

# Incorporating Topic Aspects for Online Comment Convincingness Evaluation

Yunfan Gu<sup>1</sup>, Zhongyu Wei<sup>1\*</sup>, Maoran Xu<sup>1</sup>, Hao Fu<sup>1</sup>,  
Yang Liu<sup>2</sup>, Xuanjing Huang<sup>3</sup>

<sup>1</sup>School of Data Science, Fudan University, China

<sup>2</sup>Liulishuo Company

<sup>3</sup>School of Computer Science, Fudan University, China

{17210980015, zywei, 14300180099, 14307130013}@fudan.edu.cn,  
yang.liu@liulishuo.com, xjhuang@fudan.edu.cn

## Abstract

In this paper, we propose to incorporate topic aspects information for online comments convincingness evaluation. Our model makes use of graph convolutional network to utilize implicit topic information within a discussion thread to assist the evaluation of convincingness of each single comment. In order to test the effectiveness of our proposed model, we annotate topic information on top of a public dataset for argument convincingness evaluation. Experimental results show that topic information is able to improve the performance for convincingness evaluation. We also make a move to detect topic aspects automatically.

## 1 Introduction

With the popularity of online forums such as *idebate*<sup>1</sup> and *convinceme*<sup>2</sup>, researchers have been paying increasing attention to analyzing persuasion content (Wei et al., 2016a,b). Argument convincingness assessment plays an important role in persuasion content analysis. Previous researchers attribute the convincingness of arguments to argument structure (Potash et al., 2017; Peldszus and Stede, 2015), strong evidence (Hasan and Ng, 2014; Park and Cardie, 2014), specific argument components (Habernal and Gurevych, 2016a; Persing and Ng, 2015), interactions (Ji et al., 2018), domain knowledge (Wei et al., 2017) and so on. Most efforts of convincingness evaluation focus on using explicit linguistic features, such as words (Chalaguine and Schulz, 2017) and part-of-speech (POS) (Wachsmuth et al., 2017a) etc. Considering people construct their arguments based on different topic aspects, we thus argue that topic information can be crucial for convincingness evaluation.

\*Corresponding author

<sup>1</sup><https://idebate.org/>

<sup>2</sup><http://convinceme.net/>

	Argument1	Argument2
Content	The American Water companies are Aquafina (Pepsi), Dasani (Coke), Perrier (Nestle) which provide jobs for the american citizens.	If bottled water did not exist, more people would be drinking sweetened liquids because it would be the only portable drinks! People would become fat!
Topic Aspect	Economy	Convenience and health

Table 1: Example of an argument pair and corresponding topic aspects. **Debate:** Ban plastic water bottles **Stance:** No; **Label:** A1>A2

To illustrate this idea, Table 1 gives a brief example of an argument pair. Both arguments express opinions against the banning of plastic water bottles. *Argument 1* is expressed from the topic aspect of *economy* while *Argument 2* makes the point from the aspect of *convenience and health*. As we can see, for a specific discussion subject, different aspects might reveal various degree of convinceness. Wang et al. (2017) has already made attempt to make use of latent persuasive strengths of topic aspects for quality evaluation on a formal debate dataset. However, there is still no further research on online debating texts, which is un-structured with multiple participants.

In this paper, we propose to incorporate latent topic aspects information to evaluate the convincingness of comments in online forum. We make use of graph convolutional networks (GCN) to utilize the latent topic information of comments for a specific subject. We assume that arguments sharing the same topic aspect are more likely to have similar degree of convincingness, and GCN is able to make use of the topic similarity among arguments. Bi-directional long short-term memory (Bi-LSTM) is used to encode each argument. We annotate topic aspects information on top of a public dataset collected from online forum Habernal and Gurevych (2016b) to evaluate our proposed model.

Main contribution of this article are three folds:

(1) we annotate topic aspects for each argument in an existing dataset over 16 discussion threads (2 stances for each subject); (2) we propose a BiLSTM-GCN model and prove the effectiveness of topic aspects in convincingness evaluation; (3) we implement several baseline models to detect the topic aspect automatically.

## 2 Data Description

Our experiments are conducted on *UKPConvArg1* corpus released by Habernal and Gurevych (2016b). The source argument is from 16 debates on Web debate platforms [createdebate.com](http://createdebate.com) and [convinceme.net](http://convinceme.net). Each debate is about a specific topic and has two stances. We call each (debate, stance) tuple a "split", so there are 32 splits in total. The dataset includes sets of argument pairs according to 32 splits, and each argument pair is annotated with a binary relation (0 and 1) which represents the pairwise convincingness relationship (more/less convincing), 11,650 in total. Since we take the *UKPConvArg1Strict* version as our dataset, the equal instances are removed. The topic aspect annotations are not from the initial dataset, but from our own annotations.

In order to extract topic aspects from each topic, we manually annotate each argument by two annotators. The annotation process is as follows. First, two annotators have a discussion and then determine possible topic aspect candidates for each split. Second, two annotators independently check every argument and summarize one main topic aspect. As for arguments carrying multiple aspects, we pick the primary topic aspect. Third, after all the annotations are made, they are asked to rank the topic aspects under a split according to topic aspect strength by discussion. Due to the quality of the corpus, some arguments have to be assigned to the aspect None if it (1) has nothing to do with the topic, or (2) has no point of view, or (3) is contradictory/ambiguous. Results of topic aspect annotations are shown in Appendix.

To clarify the annotations of topic aspects, we will take the annotation process for comments under the topic of "banning plastic water bottles" as an example. Comments from the con-side of the this debating might hold different topic aspects. Some of them concern about the economic decay after banning plastic water bottles. And some of them suggest that we can recycle plastic water bottles instead of banning them. The others care

more about the inconvenience and safety after banning plastic water bottles. After reading all these arguments, annotators conclude that there should be three main topic aspects, "economics", "bottle recycling" and "convenience and health" respectively. There are some arguments which have no point of view or seem ambiguous, and the topic aspects of these arguments will be set as "None". Since the average length of arguments is relatively short, so most of the arguments hold a single topic aspect. Therefore, each argument only has one label to simplify the problem. The dataset is available here<sup>3</sup>.

## 3 Proposed Model

In this paper, we propose a BiLSTM-GCN model to solve the convincingness evaluation task. The BiLSTM acts as the foundation to generate the representation of each argument, and GCN is built upon BiLSTM to make use of the inter-relationships of similar arguments. The architecture of our model is shown in Figure 1. In general, Our aim is predicting a binary relation (0 and 1) representing more/less convincing given an argument pair. All the arguments in the same split are considered as a batch.

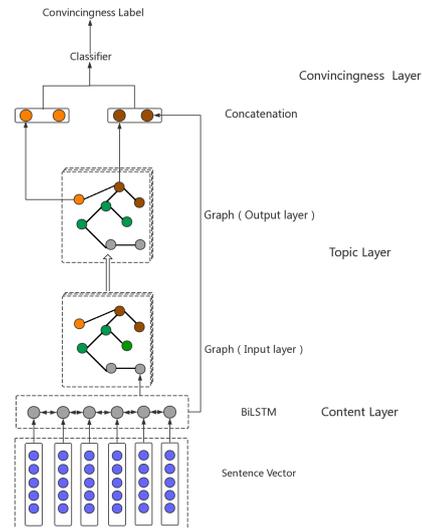


Figure 1: BiLSTM-GCN model architecture

### 3.1 Content Layer

The input of the content layer is the word embedding matrix of each argument and the output of

<sup>3</sup><http://www.sdspeople.fudan.edu.cn/zywei/data/5thargmine-topic-convince.zip>

content layer is the representation vector of each argument. BiLSTM plays the role of encoder in this layer, and it has been proved effective to encode sentences (Goodfellow et al., 2016; Dyer et al., 2015; Wang and Jiang, 2016).

We simply employ the word embeddings released by Glove (Pennington et al., 2014), and we choose the 840B tokens, 300 dimensional vector version. As to the word which is absent in the Glove vocabulary, we randomly generate a 300 dimensional vector to represent those out of vocabulary vectors. These vectors are then put into BiLSTM to get the basic representation of each argument. The dimension of argument representation is set to 64 after tuning.

### 3.2 Topic Layer

The input of topic layer is the representation vector of each argument. The output of topic layer is the updated representation vector of each argument. The core of our topic layer is GCN (Kipf and Welling). The main function of this layer is utilizing the topic aspect information.

We consider a single-layer GCN for better argument representation. The GCN layer propagates the information of a node onto its nearest neighbours.

Our GCN model simply takes the following form.  $A$  represents the adjacency matrix,  $X$  represents the feature matrix, which is a stacked version of raw argument representations generated from BiLSTM.  $W$  is a parameter matrix which can be trained in the training process. We add self connected edges to  $A$  to keep the information of the original argument itself.  $\hat{A}$  is the normalized form of  $A$ . Normalization is a compulsory part for combining information, since we have to keep the ratio of self information and neighbours' information the same for each argument.  $Z$  is the final feature matrix, and each row represents the new representation of each argument.

$$Z = f(X, A) = Relu(\hat{A}XW) \quad (1)$$

The normalization process is described below:

$$\hat{A} = D^{-\frac{1}{2}}(A + I_N)D^{\frac{1}{2}} \quad (2)$$

Here,  $I_N$  is the identity matrix, which represents the self-connections.  $D_{ii} = \sum_j A_{ij}$  is a diagonal matrix and each element on the diagonal represents the degree of argument  $i$ . The self-connections are not normalized since we think that

original argument is more useful than other arguments.

#### 3.2.1 Adjacency Matrix

The adjacency matrix represents whether there is an edge between two nodes, but the graph structure among the arguments is implicit. We can capture the implicit structure by making use of argument similarity or topic aspect.

**Argument similarity:** We can calculate the similarity of two arguments by distance under embedding setting and use threshold to decide whether there is an edge.

**Topic aspects:** When two arguments share the same topic aspect, they are supposed to have an edge between corresponding nodes.

#### 3.2.2 Feature Matrix

Feature matrix is built upon BiLSTM, and is a stacked version of argument representation. The argument representation is the mean pooling of the BiLSTM result. In fact, the result of graph convolutional network is still a feature matrix which absorbs information from neighbours. The feature dimension is set to 64, and the dimension of each matrix is fixed over 32 splits by setting the maximum argument quantities as the row dimension. Default arguments are filled by zero vectors. The result of graph convolutional network is still a feature matrix, which involves related nodes' information.

### 3.3 Convincingness Layer

The input of convincingness layer is the updated representation vector of each argument. The output of convincingness layer is a binary label (0 and 1) representing more/less convincing given an argument pair. As a result the core of this convincingness layer is a classifier.

We can simply apply the softmax classifier. However, inspired by DistMult (Yang et al., 2014), we design a classifier which will perform better. Our classifier takes the following form.  $e_s$  represents the representation of argument 1, and  $e_o$  represents the representation of argument 2, and  $W$  is a parameter matrix. We set the parameter matrix as a real antisymmetric matrix. Therefore, the result of comparing argument 1 and argument 2 will be opposite to the result of comparing argument 2 and argument 1.

$$f(s, r, o) = e_s^T R e_o \quad (3)$$

Our parameter matrix is better than the normal linear layer mainly because it can capture the interactive relationship between different feature dimensions.

## 4 Experiments

### 4.1 Experiment Setup

We test our model on the dataset depicted in Section 2 to evaluate convincingness of arguments. To compare with the algorithms applied in the initial task (Habernal and Gurevych, 2016b), we still use 32-fold cross-split cross-validation, which means 31 splits are training data and the other one is test data. The preprocessing part is the same as the original task as well. And we train our BiLSTM-GCN model as described in Section 3 and evaluate prediction accuracy on the test split.

In this paper, we implement our BiLSTM-GCN model by Pytorch. Each split is considered as a batch to train and test. The loss function we use is simply the quadratic loss function. We have tried the cross entropy loss and the quadratic loss, and latter performs better when using our classifier. The batch loss is calculated by summing the loss of each argument pair. The weights of the parameter matrix classifier are initialized randomly from the normal distribution, and the initial hidden state of our BiLSTM is set to zero. And we take topic aspects to build adjacency matrix in convincingness evaluation task.

### 4.2 Baselines

The baselines for convincingness evaluation include: (1) **SVM** (Habernal and Gurevych, 2016b): The SVM with RBF kernel is based on a number of linguistic features. (2) **BiLSTM** (Habernal and Gurevych, 2016b): The input word embedding is as depicted in Section 3.1, and hidden dimension is 64. (3) **BiLSTM+argument length**: Since BiLSTM will transform all the arguments into same dimension, the information of argument length will be lost to some extent. Here we mention argument length since it can handle this task quite well. The accuracy of judging convincingness by its length can be 0.73, which is not mentioned in the original work. As a result, we take it as a useful feature to help our model work better.

Here, we don't list the baseline with our topic aspects annotations because the topic aspects among different debates are not the same, so it can't play the role of an explicit feature. We have

to encode the topic aspect information, that is also a reason of applying Graph Convolutional Network in our work.

We also try different methods to build the adjacency matrix as depicted in Section 3.2.1, including unit matrix, Jaccard similarity, cosine similarity, word mover's distance and topic aspects. (1) **Unit matrix**: Unit matrix means that the adjacency matrix is just a unit matrix, and the GCN here act as a single linear layer. (2) **Jaccard similarity**: Jaccard similarity is calculated by the following form. A and B means the argument 1 and argument 2 in an argument pair. In our experiments we find that use threshold to build adjacency is better than use weighted adjacency matrix. As a result, we all use the threshold to build a zero-one adjacency matrix in the following building methods. The threshold of Jaccard similarity is set to 0.5 after normalizing.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

(3) **Cosine similarity**: Cosine similarity is represented using a dot product and magnitude as the following form. The threshold of cosine similarity is set to 0.95 after normalizing.

$$C(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (5)$$

(4) **Word mover's distance** (Kusner et al., 2015): The word mover's distance calculating is reproduced by reading the original paper. And the threshold is set to 0.35 after adjusting.

### 4.3 Experimental Results

This part, we compare our BiLSTM-GCN model with baselines mentioned above. Table 2 lists the results of convincingness evaluation task. The adjacency matrix of our BiLSTM-GCN model in Table 2 is based on our topic aspects annotations. In Table 3, We test other adjacency matrix building methods as described above and analyze the results.

The result shows that our BiLSTM-GCN model performs better than best baseline model, and obviously better than models utilized in the initial task. What's more, we have proved that the interrelationships of arguments can help us evaluate the convincingness better by using GCN.

The results in Table 3 show that our annotations perform the best among all the metrics, which

Model	Accuracy
SVM	0.78
BiLSTM	0.76
BiLSTM+argument length	0.797
BiLSTM-GCN+argument length (topic aspects)	<b>0.811</b>

Table 2: Results of convincingness evaluation task

Adjacency metric	Accuracy
Unit matrix	0.800
Jaccard similarity	0.799
Cosine similarity	0.793
Word mover’s distance	0.808
Topic aspects	<b>0.811</b>

Table 3: Results of BiLSTM-GCN model with various adjacency matrix building methods

means topic aspect is an excellent way to evaluate the relationship between arguments. And we can know that some state-of-the-art text similarity metric like word mover’s distance performs better than classical text similarity metrics like Jaccard similarity and cosine similarity.

#### 4.4 Further analysis of topic detection

We know that the topic aspects are not labeled in most data. Since we already have the annotations of topic aspects, so we can set a classification task, which can be further used for automatic annotation. The training data and test data are the same as convincingness evaluation. However, the labels will change. Here, if two arguments in an argument pair share the same topic aspect, the label will be set to 1, or it will be set to 0. We also suggest that None type is different from all other types including None type itself. We test some baseline models and our BiLSTM-GCN model on this task and evaluate F-score on the test split. We don’t use accuracy since the labels are imbalanced here, only about thirty percent of argument pairs have positive labels.

The baselines for topic aspect detection include: (1) **Random classification**: Select zero or one randomly. (2) **LDA clustering** (Blei et al., 2003): Automatically cluster the arguments by LDA over each split into the number of topic aspects we annotate. (3) **SVM**: The SVM with RBF kernel is based on a number of linguistic features.

We choose word mover’s distance to build adjacency matrix instead of our annotations, since it

is the kind of the latest metric of text similarity calculation. The word mover’s distance calculation is reproduced by reading the original paper (Kusner et al., 2015). The threshold is set to 0.35 after adjusting. Table 4 lists the results of topic aspect detection task.

Model	F-score
Random classification	0.425
LDA clustering	0.524
SVM	0.589
BiLSTM-GCN	<b>0.612</b>

Table 4: Results of topic aspect detection task

The result shows that our BiLSTM-GCN model performs the best among all the methods, and supervised training methods like SVM performs better than unsupervised methods like LDA clustering. All these models perform significantly better than the lower bound given by random classification.

## 5 Conclusions and Future works

In this paper, we apply BiLSTM-GCN model on a convincingness evaluation task and the model performs 3-5% better than the original method on the online debate dataset. Our model utilizes not only the explicit argument features like length but also topic aspects which are latent features. Our experiments proves that topic information is able to improve the performance for convincingness evaluation. In future, we will consider to utilize external knowledge to further improve the performance of convincingness evaluation. The possible external knowledge can be similar arguments from other websites, or argument search engine (Wachsmuth et al., 2017b).

## Acknowledgements

The work is partially supported by National Natural Science Foundation of China (Grant No. 61702106), Shanghai Science and Technology Commission (Grant No. 17JC1420200, Grant No. 17YF1427600 and Grant No.16JC1420401).

## References

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

- Lisa Andreevna Chalaguine and Claudia Schulz. 2017. Assessing convincingness of arguments in online debates with limited number of features. In *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 75–83.
- Chris Dyer, Miguel Ballesteros, Wang Ling, Austin Matthews, and Noah A Smith. 2015. Transition-based dependency parsing with stack long short-term memory. *arXiv preprint arXiv:1505.08075*.
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*, volume 1. MIT press Cambridge.
- Ivan Habernal and Iryna Gurevych. 2016a. What makes a convincing argument? empirical analysis and detecting attributes of convincingness in web argumentation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1214–1223. Association for Computational Linguistics.
- Ivan Habernal and Iryna Gurevych. 2016b. Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional LSTM. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1589–1599. Association for Computational Linguistics.
- Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 751–762. Association for Computational Linguistics.
- Lu Ji, Zhongyu Wei, Xiangkun Hu, Yang Liu, Qi Zhang, and Xuanjing Huang. 2018. Incorporating argument-level interactions for persuasion comments evaluation using co-attention model. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3703–3714. Association for Computational Linguistics.
- Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *arXiv:1609.02907*.
- Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. 2015. From word embeddings to document distances. In *International Conference on Machine Learning*, pages 957–966.
- Joonsuk Park and Claire Cardie. 2014. Identifying appropriate support for propositions in online user comments. In *Proceedings of the First Workshop on Argumentation Mining*, pages 29–38. Association for Computational Linguistics.
- Andreas Peldszus and Manfred Stede. 2015. Joint prediction in MST-style discourse parsing for argumentation mining. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 938–948. Association for Computational Linguistics.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. Association for Computational Linguistics.
- Isaac Persing and Vincent Ng. 2015. Modeling argument strength in student essays. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 543–552. Association for Computational Linguistics.
- Peter Potash, Alexey Romanov, and Anna Rumshisky. 2017. Here’s my point: Joint pointer architecture for argument mining. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1364–1373. Association for Computational Linguistics.
- Henning Wachsmuth, Nona Naderi, Ivan Habernal, Yufang Hou, Graeme Hirst, Iryna Gurevych, and Benno Stein. 2017a. Argumentation quality assessment: Theory vs. practice. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 250–255. Association for Computational Linguistics.
- Henning Wachsmuth, Martin Potthast, Khalid Al Khatib, Yamen Ajjour, Jana Puschmann, Jiani Qu, Jonas Dorsch, Viorel Morari, Janek Bevendorff, and Benno Stein. 2017b. Building an argument search engine for the web. In *Proceedings of the 4th Workshop on Argument Mining*, pages 49–59. Association for Computational Linguistics.
- Lu Wang, Nick Beauchamp, Sarah Shugars, and Kechen Qin. 2017. Winning on the merits: The joint effects of content and style on debate outcomes. *Transactions of the Association of Computational Linguistics*, 5:219–232.
- Shuohang Wang and Jing Jiang. 2016. Learning natural language inference with LSTM. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1442–1451. Association for Computational Linguistics.
- Zhongyu Wei, Chen Li, and Yang Liu. 2017. A joint framework for argumentative text analysis incorporating domain knowledge. In *arXiv preprint arXiv:1701.05343*.
- Zhongyu Wei, Yang Liu, and Yi Li. 2016a. Is this post persuasive? ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 195–200.

Zhongyu Wei, Yandi Xia, Chen Li, Yang Liu, Zachary Stallbohm, Yi Li, and Yang Jin. 2016b. A preliminary study of disputation behavior in online debating forum. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 166–171.

Bishan Yang, Wen Tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. In *arXiv:1412.6575*.

## A Topic aspect annotations

Debate	Stance	Aspects
BAN	No	[A1] Economy; [A2] Bottle recycling; [A3] Convenience and health; [A4] None
	Yes	[A1] Environment; [A2] High price of bottled water; [A3] Health; [A4] None
CHR	Atheism	[A1] Challenging Christianity/God; [A2] Atheism is more scientific; [A3] None
	Christianity	[A1] Faith and belief; [A2] Attacking Atheism; [A3] None
EVO	Creation	[A1] Paradox of evolution; [A2] Religion; [A3] None
	Evolution	[A1] Evidence of evolution; [A2] Challenging Creation/God; [A3] None
FIR	Firefox	[A1] Extensions; [A2] Poor performance of IE; [A3] IE copies other browsers; [A4] None
	IE	[A1] Universality; [A2] Useful functions; [A3] Poor performance of Firefox; [A4] None
GAY	Right	[A1] Gay marriage is ones right; [A2] It is none of others business; [A3] None
	Wrong	[A1] Gay marriage is unnatural; [A2] Religion; [A3] Ethics; [A4] Law
SPA	No	[A1] Spanking makes children behave; [A2] Only last resort
	Yes	[A1] Parents can use other ways; [A2] It will hurt children; [A3] Children will think hitting others is alright; [A4] Bad for relationship; [A5] Children are too young to know what is right; [A5] None
SPO	No	[A1] Love/relationship; [A2] It depends on what happened; [A3] People make mistakes; [A4] None
	Yes	[A1] Murder is wrong; [A2] Self security; [A3] Unwillingness to live with a murderer
IND	No	[A1] Politics; [A2] Economy; [A3] Diplomacy; [A4] Population; [A5] None
	Yes	[A1] Economy; [A2] Education and talents; [A3] Politics; [A4] Culture; [A5] Patriotism; [A6] None
LOU	Fatherless	[A1] Lousy father is harmful to kids growth; [A2] Lousy father is a stain ; [A3] Fatherless kids can still be fine
	Lousy Father	[A1] Lousy father still helps; [A2] Lousy father can get better; [A3]Lousy father motivates one to be better; [A4] None
POR	No	[A1] Sexual desires; [A2] Entertainment; [A3] Porn does no harm; [A4] None
	Yes	[A1] Health and addiction; [A2] Porn does no good; [A3] Women rights; [A4] Religion; [A5] None
SCH	Bad	[A1] Creativity; [A2] Futility; [A3]Low quality; [A4] None
	Good	[A1] Equality; [A2] Uniforms make students concentrate on study; [A3] Logo of school; [A4] None
PER	Common Good	[A1] Helping others is better; [A2] It also helps oneself; [A3] None
	Personal Pursuit	[A1] Helping oneself is prerequisite; [A2] Human nature; [A3] Personal fulfillment
PRO	Pro-Choice	[A1] Right of decision; [A2] Unwanted pregnancy; [A3] None
	Pro-Life	[A1] No right to kill; [A2] Adoption;[A3] None
PHY	No	[A1] PE is useless; [A2] Studies; [A3]PE is tiring; [A4] PE causes bullying; [A5] None
	Yes	[A1] Health and obesity; [A2] Students can develop good habits, attitudes, etc.; [A3] None
TVB	Books	[A1] Books are better for learning; [A2] TV does harms to health; [A3] Books save money; [A4] None
	TV	[A1] TV provides richer information; [A2] Convenience; [A3] None
WIL	No	[A1] Raffles contributions; [A2] Farquhar was a subordinate
	Yes	[A1] Farquhar solved problems; [A2] Reputation; [A3] None

Table 5: (1) BAN: Banning plastic water bottles; (2) CHR: Christianity or Atheism; (3) EVO: Evolution vs. Creation; (4) FIR: Firefox vs. Internet Explorer; (5) GAY: Gay marriage: Right or wrong; (6) SPA: Should parents use spanking as an option to discipline? (7) SPO: If your spouse committed murder and he or she confided in you, would you turn them in? (8) IND: India has the potential to lead the world; (9) LOU: Is it better to have a lousy father or to be fatherless? (10) POR: Is porn wrong? (11) SCH: Is the school uniform a good or bad idea? (12) PER: Which type of endeavor is better, a personal pursuit or advancing the common good? (13) PRO: Pro-Choice vs. Pro-Life; (14) PHY: Should physical education be mandatory in schools? (15) TVB: TV is better than Books; (16) WIL: William Farquhar ought to be honored as the rightful founder of Singapore.