Abstract

We explore how to detect people’s perspectives that occupy a certain proposition. We propose a Bayesian modelling approach where topics (or propositions) and their associated perspectives (or viewpoints) are modeled as latent variables. Words associated with topics or perspectives follow different generative routes. Based on the extracted perspectives, we can extract the top associated sentences from text to generate a succinct summary which allows a quick glimpse of the main viewpoints in a document. The model is evaluated on debates from the House of Commons of the UK Parliament, revealing perspectives from the debates without the use of labelled data and obtaining better results than previous related solutions under a variety of evaluations.

1 Introduction

Stance classification is binary classification to detect whether people is supporting or against a topic. Existing approaches largely rely on labelled data collected under specific topics for learning supervised classifiers for stance classification (Mohammad et al., 2016a). At most time, apart from detecting one’s stance, we are interested in finding out the arguments behind the person’s position. Perspectives, that state people’s ideas or the facts known to one, can be contrastive, i.e. to be in favour of or against something (e.g. Brexit vs Remain), or non-contrastive, i.e. independent discussions that share a common topic (e.g. unemployment and migration in the context of economy).

Recent years have seen increasing interests in argumentation mining which involves the automatic identification of argumentative structures, e.g., the claims and premises, and detection of argumentative relations between claims and premises or evidences. However, learning models for argumentation mining often require text labelled with components within argumentative structures and detailed indication of argumentative relations among them. Such labelled data is expensive to obtain in practice and it is also difficult to port models trained on one domain to another.

We are particularly interested in detecting different perspectives in political debates. Essentially, we would like to achieve somewhere in between stance classification and argumentation mining. Given a text document, we want to identify a speaker’s key arguments, without the use of any labelled data. For example, in debates about ‘Education’, we want to automatically extract sentences summarising the key perspectives and their arguments, e.g. ‘our education system needs to promote excellence in stem subjects’, ‘teenagers need to be taught with sexual and health education’ or ‘grammar schools promote inequality’. Similarly, if ‘Brexit’ is being discussed in terms of leaving or remaining, we want to cluster arguments into those two viewpoints.

To do this, we introduce a Latent Argument Model (LAM) which assumes that words can be separated as topic words and argument words and follow different generative routes. While topic words only involve a sampling of topics, argument words involve a joint sampling of both topics and arguments. The model does not rely on labelled data as opposed to most existing approaches to stance classification or argument recognition. It is also different from cross-perspective topic models which assume the perspectives are observed (Fang et al., 2012). Quantitative and qualitative evaluations on debates from the House of Commons of United Kingdom show the utility of the approach and provide a comparison against related models.
2 Related work

Our research is related to stance classification, argument recognition and topic modelling for sentiment/perspective detection.

2.1 Stance Classification

Stance detection aims to automatically detect from text whether the author is in favour of, against, or neutral towards a target. As previously reported in (Mohammad et al., 2016b), a person may express the same stance towards a target by using negative or positive language. Hence, stance detection is different from sentiment classification and sentiment features alone are not sufficient for stance detection. With the introduction of the shared task of stance detection in tweets in SemEval 2016 (Mohammad et al., 2016a), there have been increasing interests of developing various approaches for stance detection. But most of them focused on building supervised classifiers from labelled data. The best performing system (Zarrella and Marsh, 2016) made use of large unlabelled data by first learning sentence representations via a hashtag prediction auxiliary task and then fine-tuning these sentence representations for stance detection on several hundred labelled examples. Nevertheless, labelled data are expensive to obtain and there is a lack of portability of classifiers trained on one domain to move to another domain.

2.2 Argument Recognition

Closely related to stance detection is argument recognition which can be considered as a more fine-grained task that it aims to identify text segments that contain premises that are against or in support of a claim. Cabrio and Villata (2012) combined textual entailment with argumentation theory to automatically extract the arguments from online debates. Boltuzic and Šnajder (2014) trained supervised classifiers for argument extraction from their manually annotated corpus by collecting comments from online discussions about two specific topics. Sardianos et al. (2015) proposed a supervised approach based on Conditional Random Fields for argument extraction from Greek news. Nguyen and Litman (2015) run an LDA model and post-processed the output, computing argument and domain weights for each of the topics, which were then used to extract argument and domain words. Their model outperformed traditional n-grams and lexical/syntactic rules on a collection of persuasive essays. Lippi and Torroni (2016a) hypothesized that vocal features of speech can improve argument mining and proposed to train supervised classifiers by combining features from both text and speech for claim detection from annotated political debates. Apart from claim/evidence detection, there has also been work focusing on identification of argument discourse structures such as the prediction of relations among arguments or argument components (Stab and Gurevych, 2014; Peldszus and Stede, 2015). A more recent survey of various machine learning approaches used for argumentation mining can be found in (Lippi and Torroni, 2016b). All these approaches have been largely domain-specific and rely on a small set of labelled data for supervised model learning.

2.3 Topic Modeling for Sentiment/Perspective Detection

Topic models can be modified to detect sentiments or perspectives. Lin and He (2009) introduced a joint sentiment topic (JST) model, which simultaneously extracts topics and topic-associated sentiments from text. Trabelsi and Zaıane (2014) proposed a joint topic viewpoint (JTV) model for the detection of latent viewpoints under a certain topic. This is essentially equivalent to the reparameterized version of the JST model called REVERSE-JST (Lin et al., 2012) in which sentiment label (or viewpoint) generation is dependent on topics, as opposed to JST where topic generation is conditioned on sentiment labels.

Fang et al. (2012) proposed a Cross-Perspective Topic Model (CPT) in which the generative processes for topic words (nouns) and opinion words (adjectives, adverbs and verbs) are different, as the opinion words are sampled independently from the topic. Also, CPT assumed perspectives are observed, which implies texts need to be annotated with the viewpoint they belong to. Awadallah et al. (2012) detected politically controversial topics by creating an opinion-base of opinion holders and their views. Das and Lavoie (2014) observed the editions and interactions of a user in Wikipedia pages to infer topics and points of view at the same time. Qiu et al. (2015) proposed a regression-based latent factor model which jointly models user arguments, interactions, and attributes for user stance prediction in online debates.
3 Latent Argument Model (LAM)

We assume that in a political debate, the speaker first decides on which topic she wants to comment on (e.g. Education). She then takes a stance (e.g. remark the importance about stem subjects) and elaborates her stance with arguments. It is worth noting that we do not consider the temporal dimension of documents here, i.e., our model is fed with a collection of unlabeled documents without temporal order.

We use a switch variable $x$ to denote whether a word is a background word (shared across multiple topics), a topic word (relating to a certain topic) or an argument word (expressing arguments under a specific topic). Depending on the type of word, we follow a different generative process. For each word in a document, if it is a background word, we simply sample it from the background word distribution $\phi^b$; if it is a topic word, we first sample a topic $z$ from the document-specific topic distribution $\theta_d$ and then sample the word from the topic-word multinomial distribution $\psi_d(z)$ shared across all documents; if it is an argument word, we need to first jointly sample the topic-argument pair, $z, a$, where $z$ comes from the existing topics already sampled for the topic words in the document and $a$ is sampled from the topic-specific argument distribution $\omega_d$; and finally the word is drawn from the multinomial word distribution for the topic-specific argument $\psi_d(z,a)$. The argument indicator here is a latent categorical variable. It can take a binary value to denote pro/con or positive/negative towards a certain topic. More generally, it could also take a value from multiple stance or perspective categories. We thus propose a Latent Argument Model (LAM) shown in Figure 1.

Formally, the generative process is as follows:

- Draw a distribution over the word switch variable, $\phi \sim \text{Dirichlet}(\gamma)$, and background word distribution, $\psi^b \sim \text{Dirichlet}(\beta^b)$.
- For each topic $z \in \{1...T\}$, draw a multinomial topic-word distribution $\psi_d(z) \sim \text{Dirichlet}(\beta^z)$.
  - For each argument $a \in \{1...A\}$ draw a multinomial topic-argument distribution $\omega_d \sim \text{Dirichlet}(\delta)$ as well as a multinomial topic-argument-word distribution $\psi_d(z,a) \sim \text{Dirichlet}(\beta^a)$.
- For each document $d \in \{1...D\}$:
  - Draw a multinomial topic distribution, $\theta_d \sim \text{Dirichlet}(\alpha)$.
  - For each word $n \in \{1,...,N_d\}$ in $d$:
    * Choose $x_{d,n} \sim \text{Multinomial}(\phi)$.
    * If $x_{d,n} = 0$, draw a background word $w_{d,n} \sim \psi^b$.
    * If $x_{d,n} = 1$, draw a topic $z \sim \text{Multinomial}(\theta_d)$ and a word $w_{d,n} \sim \text{Multinomial}(\psi_d(z))$.
    * If $x_{d,n} = 2$, draw a topic $z \sim \text{Multinomial}(\theta_d)$, an argument $a \sim \text{Multinomial}(\omega_d)$ and a word $w_{d,n} \sim \text{Multinomial}(\psi_d(z,a))$.

Figure 1 shows its plate representation.

![Figure 1: The plate notation for the LAM model. Shadowed elements represent the observed variables (words and prior distributions).](image)

3.1 Inference and Parameter Estimation

We use Collapsed Gibbs Sampling (Casella and George, 1992) to infer the model parameters and the latent assignments of topics and arguments, given the observed data. Gibbs sampling is a Markov chain Monte Carlo method to iterative estimate latent parameters. In each iteration, a new sample of the hidden parameters is made based on the distribution of the previous epoch. Letting the index $t = (d, n)$ denote the $n$th word in document $d$ and the subscript $-t$ denote a quantity that excludes data from the $n$th word position in document $d$, $\Lambda = \{\alpha, \beta^b, \beta^z, \beta^a, \gamma, \delta\}$, the conditional posterior for $x_t$ is:

$$P(x_t = r| x_{-t}, z, a, w, \Lambda) \propto \frac{N^r_d}{N_d} + \gamma \cdot \frac{N^r_{w_t}}{N_{w_t}} + \beta^r \cdot \sum_{w'} \frac{N^r_{w_t}}{N_{w_t}} - t + W \beta^r,$$

where $r$ denotes different word types, either background word, topic word or argument word. $N^r_d$ denotes the number of words in document $d$ assigned to the word type $r$, $N_d$ is the total number of words in the document $d$, $N^r_{w_t}$ is the number of
times word \( w_t \) is sampled from the distribution for the word type \( r \), \( W \) is the vocabulary size.

For topic words, the conditional posterior for \( z_t \) is:

\[
P(z_t = k | z_{-t}, w, \Lambda) \propto \frac{N_{d,k}^{-t} + \alpha_k}{N_{d}^{-t} + \sum_k \alpha_k} \cdot \frac{N_{k,u_k}^{-t} + \beta^z}{N_{k}^{-t} + W \beta^z},
\]

(2)

where \( N_{d,k} \) is the number of times topic \( k \) was assigned to some word tokens in document \( d \), \( N_d \) is the total number of words in document \( d \), \( N_{k,u_k} \) is the number of times word \( w_t \) appeared in topic \( k \).

For argument words, the conditional posterior for \( z_t \) and \( a_t \) is:

\[
P(z_t = k, a_t = j | z_{-t}, a_{-t}, w, \Lambda) \propto \frac{N_{d,k}^{-t} + \alpha_k}{N_{d}^{-t} + \sum_k \alpha_k} \cdot \frac{N_{k,j}^{-t} + \delta_{k,j}}{N_{k}^{-t} + \sum_j \delta_{k,j}} \cdot \frac{N_{k,j,u_t}^{-t} + \beta^a}{N_{k,j}^{-t} + W \beta^a},
\]

(3)

where \( N_{k,j} \) is the number of times a word has been associated with the topic \( k \) and argument \( j \), \( N_{k,j,u_t} \) is the number of times word \( w_t \) appeared in topic \( k \) and with argument \( j \), and \( N_{k,j} \) is the number of words assigned to topic \( k \) and argument \( j \).

Once the assignments for all the latent variables are known, we can easily estimate the model parameters \( \{ \theta, \phi, \rho, \psi^b, \psi^z, \psi^a, \omega \} \). We set the symmetric prior \( \gamma = 0.3, \epsilon = 0.01, \beta^b = \beta^z = 0.01, \delta = (0.05 \times L)/A \), where \( L \) is the average document length, \( A \) is the total number of arguments, and the value of 0.05 on average allocates 5% of probability mass for mixing. The asymmetric prior \( \alpha \) is learned directly from data using maximum-likelihood estimation (Minka, 2003) and updated every 40 iterations during the Gibbs sampling procedure. In this paper we only consider two possible stances, hence, \( A = 2 \). But the model can be easily extended to accommodate more than two stances or perspectives. We set the asymmetric prior \( \beta^a \) for the topic-argument-word distribution based on a subjectivity lexicon in hoping that contrastive perspectives can be identified based on the use of positive and negative words. We run Gibbs sampler for 1 000 iterations and stop the iteration once the log-likelihood of the training data converges.

### 3.2 Separating Topic and Perspective Words

Using the word type switch variable \( x \), we could separate topic and argument words in LAM based solely on the statistics gathered from data. We also explore another two methods to separate topic and argument words based on Part-of-Speech (POS) tags and with the incorporation of a subjectivity lexicon. For the first variant, we adopt the similar strategy as in (Fang et al., 2012) that nouns (NOUN) are topic words; adjectives (ADJ), adverbs (ADV) and verbs (VERB) are argument words; words with other POS tags are background words. Essentially, \( x \) is observed. We call this model LAM.POS.

For the second variant, instead of assuming \( x \) is observed, we incorporate the POS tags as prior information to modify the Dirichlet prior \( \gamma \) for the word type switch variable at the initialization step. In addition, we also consider a subjective lexicon\(^2\), \( L \), that if a word can be found in the lexicon, then it is very likely the word is used to convey an opinion or argument, although there is still a small probability that word could be either background or topic word. Assuming an asymmetric Dirichlet prior for \( x \) is parametrized by \( \gamma^T = [\gamma_b, \gamma^z, \gamma^a] \) for background, topic and argument words, it is modified by a transformation matrix \( \lambda, \gamma^\text{new} = \lambda \times \gamma^T \), where \( \lambda \) is defined by:

- If word \( w \in L \land \text{POSTAG}(w) \neq \text{NOUN} \) then \( \lambda^T = [0.05, 0.05, 0.9] \)
- else if \( \text{POSTAG}(w) = \text{NOUN} \) then \( \lambda^T = [0.05, 0.9, 0.05] \)
- else if \( \text{POSTAG}(w) \in \{ \text{ADJ,ADV,VERB} \} \) then \( \lambda^T = [0.05, 0.05, 0.9] \)
- else \( \lambda^T = [0.9, 0.05, 0.05] \)

The conditional probability for the switch variable \( x \) is modified by simultaneously considering the POS tag \( g \) for the word at position \( t \):

\[
P(x_t = r, y_t = g | x_{-t}, z, a, w, \Lambda) \propto \frac{\{N_d^g\}^{-t} + \gamma \cdot \{N_{w^r}^g\}^{-t} + \beta^r}{\{N_d\}^{-t} + 3\gamma \cdot \sum_{w^r} \{N_{w^r}^g\}^{-t} + W \beta^r} \cdot \frac{\{N_g^r\} + \epsilon_g^r}{\{N_g\} + \sum_g \epsilon_g^r},
\]

(4)

where an additional term is added to the RHS of Equation 1. Here, \( N_{w^r}^g \) denotes the number of words with POS tag \( g \) assigned to the word type \( w \).

\(^{1}\)This equation had a typo in the original submitted paper that has been corrected.

\(^{2}\)In this work, we use the subjectivity lexicon presented at (Wiebe et al., 2005).
was applied, and a naive negation treatment was considered for the particle ‘not’, by creating bigrams for words occurs in the subjectivity lexicon (e.g., ‘not good’ becomes ‘not_good’). As topic models suffer from lack of robustness if large outliers are present, we also removed very frequent (above 99%) and rare words (below percentile 65%), assuming that word occurrences of the collection follow a Zip’s law distribution. Similar strategy was carried out for texts, in order to just consider texts of similar length. The preprocessed HCD contains a total of 1,598 speeches.

5 Experiments

This section evaluates LAM and its variants qualitatively and quantitatively (averaged over 5 runs). The models for comparison are listed below:

- **LDA.** Latent Dirichlet Allocation (Blei et al., 2003).
- **CPT.** The Cross-perspective Topic Model (Fang et al., 2012) assumes perspectives are observed. To be able to run this model on the political speeches, we implemented a version that can manage latent perspectives and separately sample topics and viewpoints.
- **JTV.** Joint Topic-Viewpoint Model (Trabelsi and Zaıane, 2014) is essentially the parameterized version of the Joint Sentiment-Topic (JST) model (Lin and He, 2009) called **REVERSE-JST** (Lin et al., 2012) in which sentiment label (or viewpoint) generation is dependent on topics. We implemented **JTV** as the reversed **JST** model.\(^7\)
- **LAM.** Latent Argument Model from §3.
- **LAM_POS.** LAM with topic, argument or background words separated by POS tags.
- **LAM_LEX.** Both POS tags and a subjective lexicon are used to initialise the Dirichlet prior \(\gamma\) for the word type switch variable as described in §3.2.

5.1 Experimental Results

Results are evaluated in terms of both topic coherence and the quality of the extracted perspectives.

5.1.1 Topic Coherence

The CV metric\(^8\) is used to measure the coherence of the topics generated by the models as it has been

\(^{3}\)https://hansard.parliament.uk

\(^{4}\)Period of time what selected on a basis of existen of a large number of major topics.

\(^{5}\)https://github.com/aghie/lam/blob/master/hcd.tsv

\(^{6}\)Percentiles were selected on an empirical basis.

\(^{7}\)We were not able to find a publicly available code of the **JTV** implementation.

\(^{8}\)https://github.com/AKSW/Palmetto/
shown to give the results closest to human evaluation compared to other topic coherence metrics (Röder et al., 2015). In brief, given a set of words, it gives an intuition of how likely those words co-occur compared to expected by chance.

Figure 2 plots the CV results\(^9\) versus the number of topics on HCD for various models. For each topic \(z\), we extract the top ten most representative words ranked by their respective normalised discriminative score defined by \(\text{DS}(w, z) = P(w|z)/[\max_{z'\neq z} P(w|z')]\). We chose this approach instead of simple \(P(w|z)\) as it was observed to turn into higher quality topics. It is clear that LAM_LEX models outperform baselines and that all variants are learning well the topics from the data, showing that the three different mechanisms for the switch variable are effective to generate coherent topics. Also, our models work robustly under different number of topics. Moreover, LAM_LEX achieve better coherence scores than the original LAM and LAM_POS. This shows that it is more effective to use POS tags and a subjectivity lexicon to initialise the Dirichlet prior for the word type switch variable rather than simply relying on POS tags or subjectivity lexica to give hard discrimination between topic and argument words.

![Figure 2: CV coherence vs the number of topics for different modeling approaches.](image)

In this section we evaluate the quality of the relation of the arguments with respect to their topics.

5.1.2 Perspectiven Summarisation

In terms of a quantitative evaluation, we are interested in knowing how strongly the perspectives are related to their topic: it might be the case that the separate CV coherence for the topic and viewpoints is high, but there is no actual relation between them, which would be an undesirable behaviour. To determine whether this is happening or not in the studied models, for each perspective we compute a mixed topic-perspective \(CV\) value, by extracting the top 5 perspective words, concatenating them with the top 5 words of the corresponding topic and running Palmetto as in the previous section.\(^{10}\) We then average the computed mixed topic-perspective \(CV\) values by \(T \times A\). Following this methodology, a high average \(CV\) value means that the perspective words are likely to occur when discussing about that particular topic, and therefore a test of whether the model is learning perspectives that have to do with it. Figure 3 compares topic-perspective models evaluated following this methodology, showing that LAM_LEX gives the best overall coherence.

For a better understanding of what perspec-

---

\(^9\)The CV results were calculated based on the top 10 words from each topic.

\(^{10}\)Palmetto does not accept more than 10 words.
Table 1: Example sentences, belonging to speeches that were assigned in Hansard different major topics labels, were clustered together by LAM (and it is sensible to do so as they are both about “farmers”).

Table 2: Ratio of topics where ≥ 0 or more than x out of top 10 topic sentences belong to the same major topic.

Table 3: Averaged LA measure across all topic-perspectives for different models.

Figure 3: Average mixed topic-perspective CV coherence, across different number of topics.

We can also rank sentences for an argument a under a topic z based on the generative probability of sentences. But this consistently produce worse results.

11We can also rank sentences for an argument a under a topic z based on the generative probability of sentences. But this consistently produce worse results.

To compare the quality of perspectives inferred by LAM_LEX and CPT (over 30 topics) we also conducted human evaluation. To do this, topics and perspectives were represented as bag-of-words. Each perspective was also represented with its three most representative sentences. The outputs from the two models was first merged
and shuffled. Two external annotators were then asked to answer (‘yes’ or ‘no’) for each topic if they could differentiate two perspectives. Cohen’s Kappa coefficient (Cohen, 1968) for inter-annotator agreement was 0.421. Table 4 shows the results of the evaluation and it is clear that LAM_Lex outperforms CPT.

Table 4: Accuracy on detecting perspectives according to the human outputs. In 1&2 a ‘yes’ answer is only valid if marked by both annotators.

<table>
<thead>
<tr>
<th>Annotator</th>
<th>LAM_Lex</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>0.67</td>
<td>0.34</td>
</tr>
<tr>
<td>1&amp;2</td>
<td>0.53</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 5 shows the three most representative perspective sentences for some of the extracted topics by LAM_Lex and CPT, to illustrate how LAM_Lex obtains more coherent sentences.

Table 5 shows the three most representative perspective sentences for some of the extracted topics by LAM_Lex and CPT, to illustrate how LAM_Lex obtains more coherent sentences. The example involving the first topic shows a case where LAM_Lex learned non-contrastive perspectives: both deal with Palestina, but focusing in different aspects (illegal settlements vs. Israel actions). In contrast, CPT mixed perspectives about Israel/Palestina and other viewpoints about GCSE and mortgages. In the second topic, LAM_Lex ranks at the top sentences relating to Sinn Fein & Northern Ireland, that show two different stances (positive vs negative) meanwhile in CPT it is not possible to infer any clear perspective despite sentences contain semantically related terms.

Table 6 shows cases where LAM_Lex obtained a less-coherent output according to the annotators. The first topic deals with Shaker Aamer and the legality of its imprisonment in Guantanamo. Perspective 2 reflects this issue, but Perspective 1 includes other types of crimes. The second example discusses issues relating to transports. While Perspective 1 is all about the negotiation with the train company, First Great Western, on its franchise extension proposal, Perspective 2 contains sentences relating to a number of different issues under transports. To alleviate this problem, we hypothesise that additional levels of information (in addition to the topic and perspective levels), such as a Bill or a speaker, might be needed to better distinguish different topics and perspectives that share a significant proportion of vocabulary.

5.1.3 Discussion

LAM_Lex gave a glimpse of the perspectives that occupy a topic. However, in many cases those differ from the initial expectation given the priors used in our model. Despite of the use of the subjectivity lexicon to initialise the Dirichlet prior $\beta^a$ for the topic-argument-word distribution, after a few iterations the initial distribution changes radically and turns instead into contrastive and non-contrastive perspectives, with the latter group being the most common one. We think this is due to factors that involve: (1) lack of contrastive speeches about very specific topics; and (2) jargon from the House of Commons that makes the task more challenging as stances are showed in subtle and polite way. This is also in line with what has been previously observed in (Mohammad et al., 2016b) that a person may express the same stance towards a target by using negative or positive language. This shows that LAM_Lex can infer perspectives from raw data, but we have little control on guiding the model on what perspectives to extract.

6 Conclusion and Future Work

We have presented LAM, a model able to provide a glimpse of what is going on in political debates, without relying on any labelled data and assuming the perspectives of a topic to be latent. It is implemented through a hierarchical Bayesian model considering that words can be separated as topic, argument or background words and follow different generative routes. Experiments show that our model obtains more coherent topics than related approaches and also extracts more interpretable perspectives. The code is made available at https://github.com/aghie/lam.

Although LAM can extract perspectives under a certain topic, there is little control in what kind of information to extract (e.g. we might want only contrastive or non-contrastive arguments). In future work, we plan to improve the model through complex priors or semantic similarity strategies. Also, adding a ‘Bill’ level could be beneficial as speeches about the same Bill should share the same high-level topic. But we need labels indicating to which Bill the text belongs to. Including a ‘speaker’ level to know which parliamentarians discuss which topics is another interesting path to follow.

---

12The examples were identified as two perspectives by at least one annotator. Its selection was made based on an existence of a similar topic both on LAM_Lex and CPT outputs.
## Table 5: Output sample for representative perspective sentences in non-contrastive and contrastive topics.

<table>
<thead>
<tr>
<th>LAM_LEX</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1</td>
<td>israel, iran, syria, settlement, relocation, counter-terrorism gaza, ipms, airline, metropolitan, passenger, franchise, fare, coast, connectivity, journey, gloucester, user, anglia, stagecoach</td>
</tr>
<tr>
<td>Perspective 1</td>
<td>it is contrary to international law in that sense, and any nation has obligations when dealing with occupied territories and their occupants.</td>
</tr>
<tr>
<td>Perspective 2</td>
<td>a) I do not want any young people to feel frightened of attending school or of their journey to and from school, and, sadly, that applies to the illegal settlements that have been put forward, but nevertheless we are concerned and are having a dialogue with Israel about that.</td>
</tr>
</tbody>
</table>

## Table 6: Output sample for non-representative perspective sentences in the LAM_LEX model.

<table>
<thead>
<tr>
<th>LAM_LEX</th>
<th>CPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1</td>
<td>aamer, shaker, bay, guantanamo, america, obama, american, timetable, embassy, harlington</td>
</tr>
<tr>
<td>Perspective 1</td>
<td>a) It is a slightly separate debate or concern if I can put it that way so the illegal settlements that have been put forward, but nevertheless we are concerned and are having a dialogue with Israel about that.</td>
</tr>
<tr>
<td>Perspective 2</td>
<td>a) I share the hon. Lady's desire that every school should offer three separate sciences at GCSE; that is very important.</td>
</tr>
</tbody>
</table>

## Acknowledgments

We thank Charles Marshall for crawling the HCD data. DV was funded by MECID (FP13/01180), MINECO (FFI2014-51978-C2-2-R) and Inditex-UDC grants for research stays. YH is partly funded by the Natural Science Foundation of China (61528302).
References


Jacob Cohen. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin, 70(4):213.


