International Journal of Engineering & Scientific Research
Vol.6 Issue 9, September2018,
ISSN: 2347-6532 Impact Factor: 6.660
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

AN APPROACH TO OPINION EXTRACTION USING NATURAL LANGUAGE PROCESSING

Ram Chandra Pal
Research Scholar
Dr A. P. J. Abdul Kalam University

Dr. Sunita Gond
Ass. Professor
Dr A. P. J. Abdul Kalam University

ABSTRACT

Electronic commerce, known as e-commerce, is a system that consists in buying and selling products/services over the internet. The internet is used by millions of people, making the management of the available information (e.g., competitor analysis market) a very difficult task for those operating an e-commerce business. So that the managers can better position their companies against competitors, comes the need to create automatic mechanisms to extract information from various web sources (websites). The hotel business is a market where e-commerce is essential since the internet is their biggest selling point, either through sales channels or through their own websites.

It has been said that there is only one opportunity to make a first impression. In the competitive world of ecommerce, attracting customers to a Web site is expensive; keeping them is a business imperative.

Research methods in information extraction, automatic categorization and clustering, automatic summarization and indexing, and statistical machine translation need to be adapted to a new kind of data. This book reviews the current research on NLP tools and methods for processing the non-traditional information from social media data that is available in large amounts (big data), and shows how innovative NLP approaches can integrate appropriate linguistic information in various fields such as social media monitoring, healthcare, business intelligence, industry, marketing, and security and defense.

INTRODUCTION

NLP (Natural Language Processing) is one of the most prolific areas of AI application. This is the branch of Artificial Intelligence which consists of understanding and processing human language. NLP, or Natural Language Processing**,** refers to all the tasks that allow a computer to process the data in human language. It is therefore a computer discipline in its own right, covering many subjects and methods, which are at the origin of search engines in particular. Some authors distinguish between so-called "low-level" tasks allowing a simple representation of the text by a computer, and "high-level" tasks allowing the machine to "understand" the text.

Online Shopping has become significantly important for people nowadays as they can save time and effort to buy a product. E-commerce has grown greatly and customer feedback has become crucial to determine their interest and activities. Sentiment analysis is used to determine to know what consumers think about the product. This helps other customers in making buying decisions about the product. A recommender system built on this can provide suggestions to other customers or show them related products while shopping. In recent years, sentiment analysis has drawn great interest, as it does the text classification based on consumer reviews. The reviews are given in form of textual

Nowadays a lot of information are in the text format (books, documents, articles, social media posts, messages, reviews, chat's conversation, description, website info etc.). Those files contain a lot of valuable information that can support business activities. Insights from text data could be extracted using NLP applications. NLP applications in enterprises range from document management support to customer support via chatbots or dialog bots by creating automated answers to the most common questions.

The advent of social media has changed traditional business methods drastically. Product development, marketing and support are increasingly conducted via social sites. Social Media allows individuals around the world to freely express their opinions without disclosing too much identity. This allows customers to provide genuine opinions about products and services in social media sites. Many business organizations interact with their customers and colleagues through social media.

As the businesses increasingly use social media, the need for better understanding of the trends and communication arises. Social media monitoring is a growing technology area in which sophisticated tools are used to monitor social interactions towards certain text terms. Applications of this technology are huge. Stock market analysis, Political campaigns, Crime monitoring, and Product research are a few of them. Competitive Intelligence is another application of social media monitoring that is receiving a lot of momentum recently. Competitive Intelligence is systematic process of gathering information about an organization's external environment.

Businesses use massive quantities of unstructured, text-heavy data and need a way to efficiently process it. A lot of the information created online and stored in databases is natural human language, and until recently, businesses could not effectively analyze this data. This is where natural language processing is useful.

Data analysis is especially relevant in the e-commerce and retail industry. They can predict the purchases, profits, losses and even manipulate customers into buying things by tracking their behavior. Retail brands analyse data to create customer profiles and learn his/her sore points and market their product accordingly to push the customer towards purchasing.

If you are running an ecommerce business, and you're serious about providing a flawless on-site search experience, then natural language processing (NLP) is definitely something that you should be considering. No longer reserved for the search engine giants, NLP is making dramatic inroads into ecommerce site search, and bringing with it significant benefits for switched-on merchants.

REVIEW OF LITERATURE

A Model to Enhance Governance Issues through Opinion Extraction

Kamran Shaukat; Talha Mahboob Alam; Muhammad Ahmed; Suhuai Luo; Ibrahim A. Hameed; Muhammad Shahid Iqbal

We live in a world where data is expanding exponentially. Most of the data is unstructured when obtained through the web. Many organizations, institutes, and governments worldwide gather public views regarding their products, services, or policies. With thousands of reviews about some product, service, or policy, it is impossible to conclude some kind of final thought from it. To handle this, there is a desperate need for a model that can extract meaningful information from data to make correct and timely decisions for the efficient growth of business and smooth running of an organization or government. Otherwise, the practice of collecting and storing data will be ineffective. In this study, we focused on conducting an extensive public survey on issues of Southern Punjab, carry out appropriate processing on collected data and predict trends in public opinion for decision-making. Natural Language Processing (NLP) and Machine Learning (ML) have dealt with this problem. Different data preprocessing techniques have been utilized to remove the noise from data. Our experiments stated that unemployment, poverty, education, and corruption are the major issues of the targeted region. This study will help government officials and non-governmental organizations to be focused on the extracted issues in the specific region.

Sentiment Analysis: an approach in Natural Language Processing for Data Extraction

Shabina Dhuria

Sentiment analysis and opinion mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. It is one of the most active research areas in natural language processing and is also widely studied in data mining, web mining, and text mining. Sentiment analysis has been used in several applications including analysis of the repercussions of events in social networks, analysis of opinions about products and services. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. Methods like supervised machine learning and lexical-based approaches are available for measuring sentiments that have a huge volume of opinionated data recorded in digital form for analysis

FST-Based Natural Language Processing Method for Opinion Extraction

Delin Liu1, a, Haopeng Chen2, b 1 Shanghai Jiao Tong University 2 Shanghai Jiao Tong University

This paper proposes rule-based and Finite State Transducers (FST) based NLP method for extracting information from massive text. The method differs from n-gram based popular method which relies on probability statistics and machine learning. In our method, the rules are grammars of a language, summarized by people. FST is the implementation tool of rules. It can process natural language and generate a syntax tree for each sentence. To support applying the rules, we tokenize and generate the stem of words, and find many word features which are recorded in a dictionary. After generating a syntax tree, we extract useful information on many aspects, such as subject-verb-object matches and opinion matches. We evaluate our system on the accuracy rate of the syntax trees, and show that the result is satisfactory

METHODS AND TECHNIQUES:

The NLP pipeline used to parse and extract phenotypic characters from taxonomic descriptions includes the Explorer of Taxon Concepts (ETC; Cui et al., 2016) and Matrix Converter (Liu et al., 2015). ETC is an online application (http://etc.cs.umb. edu/etcsite/) that contains the Text Capture and Matrix Generation tools, which are used to parse text and assemble a character matrix. Matrix Converter (Fig. 1B, Appendix 1) is a Java application (available on GitHub at https://github.com/gburleigh/MatrixConverter/ tree/master/distribution) that facilitates the evaluation and discretization of the characters extracted by ETC and the formatting of the resulting character matrices. The initial input for the ETC's Text Capture Tool consists of text contained in the body of taxonomic descriptions written in English using a telegraphic syntax. In botanical descriptive literature, the most common telegraphic syntax format is characterized by its abbreviated format that drops auxiliary verbs and unnecessary terms (Fig. 2A, step 1). Taxonomic descriptions document a taxon's phenotypic traits and its variation; thus, the traits extracted represent summary information of a taxon and not of an individual. Ideally, users should select descriptions that represent the most up-to-date or credible circumscription of taxa. Furthermore, using descriptions written for the same taxonomic treatment (e.g., floras, monographs) by one or a few authors during the same time period may facilitate the

analysis and result in a more complete matrix, as they are more likely to be parallel and have a more consistent use of language. Extended sections of descriptions, which often include descriptions of habitat and discussion of diagnostic characters, should not be included in the analysis because they are written in natural (complete) language and will not work well with the ETC parser. Parenthetical remarks that explain a trait or compare it to other taxa (e.gdistal cells quadrate [including rhombic] to hexagonal; … in cross-section [mid-limb], cone [larger than other species of the genus] 10 cm long…) should also be excluded from the analysis because they often violate the rules of telegraphic syntax and hinder the ETC parsing analysis
Stemming and lemmatization are probably the first two steps to build an NLP project — you often use one of the two. They represent the field's core concepts and are often the first techniques you will implement on your journey to be an NLP master.

Often, beginners tend to confuse the two techniques. Although they have their similarities, they are quite different.

Stemming
Stemming is a collection of algorithms that work by clipping off the end of the beginning of the word to reach its infinitive form. These algorithms do that by considering the common prefixes and suffixes of the language being analyzed. Clipping off the words can lead to the correct infinitive form, but that's not always the case. There are many algorithms to perform stemming; the most common one used in English is the Porter stemmer. This algorithm contains 5 phases that work sequentially to obtain the word's root.

Lemmatization:
To overcome the flaws of stemming, lemmatization algorithms were designed.In these types of algorithms, some linguistic and grammar knowledge needs to be fed to the algorithm to make better decisions when extracting a word's infinitive form. For lemmatization algorithms to perform accurately, they need to extract the correct lemma of each word. So, they often require a dictionary of the language to be able to categorize each word correctly.
The proposed architecture of four modules: user interface, log pre-processing, Feature Clustering using Modified K-means, Naïve Bayes Classification, Training and testing using KNN for more accurate categorization of opinion. This system can solve irrelevant data and more accuracy by associating Modified K means with Naïve Bayes Classification algorithm.

Naive Bayes (NB):
Naive Bayes Classifier uses Bayes Theorem, which finds the probability of an event given the probability of another event that has already occurred. Naive Bayes classifier performs extremely well for problems which are linearly separable and even for problems which are non-linearly separable it performs reasonably well [3]. We used the already implemented Naive Bayes implementation in Weka2 toolkit
The following databases were searched from January 2000 and February 2018; MEDLINE, EMBASE, PsycINFO, The Cochrane Library (Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register), Global Health, Health Management Information Consortium, CINAHL and Web of Science. Grey literature and Google Scholar were used to extract articles that were not retrieved in the databases searched. Owing to the diversity of terms used inferring patient experience, combinations of search terms were used. The search terms, derived from the Medical

Subject Headings vocabulary (US National Library of Medicine) for the database queries that were used can be found below. A review of the protocol was not published.

“natural language processing” OR “NLP” OR “text mining” OR “sentiment analysis” OR “opinion mining” OR “text classification” OR “document classification” OR “topic modeling” OR “machine learning” “supervised machine learning” OR “unsupervised machine learning” AND “feedback” OR “surveys and questionnaires” OR “data collection” OR “health care surveys” OR “assessment” OR “evaluation” AND “patient centred care” OR “patient satisfaction” OR “patient experience”.

Inclusion criteria

To be eligible for inclusion in the review, the primary requirement was that the article needed to focus on the description, evaluation or use of NLP algorithm or pipeline to process or analyse patient experience data. The review included randomised controlled trials, non-randomised controlled trials, case–control studies, prospective and retrospective cohort studies and qualitative studies. Queries were limited to English language but not date constraints. We excluded studies that gathered patient-reported outcome measurements, symptom monitoring, symptom information, quality of life measures and ecological momentary assessment without patient experience data. Conference abstracts were excluded, as there was limited detail in the methodology to score against quality indicators.

Study selection

The research adhered to the guideline presented in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 checklist. The initial search returned 1007 papers; after removing duplicates 241 papers were retained. The titles and abstract were screened by two reviewers (MK and PA) independently, and discrepancies were resolved by a third reviewer (EM). Thirty-one articles were identified as potentially eligible for inclusion. Full-text articles were retrieved and assessed for inclusion by the same reviewers, of which 19 were retained for final inclusion. The main reason for exclusion was the articles reported other patient-reported feedback and not patient experience.

Data collection process

We developed a data collection tool with the following data fields: department of corresponding authors, country of study, study purpose, data source, solicited feedback, time period, information extraction method, data processing, ML classifiers, text analysis approach, software, performance, key findings and limitations. Two reviewers (MK and PA) independently completed the data collection, and met to compare the results, and discrepancies were resolved by a third reviewer (EM).

Data used in this study were the linguistic corpora obtained from conference proceedings whereby research articles and peer-reviewed papers were evaluated. Data were gathered from the Neural Information Processing Systems (NIPS), with the target being a range of research publications from 1987 to 2020. Additionally, research papers from the Cognitive Science Society from between 1900 and 2020 (accessed via an open-source managed by the University of California) were obtained for analysis. These two ML communities were chosen for their richness in AI-based information [21]. Moreover, the wide information available makes it possible to analyze the trends in using the term explainability and the related concepts.

Data synthesis

Due to the heterogeneous nature of the studies, a narrative synthesis was deemed most appropriate. A formal quality assessment was not conducted, as relevant reporting standards have not been established for NLP articles. Instead, we report indicators of quality guided by elements reported in previous NLP-focused systematic reviews.We included information related to the study purpose, corpus (e.g, data source and number of comments), NLP (e.g, methodology and software used and performance metrics). Two reviewers (MK and PA) independently evaluated indicators of quality in each study; disagreements in evaluation were resolved by discussion with a third reviewer (EM). Inter-rater agreement Cohen's kappa was calculated. In the reviewed studies, we assessed the NLP methodology and the rationale for its use. The key NLP approaches were summarised based on text analysis incorporating either text classification or topic modelling depending on the corpus available and evaluation was done as to whether sentiment analysis was performed using existing or bespoke software.

Performance metrics

To understand how well an automated ML algorithm performs, there are a number of statistical values that help determine its performance with the given data Algorithm performance is measured as recall (proportion of all true positive observations that are correct, that is, true positives/(true positives false negatives)), precision (ratio of correctly predicted positive observations to the total predicted positive observations) and by the F-score which describes overall performance, representing the harmonic mean of precision and recall.K-fold cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. This ensures that the results are not by chance, and therefore ensures the validity of the algorithms performance. We look all the recorded performance metrics in each of the included studies in order to gain a better understanding of how the data and ML approach can influence the performance.

Data Analysis

The data analysis involved conducting a shallow assessment for the term 'explanation'. The frequencies were normalized between conferences and years, after which the plots in Figure 2a,b were derived. Figure 2a is a frequency plot for the term explainability obtained from the Cognitive Science Society database. Figure 2b is the frequency plot obtained from the Neural Information Processing Systems (NIPS) database searches. The voyant.tools.org program was used to analyze the corpus, generated through the word cloud analysis, to establish a connection between the words and, consequently, generate meaning. The researcher also utilized the sketchengine.eu open-source program to generate the concordance of the term 'explainability'.

Our System mainly contains two modules one is Social media and another is Business User. Social Media: In this Module we are creating a social media like Facebook or twitter. we are added a functionality like Adding profile picture, posting tweets, adding friends and removing friend from there profile. Social Media user can view the posts of other users. Business User: In this module Business user can suggest or make Advertisement of their project directly to the Social user for their Business by Notification. System Architecture: In this user can post text as an input. Using core NLP technique, given text file or code file will be processed. Proposed system is going to perform operation like stemming, stop words removal and parsing technique. Core NLP Technique: Tokenization –The process of converting a text into tokens.

Stemming:

Stemming is a rudimentary rule-based process of stripping the suffixes (“ing”, “ly”, “es”, “s” etc) from a word. Stop word removal: Language stop words (commonly used words of a language – is, am, the, of, in etc.), URLs or links, social media entities (mentions, hash tags), punctuations and industry-specific words. This step deals with removal of all types of noisy entities present in the text. Entity Extraction: Entities are defined as the most important chunks of a sentence – noun phrases, verb phrases or both. Entity Detection algorithms are generally ensemble models of rule-based parsing, dictionary lookups, post tagging, and dependency parsing. The applicability of entity detection can be seen in the automated chat bots, content analyser’s, and consumer insight.

SVM Classifiers

An SVM classifier maps the feature vectors into high dimensional vector space and computes the dot product of the two vectors inside the kernel. Its applicability to both linear and non-linear systems has been proven for different NLP applications. We used SVM implementations from scikit-learn (Pedregosa et al., 2011) and experimented with a number of classifiers. We report here on results obtained using SVC adapted from libsvm (Chang & Lin, 2011) by embedding different kernels.

RESULT

Data mining can unintentionally be misused, and can then produce results that appear to be significant; but which do not actually predict future behavior and cannot be reproduced on a new sample of data and bear little use. Often this results from investigating too many hypotheses and not performing proper statistical hypothesis testing. A simple version of this problem in machine learning is known as overfitting, but the same problem can arise at different phases of the process and thus a train/test split—when applicable at all—may not be sufficient to prevent this from happening.

The final step of knowledge discovery from data is to verify that the patterns produced by the data mining algorithms occur in the wider data set. Not all patterns found by data mining algorithms are necessarily valid. It is common for data mining algorithms to find patterns in the training set which are not present in the general data set. This is called over fitting. To overcome this, the evaluation uses a test set of data on which the data mining algorithm was not trained. The learned patterns are applied to this test set, and the resulting output is compared to the desired output. For example, a data mining algorithm trying to distinguish "spam" from "legitimate" emails would be trained on a training set of sample e-mails. Once trained, the learned patterns would be applied to the test set of e-mails on which it had *not* been trained. The accuracy of the patterns can then be measured from how many e-mails they correctly classify. Several statistical methods may be used to evaluate the algorithm, such as ROC curves.

If the learned patterns do not meet the desired standards, subsequently it is necessary to re-evaluate and change the pre-processing and data mining steps. If the learned patterns do meet the desired standards, then the final step is to interpret the learned patterns and turn them into knowledge.

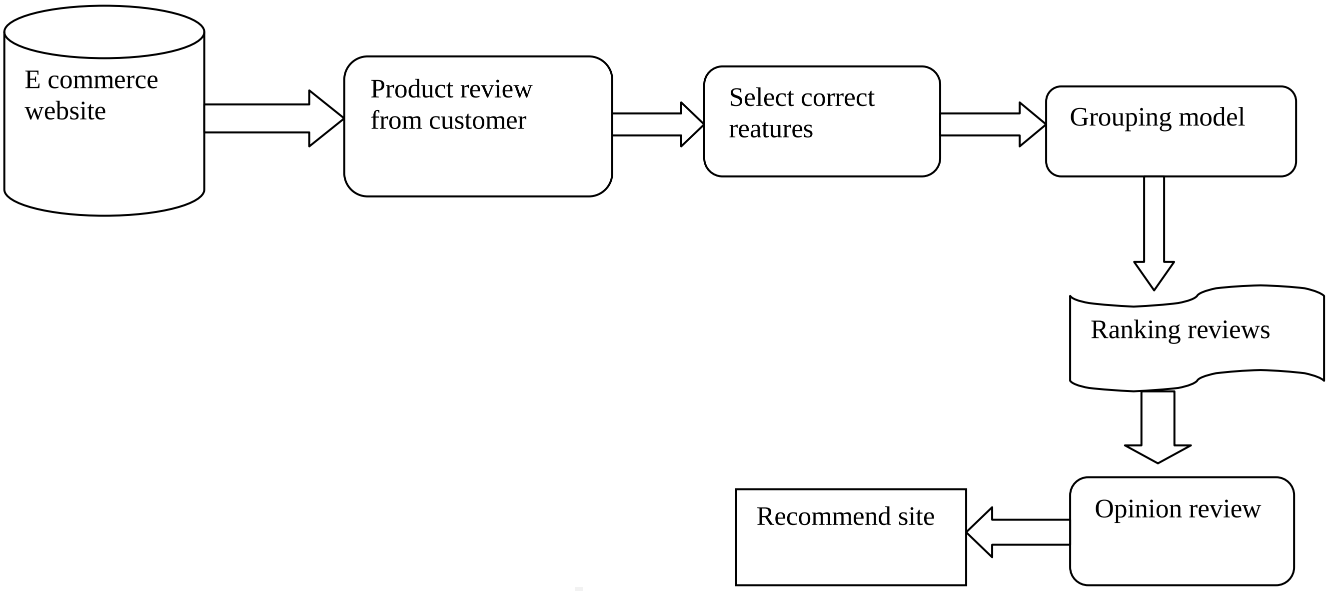


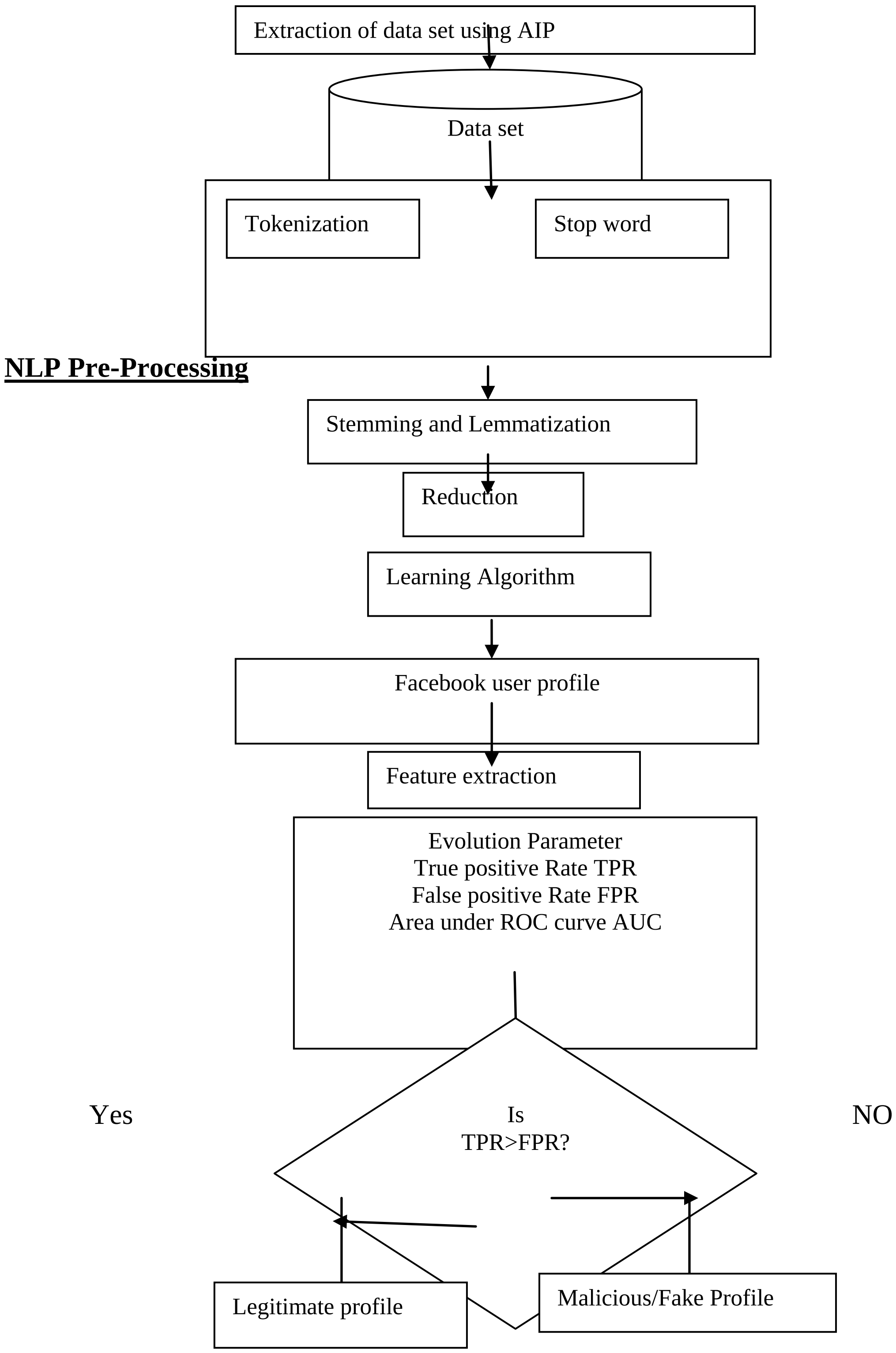
Diagram of the opinion mining process.

E-commerce categorization

Tag weighting method performs better than pure TF-IDF. But in this case the most efficient was extracting the whole pages from hyperlinks found on main pages. This is due to the fact that the dataset for categorization contains only ecommerce web sites. This reduces the volume of noisy information on the pages of the website and thus decreases the chances of topic drift on different pages. Most highly specialized e-shops place hyperlinks to catalogue page or to some product categories pages or to certain product pages where they describe them thoroughly. This information is useful for classification by product type in most cases. Exceptions here are universal e-shops for which detailed descriptions of some goods may lead to misclassification. This can be seen on Table 6 which presents average F-score for each category for the case when Tag Weighting is used and main page is concatenated with whole other pages.

Category Id	Category Name	Average F-Score
1	Auto products	0.89
2	Medical goods	0.98
3	Health and beauty products	0.82
4	Appliances and electronics	0.79
5	Household goods	0.94
6	furniture	0.92
7	Souvenirs and presents	0.69
8	Media (books, disks and concert tickets)	0.76
9	jewelry and clocks	0.73
10	Technical and industrial equipment	0.79
11	Food and kindred products	0.79
12	Pet supplies	0.85
13	Sport equipment and hobbies	0.73
14	Clothing and footwear	0.78
15	General stores	0.63

The scope of this project is to cover the political news data, of a dataset known as Liar-dataset, it is a New Benchmark Dataset for Fake News Detection and labeled by fake or trust news. We have performed analysis on "Liar" dataset . The results of the analysis of the datasets using the six algorithms have been depicted using the confusion matrix. The six algorithms used for the detection are as: • XGboost. • Random Forests. • Naive Bayes. • K-Nearest Neighbors (KNN). • Decision Tree. • SVM



Accuracy is often the most used metric representing the percentage of correctly predicted observations, either true or false. To calculate the accuracy of a model performance, the following equation can be used:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FB+FN}$$

In most cases, high accuracy value represents a good model, but considering the fact that we are training a classification model in our case, an article that was predicted as true while it was actually false (false positive) can have negative consequences; similarly, if an article was predicted as false while it contained factual data, this can create trust issues. Therefore, we have used three other metrics that take into account the incorrectly classified observation, i.e., precision, recall, and F1-score.

Discussion

Even if BERT paired many state-of-the-art results in different NLP tasks, Fast Text suits better in our pipeline - We observe that BERT fails mainly with the non-controversial datasets - Since BERT is a bigger and more complex model, it is able to separate the two communities' ways of speaking even when they are not opposite sides of a controversy, exploiting differences that we are not able to perceive - To check this behavior we plot the embeddings by t-SNE reductions.

From digital transformation and software-as-a-service to virtual reality and artificial intelligence, technology keeps pushing the limits of what ecommerce can do.

With compounding advancements in technology, there's something new competing for online retailers' attention every day. You'll never find yourself at a loss for something new and different to try — the real task is identifying the best opportunities for your ecommerce business.

Artificial intelligence isn't just a novel technology implemented for its "cool factor." Implementing AI has the potential to impact any number of business functions across your organization.

To understand how it could impact your business, it helps to have an understanding of the components of artificial intelligence.

The definition of AI is broad, and encompasses data mining, natural language processing, and machine learning.

Data mining refers to the gathering of both current and historical data to inform predictions.

Natural language processing focuses on human-computer interaction and how computers interpret natural human language.

Machine learning concern using a collection of algorithms to apply past experience or provide examples to solve problem. Deep learning "involves layering algorithms in an effort to gain greater understanding of the data."

Over the past couple of years, AI technology has matured and become a powerful tool to boost sales and optimize operations. Even many small ecommerce businesses are using technology with some kind of AI capability.

With virtual assistants and chatbot technology, you can deliver the appearance of higher touch customer support. While these bots aren't completely self-reliant, they can facilitate simple transactions, leaving live support agents able to focus on more complex issues.

Virtual agents also have the advantage of being available 24/7, so low-level questions and issues can be addressed at any time of day, without making your customer wait.

CONCLUSION

In this survey, we have elaborated on the main approaches to natural language processing. Each of the approaches has its advantages and disadvantages, and thus we can define guidelines regarding the selection of a proper NLP approach. If one is less concerned with semantics and assumes that knowledge lies within statistical facts on a specific corpus, it is advised to use statistics-based approach. Otherwise, if one is concerned with the semantics of discovered information, or it is desired to be able to easily explain and control the results, a pattern-based approach is more suitable. However, if one needs to bootstrap a pattern-based approach using statistics (for instance when there is insufficient expert knowledge available) or the other way around (e.g., when there is a need for a priori knowledge), a hybrid approach is more appropriate.

The NLP system we developed did not perform sufficiently for abstracting entire breast pathology reports. The all-or-nothing approach resulted in too broad of a scope of work and limited our flexibility to identify breast pathology procedures and results. Our NLP system was also limited by the lack of the gold standard data on rare findings and wide variation in pathology text. Focusing on individual, common elements and improving pathology text report standardization may improve performance.

Big data analytics (BDA) has emerged as the new frontier of innovation and competition in the wide spectrum of the e-commerce landscape due to the challenges and opportunities created by the information revolution. Big data analytics (BDA) increasingly provides value to e-commerce firms by using the dynamics of people, processes, and technologies to transform data into insights for robust decision making and solutions to business problems. This is a holistic process which deals with data, sources, skills, and systems in order to create a competitive advantage. Leading e-commerce firms such as Google, Amazon, eBay, ASOS, Netflix and Facebook have already embraced BDA and experienced enormous growth. Through its systematic review and creation of taxonomy of the key aspects of BDA, this study presents a useful starting point for the application of BDA in emerging e-commerce research. The study presents an approach for encapsulating all the best practices that build and shape BDA capabilities. In addition, the study reflects that once BDA and its scope are well defined; distinctive characteristics and types of big data are well understood; and challenges are properly addressed, the BDA application will maximize business value through facilitating the pervasive usage and speedy delivery of insights across organizations.

REFERENCES

1]Ajay Rastogi and Monica Malhotra, Article in Journal of Information & Knowledge Management, September 2017.

[2] Nitin Jindal and Bing Liu., "Opinion Spam and Analysis" by ACM-2008.

[3] Lim, E., Nguyen, V., Jindal, N., Liu, B. Lauw, H. 2010. Detecting product review spammers using rating behavior. CIKM.

[4] Mukherjee, A., Liu, B. and Glance, N. 2012. Spotting fake reviewer groups in consumer reviews. WWW.

[5] Poonam et.al," A Tour towards the Various Knowledge Representation Techniques for Cognitive Hybrid Sentence Modeling and Analyzer, International Journal of Informatics and Communication Technology (IJ-ICT), Vol.7, No.3, pp. 124-134, December 2018.

[6] Tanwar Poonam, Priyanka, "Spam Diffusion in Social Networking Media using Latent Dirichlet Allocation", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol.8, No.12, October, 2019.

[7] Al-Najada, H and X Zhu (2014). ISRD: Spam review detection with imbalanced data distributions. In Proceedings of the 15th IEEE International Conference on InformationReuse and Integration (IRI), pp. 553-560. New York, IEEE.

[8] Banerjee, S and AYK Chua (2014). Understanding the process of writing fake online reviews. In Proceedings of the 9th International Conference on Digital Information Management (ICDIM), pp. 68-73. New York, IEEE.

[9] S. Ghosh, B. Viswanath, F. Kooti, N. K. Sharma, G. Korlam, F. Benevenuto, N.Ganguly, and K. P. Gummadi. "Understanding and combating link farming in the twitter social network".

[10] Allahbakhsh, M and A Ignjatovic (2015). An iterative method for calculating robust ratingscores. IEEE Transactions on Parallel and Distributed Systems, 26(2), 340-350.