TEC: A Time Evolving Contextual Graph Model for Speaker State Analysis in Political Debates

Ramit Sawhney¹*, Shivam Agarwal²*, Arnav Wadhwa¹ and Rajiv Ratn Shah¹

¹IIIT Delhi
²Manipal Institute of Technology
{ramits, rajivrtn}@iiitd.ac.in, shivamag99@gmail.com

Abstract
Political discourses provide a forum for representatives to express their opinions and contribute towards policy making. Analyzing these discussions is crucial for recognizing possible delegates and making better voting choices in an independent nation. A politician’s vote on a proposition is usually associated with their past discourses and impacted by cohesion forces in political parties. We focus on predicting a speaker’s vote on a bill by augmenting linguistic models with temporal and cohesion contexts. We propose TEC, a time evolving graph based model that jointly employs links between motions, speakers, and temporal politician states. TEC outperforms competitive models, illustrating the benefit of temporal and contextual signals for predicting a politician’s stance.

1 Introduction
Parliamentary debates discuss affairs affecting the future development of a country, such as budget revisions and policy reforms. The outcome of such debates entail wide-ranging consequences; for instance, analysts show that withdrawal of the UK from the European Union without a free trade agreement can cost UK’s GDP £140 billion within 10 years of the agreement [Ries et al., 2017]. Recent studies [Davoodi et al., 2020] show that analyzing the records of such debates provide information regarding viewpoints of politicians over critical societal factors. These records also aid assessing political candidates and basing voting decisions [Utych, 2019]. However, the esoteric language and opaque parliamentary jargon form a barrier to ordinary citizen’s insights into standings of politicians. Recent advances in computational social sciences [Abercrombie, 2018a] have made progress in assessing political stance by using language features of voluminous parliamentary debates. However, analyzing the linguistic traits of a speech is not sufficient to accurately predict a politicians stance on a motion [Bhavan et al., 2019]. Political debates involves multiple contextual elements beyond language, such as party affiliations, topic of debates [Van Dijk, 1977] and historic debates [van Dijk, 2004]. Consider Figure 1, where we present speech transcripts of two speakers who belong to the Conservative party at two different times. We observe that in the first two transcripts, T1 and T2, the speakers support each other and express similar outlooks on the motion. Such partisan forces [Owens, 2003] within political parties is indicative of intra-party contexts. Next, amongst transcripts T1 and T3, we observe remarkable similarity between two different speeches from the same speaker, debating over two distinct motions, almost four years apart. Such temporal similarities [Curato, 2012] reflect unique characteristics of a speaker, which is reflected across their speeches and plays an essential role for modeling the temporal state of a speaker. We also note that the debate’s motion governs the overall content of the speech, indicating the presence of motion-level contexts in debates. Identifying such similarities in the political ecosystem unfolds the possibility to learn latent patterns among speakers, their political affiliations, and how they target various motions.

*Equal contribution.
Building on existing work [Sawhney et al., 2020] and political theories [van Dijk, 2004; Van Dijk, 1977], we propose TEC: **Time Evolving Contextual Graph model**, the first dynamic-graph based model for analysing parliamentary debates. Our model enhances prior solutions for political stance analysis [Bhavan et al., 2020; Abercrombie, 2020] by extending the political cohesion based model [Sawhney et al., 2020] to incorporate speaker’s temporal states (§3.2, §3.4). We propose a dynamic graph attention networks to jointly learn from political cohesion (§3.3) and the temporal states of a speaker (§3.5). We demonstrate practical applicability through qualitative analysis (§5.5, §5.6). Our contributions are:

1. We identify three types of debate contexts: intra-party, motion-level, and temporal contexts in parliamentary discourse and represent these contexts via dynamic graphs.
2. We propose the first dynamic-graph attention network for jointly learning from language, cohesion context and temporal speaker states in parliamentary debates.
3. Through experiments on more than 30K parliamentary speeches of the UK House of Commons spanning over 20 years, we demonstrate TEC’s applicability to provide non-expert citizens insights into political ideologies.

## 2 Related Work

### Politics and Linguistics

Analyzing political data acts as a knowledge source that provides insights into cohesion within political parties, stances of politicians towards critical motions for both the general public and across domains including humanities and linguistics [Slémbrouck, 1992]. Research at the intersection of Politics and Linguistics spans agreement detection [Duthie and Budzynska, 2018], topic-opinion analysis [Abercrombie, 2018b] and, debate stance classification [Bhavan et al., 2020]. Existing work focuses on these tasks via legislative speeches from the US Congress [Chen et al., 2017], and EU Parliament [Frid-Nielsen, 2018].

### Political Stance and Sentiment Analysis

NLP has seen a growth in analyzing and mining opinions from political discourse [Abercrombie, 2018a]. Recent approaches [Abercrombie, 2020] show the ability of transformers such as BERT in capturing domain-specific jargon better for debate feature extraction. A promising new direction at the intersection of Politics and NLP is temporal-speaker state modeling and inclusion of context such as engagement in social circles. Recent works [Bhavan et al., 2020] have shown the presence of herd mentality in political stances through shallow graph embeddings by identifying the linguistic similarities between members of the same party. [Sawhney et al., 2020] uses a static graph to encode the cohesion contexts in parties for stance prediction. Despite their success, a common limitation is that they do not consider the dynamic nature of debates and the temporal evolution of speaker states.

## 3 Time Evolving Contextual Graph Model

### Problem Formulation

For a debate $d_t \in D = \{d_1, d_2, \ldots, d_{|D|}\}$ at time $t$, we denote the transcript as $r_{ij} \in T$ corresponding to the speech made by a MP (speaker) $s_i \in S = \{s_1, s_2, \ldots, s_{|S|}\}$ on one specific motion $m_j \in M = \{m_1, m_2, \ldots, m_{|M|}\}$. Where, $T$ is a set of transcripts, and $\cdot$ denotes the cardinality of the set. Each speaker $s_i$ is affiliated to only one political party $p \in \{p_1, p_2, \ldots, p_{|P|}\}$. Given parliamentary debates for a historical lookback window of $L$ (i.e., $[t - L, t - 1]$), the task is to classify the stance $Y \in \{‘Aye’, ‘No’\}$ of the MP $s_i$ on the motion $m_j$ based on the transcript $r_{ij}$ on day $t$. ‘Aye’ and ‘No’ denote positive and negative stance, respectively.

TEC consists of two integral components: Cohesion Context Encoder and Temporal Context Encoder to model intra-party and motion contexts across a time-series of parliamentary debates shown in Figure 2.

### 3.1 Encoding Parliamentary Texts via BERT

A debate comprises of Motions, i.e., expressions over policy positions taken by the government, MPs, etc. The mo-
tion is followed by responses from other MPs in the form of
Speeches. We adopt Bidirectional Encoder Representations
from Transformers (BERT) [Devlin et al., 2019] to encode
debate transcripts r and motion descriptions m. The domain-
specific nature of political speeches and political jargon moti-
vate us to fine-tune pre-trained BERT [Sawhney et al., 2020].

We fine-tune BERT on the ParlVote dataset [Abercrombie,
2020], a corpus of debates from the UK Parliament’s House
of Commons. For each transcript r and motion m, we
obtain feature vectors hr, hm ∈ R F , where, F = 768 obtained
as the output of the [CLS] token from the final BERT layer.

### 3.2 Cohesion Contexts in Debates

Decision-makers are subtly influenced by the environment
around them [Bode et al., 2014]. Following [Sawhney et al.,
2020] we identify two types of cohesion contexts in para-
liamentary debates— intra-party context, and motion context.
The first, Intra-Party Context, captures the influence of
the same political affiliations and fellow party members over a
speaker’s speeches. During debates, some speakers tend to
express their raw individual opinions, while some may ex-
hibit homophily: the likeliness of associated individuals to
adopt similar viewpoints [Boucek, 2002]. Next, we present
a Motion Context, that captures the relationship between a
speaker’s speech and the motion of the debate.

For each debate d, we represent the relations between
speakers and motions in the form of a graph G = (V, E).
V and E represent the nodes and edges in the graph. The
nodes V consists of the set of speakers S and motions M
in the debate d. The edges E are of two types: Speaker-Speaker
edges Eds based on intra-party context, and Speaker-Motion
edges Edm based on motion context. G is a heterogeneous
graph as it has different types of nodes and edges. We now
describe the two relations that capture different contexts.

**Intra-party Context** models the relationship between a
MP and the party they belong to. We build on the hypoth-
thesis that speakers are influenced by other party members, and
there exists a partisan mentality like political cohesion within
parties [Owens, 2003]. We represent intra-party context by
a Speaker-Speaker edge Eds between two MPs si, sj ∈ S if
both si and sj are affiliated to the same political party p ∈ P.

**Motion-level Context** encodes the relation between a
speaker’s speech and the motion. In debates, a speech is based
on the current motion of discussion. We represent motion
context via Speaker-Motion edge Edm between a speaker s
and a motion m corresponding to that debate transcript.

We show the statistics of the graph G in Table 1.

### 3.3 Cohesion Context Encoder

To capture the cohesion context among speakers and the motions
they speak on, we use the cohesion context encoder over the
constructed graph G for each debate d. The cohesion context
encoder is a graph neural network that propagates information
between different speakers and motions. Following
[Sawhney et al., 2020] we use we use Graph Attention Net-
works (GATs) [Velickovic et al., 2018], as each context has a
different degree of influence on a speaker’s speech.

Here, we define a GAT layer that is applied to one de-
bate d. The input to this GAT is a set of node features
h — {x1, x2, . . . , x|V|} For each speaker, and motion m,
we set the node features to hr and hm extracted from the
speech and motion transcript using BERT, respectively.
The node features are transformed to context dependent features,
h′ — {q1, q2, . . . , q|V|} based on the influence by its neighbors.
We first apply a shared linear transform parameterized by W
to all the nodes. Then, we apply a shared self-attention mech-
anism to each node i in its neighborhood Ni. For each node
j ∈ Ni, we compute attention coefficients αij which shows the
importance of context between nodes i and j. Formally,
the attention weights αij is given as:

\[
α_{ij} = \frac{\exp(\text{LeakyReLU}(a_{ij}^T [W x_i \oplus W x_j]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(a_{ik}^T [W x_i \oplus W x_k]))}
\]  

where, ′ ⊕ represent transposition and concatenation, re-
spectively and a_w is the parameter matrix of a fully con-
ected layer. We use multi-head attention to stabilize training.
Formally, U independent executors apply the above attention
mechanism. Their outputs are concatenated to yield:

\[
q_i = \bigoplus_{k=1}^{U} \text{LeakyReLU} \left( \sum_{j \in N_i} \alpha_{ij}^k W^k x_j \right)
\]  

where α_{ik}^k and W^k denote the attention coefficients and linear
transform weight matrix computed by the kth attention head.

### 3.4 Temporal Speaker State Modelling

Parliamentary debates occur over time, and historic debates
often set a stage for upcoming discourses [Curato, 2012].
Furthermore, the sentiment of some MPs towards a motion
can evolve over time while, some MPs tend to reflect similar
ideologies and linguistic styles all the time [Hampsher-Monk,
1984]. The inherent temporal nature of parliamentary debates
suggests the presence of contextual information across a time
series of debates [van Dijk, 2004]. We identify a speaker tem-
poral context in Political debates.

**Dynamic Graph Construction** We model the speaker
temporal context by creating a sequence of graphs G =
{G1, G2, · · · , GL} corresponding to each time-step in the
lookback L. For two consecutive debates d_i and d_{i+1} oc-
curring on time τ_i and τ_{i+1}, respectively with τ_{i+1} > τ_i.
We construct heterogeneous graphs G_i(V_i, E_i) and
G_{i+1}(V_{i+1}, E_{i+1}) corresponding to the debates d_i and d_{i+1}.
The nodes V_i of the graph G_i(V_i, E_i) contains a set of speaker
nodes S_i and motion nodes M_i, and the edge set E_i contains

<table>
<thead>
<tr>
<th>Dataset Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Speakers</td>
</tr>
<tr>
<td>No. of Motions</td>
</tr>
<tr>
<td>No. of Transcripts</td>
</tr>
<tr>
<td>Maximum Node Degree</td>
</tr>
<tr>
<td>Average Node Degree</td>
</tr>
</tbody>
</table>

Table 1: Graph construction statistics
speaker-speaker edges $E_{ss_i}$ and speaker-motion edges $E_{sm_{i+1}}$. As shown in Figure 3, we construct graph $G_{i+1}(V_{i+1}, E_{i+1})$ by updating the graph $G_i$. The set of nodes $V_{i+1}$ consists of new speakers $S_{i+1}$ that join the debate $d_{i+1}$ along with speaker nodes $S_i$ of debate $d_i$. A single motion node $m_{i+1}$ (since there is only one motion per debate) is added to the motion nodes $M_i$ of the debate $d_i$. Similarly, the set of edges $E_{i+1}$ consists of speaker-speaker edges $E_{ss_{i+1}}$ along with the speaker-speaker edges $E_{ss_{i}}$ of the debate $d_i$ and speaker-motion edges $E_{sm_{i+1}}$ along with the speaker-motion edges $E_{sm_{i}}$ of the debate $d_i$. We obtain graph $G_{i+1}$ as:

$$G_{i+1}(V_{i+1}, E_{i+1}) = G_i(V_i \cup V_{i+1}, E_i \cup E_{i+1})$$

(3)

where, $\cup$ denotes union operator. When $i = 1$ i.e., for the first debate $d_1$ in the lookback, we create the graph $G_1(V_1, E_1)$ with nodes and edges of the current debate $d_1$ only.

3.5 Temporal Context Encoder

To capture the cohesion as well as speaker temporal contexts in Parliamentary debates, we use the temporal context encoder over the dynamic graph $G$. We integrate the GAT into a gated recurrent unit which can encode cohesion contexts as well as model temporal speaker states. The temporal context encoder first applies the GAT ($\S3.3$) on each time step of the lookback. As shown in Figure 2, at time $\tau$ we obtain the input node features for the GAT $h^{\tau} = \{x_1^{\tau}, \ldots, x_{\tau}^{\tau}\}$. We update these node features to cohesion context dependent features, $h^{\tau'} = \{q_1^{\tau'}, \ldots, q_{\tau'}^{\tau'}\}$ via Equation 2 given by:

$$q_i^{\tau} = \bigoplus_{k=1}^{U} \text{LeakyReLU} \left( \sum_{j \in N_i^\tau} \alpha_i^{\tau} W_k x_j^\tau \right)$$

(4)

where, $N_i^\tau$ denotes the neighborhood of node $i$ on time $\tau$ and $\alpha_i^{\tau}$ is the attention weight between nodes $i$ and $j$ on time $\tau$.

We now use a GRU to capture the temporal dependencies in the node features $q_i^{\tau}$. We feed the features $q_i^{\tau}$ of each node $i$ in the graph to obtain the hidden states $z_i^{\tau} \text{ for day } \tau$ as:

$$z_i^{\tau} = \text{GRU}(q_i^{\tau}, z_i^{\tau-1}) \quad t - L \leq \tau \leq t$$

(5)

Temporal Attention Political studies have shown that some debates have more influence on the upcoming discussions [Box-Steppensmeier and Jones, 1997]. To this end, we apply a temporal attention mechanism which learns to weigh more critical debates. The temporal attention mechanism aggregates the hidden states $z_i^{t-L}, \ldots, z_i^{t}$ from different days into an overall representation $o_i$ using learnt attention weights $\beta_i^\tau$ for each day $\tau$. We formulate this mechanism as:

$$o_i = \sum_{\tau} \beta_i^\tau z_i^{\tau}, \quad \beta_i^\tau = \frac{\exp (\tau_i^T W z_i)}{\sum_{j=1-L} \exp (\tau_i^T W z_i)}$$

(6)

where, $W$ is a learned linear transform. We now feed the outputs $o_i$ to a fully connected layer followed by softmax activation which outputs the stance of each MP $y_i \in \{\text{Aye}/\text{No}\}$. We optimize TEC using the cross-entropy loss function.

4 Experimental Setup

4.1 Dataset and Preprocessing

We evaluate TEC on the ParlVote dataset consisting of 33, 461 debate transcripts. On average, a speech in ParlVote has 760.2 $\pm$ 901.3 tokens. Following [Abercrombie, 2020], we remove non-speech elements, tokenize motions and transcripts and, preserve the texts’ original casing. The dataset is fairly balanced with 53.57/46.43% Aye/No labels. The transcripts are labeled based on a speaker’s vote to their speech, with votes for ‘Aye’ and ‘No’ representing positive and negative stance. We split the dataset temporally and obtain 1396 debates from 7/5/1997 and 16/1/2012 for training, 200 debates from 16/1/2012 to 6/1/2014 for validation, and 398 debates with from 6/1/2014 to 5/11/2019 for test.\(^1\)

4.2 Training Setup and Evaluation

We perform all experiments on Tesla T4 GPU. We use grid search for hyperparameter selection for all models and select optimal values based on validation accuracy. We fine-tune BERT using learning rate $= 5e-5$, batch size 8, with the AdamW optimizer, for 7 epochs. The GAT layer’s output size is set to 16, and number of attention heads to 4. The output space of the GRU is 16 and that of the last fully connected layer to 2 in order to classify the stance into two classes. We use Adam optimizer, which is set with learning rate of 1e$-4$.

We compare TEC with the following baselines on accuracy (Acc) & Matthews correlation coefficient (MCC):\(^2\)

<table>
<thead>
<tr>
<th>Model</th>
<th>Baseline Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority class</td>
<td>Does not use any features. The majority class in the training set as the predictions for test set.</td>
</tr>
<tr>
<td>Multi-layer Perceptron (MLP)</td>
<td>A BoW model that utilises only unigram textual features from transcripts as input with TF-IDF selection [Abercrombie, 2018a].</td>
</tr>
<tr>
<td>BERT+MLP</td>
<td>BERT embeddings fed into a MLP. It is a textual model without additional contexts [Abercrombie, 2020].</td>
</tr>
<tr>
<td>Deepwalk</td>
<td>Concatenates intra-party graph embeddings with TF-IDF based BoW text features [Bhavan et al., 2020].</td>
</tr>
<tr>
<td>GPolS</td>
<td>BERT embeddings are fed into the GAT to encode cohesion context. However, they do not use an MPs past speeches for making predictions [Sawhney et al., 2020].</td>
</tr>
</tbody>
</table>

\(^1\)Code at: https://github.com/midas-research/tec-ijcai
\(^2\)All baselines are recreated via the creator’s settings.
5 Results and Analysis

5.1 Performance Comparison with Baselines

We compare TEC with baseline methods in Table 2. We note that BERT+MLP significantly \((p < 0.01)\) outperforms Majority class and BoW based approaches (MLP). We postulate this improvement to fine-tuning BERT to obtain rich embeddings that better represent the language used in the political realm. We observe that graph-based models (GPolS, TEC, Deepwalk) outperform text-only models (MLP, BERT-MLP) likely because graph-based learning captures intra-party and motion level contexts in parliamentary debates. Additionally, TEC and GPolS perform better than featureless embedding method, Deepwalk, potentially because the GAT component augments language features of different nodes for propagating contextual information across nodes. Our work is in line with political studies [van Dijk, 2004] who find that a contextual approach benefits discourse analysis where MPs take up different roles, such as communicative, interactional, and social roles, and contexts are identified when MPs account for group memberships or speak for/against motions. Finally, social roles, and contexts are identified when MPs account for group memberships or speak for/against motions. Consequently, TEC performs better than BERT+MLP, GPolS, and Deepwalk.

5.2 Ablation Study Over Model Components

Table 3 shows how TEC’s performance benefits from each of its components. We observe that modeling the speaker temporal context by feeding the text representation of their historical debates to a GRU (T-BERT) leads to significant \((p < 0.01)\) gains likely because MPs tend to reflect similar opinions across time [Hampshire-Monk, 1984]. Next, we note large improvements on adding the temporal attention mechanism. We postulate that this improvement stems from the ability of TEC to identify more salient trends in MPs speech across time, similar to observations in [Fort, 2019] who show that there are certain historic speeches that MPs keep referring to. We note a significant improvement \((p < 0.01)\) on augmenting speaker-temporal context with cohesion contexts to encode correlations between speakers via GCNs, likely because of its ability to capture the herd mentality in political parties. We also note that adding graph attention mechanism leads to gains, likely because the GAT weighs more critical relations between MPs. Our observation is in line with other studies [Bawn et al., 2012] who find that there exists variance in relationship strength between MPs due to the presence of interest groups and key members of a party. Finally, we observe that the GAT and speaker-temporal context encoder complement each other and capture the dynamic structure [van Dijk, 2004] of political debates.

5.3 Ablation Study Over Cohesion Contexts

We perform an ablation study over the different kinds of cohesion contexts that TEC models in Table 4. We remove the contexts one by one and find that adding intra-party and motion context to speaker-temporal context leads to significant improvements \((p < 0.01)\) individually. Interestingly, we note that the performance drop on removing the intra-party context is significantly \((p < 0.01)\) larger compared to the removal of motion context, suggesting that while the motion context models the global similarities between speeches on the same motion, the intra-party context is more important to capture the actual decision taken by the MP. We postulate this observation to the existence of herd mentality in political parties, wherein MPs support their party via their votes to make one collective decision [Sawhney et al., 2020].

5.4 Sensitivity to Lookback Period

Figure 5 shows TEC’s performance with varying lengths of lookback periods. First, we observe that TEC using temporal attention outperforms GAT+GRU model due to its ability to capture critical historic debates that influence the speaker’s current state. Next, we note that using shorter lookbacks leads to poorer performance, likely because of fewer speeches for a coherent speaker state modeling [AJ Willingham, 2017]. As we increase the length of lookback, we note deterioration potentially due to inclusion of old speeches which may not contribute to the speaker’s current state. We note that TEC using...
temporal attention can filter crucial debates from large windows to an extent and works best with mid-sized windows.

5.5 Effectiveness of Cohesion Context

We first calculate the attention scores amongst MPs of two parties in Figure 4 (left) and observe high political cohesion in the Conservative party, which is further consolidated by the old age and large size of the party, as such parties show better party discipline [Hayton, 2012]. On the other hand, the Liberal-Democrats, a relatively newer and smaller party, show high attention scores along the diagonal, indicating a large self-dependency of speakers. Next, we perform a qualitative analysis in Figure 4 (middle), and observe that the Conservatives strongly oppose the Labour party’s plans to raise £2.5 billion for the national health service. We note higher attention weights between speakers that support each other than those who present diverging arguments, suggesting that the GAT accounts for different degrees of interactions between speakers. However, we note an overall unity since the MPs have a uniform stance on the motion despite their diverging speeches, indicating the presence of partisan forces within these parties. Interestingly, we also note inter-party contexts when Lefroy, Jeremy of the conservative party, acknowledges the hard work of health workers similar to Qureshi, Yasmin of the Labour Party. This observation ties up with [Hix et al., 2005] who show that MPs engage in negotiations of inter-party boundaries to create realities that fit their political ends.

5.6 Effectiveness of Temporal Context

We now analyze how an MP’s speech evolves over time via a qualitative study in Figure 4 (right). We take an example of John Redwood a member of the Conservative party and a long-term critic of the European Union [BBC, 2016]. We observe that John consistently supports the Brexit referendum in his speeches and often makes references to his previous debates, suggesting the presence of temporal patterns in his speeches. Through temporal speaker state modeling TEC is able to capture some of these cues by lending higher attention to debates that influence his current state.

6 Conclusion

Motivated by political studies [van Dijk, 2004] we present TEC, that enriches linguistic features with cohesion contexts and temporal speaker state modeling for accurate political stance estimation. In the future, we plan to adopt such a contextual approach to other political tasks such as roll-call prediction. We also wish to add speaker-specific features such as religion, the constituency they represent, etc. We would also explore other data sources, such as news and social media.

3 John Redwood: https://en.wikipedia.org/wiki/John_Redwood
References


