} \cap D_{Exp}|}{|D_{UXWE-LDA} \cup D_{Exp}|} \tag{6.2}$$

Where $D_{UXWE-LDA}$ dimensions are extracted using automatic UXWE-LDA analysis and D_{EXP} are dimensions extracted by human experts through manually rigors process. The higher the Jaccard coefficient's value, the higher the degree of overlap between the two sets of dimensions, as shown in Table 6.7. Three researchers were invited having hands-on NLP and text mining to extracts the UXDs from the randomly selected online reviews. Each researcher selected 50 reviews randomly; finally, a total of 150 reviews selected for UXWE-LDA validation. We compared the UXDs extracted from UXWE-LDA with the UXDs extracted by the human experts for checking the reliability of the result generated by UXWE-LDA. The Jaccard coefficient for both researchers and UXDs extracted by UXWE-LDA model are 0.3, 0.5, and 0.4, respectively. This concludes that our study inferred new latent variables or dimensions from the online reviews. We claim that our study outcomes are more reliable for generalization due to large corpus textual data. Due to complexity and ambiguity involves in UXDs extraction task from online reviews, the results show that UXWE-LDA is a reliable and suitable approach for UXDs extraction from online reviews.

Table 6.7: A comparison of UXDs between UXWE-LDA model and human experts.

Dimensions	**UXWE-LDA**	**Human Expert 1**	**Human Expert 2**	**Human Expert 3**
Attractiveness	✓	X	✓	✓
Dependability	✓	✓	✓	✓
Efficiency	✓	X	✓	X
Perspicuity	✓	X	✓	✓
Novelty	✓	✓	✓	✓
Stimulation	✓	✓	X	X
Aesthetics	X	✓	X	✓
Complexity	X	✓	✓	✓
Affect and emotion	X	X	✓	X

6.4 Case Study 1 - Video Games Reviews

We used the publically available amazon data [112] of user reviews related to games reviews. The online reviews contain different words used by the different users to express their opinion; some words form the long tail as depicted in Figure 6.10. In total, 122502 numbers of words were considered after applying the preprocess for UX dimensions extraction.

Figure 6.11 shows the frequency of user satisfaction score in the term of rating in the used dataset.

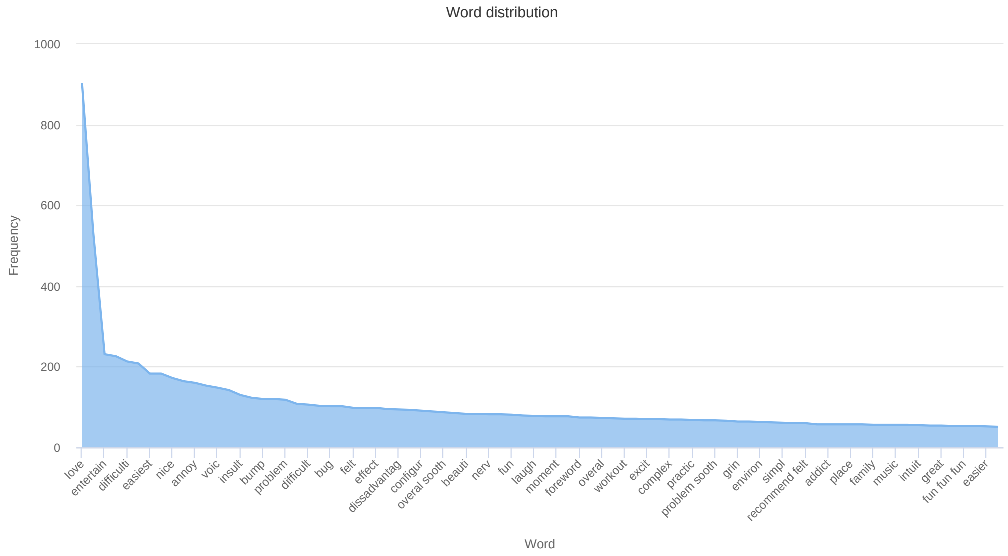


Figure 6.10: Word distribution for games reviews.

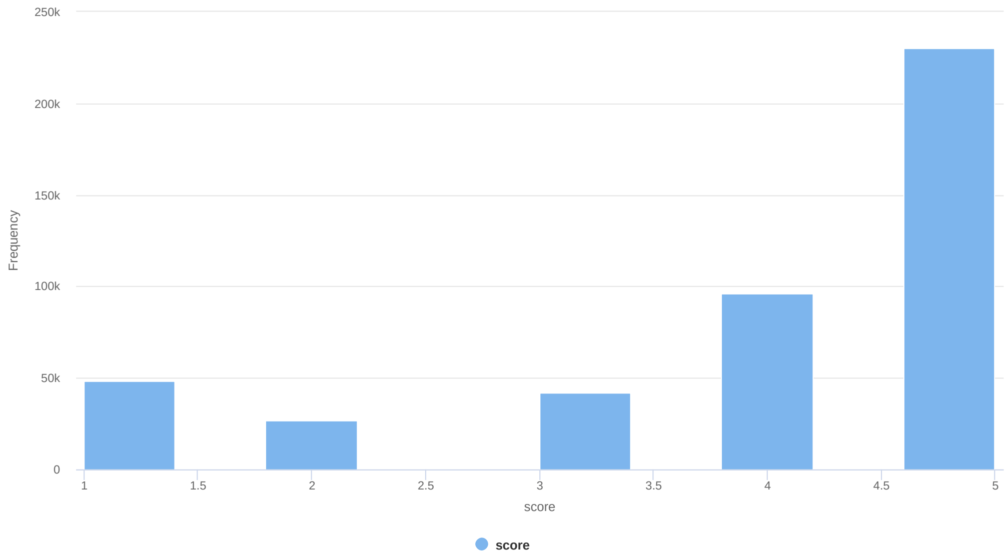


Figure 6.11: Overall user rating on reviews.

First, we applied the UXMCQ model for filtering out the unrelated reviews. Then applied the UXWE-LDA model for the dimension extraction. Figure 6.12 shows the extracted dimensions from the online reviews.

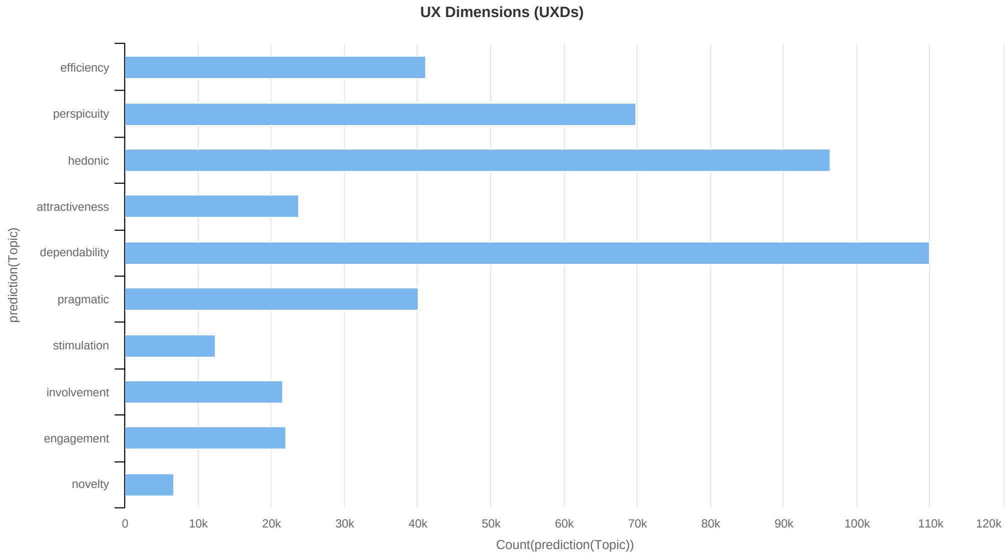


Figure 6.12: Extracted UX dimensions from user reviews.

The user sentiment orientations towards each UXD of online user reviews are shown in Figure 6.13. The results show that the users have positive opinions towards the extracts UX dimensions as compared to negative.

We used the structured data having W_i^{pos} and W_i^{neg} vectors generated by part-2 of the proposed methodology to train the ENNM Model as shown in Table 6.8.

Table 6.8: The values of positive and nagative vectors generated by ENNM

UXD	W_i^{pos}	W_i^{neg}
attractiveness	0.14	-0.19
dependability	0.19	-0.14
efficiency	-0.19	-0.17
engagement	-0.25	-0.27
hedonic	0.08	0.25
involvement	-0.37	-0.26
perspicuity	0.14	0.15
pragmatic	-0.18	0.04
stimulation	0.03	-0.08

According to Table 6.8 generated by the ENNM model, the category of each UXDs of game reviews can be identified, and maps in the Kano Model, as shown in Figure 8.

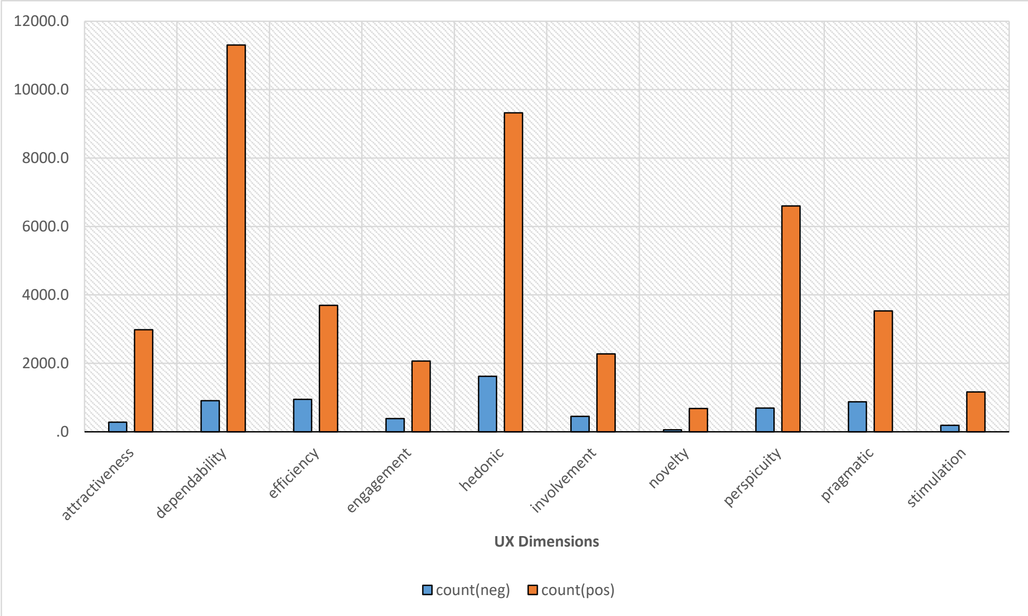


Figure 6.13: The sentiment orientations result towards each extracted UXD.

In Figure 6.14, is the threshold for determining whether a UXD is an indifferent UXD. As it can be seen from Figure 6.14, the UXD that identified as excitement UXDs includes: hedonic and perspicuity; pragmatic identified as reversed UXD; must-be UXD includes involvement and efficiency; finally, performance UXD consists of three dimensions (stimulation, attractiveness, and dependability).

6.5 Case Study 2 - Google App Reviews

We used the publically available google play store dataset [112] of user reviews related to application reviews. The online rating distribution of the data sets is shown in Figure 6.15.

This dataset consists of different categories of apps both paid and free. The overall statistics of the app categories are shown in Figure 6.16.

First, we applied the UXMCQ model for filtering out the unrelated reviews. Then applied the UXWE-LDA model for the dimension extraction.

The user sentiment orientations towards each UXD of online user reviews are shown in Figure

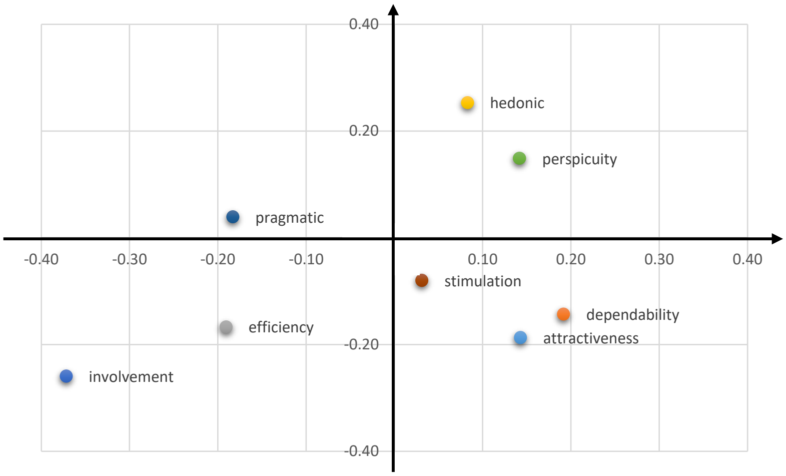


Figure 6.14: Mapping the extracted dimensions on Kano Model.

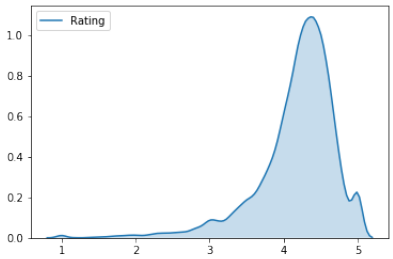


Figure 6.15: overall rating of user on online reviews

6.17. The results show that the users have positive opinions towards the extracts UX dimensions as compared to negative.

We used the structured data having W_i^{pos} and W_i^{neg} vectors generated by part-2 of the proposed methodology to train the ENNM Model as shown in Table 6.9.

According to Table 6.8 generated by the ENNM model, the category of each UXDs of game reviews can be identified, and maps in the Kano Model, as shown in Figure 8.

In Figure 6.18, is the threshold for determining whether a UXD is an indifferent UXD. As it can be seen from Figure 6.18, the UXD that identified as excitement UXDs includes: hedonic and perspicuity; pragmatic identified as reversed UXD; must-be UXD includes involvement and

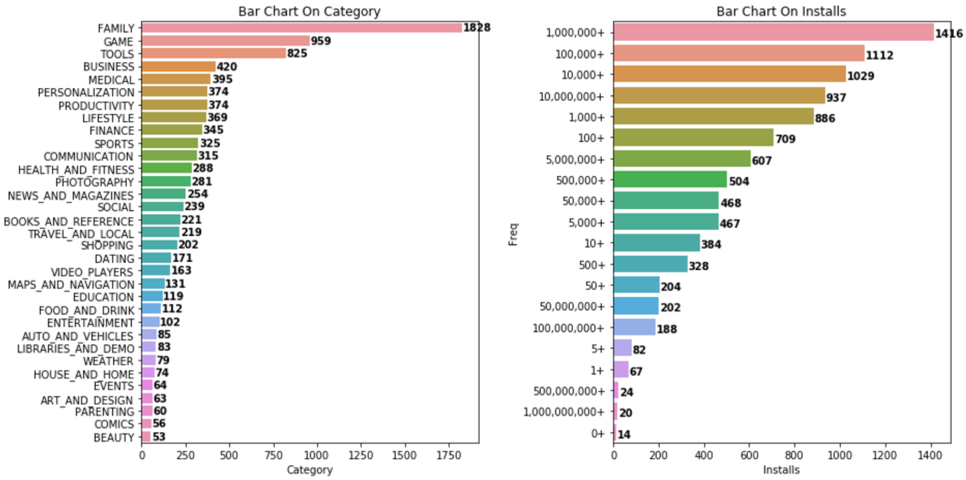


Figure 6.16: The apps distributions.

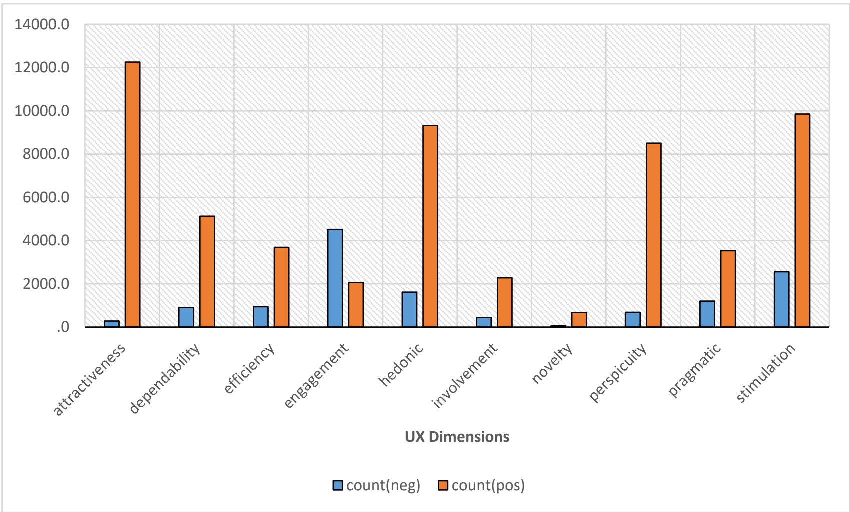


Figure 6.17: The sentiment orientations result towards each extracted UXD.

efficiency; finally, performance UXD consists of three dimensions (stimulation, attractiveness, and dependability).

Table 6.9: The values of positive and nagative vectors generated by ENNM

UXD	W_i^{pos}	W_i^{neg}
attractiveness	0.14	-0.19
dependability	0.19	-0.14
efficiency	-0.19	-0.17
engagement	-0.25	-0.27
hedonic	0.08	0.25
involvement	-0.37	-0.26
perspicuity	0.14	0.15
pragmatic	-0.18	0.04
stimulation	0.03	-0.08

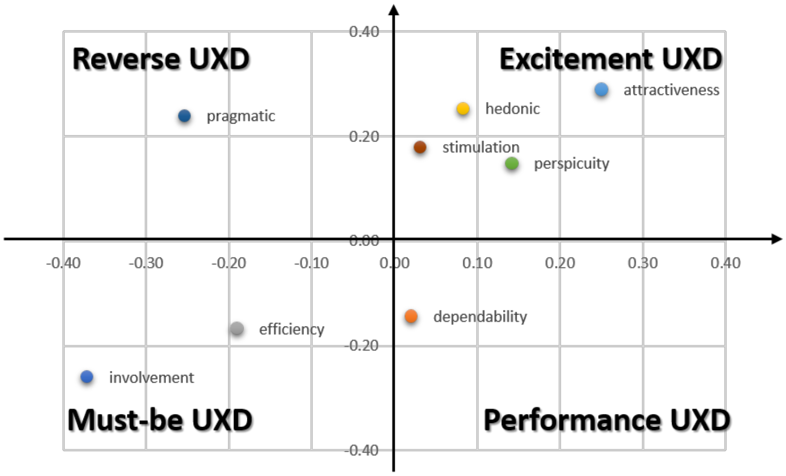


Figure 6.18: Mapping the extracted dimensions on Kano Model.

Chapter 7

Conclusion and Future Directions

7.1 Conclusion

Due to advancement in social media platforms, user daily posted their opinions in the form of online reviews. These online reviews contain beneficial information related to UX. These reviews can be used for understanding the UX and UX modeling. The main goal of this thesis to mine UX related information from these substantial online reviews automatically. The automatic approach overcomes the problems of a manual analysis of those vast data. To this end, we designed a comprehensive framework for modeling UX from online reviews. In the method, first, we filter those reviews unrelated to UX domain using UX multi-criteria qualifiers (UXMCQ). Then, we extract the UXDs from the filtered reviews using enhanced topic extraction methodology called UXWE-LDA. UXWE-LDA improve the existing knowledge-based topic models by for the extractions of more domain dependent dimensions in the UX area through UGC. Finally, the method for modeling user satisfaction on kano model by mapping the UX dimensions. The presented study has potential implication in product design.

- It can extract those UX aspects from online reviews that customers are most concerned about.
- It can be used to mine the user opinion toward each UX aspect so that that product designers can make a better decision to improve the positive UX of their customers. Additionally, they can further know the strengths and weaknesses of the product.
- This method allows the product designer to understand the different categories of UDXs in term of the Kano model, which is essential for product enhancement. According to the clas-

sification results of UXDs, the priority order of UXDs for developing product enhancement plans can be determined.

7.2 Future Directions

There exist some significant limitations in terms of computation cost and processing time in the study. However, over time, the advancement in technology and computing techniques can overcome these limitations. Also, the study neglected some of the words like infrequent words, which helps in indicating user preference and needs for a product or services. Therefore, we need to examine effective solution for incorporating word embedding. Furthermore, the experiment also needs to be extended with other settings as well as datasets to overcome the existing limitations.

Bibliography

[1] J.-W. Bi, Y. Liu, Z.-P. Fan, and E. Cambria, "Modelling customer satisfaction from online reviews using ensemble neural network and effect-based kano model," *International Journal of Production Research*, pp. 1–21, 2019.

[2] C. E. Lin and Y.-H. Lai, "Quasi-ads-b based uav conflict detection and resolution to manned aircraft," *Journal of Electrical and Computer Engineering*, vol. 2015, p. 14, 2015.

[3] F. Pucillo and G. Cascini, "A framework for user experience, needs and affordances," *Design Studies*, vol. 35, no. 2, pp. 160–179, 2014.

[4] E. L.-C. Law, V. Roto, M. Hassenzahl, A. P. Vermeeren, and J. Kort, "Understanding, scoping and defining user experience: a survey approach," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2009, pp. 719–728.

[5] M. Hassenzahl, "User experience (ux): towards an experiential perspective on product quality." in *IHM*, vol. 8, 2008, pp. 11–15.

[6] "International organization for standardization, iso 9241-11:2018. ergonomics of human-system interaction – part 11: Usability: Definitions and concepts," ISO, 2018.

[7] J. Hussain, W. A. Khan, T. Hur, H. S. M. Bilal, J. Bang, A. U. Hassan, M. Afzal, and S. Lee, "A multimodal deep log-based user experience (ux) platform for ux evaluation," *Sensors*, vol. 18, no. 5, p. 1622, 2018.

[8] Z. Kosmadoudi, T. Lim, J. Ritchie, S. Louchart, Y. Liu, and R. Sung, "Engineering design using game-enhanced cad: The potential to augment the user experience with game elements," *Computer-Aided Design*, vol. 45, no. 3, pp. 777–795, 2013.

[9] E. L.-C. Law, P. Van Schaik, and V. Roto, "Attitudes towards user experience (ux) measurement," *International Journal of Human-Computer Studies*, vol. 72, no. 6, pp. 526–541, 2014.

[10] F. Xu, L. La, F. Zhen, T. Lobsang, and C. Huang, "A data-driven approach to guest experiences and satisfaction in sharing," *Journal of Travel & Tourism Marketing*, vol. 36, no. 4, pp. 484–496, 2019.

[11] L. Zhang, X. Chu, H. Chen, and B. Yan, "A data-driven approach for the optimisation of product specifications," *International Journal of Production Research*, pp. 1–19, 2018.

[12] B. Yang, Y. Liu, Y. Liang, and M. Tang, "Exploiting user experience from online customer reviews for product design," *International Journal of Information Management*, vol. 46, pp. 173–186, 2019.

[13] S. Hedegaard and J. G. Simonsen, "Extracting usability and user experience information from online user reviews," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013, pp. 2089–2098.

[14] Y. Guo, S. J. Barnes, and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," *Tourism Management*, vol. 59, pp. 467–483, 2017.

[15] H. Q. Vu, G. Li, R. Law, and Y. Zhang, "Exploring tourist dining preferences based on restaurant reviews," *Journal of Travel Research*, vol. 58, no. 1, pp. 149–167, 2019.

[16] Z. Chen, N. Ma, and B. Liu, "Lifelong learning for sentiment classification," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, vol. 2, 2015, pp. 750–756.

[17] S. N. Ahmad and M. Laroche, "Analyzing electronic word of mouth: A social commerce construct," *International Journal of Information Management*, vol. 37, no. 3, pp. 202–213, 2017.

[18] J. Khan, A. Alam, J. Hussain, and Y.-K. Lee, “Enswf: effective features extraction and selection in conjunction with ensemble learning methods for document sentiment classification,” *Applied Intelligence*, pp. 1–23, 2019.

[19] M. Farhadloo, R. A. Patterson, and E. Rolland, “Modeling customer satisfaction from unstructured data using a bayesian approach,” *Decision Support Systems*, vol. 90, pp. 1–11, 2016.

[20] S. Tirunillai and G. J. Tellis, “Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation,” *Journal of Marketing Research*, vol. 51, no. 4, pp. 463–479, 2014.

[21] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-Graber, and D. M. Blei, “Reading tea leaves: How humans interpret topic models,” in *Advances in neural information processing systems*, 2009, pp. 288–296.

[22] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum, “Optimizing semantic coherence in topic models,” in *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2011, pp. 262–272.

[23] L. Yao, Y. Zhang, Q. Chen, H. Qian, B. Wei, and Z. Hu, “Mining coherent topics in documents using word embeddings and large-scale text data,” *Engineering Applications of Artificial Intelligence*, vol. 64, pp. 432–439, 2017.

[24] M. Hassenzahl and N. Tractinsky, “User experience-a research agenda,” *Behaviour & information technology*, vol. 25, no. 2, pp. 91–97, 2006.

[25] Y. Liang, Y. Liu, H. T. Loh *et al.*, “Exploring online reviews for user experience modeling,” in *DS 75-7: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 7: Human Behaviour in Design, Seoul, Korea, 19-22.08. 2013*, 2013, pp. 237–246.

[26] I. Kula, R. J. Branaghan, R. K. Atkinson, and R. D. Roscoe, "Assessing user experience via biometric sensor affect detection," in *Human Performance Technology: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2019, pp. 698–714.

[27] E. L.-C. Law and P. Van Schaik, "Modelling user experience–an agenda for research and practice," *Interacting with computers*, vol. 22, no. 5, pp. 313–322, 2010.

[28] V. Roto, E. Law, A. Vermeeren, and J. Hoonhout, "User experience white paper. bringing clarity to the concept of user experience. result from dagstuhl seminar on demarcating user experience, september 15-18 (2010)," *Disponible en ligne le*, vol. 22, pp. 06–15, 2011.

[29] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Symposium of the Austrian HCI and Usability Engineering Group*. Springer, 2008, pp. 63–76.

[30] N. Bolger, A. Davis, and E. Rafaeli, "Diary methods: Capturing life as it is lived," *Annual review of psychology*, vol. 54, no. 1, pp. 579–616, 2003.

[31] E. Karapanos, J. Zimmerman, J. Forlizzi, and J.-B. Martens, "User experience over time: an initial framework," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2009, pp. 729–738.

[32] D. Fallman and J. Waterworth, "Dealing with user experience and affective evaluation in hci design: A repertory grid approach," in *Workshop Paper, CHI*, 2005, pp. 2–7.

[33] C. N. Scollon, C.-K. Prieto, and E. Diener, "Experience sampling: promises and pitfalls, strength and weaknesses," in *Assessing well-being*. Springer, 2009, pp. 157–180.

[34] J. R. Edwards and R. P. Bagozzi, "On the nature and direction of relationships between constructs and measures." *Psychological methods*, vol. 5, no. 2, p. 155, 2000.

[35] D. E. Pournarakis, D. N. Sotiropoulos, and G. M. Giaglis, "A computational model for mining consumer perceptions in social media," *Decision Support Systems*, vol. 93, pp. 98–110, 2017.

[36] R. M. Groves, “Nonresponse rates and nonresponse bias in household surveys,” *Public opinion quarterly*, vol. 70, no. 5, pp. 646–675, 2006.

[37] A. Culotta and J. Cutler, “Mining brand perceptions from twitter social networks,” *Marketing science*, vol. 35, no. 3, pp. 343–362, 2016.

[38] A. Y. L. Chong, E. Chng, M. J. Liu, and B. Li, “Predicting consumer product demands via big data: the roles of online promotional marketing and online reviews,” *International Journal of Production Research*, vol. 55, no. 17, pp. 5142–5156, 2017.

[39] K. Hornbæk, “Current practice in measuring usability: Challenges to usability studies and research,” *International journal of human-computer studies*, vol. 64, no. 2, pp. 79–102, 2006.

[40] N. Bevan, “What is the difference between the purpose of usability and user experience evaluation methods,” in *Proceedings of the Workshop UXEM*, vol. 9, 2009, pp. 1–4.

[41] P. Ketola and V. Roto, “Exploring user experience measurement needs,” in *Proc. of the 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM). Reykjavik, Island*, 2008, pp. 23–26.

[42] E. Folmer, J. Van Gurp, and J. Bosch, “A framework for capturing the relationship between usability and software architecture,” *Software Process: Improvement and Practice*, vol. 8, no. 2, pp. 67–87, 2003.

[43] A. Seffah, M. Donyaee, R. B. Kline, and H. K. Padda, “Usability measurement and metrics: A consolidated model,” *Software quality journal*, vol. 14, no. 2, pp. 159–178, 2006.

[44] N. Bevan, “Classifying and selecting ux and usability measures,” in *International Workshop on Meaningful Measures: Valid Useful User Experience Measurement*, vol. 11, 2008, pp. 13–18.

[45] D. Andrzejewski and X. Zhu, “Latent dirichlet allocation with topic-in-set knowledge,” in *Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing*. Association for Computational Linguistics, 2009, pp. 43–48.

[46] D. Andrzejewski, X. Zhu, and M. Craven, “Incorporating domain knowledge into topic modeling via dirichlet forest priors,” in *Proceedings of the 26th annual international conference on machine learning*. ACM, 2009, pp. 25–32.

[47] C. Chemudugunta, A. Holloway, P. Smyth, and M. Steyvers, “Modeling documents by combining semantic concepts with unsupervised statistical learning,” in *International Semantic Web Conference*. Springer, 2008, pp. 229–244.

[48] D. Newman, E. V. Bonilla, and W. Buntine, “Improving topic coherence with regularized topic models,” in *Advances in neural information processing systems*, 2011, pp. 496–504.

[49] Z. Chen, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, “Leveraging multi-domain prior knowledge in topic models,” in *Twenty-Third International Joint Conference on Artificial Intelligence*, 2013.

[50] ——, “Discovering coherent topics using general knowledge,” in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. ACM, 2013, pp. 209–218.

[51] L. Yao, Y. Zhang, B. Wei, H. Qian, and Y. Wang, “Incorporating probabilistic knowledge into topic models,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2015, pp. 586–597.

[52] Y. Yang, D. Downey, and J. Boyd-Graber, “Efficient methods for incorporating knowledge into topic models,” in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 308–317.

[53] Z. Chen, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, “Exploiting domain knowledge in aspect extraction,” in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 1655–1667.

[54] A. Fang, C. Macdonald, I. Ounis, and P. Habel, “Using word embedding to evaluate the coherence of topics from twitter data,” in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 2016, pp. 1057–1060.

[55] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[56] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. Association for Computational Linguistics, 2002, pp. 79–86.

[57] R. Moraes, J. F. Valiati, and W. P. G. Neto, "Document-level sentiment classification: An empirical comparison between svm and ann," *Expert Systems with Applications*, vol. 40, no. 2, pp. 621–633, 2013.

[58] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2002, pp. 417–424.

[59] J. Khan, B. S. Jeong, Y.-K. Lee, and A. Alam, "Sentiment analysis at sentence level for heterogeneous datasets," in *Proceedings of the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory*. ACM, 2016, pp. 159–163.

[60] J. Zhu, H. Wang, M. Zhu, B. K. Tsou, and M. Ma, "Aspect-based opinion polling from customer reviews," *IEEE Transactions on Affective Computing*, vol. 2, no. 1, pp. 37–49, 2011.

[61] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[62] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proceedings of the 12th international conference on World Wide Web*. ACM, 2003, pp. 519–528.

[63] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," *Expert Systems with Applications*, vol. 57, pp. 117–126, 2016.

[64] M. R. Saleh, M. T. Martín-Valdivia, A. Montejo-Ráez, and L. Ureña-López, "Experiments with svm to classify opinions in different domains," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14 799–14 804, 2011.

[65] V. Ng, S. Dasgupta, and S. Arifin, "Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews," in *Proceedings of the COLING/ACL on Main conference poster sessions*. Association for Computational Linguistics, 2006, pp. 611–618.

[66] Z. Zhai, H. Xu, B. Kang, and P. Jia, "Exploiting effective features for chinese sentiment classification," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9139–9146, 2011.

[67] S. Tan and J. Zhang, "An empirical study of sentiment analysis for chinese documents," *Expert Systems with applications*, vol. 34, no. 4, pp. 2622–2629, 2008.

[68] A. Yousefpour, R. Ibrahim, and H. N. A. Hamed, "Ordinal-based and frequency-based integration of feature selection methods for sentiment analysis," *Expert Systems with Applications*, vol. 75, pp. 80–93, 2017.

[69] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, "Sentiment classification: The contribution of ensemble learning," *Decision support systems*, vol. 57, pp. 77–93, 2014.

[70] Y. Su, Y. Zhang, D. Ji, Y. Wang, and H. Wu, "Ensemble learning for sentiment classification," in *Workshop on Chinese Lexical Semantics*. Springer, 2012, pp. 84–93.

[71] N. Chawla, S. Eschrich, and L. O. Hall, "Creating ensembles of classifiers," in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*. IEEE, 2001, pp. 580–581.

[72] C. Catal and M. Nangir, "A sentiment classification model based on multiple classifiers," *Applied Soft Computing*, vol. 50, pp. 135–141, 2017.

[73] S. Li, C. Zong, and X. Wang, "Sentiment classification through combining classifiers with multiple feature sets," in *Natural Language Processing and Knowledge Engineering, 2007. NLP-KE 2007. International Conference on*. IEEE, 2007, pp. 135–140.

[74] K. Tsutsumi, K. Shimada, and T. Endo, "Movie review classification based on a multiple classifier," in *Proceedings of the 21st Pacific Asia Conference on Language, Information and Computation*, 2007, pp. 481–488.

[75] Y. Wan and Q. Gao, "An ensemble sentiment classification system of twitter data for airline services analysis," in *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1318–1325.

[76] R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," *Information Sciences*, vol. 181, no. 6, pp. 1138–1152, 2011.

[77] A. Onan and S. Korukoğlu, "A feature selection model based on genetic rank aggregation for text sentiment classification," *Journal of Information Science*, vol. 43, no. 1, pp. 25–38, 2017.

[78] E. Guldogan and M. Gabbouj, "Feature selection for content-based image retrieval," *Signal, Image and Video Processing*, vol. 2, no. 3, pp. 241–250, 2008.

[79] W. Bouaguel, G. B. Mufti, and M. Limam, "A fusion approach based on wrapper and filter feature selection methods using majority vote and feature weighting," in *Computer applications technology (ICCAT), 2013 International conference on*. IEEE, 2013, pp. 1–6.

[80] S. Dahiya, S. Handa, and N. Singh, "A rank aggregation algorithm for ensemble of multiple feature selection techniques in credit risk evaluation," *International Journal of Advanced Research in Artificial Intelligence*, vol. 5, pp. 1–8, 2016.

[81] B. Seijo-Pardo, V. Bolón-Canedo, and A. Alonso-Betanzos, "Using a feature selection ensemble on dna microarray datasets," in *Proceeding of 24th European Symposium on Artificial Neural Networks*, 2016, pp. 277–282.

[82] O. Osanaiye, H. Cai, K.-K. R. Choo, A. Dehghantanha, Z. Xu, and M. Dlodlo, "Ensemble-based multi-filter feature selection method for ddos detection in cloud computing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 130, 2016.

[83] R. Decker and M. Trusov, "Estimating aggregate consumer preferences from online product reviews," *International Journal of Research in Marketing*, vol. 27, no. 4, pp. 293–307, 2010.

[84] S. Xiao, C.-P. Wei, and M. Dong, "Crowd intelligence: Analyzing online product reviews for preference measurement," *Information & Management*, vol. 53, no. 2, pp. 169–182, 2016.

[85] J. Qi, Z. Zhang, S. Jeon, and Y. Zhou, "Mining customer requirements from online reviews: A product improvement perspective," *Information & Management*, vol. 53, no. 8, pp. 951–963, 2016.

[86] N. Hu, P. A. Pavlou, and J. J. Zhang, "Why do online product reviews have a j-shaped distribution? overcoming biases in online word-of-mouth communication," *Communications of the ACM*, vol. 52, no. 10, pp. 144–147, 2009.

[87] E. Sauerwein, F. Bailom, K. Matzler, and H. H. Hinterhuber, "The kano model: How to delight your customers," in *International Working Seminar on Production Economics*, vol. 1, no. 4. Innsbruck, 1996, pp. 313–327.

[88] Q. Su, K. Xiang, H. Wang, B. Sun, and S. Yu, "Using pointwise mutual information to identify implicit features in customer reviews," in *International Conference on Computer Processing of Oriental Languages*. Springer, 2006, pp. 22–30.

[89] A. Taylor, M. Marcus, and B. Santorini, "The penn treebank: an overview," in *Treebanks*. Springer, 2003, pp. 5–22.

[90] J. Hussain and S. Lee, "Review-based user experience (ux) modeling," , pp. 980–982, 2016.

[91] A. Abeillé, *Treebanks: Building and using parsed corpora*. Springer Science & Business Media, 2012, vol. 20.

[92] P. Willett, "The porter stemming algorithm: then and now," *Program*, vol. 40, no. 3, pp. 219–223, 2006.

[93] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization."

[94] H. Wang, T. M. Khoshgoftaar, and J. Van Hulse, "A comparative study of threshold-based feature selection techniques," in *Granular Computing (GrC), 2010 IEEE International Conference on*. IEEE, 2010, pp. 499–504.

[95] R. P. Priyadarsini, M. Valarmathi, and S. Sivakumari, "Gain ratio based feature selection method for privacy preservation," *ICTACT J Soft Comput*, vol. 1, no. 04, pp. 2229–6956, 2011.

[96] 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.

[97] J. Jagarlamudi, R. Udupa, and H. Daumé III, "Incorporating prior lexical information in topic models."

[98] W. L. Hamilton, K. Clark, J. Leskovec, and D. Jurafsky, "Inducing domain-specific sentiment lexicons from unlabeled corpora," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, vol. 2016. NIH Public Access, 2016, p. 595.

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

[100] T. Shi, K. Kang, J. Choo, and C. K. Reddy, "Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2018, pp. 1105–1114.

[101] F. Viegas, S. Canuto, C. Gomes, W. Luiz, T. Rosa, S. Ribas, L. Rocha, and M. A. Gonçalves, "Cluwords: Exploiting semantic word clustering representation for enhanced topic modeling," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, 2019, pp. 753–761.

[102] X. Rong, "word2vec parameter learning explained," *arXiv preprint arXiv:1411.2738*, 2014.

[103] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[104] M. Hassenzahl, M. Burmester, and F. Koller, "Attrakdiff: A questionnaire to measure perceived hedonic and pragmatic quality," in *Mensch & Computer*, 2003, pp. 187–196.

[105] T. Lavie and N. Tractinsky, "Assessing dimensions of perceived visual aesthetics of web sites," *International journal of human-computer studies*, vol. 60, no. 3, pp. 269–298, 2004.

[106] H. C. Kraemer, "Kappa coefficient," *Wiley StatsRef: Statistics Reference Online*, pp. 1–4, 2014.

[107] G. Ganu, N. Elhadad, and A. Marian, "Beyond the stars: improving rating predictions using review text content." in *WebDB*, vol. 9. Citeseer, 2009, pp. 1–6.

[108] S. Brody and N. Elhadad, "An unsupervised aspect-sentiment model for online reviews," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010, pp. 804–812.

[109] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a maxent-lda hybrid," in *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2010, pp. 56–65.

[110] Z. Chen and B. Liu, "Mining topics in documents: standing on the shoulders of big data," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 1116–1125.

[111] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of jaccard coefficient for keywords similarity," in *Proceedings of the international multiconference of engineers and computer scientists*, vol. 1, no. 6, 2013, pp. 380–384.

[112] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: understanding rating dimensions with review text," in *Proceedings of the 7th ACM conference on Recommender systems*. ACM, 2013, pp. 165–172.

Chapter 8

List of Publications

[1] Jamil Hussain, Fahad Ahmed Satti, Wajahat Ali Khan, Muhammad Afzal, Hafiz Syed Muhammad Bilal, Muhammad Zaki Ansaar, Hafiz Farooq Ahmad, Taeho Hur, Jaehun Bang, Jee-In Kim, Gwang Hoon Park, Hyonwoo Seung, and Sungyoung Lee, ”Exploring the dominant features of social media for depression detection”, Journal of Information Science (SCIE, IF: 1.939), 2019

[2] Jamil Hussain, Anees Ul Hassan, Hafiz Syed Muhammad Bilal,Muhammad Afzal,Shujaat Hussain, Jaehun Bang,Oresti Banos, and Sungyoung Lee, ”Model-based adaptive user interface based on context and user experience evaluation”, Journal on Multimodal User Interfaces (SCIE, IF: 1.140), Vol.12, Issue 1, pp.1-16, 2018

[3] Jamil Hussain, Fahad Ahmed Satti, Wajahat Ali Khan, Muhammad Afzal, Hafiz Syed Muhammad Bilal, Muhammad Zaki Ansaar, Hafiz Farooq Ahmad, Taeho Hur, Jaehun Bang, Jee-In Kim, Gwang Hoon Park, Hyonwoo Seung, and Sungyoung Lee, ”Exploring the dominant features of social media for depression detection”, Journal of Information Science, 2019

[4] Jamil Hussain, Wajahat Ali Khan, Taeho Hur, Hafiz Syed Muhammad Bilal, Jaehun Bang, Anees Ul Hassan, Muhammad Afzal and Sungyoung Lee, ”A Multimodal Deep Log-Based User Experience (UX) Platform for UX Evaluation”, Sensors, Vol.18, Issue 5, pp.1-31, 2018

[5] Jamil Hussain, and Sungyoung Lee, ”Mining user experience dimensions from mental illness apps”, 15th International Conference On Smart homes and health Telematics (ICOST 2017), Paris, France, Aug 29-31, 2017

[6] Jamil Hussain, and Sungyoung Lee, ”Model-Driven Adaptive User Interface based on Context and User Experience Evaluation”, 2016 International Symposium on Perception, Action

and Cognitive Systems (PACS), Seoul, Korea, Oct 27-28, 2016

[7] Jamil Hussain, Maqbool Ali, Hafiz Syed Muhammad Bilal, Muhammad Afzal, Hafiz Farooq Ahmad, Oresti Banos and Sungyoung Lee, ”SNS Based Predictive Model for Depression”, International Conference On Smart homes and Health Telematics (ICOST 2015), Geneva, Switzerland, June 10-12, (2015).

[8] Jamil Hussain, Wajahat Ali Khan, Muhammad Afzal, Maqbool Hussain, Byeong Ho Kang and Sungyoung Lee, ”Adaptive User Interface and User Experience based Authoring Tool for Recommendation Systems”, UCAmI 2014, Belfast, Ireland, Dec 2-5, 2014

[9] Taqdir Ali, Jamil Hussain, Muhammad Bilal Amin, Musarrat Hussain, Usman Akhtar, Wajahat Ali Khan, Ubaid Ur Rehman, Sungyoung Lee, Byeong Ho Kang, Maqbool Hussain, Muhammad Afzal, Ho-Seong Han, June Young Choi, Hyeong Won Yu and Arif Jamshed, ”Intelligent Medical Platform (IMP): A novel dialogue-based platform for healthcare services”, IEEE Magazine: Computer

[10] Shujaat Hussain, Maqbool Hussain, Muhammad Afzal, Jamil Hussain, Jaehun Bang, Hyonwoo Seung and Sungyoung Lee, ”Semantic Preservation of standardized healthcare documents in big data”, International Journal of Medical Informatics, 2019

[11] Maqbool Ali, Jamil Hussain, Sungyoung Lee, Byeong Ho Kang and Kashif Sattar, ””A knowledge construction methodology to automate case-based learning using clinical documents”, Expert Systems (SCIE, IF: 1.43), DOI: 10.1111/exsy.12401, 2019.

[12] Jawad KhanAftab AlamJamil HussainYoung-Koo Lee ”EnSWF: effective features extraction and selection in conjunction with ensemble learning methods for document sentiment classification.” Applied Intelligence (2019): 1-23.

[13] Taqdir Ali, Maqbool Hussain, Wajahat Ali Khan, Muhammad Afzal, Jamil Hussain, Rahman Ali, Waseem Hassan, Arif Jamshed, Byeong Ho Kang and Sungyoung Lee, ”Multi-model-based interactive authoring environment for creating shareable medical knowledge”, Computer methods and programs in biomedicine, (SCI, IF:2.503), 150 (2017): 41-72

[14] Rahman Ali, Jamil Hussain, Muhammad Hameed Siddiqi, Maqbool Hussain and Sungyoung Lee, ”H2RM: A Hybrid Rough Set Reasoning Model for Prediction and Management of Diabetes Mellitus”, Sensors (SCIE, IF: 2.245), Vol.15, No.7, pp.15921-15951, 2015

[15] Muhammad Afzal, Maqbool Hussain, Taqdir Ali, Jamil Hussain,Wajahat Ali Khan, Sungyoung Lee and Byeong Ho Kang, ”Knowledge-Based Query Construction Using the CDSS Knowledge Base for Efficient Evidence Retrieval”, Sensors (SCIE, IF:2.245), Vol.15, Issue 9, pp.21294-21314, 2015

[16] Muhammad Zaki Ansaar, Jamil Hussain, Asim Abass, Musarrat Hussain, and Sungyoung Lee ”Users Emotional eXperience Analysis of Wizard Form Pattern Using Objective and Subjective Measures”,Web Engineering, 19th International Conference, ICWE 2019, Daejeon, South Korea, June 1114, 2019.

[17] Musarrat Hussain, Jamil Hussain, Muhammad Sadiq, Anees Ul Hassan and Sungyoung Lee, ”Recommendation Statements Identification in Clinical Practice Guidelines Using Heuristic Patterns”, 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2018), Busan, Korea, June 26-29, 2018 Intl. Conference

[18] Usman Akhtar, Jamil Hussain and Sungyoung Lee, ”Medical Semantic Question Answering Framework on RDF Data Cubes”, 15th International Conference On Smart homes and health Telematics (ICOST 2017), Paris, France, Aug 29-31, 2017

[19] Anees Ul Hassan, Jamil Hussain, Musarrat Hussain, Muhammad Sadiq, Sungyoung Lee Sentiment Analysis of Social Networking Sites (SNS) Data using Machine Learning Approach for the Measurement of Depression 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, South Korea 2017

[20] Maqbool Ali, Jamil Hussain, Sungyoung Lee and Byeong Ho Kang, ”X-UDeKAM: An Intelligent Method for Acquiring Declarative Structured Knowledge using Chatterbot”, 2016 International Symposium on Perception, Action and Cognitive Systems (PACS), Seoul, Korea, Oct 27-28, 2016

[21] Wajahat Ali Khan, Muhammad Bilal Amin, Oresti Banos, Taqdir Ali, Maqbool Hussain, Muhammad Afzal, Shujaat Hussain, Jamil Hussain, Rahman Ali, Maqbool Ali, Dongwook Kang, Jaehun Bang, Tae Ho Hur, Bilal Ali, Muhammad Idris, Asif Razzaq, Sungyoung Lee and Byeong Ho Kang, ”Mining Minds: Journey of Evolutionary Platform for Ubiquitous Wellness”, 12th International Conference on Ubiquitous Healthcare (u-Healthcare 2015), Osaka, Japan, Nov 30- Dec 2, 2015

[22] Oresti Banos, Muhammad Bilal Amin, Wajahat Ali Khan, Muhammad Afzel, Mahmood Ahmad, Maqbool Ali, Taqdir Ali, Rahman Ali, Muhammad Bilal, Manhyung Han, Jamil Hussain, Maqbool Hussain, Shujaat Hussain, Tae Ho Hur, Jae Hun Bang, Thien Huynh-The, Muhammad Idris, Dong Wook Kang, Sang Beom Park, Hameed Siddiqui, Le-Ba Vui, Muhammad Fahim, Asad Masood Khattak, Byeong Ho Kang and Sungyoung Lee, ”An Innovative Platform for Person-Centric Health and Wellness Support”, 3rd International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2015), Granada, Spain, April 15-17, 2015

[23] Maqbool Ali, Haflz Syed Muhammad Bilal, Jamil Hussain, Sungyoung Lee and Byeong Ho Kang, ”An Interactive Case-Based Flip Learning Tool for Medical Education”, 13th Internation Conference On Smart homes and health Telematics: Inclusive Smart Cities and e-Health (ICOST 2015), Geneva, Switzerland, June 10-12, 2015