www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3, Issue 11 November, 2014 Page No. 9079-9083

Survey on Sentiment-Topic Detection from Text

Teenumol P D [1], Jeena Joy [2]

[1]Department Computer Science, Govt. Engineering College, Idukki
teenupdevassy@gmail.com

[2] Department Computer Science, Govt. Engineering College, Idukki
Jeenajoy38@gmail.com

Abstract: ***Sentiment analysis or opinion mining aims to use automated tools to detect subjective information such as opinions, attitudes, and feelings expressed in text. An important part of our information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as personal blogs and online review sites, new challenges and opportunities arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. This survey covers techniques and approaches that are used for sentiment analysis.***

Keywords: Opinion mining, Sentiment analysis, latent Dirichlet allocation (LDA), joint sentiment-topic (JST) model.

1. Introduction

Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. An alternative term for sentiment analysis is opinion mining [10], as it derives the opinion, or the attitude of a speaker. A common use case for this technology is to discover how people feel about a particular topic. Sentiment Analysis can be used to determine sentiment on a variety of levels. It will score the entire document as negative or positive, and it will also score the sentiment of individual words or phrases in the document.

Sentiment Analysis can track a particular topic, many companies use it to track or monitor their products, services or reputation in general. For example, if someone is attacking your brand on social media, sentiment analysis will score the post as extremely negative, and you can create alerts for posts with hyper-negative sentiment scores. A fundamental technology in many current opinion-mining and sentiment-analysis applications is classification. The reason that classification is so important is that many problems of interest can be formulated as applying classification/ regression/ ranking to given textual units.

This paper presents a survey of sentiment analysis techniques. Here we discuss five different method used for sentiment analysis. The first method describes *latent Dirichlet allocation* (LDA) [1], 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 set is modeled as a finite mixture over an underlying set of topics. Each topic is modeled as an infinite mixture over an underlying set of topic probabilities. In the perspective of text modeling, the topic probabilities provide a clear representation of a document.

Second method is DiscLDA, a discriminative variation on Latent Dirichlet Allocation (LDA) [3] in which a class-dependent linear transformation is introduced on the topic mixture proportions. This parameter is estimated by maximizing the conditional probability. By using the transformed topic mixture proportions as a new representation of documents, a supervised dimensionality reduction algorithm is obtained that uncovers the latent structure in a document collection while preserving predictive power for the task of classification.

In the third method a new framework for extracting the ratable aspects of objects from online user reviews is presented. Extracting such aspects is an important challenge in automatically mining product opinions from the web and in generating opinion-based summaries of user reviews. Multi-grain models are more suitable for standard models that tend to produce topics correspond to global properties of objects relatively than the aspects of an object that tend to be rated by a user. This model extracts not only ratable aspects, but also gather s them into coherent topics. This differentiates it from much of the previous work which extracts aspects through term frequency analysis with minimal clustering.

In next method a statistical model which is able to determine corresponding topics in text and extract textual evidence from reviews supporting each of these aspect ratings is discussed, that is a fundamental problem in aspect-based sentiment summarization. This model [4] attains high accuracy, without any explicitly labeled data except the user provided opinion ratings. The approach is general and can be used for segmentation in other applications where sequential data is accompanied with correlated signals.

The last method is a novel probabilistic modeling framework called joint sentiment-topic (JST) model based on latent Dirichlet allocation (LDA), which detects sentiment and topic concurrently from text. This method [5] focuses on document-level sentiment classification for general domains in conjunction with topic detection and topic sentiment analysis. This model extends the state-of-the-art topic model latent

Dirichlet allocation (LDA), by composing an additional sentiment layer, assuming that topics are generated dependent on sentiment distributions and words are generated conditioned on the sentiment-topic pairs.

2. Sentiment Analysis Techniques

Here we discuss five methods used for sentiment analysis. They are Latent Dirichlet allocation (LDA), Discriminative variation on Latent Dirichlet Allocation, Multi-grain LDA, Multi-Aspect Sentiment Model and Weakly Supervised Joint Sentiment-Topic Detection.

3.1 Latent Dirichlet Allocation (LDA)

Latent Dirichlet allocation (LDA) is a generative probabilistic model of a corpus. The key idea is that documents are characterized 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:

- Choose N ~ Poisson(x).
- Choose θ ~ Dir(a).
- For each of the N words w_n:

(a) Choose a topic z_n ~Multinomial(θ).
(b) Choose a word w_n from $p(w_n| z_n,\beta)$, a multinomial probability conditioned on the topic z_n.

Numerous simplifying assumptions are made in this basic model, some of which remove in subsequent sections. First, the dimensionality k of the Dirichlet distribution is assumed known and fixed. Second, the word probabilities are parameterized by a $k\times V$ matrix β where $\beta_{ij} = p(w_j = 1| z_i = 1)$, which will treat as a fixed quantity that is to be estimated. Finally, the Poisson assumption is not crucial 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 ignores its randomness in the subsequent development.
A k-dimensional Dirichlet random variable θ can get values in the $(k-1)$-simplex (a k-vector θ lies in the $(k-1)$-simplex if $\theta_i \geq 0$, $\sum^{k}_{i=1}\theta_i=1$). The Dirichlet is a suitable distribution on the simplex—it is in the exponential family, has finite dimensional sufficient statistics, and is conjugate to the multinomial distribution.

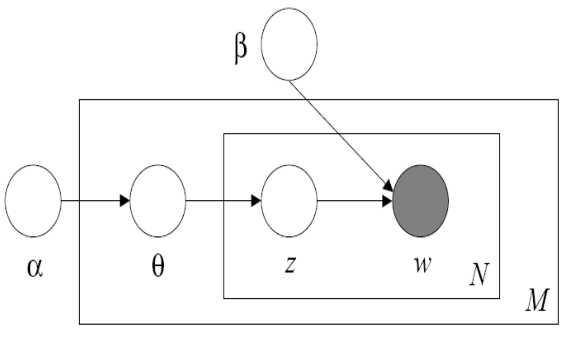

Fig. 1 Graphical model representation of LDA.

The LDA model is represented as a probabilistic graphical model in Figure 1. The boxes are “plates” representing replicates. The outer plate corresponds to documents, while the inner plate corresponds to the repeated choice of topics and words within a document. As the figure shows clear, there are three levels to the LDA representation. The parameters α and β are corpus level parameters, that is assumed to be sampled once in the process of generating a corpus. The variables θ_d are document-level variables, sampled once per document. At last, the variables z_{dn} and w_{dn} are word-level variables and are sampled once for each word in each document.

3.2 Discriminative variation on Latent Dirichlet Allocation (DiscLDA)

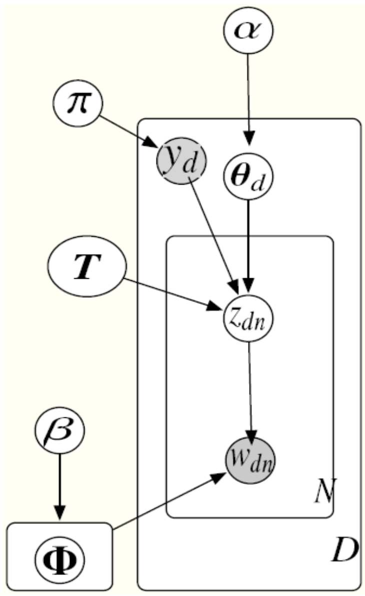

Fig. 2 DiscLDA

In this setting each document is additionally associated with a categorical variable or class label $y_d \in \{1, 2, . . . ,C\}$.To model this labeling information, a simple extension to the standard LDA model is introduced. In particular, for each class label y, a linear transformation $T^y: \Re^K \rightarrow \Re^L_+$ introduced, which transforms a K-dimensional Dirichlet variable θ_d to a mixture of Dirichlet distributions: $T^y\theta_d \in \Re^L$. To generate a word w_{dn}, its topic z_{dn} from $T^{yd}\theta_d$ is drawn. Note that T^y is constrained to have its columns sum to one to ensure the normalization of the transformed variable $T^y\theta_d$ and is thus a stochastic matrix. Intuitively, every document in the text corpus is represented through θ_d as a point in the topic simplex $\{\theta \mid \sum_k \theta_k = 1\}$, and hope that the linear transformation $\{T^y\}$ will be able to *reposition* these points such that documents with the same class labels are represented by points nearby to each other. Note that these points cannot be placed randomly, as all documents - whether they have the same class labels or they do not- share the parameter $\Phi \in \Re^{V\times L}$. The graphical model in Figure 2 shows the new generative process. Compared to standard LDA, the nodes for the variable y_d is added, the transformation matrices T^y and the corresponding edges.

3.3 Multi-grain LDA (MG-LDA)

Multi-grain LDA (MG-LDA) [2] models two distinct topics: global topics and local topics. As in Probabilistic Latent Semantic Analysis (PLSA) [6] and LDA [1], the distribution of global topics is fixed for a document. However, the distribution of local topics is allowed to vary across the document. A word in the document is sampled either from the mixture of global topics or from the mixture of local topics specific for the local context of the word. The hypothesis is that ratable aspects will be captured by local topics and global topics will capture properties of reviewed items. Local topics are expected to be reused between very different types of items, whereas global topics will correspond only to particular types of items. In order to capture only genuine local topics, a large number of global topics allowed, effectively, creating a bottleneck at the level of local topics. Of course, this bottleneck is specific to our purposes. Other applications of multi-grain topic models conceivably might even prefer the bottleneck reversed. Finally, the definition of multi-grain is simply for two-levels of granularity, global and local. However, there is nothing preventing the model described in this section from extending beyond two levels. One might expect that for other tasks even more levels of granularity could be beneficial. In Figure 3 the corresponding graphical model is presented.

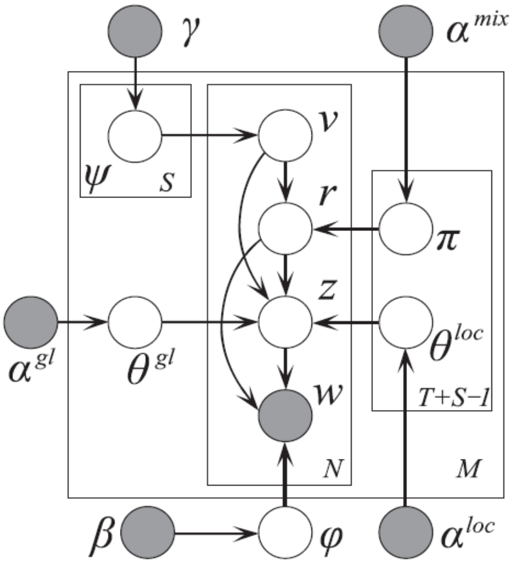

Fig. 3 MG-LDA

A document is represented as a set of sliding windows, each covering T adjacent sentences within a document. Each window v in document d has an associated distribution over local topics $\theta_{d,v}{}^{loc}$ and a distribution defining preference for local topics versus global topics $\pi_{d,v}$. A word can be sampled using any window covering its sentence s, where the window is chosen according to a categorical distribution Ψ_s. Significantly, the fact that the windows overlap, permits to exploit a larger co-occurrence domain.

The formal definition of the model with K^{gl} global and K^{loc} local topics is the following. First, draw K^{gl} word distributions for global topics $\varphi^{gl}{}_z$ from a Dirichlet prior $Dir(\beta^{gl})$ and K^{loc} word distributions for local topics $\varphi^{loc}{}_{z'}$ - from $Dir(\beta^{loc})$. Then, for each document d:

- Choose a distribution of global topics $\theta^{gl}{}_d \sim Dir(\alpha^{gl})$.
- For each sentence s choose a distribution $\Psi_{d,s}(v) \sim Dir(\gamma)$.
- For each sliding window v
  – choose $\theta^{loc}{}_{d,v} \sim Dir(\alpha^{loc})$,
  – choose $\pi_{d,v} \sim$Beta(α^{mix}).
- For each word i in sentence s of document d
  – choose window $v_{d,i} \sim \Psi_{d},_{s}$,
  – choose $r_{d,i} \sim \pi_{d,vd,i}$,
  – if $r_{d,i}$ = gl choose global topic $z_{d,i} \sim \theta^{gl}{}_d$,
  – if $r_{d,i}$ = loc choose local topic $z_{d,i} \sim \theta^{loc}{}_{d,vd,i}$
  – choose word $w_{d,i}$ from the word distribution $\varphi^{rd,i}{}_{zd,i.}$

Here, Beta (α^{mix}) is a prior Beta distribution for choosing between local and global topics. Though symmetrical Beta distributions can be considered, a non-symmetrical one is used as it permits to regulate preference to either global or local topics by setting $\alpha^{mix}{}_{gl}$ and $\alpha^{mix}{}_{loc}$ accordingly.

3.4 Multi-Aspect Sentiment Model

MG-LDA constructs a set of topics that ideally correspond to ratable aspects of an entity. A major shortcoming of this model - and all other unsupervised models - is that this correspondence is not explicit, i.e., how does one say that topic X is really about aspect Y? However, the numeric aspect ratings are often included in the data by users who left the reviews. Then make the assumption that the text of the review discussing an aspect is predictive of its rating. Thus, if the prediction of aspect ratings jointly with the construction of explicitly associated topics is modeled, then such a model should benefit from both higher quality topics and a direct assignment from topics to aspects. This is the basic idea behind the Multi-Aspect Sentiment model (MAS).

This method is to estimate the distribution of possible values of an aspect rating on the basis of the overall sentiment rating and to use the words given to the corresponding topic to compute corrections for this aspect. An aspect rating is naturally correlated to the overall sentiment rating and the fragments discussing this particular aspect will help to correct the overall sentiment in the appropriate direction. The aspect sentiment ratings can often be regarded as conditionally independent given the overall rating; therefore the model will not be forced to include in an aspect topic any words from other aspect topics. The fragments discussing overall opinion will influence the aspect rating only through the overall sentiment rating. The overall sentiment is almost constantly present in the real data along with the aspect ratings, but it can be coarsely discretised and preferred to use a latent overall sentiment.

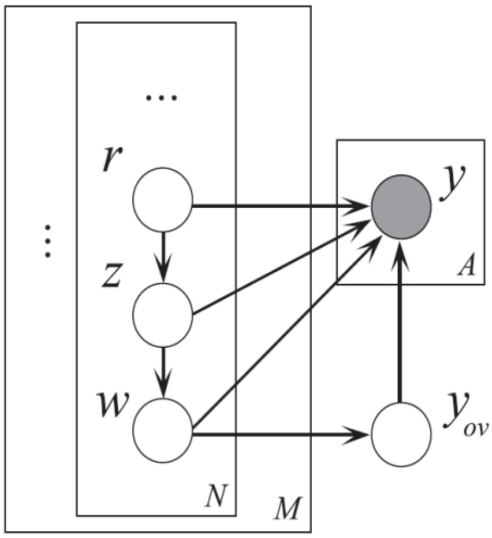

Fig. 4 MAS Model

The MAS model is presented in Figure 4. As in MG-LDA, MAS has global topics which are expected to capture topics corresponding to particular types of items. In figure 4 the aspect ratings y_a is shaded, assuming that every aspect rating is present in the data. In this model the distribution of the overall sentiment rating y_{ov} is based on all the n-gram features of a review text. Then the distribution of y_a, for every rated aspect a, can be computed from the distribution of y_{ov} and from any n-gram feature where at least one word in the n-gram is assigned to the associated aspect topic ($r = loc$, $z = a$).

Instead of having a latent variable y_{ov}, a similar model which does not have an explicit notion of y_{ov} is used. The distribution of a sentiment rating y_a for each rated aspect a is computed from two scores. The first score is computed on the basis of all the n-grams, but using a common set of weights independent of the aspect a. Another score is computed only using n-grams associated with the related topic, but an aspect-specific set of weights is used in this computation.

3.5 Weakly Supervised Joint Sentiment-Topic Detection

The existing framework of LDA has three hierarchical layers, where topics are associated with documents, and words are associated with topics. In order to model document sentiments, a joint sentiment-topic model [9] is proposed by adding an additional sentiment layer between the document and the topic layers. Hence, JST is effectively a four-layer model, where sentiment labels are associated with documents, under which topics are associated with sentiment labels and words are associated with both sentiment labels and topics. A graphical model of JST is represented in Figure 5.

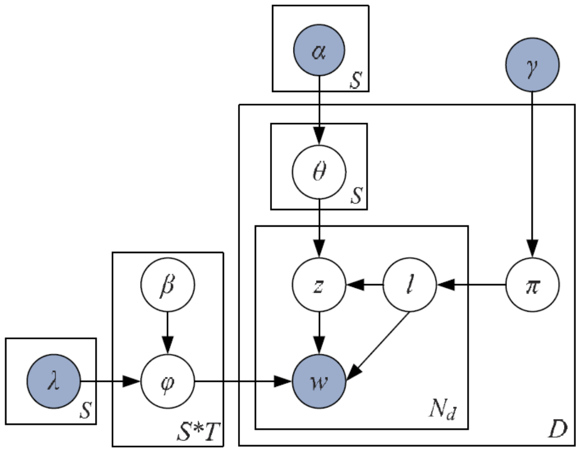

Fig. 5 JST Model

Consider a corpus with a collection of D documents denoted by $C=\{d_1,d_2,\ldots,d_D\}$; each document in the corpus is a sequence of N_d words denoted by $d=(w_1,w_2,\ldots,w_{Nd})$, and each word in the document is an item from a vocabulary index with V distinct terms denoted by $\{1, 2,\ldots,V\}$. Also, let S be the number of distinct sentiment labels, and T be the total number of topics. The procedure for generating a word w_i in document d under JST boils down to three stages. First, one select a sentiment label l from the per-document sentiment distribution π_d. Following that, one chooses a topic from the topic distribution $\theta_{d,l}$, where $\theta_{d,l}$ is conditioned on the sampled sentiment label l. It is important to note that the topic distribution of JST is different from that of LDA. In LDA, there is only one topic distribution θ for each individual document. In contrast, in JST each document is associated with S topic distributions, each of which corresponds to a sentiment label l with the same number of topics. This feature essentially provides means for the JST model to predict the sentiment associated with the extracted topics. Finally, one draws a word from the per-corpus word distribution conditioned on both topic and sentiment label. This is again different from LDA that in LDA a word is sampled from the word distribution only conditioned on topic.

The formal definition of the generative process in JST corresponding to the graphical model shown in Figure 5 is as follows:

- For each sentiment label $l\in\{1,\ldots,S\}$
  - For each topic $j\in\{1,\ldots,T\}$ draw $\phi_{lj} \sim \mathrm{Dir}(\lambda_l \times \beta_{lj}^T)$.
- For each document d, choose a distribution $\pi_d \sim \mathrm{Dir}(\gamma)$.
- For each sentiment label l under document d, choose a distribution $\theta_{d,l} \sim \mathrm{Dir}(\alpha)$.
- For each word w_i in document d
  - choose a sentiment label $l_i \sim \mathrm{Mult}(\pi_d)$,
  - choose a topic $z_i \sim \mathrm{Mult}(\theta_{d,li})$
  - choose a word w_i from ϕ_{lizi} , a multinomial distribution over words conditioned on topic z_i and sentiment label l_i.

The hyper-parameters α and β in JST can be treated as the prior observation counts for the number of times topic j associated with sentiment label l is sampled from a document and the number of times words sampled from topic j are coupled with sentiment label l, respectively, before having observed any actual words. Likewise, the hyper-parameter γ can be interpreted as the prior observation counts for the number of times sentiment label l sampled from a document before any word from the corpus is observed. In this implementation, asymmetric prior α and symmetric prior β and γ are used. In addition, there are three sets of latent variables that need to infer in JST, i.e., the per-document sentiment distribution π, the per-document sentiment label specific topic distribution θ, and the per-corpus joint sentiment-topic word distribution ϕ.

3. Conclusion

In this paper we have discussed several methods used for sentiment analysis. Sentiment detection has a wide variety of applications in information systems, including classifying reviews, summarizing review and other real time applications etc. Sentiment classifiers are dependent on domains or topics. Different types of features and classification algorithms can be combined in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment classification performance.

References

[1] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.

[2] I. Titov and R. McDonald, "Modeling Online Reviews with Multi-Grain Topic Models," Proc. 17th Int'l Conf. World Wide Web, pp. 111-120, 2008.

[3] S. Lacoste-Julien, F. Sha, and M. Jordan, "DiscLDA: Discriminative Learning for Dimensionality Reduction and Classification," Proc. Neural Information Processing Systems (NIPS), 2008.

[4] I. Titov and R. McDonald, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization," Proc. Assoc. Computational Linguistics—Human Language Technology (ACL-HLT), pp. 308-316, 2008.

[5] C. Lin, Yulan He, R. Everson ―Weakly Supervised Joint Sentiment-Topic Detection from Text, IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 6, June 2012.

[6] T. Hofmann. Unsupervised Learning by Probabilistic Latent Semantic Analysis. Machine Learning, 42(1):177–196, 2001.

[7] M. Hu and B. Liu. Mining Opinion Features in Customer Reviews. In Proceedings of Nineteenth National Conference on Artificial Intellgience, 2004.

[8] G. Carenini, R. Ng, and A. Pauls. Multi-Document Summarization of Evaluative Text. In Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics, 2006.

[9] C. Lin and Y. He, "Joint Sentiment/Topic Model for Sentiment Analysis," Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 375-384, 2009.

[10] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," J. Foundations and Trends in Information Retrieval, vol. 2, nos. 1/2, pp. 1-135, 2008.

[11] S. Li and C. Zong, "Multi-Domain Sentiment Classification," Proc. Assoc. Computational Linguistics—Human Language Technology (ACL-HLT), pp. 257-260, 2008.

[12] D. Ramage, D. Hall, R. Nallapati, and C. Manning, "Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-Labeled Corpora," Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), pp. 248-256, 2009.