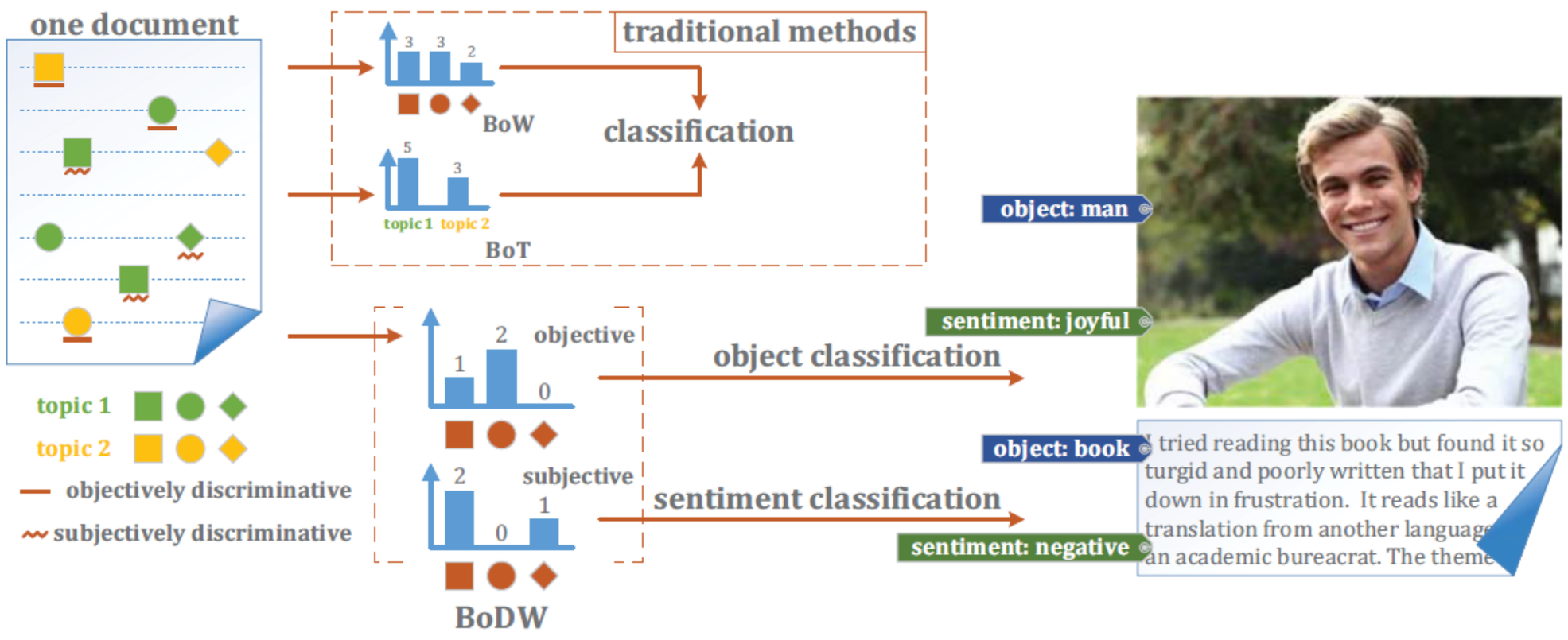
AK Tech Training and Placements

Transform Dreams into Reality

BAG-OF-DISCRIMINATIVE-WORDS (BODW) REPRESENTATION VIA TOPIC MODELING

Many of words in a given document either deliver facts (objective) or express opinions (subjective) respectively depending on the topics they are involved in. For example, given a bunch of documents, the word “bug” assigned to the topic “order Hemiptera” apparently remarks one object (i.e., one kind of insects), while the same word assigned to the topic “software” probably conveys a negative opinion. Motivated by the intuitive assumption that different words have varying degrees of discriminative power in delivering the objective sense or the subjective sense with respect to their assigned topics, a model named as discriminatively objective-subjective LDA (dosLDA) is proposed in this paper. The essential idea underlying the proposed dosLDA is that pair of objective and subjective selection variables is explicitly employed to encode the interplay between topics and discriminative power for the words in documents in a supervised manner. As a result, each document is appropriately represented as “bag-of-discriminative-words” (BoDW). The experiments reported on documents and images demonstrate that dosLDA not only performs competitively over traditional approaches in terms of topic modeling and document classification, but also has the ability to discern the discriminative power of each word in terms of its objective or subjective sense with respect to its assigned topic.

ARCHITECTURE



Mail: careers@aktechsolutions.in Phone: 7777-976-476 Web: http://aktechsolutions.in/

MODULES

There are three modules can be divided here for this project they are listed as below

- **Document Analysis**
- **Image Analysis**
- **Graphical Representation**

From the above three modules, project is implemented. Bag of discriminative words are achieved

MODULE DESCRIPTION

The modules are implemented as given in the following ways

1.Document Analysis

Users are uploading the document. The uploaded document can be analyzed and highlight the words. Every positive word in document highlighted in Green color and negative words in red color. The graph Analysis of the given document can be viewed as pie chart. The Graph has been plot for document total words, neutral words, positive and negative words.

2.Image Analysis

Admin is the one who can upload the picture for analysis. User can view the picture and rate according to their Perspective. And give comments to that image. From the comments and ratings admin can analysis the Sentiment of image. The Sentiment of the image can give to admin based on comments that are given by users.

3.Graphical Representation

Both admin and user can get the analysis respectively. The graph can be plot based on various factors that means number of word and positive and negative words count. User can get Line chart and bar chart for individual documents. Admin only gets the analysis for the image in Doughnut Chart

EXISTING SYSTEM

The two most successful and representative works in topic modeling are probabilistic latent semantic analysis (pLSA) and latent Dirichlet allocation (LDA). As the first topic model, pLSA evolves from latent semantic analysis (LSA) and is able to capture the hidden semantics conveyed by different words via a probabilistic generative process of the documents. In pLSA, documents are projected into a low-dimensional topic space by assigning each word with a latent topic, where each topic is usually represented as a multinomial distribution over a fixed vocabulary. The LDA model inherits the notion of pLSA, but it employs an extra generative process on the topic proportion of each document and models the whole corpus via a hierarchical Bayesian framework. In fact, pLSA turns out to be a special case of LDA with a uniform Dirichlet prior in a maximum a posteriori model, while LDA has a better ability of modeling large-scale documents for its welldefined a priori. In the past decade, the LDA model has been intensively studied and widely applied for many different tasks.

The BoW representation disregards the linguistic structures between the words. In such an unsupervised manner, the learned representations of documents provided by LDA are often found to be not strongly predictive. From a pure viewpoint of prediction, unsupervised LDA unfortunately ignores the nature of the discriminative task of interest such as classification, thus provides no guarantee that the extracted information will be effectual. To alleviate such limitation, many approaches attempt to exploit the useful auxiliary information (e.g., the category labels or the ratings provided by the authors) when modeling of its corresponding documents in a supervised manner. In such variants of LDA, the auxiliary information is usually considered to be a response variable predicted based on the latent representation of the document (i.e., the proportion of topics), where the assignments of topics to each word take effect instead of the words themselves. In other words, the “Bag-of-Topics” (BoT) representation has taken place of the traditional BoW representation to better characterize massive documents in predictive tasks such as regression and classification. The most representative models that proposed in the notion of BoT are the supervised LDA (sLDA)], the scene-understanding model, multi-class sLDA, and τLDA.

DISADVANTAGES

·From a pure viewpoint of prediction, unsupervised LDA unfortunately ignores the nature of the discriminative task of interest such as classification, thus provides no guarantee that the extracted information will be effectual.
The assignments of topics to each word take effect instead of the words themselves.

Mail: careers@aktechsolutions.in **Phone:** 7777-976-476 **Web:** http://aktechsolutions.in/

PROPOSED SYSTEM

The proposed work is an approach named as discriminatively objective-subjective LDA (dosLDA). The essential idea underlying it is that a pair of objective and subjective selection variables is explicitly employed to encode the interplay between topics and discriminative power with respect to the words in a supervised manner. The dosLDA possesses the attractive power in naturally selecting out those words that are discriminative in delivering either an objective or a subjective sense in one given document, and generates the novel "bag-of-discriminative words" (BoDW) representations for each document, which is illustrated in Figure. It is demonstrated via several experiments that our proposed BoDW is more predictive for discriminative tasks than the traditional BoW and BoT representations employed in the current methods.

ADVANTAGES

The bag of discriminated words is very effective when it is comes to analysis the document or image itself
For images, the system gets the comments from the user in order to involve the user and get the user view about image and from there they can find the sentiments of the image

CONCLUSION

In this paper, a supervised topic model named as dosLDA is proposed to discover the words having discriminative power to deliver either an objective or a subjective sense with regard to their assigned topics. The dosLDA model is able to obtain the BoDW representations for documents, and each document is endowed with two different BoDW representations in terms of objective and subjective senses, respectively. The results obtained on several experiments suggest that: (1) the BoDW representation is more predictive than the traditional BoT representation for discriminative tasks; (2) dosLDA boosts the performance of topic modeling via the joint discovery of latent semantic structure of the whole dataset and the different objective and subjective discrimination among the words; (3) dosLDA has lower computational complexity than sLDA, especially under an increasing number of topics; (4) the detected discriminative words or visual words are useful in topic demonstration as well as objective and sentimental region localization.

Mail: careers@aktechsolutions.in **Phone:** 7777-976-476 **Web:** http://aktechsolutions.in/