A Two-Dimensional Topic-Aspect Model for Discovering Multi-Faceted Topics

Michael Paul and Roxana Girju
Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801
{mjpaul2, girju}@illinois.edu

Abstract

This paper presents the *Topic-Aspect Model* (TAM), a Bayesian mixture model which jointly discovers *topics* and *aspects*. We broadly define an aspect of a document as a characteristic that spans the document, such as an underlying theme or perspective. Unlike previous models which cluster words by topic or aspect, our model can generate token assignments in both of these dimensions, rather than assuming words come from only one of two orthogonal models. We present two applications of the model. First, we model a corpus of computational linguistics abstracts, and find that the scientific topics identified in the data tend to include both a computational aspect and a linguistic aspect. For example, the computational aspect of GRAMMAR emphasizes parsing, whereas the linguistic aspect focuses on formal languages. Secondly, we show that the model can capture different viewpoints on a variety of topics in a corpus of editorials about the Israeli-Palestinian conflict. We show both qualitative and quantitative improvements in TAM over two other state-of-the-art topic models.

Probabilistic topic models such as LDA (Blei, Ng, and Jordan 2003) have emerged in recent years as a popular approach to uncovering hidden structures in text collections, and offer a powerful way to represent the content of documents. These models, however, typically learn distributions over words along only a single dimension of topicality, and ignore the fact that words may fall along other dimensions such as sentiment, perspective, or theme.

Some work has been done to simultaneously model both topics and other types of groupings. For example, in the topic and perspective model (Lin, Xing, and Hauptmann 2008), each word is modeled as having some weight of topicality and perspective (e.g., liberal or conservative), however, this model assumes that all documents are about the same topic. The topic-sentiment mixture model (Mei et al. 2007) models each document as both a mixture of topics and a mixture of different sentiments (i.e. negative/positive), however, words come from either the topic model or the sentiment model rather than a combination of both.

In these approaches, there is no inter-dependency of topics and perspectives, and they cannot capture how these perspectives appear in different topics. Recently we have presented a new model, cross-collection latent Dirichlet allocation (ccLDA) (Paul and Girju 2009a), which can both discover topics among multiple text collections as well as differences between them, and we used this to capture the perspectives of different cultures in blogs. Each topic is associated with a probability distribution over words that is shared across all collections, as well as a distribution that is unique to each collection. For example, the topic of FOOD found in documents from different countries might contain the words *food* and *eat* among all collections, but *curry* would be more likely to appear in the India collection.

In many settings, however, it is more realistic to assume that a document is a *mixture* of such aspects, rather than belonging to exactly one aspect. For example, a recipe for curry pizza would contain elements from both Indian and American food, rather than strictly one or the other. In this paper we introduce a novel topic model, TAM, which not only allows documents to contain multiple aspects, but it can learn these aspects automatically. Unlike ccLDA, the model can be applied to a single collection and can discover patterns without document labels.

A common application of topic models is topic discovery in scientific literature (Griffiths and Steyvers 2004), which is useful for browsing large collections of literature. Topic models can also be used to assign research papers to reviewers (Mimno and McCallum 2007). In computational linguistics, (Hall, Jurafsky, and Manning 2008) and (Paul and Girju 2009b) model topics in this field and study their history.

These studies, however, have ignored the multi-faceted and interdisciplinary nature of many scientific topics. The only work in this direction we are aware of is our recent work (Paul and Girju 2009b) where we model scientific literature from multiple disciplines such as computational linguistics and linguistics. However, in that approach the fields are modeled independently, whereas TAM incorporates this directly into the model. In this paper we show how TAM can be used for the discovery of multi-faceted scientific topics.

Additionally, we model a corpus of editorials on the Israeli-Palestinian conflict. We improve upon studies of this corpus (Lin et al. 2006; Lin, Xing, and Hauptmann 2008) by modeling how different perspectives on this issue affect multiple topics within the data.

Copyright © 2010, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

The Model

In this section we first review the principles of unsupervised topic models such as PLSI and LDA. We then introduce our *Topic-Aspect Model* (TAM).

Unsupervised Topic Modeling

Probabilistic latent semantic indexing (PLSI) (Hofmann 1999) is a probabilistic model where each word w is associated with a hidden variable z that represents the *topic* that the word belongs to. Topic models such as PLSI are elegant and flexible approaches to clustering large collections of unannotated data. Each topic is associated with a probability distribution over words that captures meaningful word co-occurrences. However, one of the main criticisms of PLSI is that each document is represented as a variable d and it is not clear how to label previously unseen documents. With PLSI the number of parameters increases with corpus size, which leads to severe overfitting. This issue is addressed by (Blei, Ng, and Jordan 2003) with latent Dirichlet allocation (LDA), a Bayesian model which is similar to PLSI, but the distributions over topics and words have Dirichlet priors.

In LDA, a document is generated by first choosing a probability distribution over topics according to the probability given by Dirichlet(α). The Dirichlet parameter α is a vector which represents the average of the respective distributions. In many applications, it is sufficient to assume that such vectors are uniform and to fix them at a value pre-defined by the user, and these values act as smoothing coefficients.

Topic-Aspect Model (TAM)

Like other probabilistic topic models, TAM decomposes each document into some mix of *topics* that are characterized by a multinomial distribution over words. Words within each topic are typically related in some way. New to TAM over other topic models, however, is a second mixture component that can affect the nature of a document's content. We broadly define an *aspect* of a document as a characteristic that spans the document such as an underlying theme or perspective. The model expects that each aspect affects all topics in a similar manner.

For example, a computational linguistics paper may have both a computational aspect and a linguistic aspect. For instance, the computational aspect of the SPEECH RECOGNITION topic might focus on *Markov models* and *error detection*, while the linguistic aspect might focus on *prosody*. Other computational linguistics topics would likewise have words that are characteristic of each aspect.

Similar to ccLDA (Paul and Girju 2009a), we use a binary switching variable x to determine if the word comes from the aspect-neutral word distribution or aspect-dependent distribution. For example, the SPEECH RECOGNITION topic would have the word *speech* in its aspect-neutral distribution, but words like *markov* and *pitch* would respectively be more probable in the computational and linguistic aspect-dependent distributions.

Unlike ccLDA, however, TAM also includes an additional mixture component to distinguish common words and function words from topical words[1]. The top level $\ell = 0$ includes common "background" words that appear independently of a document's topical content. For example, a common word like *using* would likely belong to the background level, as it is not particularly topical. In the lower level $\ell = 1$, each word is associated with a topic.

Thus, each token i in the corpus is associated with five variables: a word w_i, a topic z_i, an aspect y_i, a level ℓ_i, and a route x_i. The word w_i is observable; the values of the other variables may be unknown. According to our model, a word in a document is generated as follows: One first chooses a topic and an aspect, then decides if the word should be a background word or a topical word (corresponding to level ℓ=0 or ℓ=1, respectively), then decides if the word should depend on the aspect or not (corresponding to route x=1 or x=0, respectively). Finally, a word is chosen according to some probability depending on these four factors – thus, a word may depend on a topic, an aspect, both, or neither.

If the word is to come from the background model, the word is sampled from $P(word|\ell = 0, x = 0)$ or $P(word|\ell = 0, x = 1, aspect)$ depending on if the aspect-independent or -dependent model is used. If the word is to be topical, it is sampled from $P(word|\ell = 1, x = 0, topic)$ or $P(word|\ell = 1, x = 1, aspect, topic)$. The topic, aspect, and level ℓ are chosen independently. The probability of choosing a route x, however, depends on the level and topic. This is because we expect that this may differ between the background and topical levels, and we cannot expect that all topics have the same degree of aspectuality.

Similar to LDA and ccLDA, the prior probability of the distributions in our model are defined by Dirichlet and Beta (the bivariate analog of Dirichlet) distributions. The model takes as input the number of topics and aspects.

Formally, the generative process for a corpus D is:
(1) Draw a multinomial word distribution ϕ_0 from Dirichlet(ω) for the background, ϕ_{0y} for each aspect, ϕ_{1z} for each topic, and ϕ_{1yz} for each aspect and each topic
(2) Draw a binomial route distribution ψ_0 from Beta(γ_0, γ_1) for the background and ψ_{1z} for each topic
(3) For each document $d \in D$:
- (a) Choose a document length N
- (b) Draw a multinomial topic mixture θ from Dirichlet(α)
- (c) Draw a multinomial aspect mixture π from Dirichlet(β)
- (d) Draw a binomial level mixture σ from Beta(δ_0, δ_1)
- (e) For each token $0 \leq i < N$ in the document:
  - i. Sample a topic z_i from θ
  - ii. Sample an aspect y_i from π
  - iii. Sample a level ℓ_i from σ
  - iv. Sample a route x_i from ψ_0 if $\ell = 0$ or ψ_{1z} if $\ell = 1$
  - v. Sample a word w_i from ϕ_0 if $\ell = 0$ and $x = 0$, ϕ_{0y} if $\ell = 0$ and $x = 1$, ϕ_{1z} if $\ell = 1$ and $x = 0$, or ϕ_{1yz} if $\ell = 1$ and $x = 1$

[1] A background model for common words is also used in the topic-sentiment model (Mei et al. 2007). This technique of modeling words at different levels of granularity is utilized in (Chemudugunta, Smyth, and Steyvers 2006) and (Haghighi and Vanderwende 2009). Our model differs from previous approaches in that there are multiple distributions (aspect-neutral and -dependent) even in the background level.

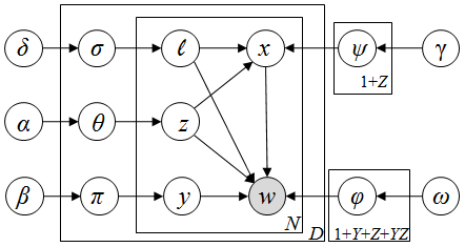

Figure 1: The graphical representation of TAM.

Note that words at either level ℓ can depend on the aspect, but only words at the topical level can depend on the topic. Since most documents will share many of the same background words regardless of their topical content, allowing the background words to be aspect-dependent gives some consistency to the aspectual word distributions. This helps enforce that all topics are affected by aspects in a similar fashion. If this were not the case, then the aspect-dependent distributions for each topic might form independently of the other topics, and the aspects would not show consistency.

Inference and Parameter Estimation

Exact inference of the posterior distribution of the hidden variables is intractable. We will instead approximate this using Gibbs sampling, a Markov chain Monte Carlo algorithm. In a Gibbs sampler, new values for z_i, y_i, ℓ_i, and x_i are iteratively sampled for each token i from the posterior probability conditioned on the previous state of the model (i.e., the current values for all other tokens) (Andrieu et al. 2003). This sampling equation is given in Figure 2.

The Dirichlet/Beta hyperparameters α, β, δ, γ and ω can be interpreted as "pseudo-counts" that are added to the observed counts. In the update equation, the values of these are assumed to be known. In many applications, it may be sufficient to leave these as pre-defined constants (Griffiths and Steyvers 2004). In some cases, it is important to obtain the optimal value for these parameters, in which case they can be estimated during the sampling procedure (Minka 2003). In our experiments, we find that we can achieve good performance by setting these parameters by hand and adjusting them until the results look reasonable.

Experimental Results

Experimental Setup

The field of computational linguistics is inherently interdisciplinary, with both computer science and linguistics aspects. In our first setup, we model computational linguistics abstracts with two aspects and hope to capture these two perspectives.

We collect our data from the openly available ACL Anthology[2]. We collected abstracts from the proceedings of the three most prominent conferences – ACL, COLING, and HLT – that were published after 1980 (the CL-only dataset). The distribution of data is shown in Table 1. The abstract

[2]http://www.aclweb.org/anthology-new/

$$P(z_i, y_i, \ell_i, x_i | \mathbf{w}, \mathbf{z}_{-i}, \mathbf{y}_{-i}, \ell_{-i}, \mathbf{x}_{-i}, \alpha, \beta, \delta, \gamma, \omega) \propto$$

$$(n^d_{z_i} + \alpha) \times (n^d_{y_i} + \beta) \times (n^d_{\ell_i} + \delta_{\ell_i}) \times \frac{n^\xi_{x_i} + \gamma_{x_i}}{n^\xi_* + \gamma_0 + \gamma_1} \times \frac{n^\zeta_{w_i} + \omega}{n^\zeta_* + W\omega}$$

$$\xi = \begin{cases} \ell = 0 & \ell_i = 0 \\ \ell = 1, z_i & \ell_i = 1 \end{cases} \qquad \zeta = \begin{cases} \ell = 0 & \ell_i = 0, x_i = 0 \\ \ell = 0, y_i & \ell_i = 0, x_i = 1 \\ \ell = 1, z_i & \ell_i = 1, x_i = 0 \\ \ell = 1, y_i, z_i & \ell_i = 1, x_i = 1 \end{cases}$$

Figure 2: The Gibbs sampling equation for new variable assignments. The notation n^a_b refers to the number of times b has been assigned to a, with $*$ being a wildcard. The counts exclude the current assignments of the token i. W is the size of the vocabulary. Note that the counts in the rightmost two terms are dependent on other variables.

Field	Venue	# Documents	Years
LING	Language	1031	78-08
LING	Linguistics, Journal of	152	97-08
LING	Linguistic Inquiry	338	98-08
LING	Ling. & Philosophy	652	77-08
CL	ACL	1,826	79-08
CL	COLING	1,549	80-08
CL	HLT	872	86-05

Table 1: The number of documents per field and publication venue. CL is Computational Linguistics; LING - Linguistics.

of each document can usually be extracted simply by grabbing the text in between the section headings "abstract" and "introduction." In other cases, we simply take the first 200 words of the document. We then remove common stop words and words with a frequency of less than 10. All punctuation is treated as a word separator.

We model this dataset with $Z = 25$ topics and $Y = 2$ aspects using $\omega = 0.01$, $\alpha = 0.1$, $\beta = 1.0$, $\gamma_0 = \gamma_1 = 10.0$ and $\delta_0 = \delta_1 = 10.0$. We run the Gibbs sampler for 5000 iterations.

In our second setup, we also include a set of abstracts from linguistics journals (the CL-LING dataset). We assume that the ACL abstracts are more likely to belong to one aspect, while the linguistics abstracts are more likely to belong to the other. Thus, we guide the model using semi-supervision by defining a prior probability for $P(y)$ based on which collection the document comes from. We do this by setting $\beta = 9.0$ for the aspect corresponding to the document collection and $\beta = 1.0$ otherwise. The other parameters are kept the same.

Finally, we consider a corpus of 594 editorials written by Israeli and Palestinian authors on the Israeli-Palestinian conflict[3] (the I-P dataset). This dataset is fully described in (Lin et al. 2006). It includes metadata that gives the perspective of each document, so we experiment with this dataset in both an unsupervised and semi-supervised (with a prior based on the document label) setting, using the same parameters as the first two setups, with $Z = 12$ topics.

[3]http://www.bitterlemons.org

Background
Neutral
paper
based
approach
information
present
language
new
using
model
analysis
different
problem
set
describes
context
work

Background
Aspect A
results
method
corpus
using
data
task
performance
learning
text
evaluation
methods
automatic
features
experiments
accuracy
algorithm

Topical		
Aspect A	**Neutral**	**Aspect B**
TOPIC 1		
similarity	semantic	ontology
patterns	relations	conceptual
clustering	lexical	verbs
words	relation	verb
classification	relationships	concepts
distributional	nouns	hierarchy
occurrence	categories	objects
TOPIC 2		
segmentation	discourse	temporal
text	relations	expressions
segment	events	tense
segments	event	theory
local	structure	aspect
coherence	descriptions	referring
cohesion	time	spatial

Background
Aspect B
natural
language
processing
structure
representation
semantic
linguistic
text
knowledge
framework
general
generation
form
computational
implemented
theory

Table 2: A sample of computational linguistics topics discovered in the CL-Only corpus with $Z = 25$ topics.

Topic and Aspect Discovery

Computational Linguistics Domain Modeling the CL-Only corpus produces results we might hope for. One aspect leans toward linguistic theory and natural language processing applications, with words like *language*, *semantic*, and *grammatical* near the top of the aspect-specific background-level distribution. The other aspect leans toward mathematical and computational problems and approaches, with top words like *automatic*, *algorithm*, and *statistical*.

Table 2 shows a sample of these topics as well as the top words in the background-level distributions. As an example, within the LEXICAL SEMANTICS topic, the computational aspect focuses on *distributional lexical semantics* and *word clustering*, while the linguistic aspect focuses more generally on *dictionaries and ontologies*. Another example (not shown) is the topic of GRAMMAR with focus on *parsing* (computational) and *formal languages* (linguistic).

Linguistics + Computational Linguistics Domains TAM indeed discovers many topics that accurately represent both collections. For example, the topic of COMMUNICATION contains words like *interaction* and *communication* in its aspect-neutral distribution. The CL aspect focuses on dialogue systems, with words like *dialogue* and *user*. The LING aspect focuses more generally on communication, with words like *communicative*, *conversation*, and *social*. In SEMANTICS, the LING aspect includes ideas less prevalent in the CL documents such as *pragmatics* and *metaphor* as well as *formal semantics*, while the CL aspect focuses on *semantic representation, inference*, and *textual entailment*.

Editorial Perspective Unsupervised modeling of the Israeli-Palestinian (I-P) data produces aspects that appear to correspond to the two perspectives of the conflict, and our quantitative results in the evaluation section seem to confirm this intuition.

It is harder to interpret the results than with the computational linguistics data, but looking at the aspect-dependent background distributions reveals some differences in the language used, such as *settlement* vs *occupation*. One aspect's distribution contains more words that might be used by Israelis such as *jewish* and the other has words more often used by Palestinians such as *palestine* and the Arabic word *intifada*. The aspects induced with the semi-supervised method are more clear, with *israel* near the top of one aspect's distribution and *palestinians* near the top of the other.

palestinian israeli israel military civilians attacks	
Aspect A	**Aspect B**
war	violence
public	palestinians
government	occupation
media	resistance
society	intifada
terrorist	violent
soldiers	non
incitement	force

state israel solution palestine palestinian states borders	
Israeli	**Palestinian**
jewish	palestinians
arab	return
israeli	right
jews	refugees
population	problem
jordan	refugee
west	rights
south	resolution

Table 3: Two topics in the I-P corpus discovered with unsupervised (left) and semi-supervised (right) methods. Aspect A and B seem to represent the Israeli and Palestinian perspectives.

Model Evaluation

Topic models are typically evaluated by measuring the likelihood of held-out data given a trained model (Wallach et al. 2009). However, it has been observed that such a likelihood measurement might not correlate with the quality of topics as interpreted by humans (Chang et al. 2009), which is important in our research. Thus, we instead evaluate TAM with human judgments of cluster coherence. We also demonstrate the representational power of aspects by applying TAM to a prediction task.

Cluster Coherence For the application of scientific topic discovery, it is important that the induced topics are semantically meaningful to humans and that the word clusters are

reasonably coherent. TAM was not designed to produce more coherent topics than other models, but rather different kinds of topics, and so it is not necessarily our goal to show that TAM exceeds other models in this regard. Instead, we simply want to show that the structure of TAM does not degrade the quality of topics, and we want to measure if TAM is at least as good as established topic models whose cluster quality has already been demonstrated.

To measure the cluster coherence, we follow the *word intrusion* methodology of (Chang et al. 2009). The idea is to give human annotators a set of words from one topic, and a randomly chosen "intruder" word from another topic. Annotators are asked to choose which word is the intruder – if the topic is coherent, then it should be easy to spot the out-of-place word. If the topic is not strongly coherent, then annotators are likely to guess and choose incorrectly.

We first evaluate TAM on the CL-LING dataset and compare against ccLDA (same experimental setup as above). We also evaluate TAM on the CL-Only dataset. Since ccLDA cannot be used on this single-collection dataset, we compare against LDA (25 topics and $\alpha = 1.0$, $\beta = 0.01$).

For TAM and ccLDA, we take the word with the highest probability in the aspect-neutral distribution as well as each aspect-specific distribution, for a total of 3 words in each topic. For LDA, we simply take the top 3 words for each topic. For each annotator, a word is randomly sampled from the top 10 words of a randomly chosen topic that is different from the topic being evaluated. It total, 4 words for each topic are shown to annotators in a random order. There are thus 100 different model/topic combinations (2 datasets with 2 models, each with 25 topics). These 100 topics are given to each annotator in a random order. The five annotators are computational linguistics graduate students and faculty.

For each topic, we define its score as the number of annotators that agreed with the model. Figure 3 shows the distribution of scores. It seems that TAM scores a little bit better than ccLDA – although it has 2 more topics on which no annotators were correct, it has 4 more topics with strong annotator agreement. On the CL-Only dataset, the interpretability of TAM seems to be better than LDA – TAM produced no topics on which all annotators were wrong, whereas LDA produced 3. One might have expected LDA to have better cluster coherence as it does not impose constraints on the topics (i.e. that each topic must fit across both aspects), but it seems that this is not the case.

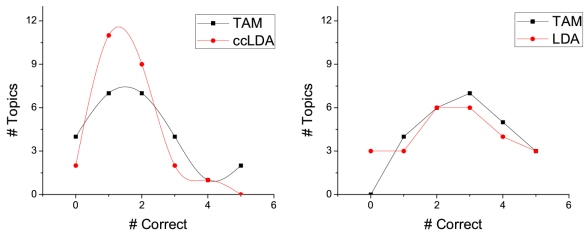

Figure 3: The annotator *word intrusion* scores on the CL-LING dataset (left) and CL-Only dataset (right).

Document Classification We would also like to quantitatively evaluate TAM's clustering abilities to see if the two aspects learned by TAM in the above experiments are meaningful. One way to do this is to use the unsupervised output of TAM as input to a supervised classifier – for example, if a classifier can learn to correctly label the aspect/perspective of a document using only the document's probability of membership to each aspect as input, then we can say that there is a strong correspondence between the aspects learned by TAM and the actual document aspects.

Specifically, we build binary classifiers to predict if an I-P document was written from the Israeli or Palestinian perspective. We train an SVM[4] using the learned real-valued topic and aspect probabilities as features.

We perform this on the I-P corpus[5] with 2 aspects and 12 topics, as done in the experiments above, using the unsupervised variant of TAM. The corpora are modeled without any reference to the document labels. As a comparison, we also test the standard LDA model with 12 topics (α=1.0). We also test LDA using only 2 topics to see if these topics are similar to the 2 aspects learned by TAM.

We experimented with five different feature spaces. LDA2 refers to the two-dimensional feature space using the documents' topic distributions learned by LDA with 2 topics; LDA12 refers to the same but with 12 topics. TAM2 refers to the documents' aspect distributions learned by TAM; TAM12 refers to the topic distributions learned by TAM; TAM2/12 refers to the concatenation of these two distributions (14 features in total).

For various values of K, we train each classifier using $\frac{1}{K}$ of the corpus and test its accuracy on the remaining $(K-1)/K$ documents. Our reported accuracies are computed using K-fold cross-validation; that is, we perform this procedure for K such partitionings and take the average accuracy. As a baseline, we simply predict a document's label as whichever class was more likely in the training data. Results are shown in Table 4. Each column shows the accuracy after training on different amounts of data.

We might expect that running LDA with two topics would produce topics that represent the two perspectives, however, TAM2 outperforms LDA2 by nearly 20%, so it is clearly not the case that the two topics induced by LDA are the same as TAM's two aspects. The fairly high classification accuracy using the simple two-dimensional feature space of TAM2, which is nearly as good as using both aspects and topics, demonstrates that the aspects learned by TAM have a strong correlation with the perspectives of the documents, despite the fact that they were modeled without supervision.

TAM12 performs about as poorly as LDA2, which shows that the topics learned by TAM are not by themselves a good determiner of the perspective. This makes sense, since the topics learned by TAM are modeled as belonging to both aspects, and thus it is less likely that the average topic distributions will greatly differ between the two perspectives.

[4]We used the SVM^{light} kit with the default C parameter and a linear kernel. (Available at http://svmlight.joachims.org)

[5]We focus on the I-P corpus because the CL-LING data is extremely easy to classify regardless of the method used.

Model / p	0.5%	1%	5%	10%	20%	80%
LDA2	50.23	53.31	58.84	62.39	64.14	64.23
LDA12	50.81	59.40	80.94	83.88	86.41	87.38
TAM2	**57.24**	**70.30**	84.19	84.77	85.02	85.27
TAM12	50.83	52.84	56.45	60.66	64.31	66.16
TAM2/12	54.44	67.15	**84.85**	**86.25**	**86.95**	**87.54**

Table 4: The average accuracy (%) of document classification in feature spaces learned by two topic models. Each classifier is trained on p% of the corpus and tested in the remaining $(1-p)$%.

We find that there is not a huge difference between LDA12 and TAM2/12 with large amounts of training data, however, there is a very large improvement in TAM over LDA when there is less training data. This suggests that aspects provide a more robust generalization of the data than topics. (On the other hand, if we were classifying documents by topic, then topics would likely be a more useful feature than aspects.) In fact, the aspect-only representation of TAM2 actually outperforms TAM2/12 with small amounts of training data, suggesting that it is hard to establish patterns of topicality from a small number of documents, whereas the aspects show a consistent pattern.

Discussion

The structure of the Topic-Aspect Model is very malleable and can be easily altered to suit the needs of a particular application. For example, the background/topical level binomial could be shared across the entire corpus rather than being drawn per-document. Conversely, the binomial distribution over x could be made to be generated per-document. The dependencies of x on z and/or ℓ could be dropped if a more rigid model is desired, or for more flexibility x could also depend on the aspect y.

We believe there are a number of applications in which TAM could potentially be used. LDA-style topic models have been shown to be very useful for document summarization (Haghighi and Vanderwende 2009), and TAM could be used similarly, for example to extract sentences to summarize the same information from different perspectives. TAM's outputs could be used to enrich the features used in certain systems. For example, if we wanted to train a system to extract the computational approaches used for a problem in a scientific paper, the aspect(s) assigned to a sequence of words might be useful features for distinguishing the method/approach from the problem. TAM could also be used for modeling sentiment and dialectical differences.

We must reconsider the notion of a "topic" and what it is that topic models uncover. When applied to text, topic models most often group words by what people would consider *topicality*, but this is clearly not the only such grouping. Furthermore, words may have a position in all of these dimensions, as has been shown in this research – instead of being associated with only a topic z or an aspect y, a word may be associated with a (topic, aspect) pair (z, y). There are interesting research possibilties in this direction of multi-dimensional latent spaces.

References

Andrieu, C.; de Freitas, N.; Doucet, A.; and Jordan, M. 2003. An introduction to mcmc for machine learning. *Machine Learning* 50(1):5–43.

Blei, D.; Ng, A.; and Jordan, M. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3.

Chang, J.; Boyd-Graber, J.; Gerrish, S.; Wang, C.; and Blei, D. 2009. Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems*.

Chemudugunta, C.; Smyth, P.; and Steyvers, M. 2006. Modeling general and specific aspects of documents with a probabilistic topic model. In *NIPS*, 241–248.

Griffiths, T., and Steyvers, M. 2004. Finding scientific topics. In *Proceedings of the National Academy of Sciences of the United States of America*.

Haghighi, A., and Vanderwende, L. 2009. Exploring content models for multi-document summarization. In *NAACL '09: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 362–370.

Hall, D.; Jurafsky, D.; and Manning, C. 2008. Studying the history of ideas using topic models. In *Empirical Natural Language Processing Conference*.

Hofmann, T. 1999. Probabilistic latent semantic indexing. In *SIGIR '99: Proceedings of the 22nd ACM SIGIR Conference*, 50–57.

Lin, W.; Wilson, T.; Wiebe, J.; and Hauptmann, A. 2006. Which side are you on? identifying perspectives at the document and sentence levels. In *Proceedings of Tenth Conference on Natural Language Learning (CoNLL)*.

Lin, W.; Xing, E.; and Hauptmann, A. 2008. A joint topic and perspective model for ideological discourse. In *ECML PKDD '08: Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases - Part II*, 17–32. Berlin, Heidelberg: Springer-Verlag.

Mei, Q.; Ling, X.; Wondra, M.; Su, H.; and Zhai, C. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, 171–180.

Mimno, D., and McCallum, A. 2007. Expertise modeling for matching papers with reviewers. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 500–509.

Minka, T. 2003. Estimating a dirichlet distribution.

Paul, M., and Girju, R. 2009a. Cross-cultural analysis of blogs and forums with mixed-collection topic models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 1408–1417.

Paul, M., and Girju, R. 2009b. Topic modeling of research fields: An interdisciplinary perspective. In *Recent Advances in Natural Language Processing (RANLP)*.

Wallach, H.; Murray, I.; Salakhutdinov, R.; and Mimno, D. 2009. Evaluation methods for topic models. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, 1105–1112.