Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Semi-supervised method for sensitivity based documents' classification for online service providers

Sharaf J. Malebary *, Shakeel Ahmad

Faculty of Computing and Information Technology in Rabigh (FCITR), King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 4 November 2019
Received in revised form
5 February 2020
Accepted 7 February 2020

Keywords:
Service
Sentiment analysis
Supervised learning method
Unsupervised learning method
Latent semantic indexing

ABSTRACT

In today's digital era, many services providing companies exist on the web whereas service is the logical product of a company, which can be utilized through the Internet. Different service providers provide these services i.e., Online counselling service, online doctor consultation, cloud service provider, web hosting service, etc. to their customers. When customers face some problems, they may text to their providers. One solution is that providers can solve these issues based on the First-Come-First-Serve formula. But there should be an option to detect sensitive issue which may need to be solved first. How can this sensitivity be determined? Already there is a lot of researched work based on text to determine the polarity as positive and negative. Besides this classification, there are also some other classification methods investigated, such as aspect, not aspect, subjective, objective, spam, not spam, etc. regarding text sensitivity, whether it is sensitive or not? This classification is not yet considered for service providers. This paper presents a strategy for sensitivity based classification using Latent Semantic Indexing (LSI). The purpose of LSI is to rank documents concerning a given query. However, in this study, a mechanism was provided to generate query automatically based on sensitive general words with the words from all documents. This is a semi-supervised approach because 4782 sensitive words have been labeled from various sources and used based on an unsupervised approach to detect the sensitivity of the document. The sorted lists of documents based on the LSI scores generated by the sensitive-query were checked manually and were proved to be highly satisfactory. The topmost document in this list was the most sensitive, and the last document in the list was least sensitive.

1. Introduction

Nowadays, the Internet is widely and publicly used through smartphones. This has paved the way for many services to be carried out through the internet to reach out to customers. Some of these services are provided for entertainment or comfort purposes, while others can be lives saving. Some of the online applications that provide services to the public are online career counseling, online doctor consultation, internet service providers, FTP providers, cloud software providers (Asfoura et al., 2018). All these providers may have several customers, and each customer may have different problems daily. It is very important to handle critical problems first. How can critical problems be determined? The defined problem is written in text format, and already there is much analysis based on text. Here the author provides a method to determine the critical problem of all customers, which assigned an LSI score to each query. LSI (Latent Semantic Indexing) method is better for searching. It has been used for the clustering of documents and concept representations with key-word and key-sentences (Ahmad et al., 2017). LSI has also been used in determining the most positive and most negative review of a product based on an automatically generated query. Researchers have also done work on the detection of spam opinion (Li et al., 2011; Teli and Biradar, 2014), co-reference resolution (Ding and Liu, 2010), detection of sense of ambiguated word (Saqib et al., 2018) and aspects grouping (different words belonging to same aspects) (Saqib et al., 2019). Besides covering classifications of text from different dimensions for a customer who purchases a physical product, there is

* Corresponding Author.
Email Address: smalebary@kau.edu.sa (S. J. Malebary)
https://doi.org/10.21833/ijaas.2020.05.004
Corresponding author's ORCID profile:
https://orcid.org/0000-0003-4339-3791

still a need for classification of text based on sensitivity for those customers who purchase a service from a provider.

Some service provider provides a method to the customer to send issue with predefined lists of priority options as high, medium, low. The selection of these options may be ironic. Instead of selecting this option manually, there should be a method which can select one of this option based on the sensitivity of written text automatically. Considering such type of application, the proposed method has been investigated. This study made the following key contributions:

- A method is proposed for generating an Automatic Query (AQ) using sensitive and critical words, which is necessary for the Latent Semantic Indexing (LSI) technique, i.e., there is no need to provide the queries (priority label) as input.
- Generating a ranked list of documents from highest to lowest sensitivity based on the LSI scores, the highest scored document will have the highest priority, and the lowest scored document will have the lowest priority.

2. Literature review

How can a company improve the quality of a product, place, etc. from the huge amount of reviews? Many studies have been carried out with regard to sentiment analysis, which is about determining the sentiment orientation of a review or comment (Chen et al., 2017). Sentiment orientation means that a positive opinion will be an exact positive, and a negative opinion will be an exact negative (Liu, 2012). The view, assessment, or feeling of a person towards a product (Jin et al., 2016), aspect (Shu et al., 2017), or service is known as a sentiment (Khan et al., 2009; Asghar et al., 2014). Such a feeling, which is either positive or negative, can be assigned a score. Most of the work in sentiment analysis is based on binary classification, which means that reviews or blogs are divided into “positive” and “negative” classes (Wang et al., 2009). The classification of text sentiments can be done in two ways, i.e. through machine learning and score-based approaches (Wang et al., 2011; Chen et al., 2011). Machine learning uses training data (Hameed et al., 2018), while the other method uses several attributes of an entity to determine the scores. In the score-based approach, opinions can be oriented as positive or negative (Kundi et al., 2014b; Saqib and Kundi, 2016). Kundi et al. (2014a) used a combined approach of SentiWordNet and lexical resources to determine the scores for slangs. A lexicon-based approach for extracting sentiment orientations of opinions has been used for scoring. Gupta and Ekbal (2014) used lists of positive and negative words to determine the polarity of a sentence by creating a training matrix and random forest classifier based on supervised learning. A sentiment analysis can be performed using different methods (Rosenthal et al., 2017), with each method having an improved accuracy with respect to the previous one. Although a lot of work is involved in sentiment orientation with the use of adjectives, frequent nouns and noun phrases, sentiment shifters, handling of ‘but’ clauses, decreased and increased quantity of an opinionated item; high, low, increased and decreased quantity of a positive or negative potential item; desirable or undesirable facts; deviations from the norm or a desired value range; and the production and consumption of resources and waste, etc., these are very important for determining the polarity of a document or sentence (Htay and Lynn, 2013). However, a large amount of online data is generated every day with unprecedented speed and size. Most of the available information on the Internet is in text and unstructured forms, i.e. online reviews, blogs, chats, and news. An aspect-based sentiment analysis, which can be carried out by using only particular aspects (Gojali and Khodra, 2016), requires less effort compared to a sentiment analysis of an object with respect to all aspects. Reviews are rated according to an object, so there should be a direct method to determine whether a review is positive or negative. LSI (Latent Semantic Indexing) is better for such a purpose (Ahmad et al., 2017; Saqib et al., 2016). LSI (Huang et al., 2009) has been used for the clustering of documents and for concept representations. An extended method based on LSI is able to filter unwanted emails in Chinese and English (Yang and Li, 2005). A hybrid approach for sentiment analysis of Arabic tweets based on two stages. Firstly, the pre-processing methods like stop-word removal, tokenization and stemming are applied, and then two features weighting algorithms (information gain and chai square) are utilized to assign high weights to the most significant features of the Arabic tweets. Secondly, the deep learning technique is employed to effectively and accurately classify the Arabic tweets either as positive or negative tweets (Altaher, 2017). To improve accuracy of sentiment analysis, lot of work has also been done on words sense disambiguation (Rios et al., 2017; Swathy, 2017). Machine learning approaches, also called corpus-based approaches, do not make use of any knowledge resources for disambiguation (Raganato et al., 2017). Most accurate WSD systems to date exploit supervised methods which automatically learn cues useful for disambiguation from manually sense-annotated data (Wang et al., 2017). All above analyses are very useful for a company from where a user can purchase a physical product as well as on line service provider. Online service provider also has different customers to handle their issues and problems. There should be a method which can detect the most critical issues, so they can be handled first.

3. Applications of proposed work

This methodology is suitable for all those online applications which provide services to many

customers and deal with issues daily. Few of them are described as following.

3.1. Online counseling service

Online therapy, also known as e-therapy, e-counseling, teletherapy, or cyber-counseling, is a relatively new development in mental health in which a therapist or counselor provides psychological advice and support over the internet. This can occur through email, chat, messaging, or internet phone (Mallen and Vogel, 2005).

3.2. Online doctor consultation

In 2000, many people came to treat the internet as a first, or at least a major source of information and communication. Health advice is now the most popular topic. In developed countries, many online doctors prescribe so-called 'lifestyle drugs, such as for weight loss, hair loss, or erectile dysfunction (Glover-Thomas and Fanning, 2010).

3.3. Cloud service provider

A cloud service is any service made available to users on-demand via the Internet from a cloud computing provider's servers as opposed to being provided from a company's on-premises servers. Cloud services are designed to provide easy, scalable access to applications, resources, and services, and are fully managed by a cloud services provider (Saqib et al., 2011).

3.4. Web hosting service

Hosting (also known as website hosting or Web hosting) is the business of housing, serving, and maintaining files for one or more websites. In a sense, you rent space on a computer to hold your website made by a web-designer (Bazsova, 2019).

4. Classification of document based on the sensitivity

Latent Semantic Indexing method will determine the closest text with Automated Query (AQ). AQ contains a list of critical words from all texts. In this method, the 1st step is to find AQ, and then the second is to find the score.

4.1. Automatically generated query (AQ)

If we generally think or experienced from daily basis issues, it may clear that sensitive issues contain those words, which are mostly negative. So, here I collected a list of 4782 negative words named SenWord (Liu et al., 2005). The flow chart for generating Automatic Query is given in Fig. 1.

In Fig. 1, C is the chunks, i.e., all words of given text document D, FC are the filtered chunks means filtered words from stop-words. SFC will contain common words from FC and SenWord (sensitive words). These words will be updated with AQ. In the end, AQ has contained all sensitive words from whole documents. Hence automatic query has been generated as AQ. Eq. 1 shows all the documents, and Eq. 2 depicts the tokens. Eq. 3 is used for filtering the tokens, i.e., removing all the stop words.

$$D = \cup_{x=1}^{n} D_x \quad (1)$$

$$T(x) = \cup_{i=1}^{n} D_i \quad (2)$$

$$FT(x) = \cup_{i=1}^{n} \{T_i , \quad if\ T_i \notin SW \quad (3)$$

where x=1, 2, 3,..., n; SW means stop words; D represents the total number of Documents; T(x) represents the tokens of the x[th] review, and FT(x) represents the filtered tokens of the x[th] review.

Now Automatic Query (AQ) can generate FT(x) and list of 4782 sensitive words SenWord as calculated in Eq. 4:

$$A_Q = \cup_{x=1}^{n} \{FT(x)_i , \quad \cup_{i=1}^{n} \quad if\ FT(x)_i \in \text{SenWord} \quad (4)$$

where x=1, 2, 3,..., n and $FT(x)_i$ means the i[th] chunk of the x[th] document. A_Q contains those words from all the documents that belong to the SenWord.

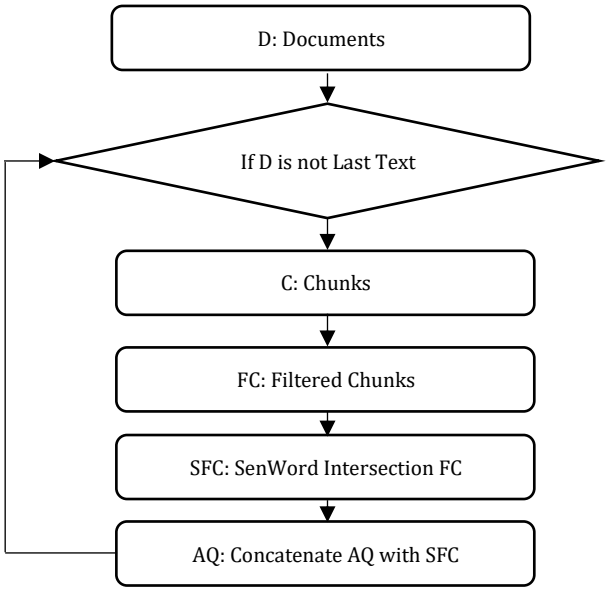

Fig. 1: Flowchart for AQ

4.2. Scoring each document with LSI

LSI is an efficient information retrieval algorithm (Phadnis and Gadge, 2014). Basically, in LSI, there is a cosine similarity measure between the coordinates of a document vector and the coordinates of a query vector. If this value is 1, it means the document is 100% closer to the query if it is 0.5, it means the document is 50% closer to the query, and if it is 0.9, it means the document is 90% closer to the query and so forth. The significant point now is finding the coordinates of each document and query. A Singular Value Decomposition (SVD) can determine the points or coordinates of a document and query. Through the SVD, three matrices, S, V, and U, which will be used for further processing, can be

determined by a matrix. To determine the values of such variables, the SVD requires a matrix. The matrix consists of rows and columns containing integers, but the inputs under consideration are the different text documents. A feature matrix can be obtained by calculating the frequencies of each word. This means that first, a feature matrix is created from all the documents, and then, the SVD is calculated. After this, the supporting variables, S, V, and U will be calculated by using NumPy (Numeric Python). The coordinates of all the documents will be determined from S, and these coordinates will be merged with the query to obtain the query coordinates. Finally, a cosine similarity function will be applied to these coordinates to find the documents that are closest to the query (Saqib et al., 2016). Here, it is clear that a document with the highest score is very close to sensitive words means that this document is most sensitive and vice versa. Fig. 2 shows the flow chart for the score of each document with AQ.

In Fig. 2, LSI (AQ, D) will calculate the score of each document with AQ as defined in the following equation Eq-5.

$$LSI(Score)_x = \cup_{x=1}^{n}(LSI_x\ (FT(x), A_Q)) \quad (5)$$

From Eq. 5, $FT(x)$ are the filtered tokens of each x^{th} document. Then, the LSI score $LSI(Score)_x$ of each document based on LSI can be found through the automated query, A_Q.

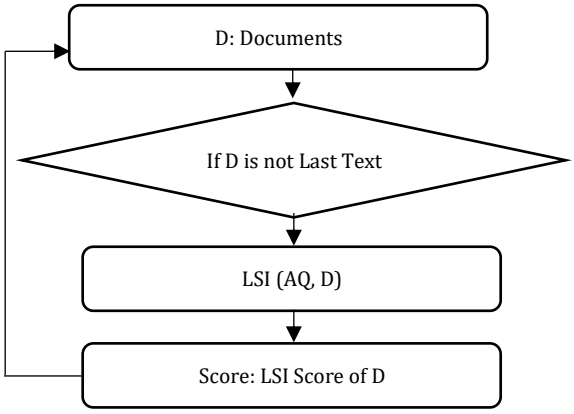

Fig. 2: Flowchart for score of each document with AQ

5. Result and discussion

Dataset is a list of 1,000 hotels and their reviews provided by Datafiniti's Business Database. The dataset includes hotel location, name, rating, review data, title, username, and more (data.world, 2018). This study selected 225 sensitive reviews about a hotel. These reviews were labeled from S [1] to S [225], and different attempts were made to determine the maturity level of this work. A sample listing of the said datasets is presented in Table 1, where S [1], S [2], S [3] were three sensitive reviews from S [1] to S [225].

Table 1: Sample sensitive document from prepared datasets

Labels	Actual Text
S [1]	"I chose the Cellini based on all the wonderful reviews I had read on Tripadvisor. I was extremely disappointed and agreed with everything said in the honest negative reviews of the hotel. This place is nothing special and needs to do a lot more to regain its #1 rating on Tripadvisor! I stayed in a Junior Suite with my family. The room is not a Suite; it is a room with a king-size bed and two single beds! It is a family room, and nothing more! It was not filthy dirty, but it could have been cleaner, particularly in the bathroom. The hotel staff did no more than they should have done for us. The breakfast was poor, and the cold meat and cheese looked as if it had been there for months. I would not stay there again. I really cannot understand all the positive reviews for this hotel on Tripadvisor. This hotel is far too expensive for what you get! I have traveled all over Italy and have stayed in Rome many times, try the Hotel Sant Anna in Borgo Pio, near the Vatican same price but 100% better!"
S [2]	"This place is too noisy. You can hear the person using the bathroom in the next room and hear every little conversation!! The doors do not shut all the way. You have to give your key when you go and come so anyone can get into your room. You can tell them a room number, and they just give you a key. I spoke with the manager about a problem that i encountered with the receptionist. When all is said and done, they did nothing and said it was a misunderstanding!!! We talked with another couple staying here, and they said the same thing, very noisy!! Avoid this place if you can. Not worth the money!!"
S [3]	"Hotel is kind of a misnomer. The reason there isn't a picture is that the "hotel" takes two buzzers to get in, and you are inside a large, nondescript building. If you weren't looking for it, you'd never find it. I found the price, relative to the reviews, exorbitant. I expected more of a hotel. As an example, every time you open your door, the owner, kind of peeks her head out to see what it is you want, nice, but annoying in the sense of, if I wanted to go to get a toothbrush or get some air outside, you have to go through the buzzer routine to get in and out of the hotel. There are a great many hotels, closer to the Termini Station that are traditional hotels and charge far less money. The hotel would be worth 50-60 euro's, not the ridiculous 120 that we paid."

First, all the known sensitive documents and automated query AQ were passed through the proposed algorithm to find the score of each document. Following is the LSI score of the first ten and last ten documents in Table 2 and Table 3, respectively.

From S [0] to S [224], S [202] had the highest score of 0.99997009638179, and S [40] had the lowest score of -0.304431581352365. It means S [202] is most sensitive, and S [40] is least sensitive from 225 texts. The most sensitive and least sensitive texts are shown in Table 4.

The first ten and last ten sensitive documents are shown in Table 5 and Table 6, respectively.

Table 2: LSI score of first 10 documents

Text-Tag	LSI-Score	Name of File
S[0]	0.66211	0001.neg
S[1]	0.308886	0002.neg
S[2]	0.347183	0003.neg
S[3]	-0.02453	0004.neg
S[4]	-0.02971	0005.neg
S[5]	0.509722	0006.neg
S[6]	0.033552	0007.neg
S[7]	0.12608	0008.neg
S[8]	0.414583	0009.neg
S[9]	0.934279	0010.neg

The sorted lists of sensitive documents based on the LSI scores generated by the automated query

(AQ) were checked manually and were proved to be highly satisfactory.

5.1. Statistical results

A confusion matrix is formed from the four outcomes produced as a result of binary classification. A binary classifier predicts all data instances of a test dataset as either positive or negative. This classification (or prediction) produces four outcomes–true sensitive (TS), true not-sensitive (TNS), false sensitive (FS), and false not-sensitive (FNS). Here we used 225 negative hotel reviews as Sensitive and 225 positive hotel reviews as Not Sensitive, i.e., means 50% sensitive and 50% Not Sensitive. After experimental results, we found, at LSI-score greater than 0.7, recall with respect to Not Sensitive is 40%, and with respect to Sensitive is 60%, only 10% Not sensitive considered as Sensitive. Obtained results based on actual Sensitive and Not Sensitive and Predicted Sensitive and Not Sensitive. Some of the samples are given in Table 7.

Table 3: LSI score of last 10 documents

Text-Tag	LSI-Score	Name of File
S[215]	0.680958	1491.neg
S[216]	0.996024	1492.neg
S[217]	0.884368	1493.neg
S[218]	0.955146	1494.neg
S[219]	0.971581	1495.neg
S[220]	0.779998	1496.neg
S[221]	0.766896	1497.neg
S[222]	0.836713	1498.neg
S[223]	0.972741	1499.neg
S[224]	0.996066	1500.neg

Table 4: Most and least sensitive texts

Text-Tag	LSI-Score	Original Text
S [202]	0.99997	"Overpriced margaritas, overpriced mediocre food. I ordered the spicy shrimp quesadilla, which was supposed to be loaded with shrimp, and maybe there were 5 tiny shrimp in each cut section. I believe the cost of that was $17.00, which is ridiculous. There was nothing in it but cheese and a few shrimp. Fantastic location is all they have going for them"
S [40]	-0.30443	"The most horrible hotel experience of my life during our honeymoon we walk into the room and the floor is filthy, there are stains on the wall, there are mold and paint peeling on the ceiling of the shower, the bed is broken, the tv remote doesn't work, the showerhead is broken and sprays water into light fixtures health hazard, safety hazard, tiny dirty room"

Table 5: First 10 sensitive documents

Text-Tag	LSI-Score	Name of File
S[202]	0.99997	1478.neg
S[144]	0.999929	1420.neg
S[182]	0.999518	1458.neg
S[147]	0.99947	1423.neg
S[142]	0.999087	1418.neg
S[159]	0.998964	1435.neg
S[121]	0.998909	1397.neg
S[194]	0.998875	1470.neg
S[102]	0.998863	1378.neg
S[120]	0.998598	1396.neg

Table 6: Last 10 sensitive documents

Text-Tag	LSI-Score	Name of File
S[60]	-0.13434	0061.neg
S[35]	-0.14494	0036.neg
S[58]	-0.16061	0059.neg
S[32]	-0.1712	0033.neg
S[75]	-0.17132	0076.neg
S[64]	-0.20583	0065.neg
S[27]	-0.22557	0028.neg
S[71]	-0.26339	0072.neg
S[72]	-0.26841	0073.neg
S[40]	-0.30443	0041.neg

Table 7: Last 10 sensitive documents

S.No	LSI Score	Name on Drive	Actual Classification	Predicted Classification	Results
1	0.787896	0001.pos	Not Sensitive	Not Sensitive	True Not Sensitive
2	0.999961	0011.neg	Sensitive	Sensitive	True Sensitive
3	0.998876	0041.pos	Not Sensitive	Sensitive	False Sensitive
4	0.141457	0043.neg	Sensitive	Not Sensitive	False Not Sensitive
5	0.944453	1387.pos	Not Sensitive	Not Sensitive	True Not Sensitive
6	0.997192	1392.neg	Sensitive	Sensitive	True Sensitive
7	0.999937	1389.neg	Sensitive	Sensitive	True Sensitive
8	0.931086	1389.pos	Not Sensitive	Not Sensitive	True Not Sensitive
9	0.919953	1390.neg	Sensitive	Not Sensitive	False Not Sensitive
10	0.999858	1391.neg	Sensitive	Sensitive	True Sensitive

6. Conclusion and future work

The major purpose of the proposed work is to separate sensitive and not sensitive text. Also, the major contribution is to find out the most sensitive text to least sensitive text. This can be very beneficial for those environments where online suggestions or solutions can be provided based on the critical condition of the customer. After experimental results using all reviews from hotel-dataset and an automatic query to LSI as input, it is observed that detection of sensitive document or reviews are very satisfactory. We selected 500 hotel reviews (225 positive, i.e., Not Sensitive and 225 negative reviews, i.e., Sensitive) and 4782 sensitive words. Further range of dataset and size of sensitive words could also be analyzed. These words have been selected from different sources according to their negative or critical meanings. Generating lexicon of sensitive words can also be considered as future work by using synonyms of a critical or sensitive word and their path distance specified in SentiWordNet.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

Ahmad S, Saqib SM, Almagrabi AO, and Alotaibi FM (2017). LSI based search technique: Using extracted keywords and key-

sentences. VAWKUM Transactions on Computer Sciences, 14(2): 1-8.
https://doi.org/10.21015/vtcs.v14i2.471

Altaher A (2017). Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting. International Journal of Advanced and Applied Sciences, 4(8): 43-49.
https://doi.org/10.21833/ijaas.2017.08.007

Asfoura E, Abdel-Haq MS, Chatti H, and Radouche T (2018). Classification of business models with focusing on characterizing "as a service" offers. International Journal of Advanced and Applied Sciences, 5(11): 16-23.
https://doi.org/10.21833/ijaas.2018.11.002

Asghar MZ, Khan A, Ahmad S, and Kundi FM (2014). A review of feature extraction in sentiment analysis. Journal of Basic and Applied Scientific Research, 4(3): 181-186.

Bazsova B (2019). How can the company choose the best web designer? Decision-making application within a company. International Journal of Advanced and Applied Sciences, 6(2): 6–11.
https://doi.org/10.21833/ijaas.2019.02.002

Chen LS, Liu CH, and Chiu HJ (2011). A neural network based approach for sentiment classification in the blogosphere. Journal of Informetrics, 5(2): 313-322.
https://doi.org/10.1016/j.joi.2011.01.003

Chen T, Xu R, He Y, and Wang X (2017). Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Systems with Applications, 72: 221-230.
https://doi.org/10.1016/j.eswa.2016.10.065

data.world (2018). Hotel-reviews. Available online at:
https://bit.ly/38AeLAW

Ding X and Liu B (2010). Resolving object and attribute coreference in opinion mining. In the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, Beijing, China: 268-276.

Glover-Thomas N and Fanning J (2010). Medicalisation: The role of e-pharmacies in iatrogenic harm. Medical Law Review, 18(1): 28-55.
https://doi.org/10.1093/medlaw/fwp026 **PMid:20133321**

Gojali S and Khodra ML (2016). Aspect based sentiment analysis for review rating prediction. In the International Conference on Advanced Informatics: Concepts, Theory and Application, IEEE, George Town, Malaysia.
https://doi.org/10.1109/ICAICTA.2016.7803110

Gupta DK and Ekbal A (2014). IITP: Supervised machine learning for aspect based sentiment analysis. In the 8th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Dublin, Ireland: 319-323.
https://doi.org/10.3115/v1/S14-2053

Hameed M, Tahir F, and Shahzad MA (2018). Empirical comparison of sentiment analysis techniques for social media. International Journal of Advanced and Applied Sciences, 5(4): 115-123.
https://doi.org/10.21833/ijaas.2018.04.015

Htay SS and Lynn KT (2013). Extracting product features and opinion words using pattern knowledge in customer reviews. The Scientific World Journal, 2013: 394758.
https://doi.org/10.1155/2013/394758
PMid:24459430 PMCid:PMC3888732

Huang A, Milne D, Frank E, and Witten IH (2009). Clustering documents using a wikipedia-based concept representation. In the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, Bangkok, Thailand: 628-636.
https://doi.org/10.1007/978-3-642-01307-2_62

Jin J, Ji P, and Gu R (2016). Identifying comparative customer requirements from product online reviews for competitor analysis. Engineering Applications of Artificial Intelligence, 49: 61-73.
https://doi.org/10.1016/j.engappai.2015.12.005

Khan K, Baharudin BB, and Khan A (2009). Mining opinion from text documents: A survey. In 3rd IEEE International Conference on Digital Ecosystems and Technologies, IEEE, Istanbul, Turkey: 217-222.
https://doi.org/10.1109/DEST.2009.5276756
PMCid:PMC2694658

Kundi FM, Ahmad S, Khan A, and Asghar MZ (2014a). Detection and scoring of internet slangs for sentiment analysis using SentiWordNet. Life Science Journal, 11(9): 66-72.

Kundi FM, Khan A, Ahmad S, and Asghar MZ (2014b). Lexicon-based sentiment analysis in the social web. Journal of Basic and Applied Scientific Research, 4(6): 238-48.

Li FH, Huang M, Yang Y, and Zhu X (2011). Learning to identify review spam. In the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Spain: 2488-2493.

Liu B (2012). Sentiment analysis and opinion mining: Synthesis lectures on human language technologies. Morgan and Claypool Publishers, San Rafael, USA.
https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Liu B, Hu M, and Cheng J (2005). Opinion observer: Analyzing and comparing opinions on the web. In the 14th International Conference on World Wide Web, Association for Computing Machinery, Chiba, Japan: 342–351.
https://doi.org/10.1145/1060745.1060797

Mallen MJ and Vogel DL (2005). Introduction to the major contribution: Counseling psychology and online counseling. The Counseling Psychologist, 33(6): 761-775.
https://doi.org/10.1177/0011000005278623

Phadnis N and Gadge J (2014). Framework for document retrieval using latent semantic indexing. International Journal of Computer Applications, 94(14): 37-41.
https://doi.org/10.5120/16414-6065

Raganato A, Camacho-Collados J, and Navigli R (2017). Word sense disambiguation: A unified evaluation framework and empirical comparison. In the 15th Conference of the European Chapter of the Association for Computational Linguistics, 1: 99-110.
https://doi.org/10.18653/v1/E17-1010

Rios A, Mascarell L, and Sennrich R (2017). Improving word sense disambiguation in neural machine translation with sense embeddings. In the 2nd Conference on Machine Translation, Association for Computational Linguistics, Copenhagen, Denmark: 11-19.

Rosenthal S, Farra N, and Nakov P (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. In the 11th International Workshop on Semantic Evaluations, Vancouver, Canada: 502–518.
https://doi.org/10.18653/v1/S17-2088

Saqib SM and Kundi FM (2016). MMO: Multiply-Minus-One rule for detecting and ranking positive and negative opinion. International Journal of Advanced Computer Science and Applications, 7(5): 122-127.
https://doi.org/10.14569/IJACSA.2016.070519

Saqib SM, Ahmad S, Syed AH, Naeem T, and Alotaibi FM (2019). Grouping of aspects into relevant category based on wordnet definitions. International Journal of Computer Science and Network Security, 19(2): 113–119.

Saqib SM, Jan MA, Ahmad B, Ahmad S, and Asghar MZ (2011). Custom software under the shade of cloud computing. International Journal of Computer Science and Information Security, 9(5): 219-223.

Saqib SM, Kundi FM, Syed AH, and Ahmad S (2018). Semi supervised method for detection of ambiguous word and creation of sense: Using WordNet. International Journal of Advanced Computer Science and Applications, 9(11): 353-359.
https://doi.org/10.14569/IJACSA.2018.091149

Saqib SM, Mahmood K, and Naeem T (2016). Comparison of LSI algorithms without and with pre-processing: Using text

document based search. Transactions on Information Security, 1(4): 44-51.

Shu L, Xu H, and Liu B (2017). Lifelong learning CRF for supervised aspect extraction. In the 55th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Vancouver, Canada, 2: 148-154. https://doi.org/10.18653/v1/P17-2023
PMCid:PMC5576273

Swathy R (2017). A survey on word sense disambiguation used in NLP. International Journal of Innovative Research in Computer and Communication Engineering, 5(3): 5116–5117.

Teli S and Biradar S (2014). Effective spam detection method for email. In the International Conference on Advances in Engineering and Technology, Singapore, Singapore: 68-72.

Wang S, Li D, Song X, Wei Y, and Li H (2011). A feature selection method based on improved fisher's discriminant ratio for text sentiment classification. Expert Systems with Applications, 38(7): 8696-8702. https://doi.org/10.1016/j.eswa.2011.01.077

Wang S, Li D, Wei Y, and Li H (2009). A feature selection method based on fisher's discriminant ratio for text sentiment classification. In the International Conference on Web Information Systems and Mining, Springer, Shanghai, China: 88-97. https://doi.org/10.1007/978-3-642-05250-7_10

Wang T, Li W, Liu F, and Hua J (2017). Sprinkled semantic diffusion kernel for word sense disambiguation. Engineering Applications of Artificial Intelligence, 64: 43-51. https://doi.org/10.1016/j.engappai.2017.05.010

Yang Q and Li FM (2005). Support vector machine for customized email filtering based on improving latent semantic indexing. In the International Conference on Machine Learning and Cybernetics, IEEE, Guangzhou, China, 6: 3787-3791. https://doi.org/10.1109/ICMLC.2005.1527599