International Research Journal of Emerging Trends in Multidisciplinary

ISSN 2395-4434

Volume 1, Issue 5 July 2015

www.irjetm.com

ONLINE TEST EVALUATION USING AUTOMATIC TEXT ANALYSIS

Mrs. A. ANGAYARKANNI

ASSISTANT PROFESSOR,
DEPT OF MCA,
SRI MANAKULA VINAYAGAR
ENGINEERING COLLEGE

K. UMAVIGNESH

MCA
SRI MANAKULA VINAYAGAR
ENGINEERING COLLEGE

BALAJI. J

MCA
SRI MANAKULA VINAYAGAR
ENGINEERING COLLEGE

ABSTRACT:

The concept of automatic text analysis can be implemented in the online examinations to evaluate the tests written by the candidate and produces the appropriate results, which reduces the time and space complexity over the traditional paper written tests. It can also be implemented in the corporate's interview process for conducting tests in the selection process like essay writing or letter writing

TEXT MINING:

Text mining, also referred to as text data mining, roughly equivalent to text analytics, refers to the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text.

Text analysis involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization, and predictive analytics. The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods. Typical text mining tasks include text categorization, text clustering, concept / entity extraction, and production of granular taxonomies, sentiment analysis, document summarization, and entity relation modelling.

CONTENT ANALYSIS:

Content analysis refers to a general set of techniques useful for analyzing and

International Research Journal of Emerging Trends in Multidisciplinary

ISSN 2395-4434

Volume 1, Issue 5 July 2015

www.irjetm.com

understanding collections of text. There is considerable work done in this area, which predates Internet research by decades. In the context of understanding the impact of digitized collections and websites, one particularly relevant type of content analysis is the analysis of news articles. These news articles may be about the collection, or they may be about the type of resource in general.

In the context of understanding impact, these news articles can help you understand several things, including:

1. How well efforts to publicize the resource are reflected in the news.
2. For articles that aren't just re-prints of press releases, how is the resource or others like it being framed in the media? "Framing" is a concept used in fields such as media studies to understand how the public discourse on a topic influences public opinion, and also further public discourse. So, for instance, if digitization efforts are being portrayed as efficient ways to make rare materials available, that frame is very different than if articles are suggesting that digitization grants are an example of wasteful government spending.
3. From a strictly quantitative perspective, even counts of articles can give you some indication of impact based on frequency of mentions in the media.

Existing System:

Conducting a Content Analysis:

According to Krippendorff, six questions must be addressed in every content analysis:

1) Which data are analyzed?

2) How are they defined?

3) What is the population from which they are drawn?

4) What is the context relative to which the data are analyzed?

5) What are the boundaries of the analysis?

6) What is the target of the inferences?

Proposed System:

It consists of online evaluation of written test and produce appropriate grades. For this it requires the analyzing of the text which uses Latent Dirichlet Allocation (LDA), a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture

International Research Journal of Emerging Trends in Multidisciplinary

ISSN 2395-4434

Volume 1, Issue 5 July 2015

www.irjetm.com

over an underlying set of topic probabilities. In the context of text modeling, the topic probabilities provide an explicit representation of a document.

We present efficient approximate inference techniques based on vibrational methods and an EM algorithm for empirical Bayes parameter estimation. We report results in document modeling, text classification, and collaborative filtering, comparing to a mixture of unigrams model and the probabilistic LSI model.

Implementation:

Latent Dirichlet allocation (LDA) is a generative probabilistic model of a corpus. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. LDA assumes the following generative process for each document w in a corpus D:

1. Choose N ~ Poisson (ξ).

2. Choose θ ~ Dir (α).

3. For each of the N words w:

(a) Choose a topic $z_n \sim$ Multinomial (θ).

(b) Choose a word w_n from $p(w_n \mid z_n; \beta)$, a multinomial probability conditioned on the topic z_n.

Several simplifying assumptions are made in this basic model, some of which we remove in subsequent sections. First, the dimensionality k of the Dirichlet distribution (and thus the dimensionality of the topic variable z) is assumed known and fixed. Second, the word probabilities are parameterized by a k X V matrix β, which for now we treat as a fixed quantity that is to be estimated. Finally, the Poisson assumption is not critical to anything that follows and more realistic document length distributions can be used as needed. Furthermore, note that N is independent of all the other data generating variables (θ and z). It is thus an ancillary variable and we will generally ignore its randomness in the subsequent development.

A k-dimensional Dirichlet random variable θ can take values in the (k −1)-simplex (a k-vector θ lies in the (k −1)-simplex and has the following probability density on this simplex:

$$p(\theta|\alpha) = \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \dots \theta_k^{\alpha_k - 1},$$

where the parameter α is a k-vector with components $\alpha_i > 0$, and where Γ(x) is the Gamma function. The Dirichlet is a convenient distribution on the simplex —it is in the exponential family, has finite dimensional sufficient statistics, and is conjugate to the multinomial distribution. In Section 5, these properties will facilitate the development of inference and parameter estimation algorithms for LDA.

Graphical representation:

International Research Journal of Emerging Trends in Multidisciplinary

ISSN 2395-4434

Volume 1, Issue 5 July 2015

www.irjetm.com

Graphical model representation of LDA. The boxes are "plates" representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document.

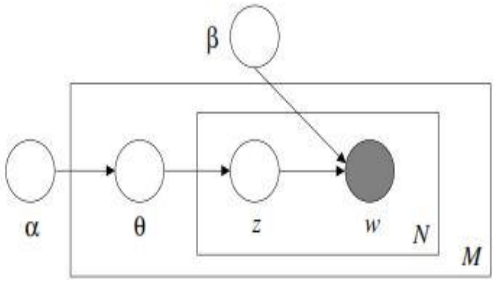

Fig 1 LDA Graphical representation

The below given is the methodology used to evaluate the content in the essay or letter writing. It holds the group of words or its alternatives. So we can easily able to match the text in the textbox with the database keywords. Number of matching is calculated and results are provided to the user.

"ARTS"	"BUDGETS"	"CHILDREN"	"EDUCATION"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT

International Research Journal of Emerging Trends in Multidisciplinary

ISSN 2395-4434

Volume 1, Issue 5 July 2015

www.irjetm.com

ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

Algorithm:

(1) $initialize\ \emptyset_{ni}^{0} := 1/k\ for\ all\ i\ and\ n$

(2) $initialize\ \gamma_i := \alpha_i + N/k\ for\ all\ i$

(3) ***repeat***

(4) ***for*** $n = 1$ ***to*** N

(5) ***for*** $i = 1$ ***to*** k

(6) $\emptyset_{ni}^{t+1} := \beta_{iw_n} exp\left(\psi(\gamma_i^t)\right)$

(7) $normalize\ \emptyset_n^{t+1}\ to\ sum\ to\ 1$

(8) $\gamma^{t+1} := \alpha + \sum_{n=1}^{N} \emptyset_n^{t+1}$

(9) ***until*** $convergence$

Fig 1.2 variation interface algorithm for LDA

We summarize the variational inference procedure in Figure1.2 with appropriate starting points for γ and φ. From the pseudo code it is clear that each iteration of variational inference for LDA requires O((N +1)k) operations. Empirically, we find that the number of iterations required for a single document is on the order of the number of words in the document. This yields a total number of operations roughly on the order of N^2k.

Applications:

As the automatic text analysis is used in the various applications, the main area we discussed is in the process of online exams valuation. Which is the key feature for the text analysis, this reduces the manual work and reduces the time and produces the accuracy. It can be used in the corporate recruitment process like essay writing and letter writing processes, which improves the performance. The model for online exams is given as follows:

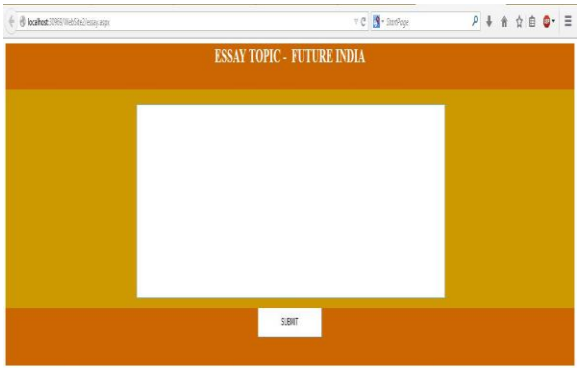

Fig 1.3 Essay Writing

Fig 1.3 states the essay writing window, which consists of the essay topic and the text box which provides space for writing the essay. The words used in the essay are compared with the key words in the database and the appropriate match ratio is collected and grades are provided accordingly

International Research Journal of Emerging Trends in Multidisciplinary

ISSN 2395-4434

Volume 1, Issue 5 July 2015

www.irjetm.com

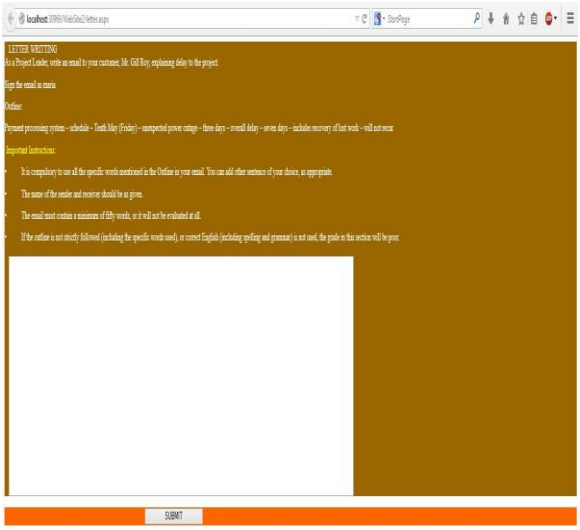

Fig 1.4 Letter Writing window

Fig. 1.4 is the letter writing window which helps to write the essay with the help of the hints given. It should also follow the appropriate rules.

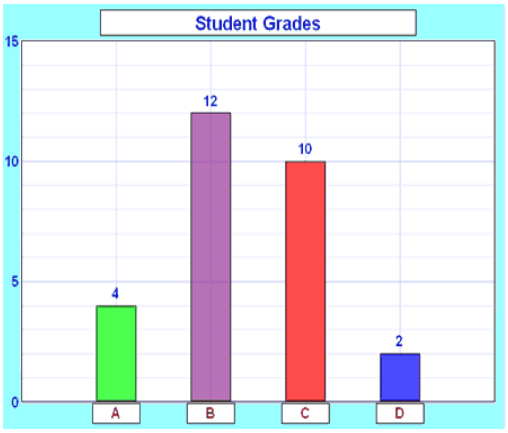

Fig1.5 Graphical representation of students result

Fig 1.5 helps to depict the user results. The result also helps to compare the mark of the candidate with others. The results may be in the form of graphical representation as follows:

We can also send the generated result to the candidate with help of e-mail; it helps for analyzing their performance with other candidate's results.

Future Enhancement:

Automatic Text Analysis by Artificial Intelligence:

Text is one of the traditional ways of communication between people. With the growing availability of text data in electronic form, handling and analysis of text by means of computers gained popularity. Handling text data with machine learning methods brought interesting challenges to the area that got further extended by incorporation of some natural language specifics. As the methods were capable of addressing more complex problems related to text data, the expectations become bigger calling for more sophisticated methods, in particular a combination of methods from different research areas including information retrieval, machine learning, statistical data analysis, data mining, natural language processing, semantic technologies.

Automatic text analysis become an integral part of many systems, pushing boundaries of research capabilities towards what one can refer to as an artificial intelligence dream - never ending learning from text aiming at mimicking ways of human learning. The paper presents

development of text analysis research in Slovenian that we have been personally involved in, pointing out interesting research problems that have been and are still addressed by the research, example tasks that have been addressed and some challenges on the way.

CONCLUSION:

Using the LDA technique for automatic text analysis simplifies the analysis of the text. So we can use the concept in the online exam valuation for calculating the candidate's performance in that particular test. We can implement in evaluating the test process such as essay writing, letter writing, etc., we can do this by using the concept of keyword comparison between the users text and the predefined keywords matching ratio. We can also able to provide the result in form of graphical representation. It also reduces the manual work and improves the performance.

References:

1. Multi dimensional analysis in political texts, elsevier journal, August 2013
2. Automatic Text Analysis by Artificial Intelligence, international journal, November 2012
3. Latent Dirichlet Allocation, Journal of Machine Learning Research 3 (2003) 993-1022