

# Political Preferences for Pork Barrel Messages under Different Electoral Systems

Yen-Chieh Liao \*

University College Dublin

February 8, 2024

## Abstract

Measuring legislator behaviours and tendencies towards constituencies under different electoral systems is important. This paper quantitatively investigates this topic using the case of Taiwan Legislative Yuan and data on written parliamentary questions through an electoral reform from multi-member districts (MMD) to single-member districts (SMD). To answer the research questions, I train deep learning models on multi-convolutional neural networks with an embedding layer extracted from Transformer BERT to detect pork-barrel features in parliamentary questions over time. With Transformers' attention mechanisms, this combination approach enables the machine to learn the condensed features of embedding representation and better handle polysemous words than traditional embedding approaches like Word2Vec. Last, I employ regression analysis to test the impact of the reform occurrence and control for differences between districts and legislators through the transition. Evidence exists to show that the transition of electoral reform incurs essential changes in legislators' behaviour. Legislators under multi-member districts are more likely to express political intention regarding pork-barrel projects in the questions. The reform subsequently demonstrates heterogeneous effects on large parties vis-à-vis small parties.

---

\*Yen-Chieh (David) Liao is a postdoctoral researcher with [Text and Policy Research Group](#) at the School of Politics and International Relations, in University College Dublin. Email: [davidycliao@gmail.com](mailto:davidycliao@gmail.com). This draft is derived from my PhD dissertation and is currently undergoing revisions for submission to a journal.

# Introduction

How does the electoral reform change legislators' preference and their intentions to bring home the bacon? Scholars have clearly explained why intraparty competition by different rules of electoral systems increases legislators' incentives to run on a personal reputation (Cox, 1990; Downs, 1990; Carey and Shugart, 1995). For example, candidates under multi-member districts (MMD) are more likely to reward small groups of supporters with particularised benefits and deviate from their party line, whereas candidates under single-member districts (SMD) prefer to adopt electoral strategies that target median voters (Cox, 1990). With abundant literature on an empirical examination of the relationship between electoral systems and political behaviours (e.g. Cox, 1990; Catalinac, 2016, 2017; Goplerud and Smith, 2021), we however know little about whether actual impacts introduced by the electoral reform through MMD to SMD reduce legislators' motives to pursue pork barrel project in the legislature. Understanding such legislative motions is very important for the process of policy making and decision.

Parliamentary activities such as debates and written parliamentary questions, play a significant role in most parliamentary democracies. For example, the floor debate functions as a major platform for Members of Parliament (MPs) to uncover information and discussion regarding government policy, proposed new laws and topical issues of the day. With different regulations limiting MPs to access to the floor, the distribution of time and topics tends to be restricted to displace salient issues (e.g., Bäck, Baumann and Debus, 2019; Martin, 2011), which increase difficulties in discovering MPs' genuine interests in preferences (Martin, 2011; Saalfeld, 2011; Russo, 2021) and their nature of substantive representation (Kolpinskaya, 2017). On the contrary, the parliamentary questions, known as "non-legislation activities" (Martin and Rozenberg, 2017), is one of the primary tools that MPs can freely use to press for action in the government and express their concerns with regards to constituency (Russo, 2021; Saalfeld, 2011; Martin, 2011).

Measuring legislator behaviours and tendencies towards constituencies under the different systems is important. This study consider the relationships between the electoral system and the pork-barrel phenomenon, focusing in particular on legislators electoral strategies and communication style in parliamentary questions. From theoretical perspective, candidates in

single non-transferable vote in multi-member districts (SNTV-MMD) were not only competing with competitors from rivalry parties but also co-partisan candidate from their party (Carey and Shugart, 1995). Hence, candidates have more incentives to run on personal votes (Cain, Ferejohn and Fiorina, 1987; Carey and Shugart, 1995). For example, Japanese candidates elected via SMD are more likely to pursue national-wised issues (such as defence and fiscal policy) in election manifesto than in SNTV-MMD (Catalinac, 2016). However, it is hard to distinguish whether the changes in electoral strategies was caused by the reform or possibly other factors such as recession and military threats (Ishima, 2020). While useful for many purposes for understating the implication of the reform, election manifesto are only distributed during election periods and is not able to instantly reflect actual preferences across times (Martin, 2011; Martin and Rozenberg, 2017).

MPs ask questions for several reasons. Generally speaking, when MPs ask written questions to the executives (ministerial officials), it may be because of their expertise or domain responsibility of delegation for question topics. Nevertheless, MPs under electoral incentives may be eager to display their attentions to constituency interests in parliamentary questions. In Martin's view (Martin (2011)), for MPs to employ parliamentary questions are primarily about involving themselves in the policymaking process, and important factors to consider are their motivations to run on personal vote and differences in the electoral rules (Martin, 2011, p.260). In particular, party leaders under SNTV-MMD have a strong incentive to nominate more than one candidate to run in each district in multi-member districts (Cain, Ferejohn and Fiorina, 1987; Carey and Shugart, 1995; Reed, 1995; Catalinac, 2017). Therefore, candidates cannot rely exclusively on their party reputation and have to find an alternative means of attracting votes by running on a personal reputation, which therefore motivate candidates to deviate their party line (Cain, Ferejohn and Fiorina, 1987; Carey and Shugart, 1995). Thus, question types may coherently reflect the orientation that MPs attempt to pursue (Saalfeld, 2011; Martin, 2011; Martin and Rozenberg, 2017).

In this paper, I introduce the case of Taiwan Legislative Yuan, where the electoral system reformed through SNTV-MMD to SMD, to evaluate how electoral motives shape legislators' tendency to pork-barrel projects under different electoral systems. In particular, parliamentary questions are the primary channel for legislators to scrutinise the government and express po-

litical intentions. Most importantly, the number of written questions in Taiwan tables to approximately one hundred and fifty thousand since 1993. These parliamentary questions allow identification of different question topics, categories and further information regarding legislators' opinions of policy interests and agenda at the individual level, which enables us to conduct a more nuanced test of theoretical expectations than previously attempted.

Concretely, this paper described here focuses on the following two research questions: are the legislators in the SNTV-MMD more likely to bring home the bacon by asking more about the provision of particularistic goods in the parliamentary questions? Dose the reform change legislators' electoral strategies and increase their attention to other public policies such as regulatory issues?

To answer the research questions, I train deep learning models on multi-convolutional neural networks with an embedding layer extracted from Transformer BERT to detect pork-barrel features in parliamentary questions over time. With Transformers' attention mechanisms, this combination approach enables the machine to learn the condensed features of embedding representation and better handle polysemous words than traditional embedding approaches like Word2Vec. Last, I employ regression analysis to test the impact of the reform occurrence and control for differences between districts and legislators through the transition. Evidence exists to show that the transition of electoral reform incurs essential changes in legislators' behaviour. Legislators under multi-member districts are more likely to express political intention regarding pork-barrel projects in the questions. The reform subsequently demonstrates heterogeneous effects on large parties vis-à-vis small parties.

## **Personal Vote and the Electoral Reform**

Since the legislative election became direct in 1992, Taiwan Legislative Yuan used the Single Non-transferable Vote in multi-member districts (SNTV-MMD) to elect 161 to 174 members, including party lists and seats for aborigines and overseas Chinese. In 1999, the total number of seats increased to 225, avoiding potential political chaos caused by abolishing Taiwan's provincial governmental (state-level) body (Wang, 1994).<sup>1</sup> Under SNTV-MMD, voters can only

---

<sup>1</sup>The National Assembly, the authoritative legislative body of the Republic of China, gradually became a dormant body since the provincial governmental body was formally abolished in 1997. Since the 1990s, the National

cast a single vote for one of the candidates whose number of seats in a district range from the minimum seat 1 to 7, and the candidates who get the most votes win the seats.

In the literature, scholars have explained why and how candidates in multi-member districts have incentives to adopt electoral strategies to target a small group of voters (Cox, 1990; Downs, 1990; Myerson, 1993). Theoretically, SNTV-MMD encourages majority-seeking parties to nominate more than one candidate in a district (Shugart and Wattenberg, 2003) and is associated with the intensified intraparty competition. Party leaders not only need to consider the process of nomination but also vote management carefully. If the leader's over-nominate candidates in a district, it would generate uncertainties for co-partisan candidates to run personal reputations against each other (Shugart and Wattenberg, 2003; Cox, 1990). For example, Taiwan's legislators in multi-member districts are more likely to move toward the extreme direction from the party line particularly when the party is mismanaged during the election (Jang and Lin, 2019). It, therefore, creates more difficulty for their party to discipline voters to spread their votes equally across all a party's candidates.

In addition, district magnitude affect the structure of the party system and inter-party relation (Duverger, 1954). In the 2000s, Taiwan's SNTV-MMD has been criticised for not only creating intraparty competition (e.g., Hirano, 2006) but also encouraging factional politics and money politics (Cox, 1996; Cox and Niou, 1994; Batto and Huang, 2016; Wu, 2003; Richardson, 1988). More candidates in a district under SNTV-MMD require lower votes, which increases candidates' motivation to target small groups of voters (Downs, 1990; Myerson, 1993). Under the circumstances, the candidate not only competes with candidates from opposite parties but avoids co-partisans carving out shared voters. In practice, candidates cannot rely exclusively on their party labels and must find an alternative means of separating themselves from different co-partisans. The most conventional approach is to attract more targeted voters by cultivating a personal reputation (Cain, Ferejohn and Fiorina, 1987; Reed, 1994).

In many democracies, the parliamentary question provides a function used to impose ministers and related ministerial officials' accountability in the public domain (Martin and Rozenberg, 2017). For example, the parliamentary question is important for Taiwan's legislators to carry messages to the ministers for the constituency's needs. In particular, each legislator has

---

Assembly's power was transferred to the Legislative Yuan.

an equal chance to ask for information about policy and activities of ministries related to any topic of public affairs. Unlike floor debates, the word usages and strategies for delivering debates are deliberately considered (Slapin and Proksch, 2014; Slapin et al., 2018). It is worth noting that during floor debate, legislators only have 15 minutes to question invited ministers in the affiliated committee. Therefore, only topical subjects are raised in the discussion for a minimal period.

In Taiwan, legislators, like other representatives in democratic countries, take a position by sponsoring legislation, whereas the ministers in the central government rely on any records related to legislative motions to understand their preferences (Luor and Hsieh, 2008). In practice, legislators attempt to request the pork-barrel project by leaving comments on the purpose of the statute (立法意旨) or recommendation section (建議事項). For example, Luor and Liao (2009) finds that the more pork-barrel legislation legislators propose, the higher grant allocation the municipalities (districts) receive.<sup>2</sup> It therefore gives rise to my first hypothesis 1 in this chapter:

**Hypothesis 1.** *Under SNTV-MMD, legislators are more likely to propose parliamentary questions regarding the provision of particularistic goods.*

## Pork Barreling in Parliamentary Questions

Parliamentary questions serve a variety of purposes. For example, MPs scrutinise the governments by asking questions without strict party discipline. As a result, the content of the question should be evident if MPs developed a tendency to run on personal reputation. While parliamentary questions as the unit of analysis provide a valuable proxy for understanding MP's representation and electoral purpose, few systematic studies have explained to what extent the reform diminishes legislators' incentive to ask questions devoted to particularistic goods.

To date, most researchers observing the effects of the electoral system on political representation in parliamentary democracies have either focused on the party-level unit such as speeches (e.g., Høyland and Søyland, 2019; Guinaudeau and Costa, 2021; Ishima, 2020) or legislator-

---

<sup>2</sup>In general, the pork-barrel items in Taiwanese political context include any projects related to such as road and bridge construction, tax breaks, and subsidies programmes designated for specific groups or areas.

level analysis, i.e., politicians' accounts on social media and manifesto (e.g., [Catalinac, 2017](#); [Schürmann and Stier, 2022](#)). For example, [Schürmann and Stier \(2022\)](#) finds that elected legislators from districts mention more territorial references on Facebook and Twitter than those from party lists, whereas [Catalinac \(2017\)](#) shows that the reform in Japan through SNTV to SMD substantially decreased LDP candidates' tendency to mention particularistic goods in the election manifesto. In similar, the literature on the same topic highlights the relationship between changes in district magnitude and pork-barrel behaviour in sponsoring the legislation ([Luor and Hsieh, 2008](#); [Luor and Liao, 2009](#); [Sheng, 2014a,b](#)).

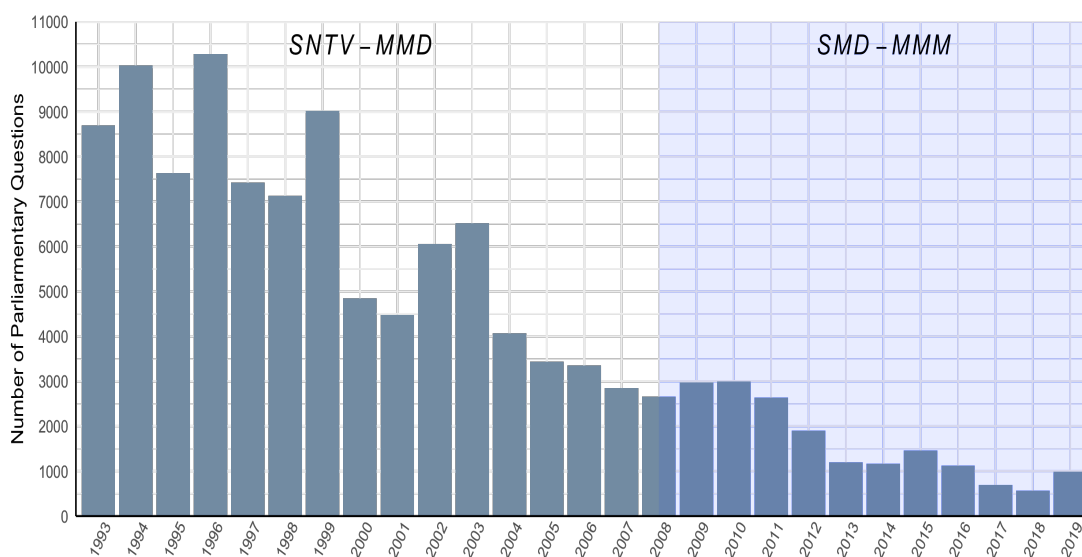
However, while these approaches in the literature provide insightful implications for studying the impact of the different electoral systems, there are fundamental limitations. First, analyses of speech data have become popular due to the advent of hands-on computational tools and are invaluable for understanding party competition. However, access to the floor for speeches tends to be generally restricted and capped ([Proksch and Slapin, 2009](#); [Martin, 2011](#)). Similarly, passing new legislation is hugely time-consuming, and the cost is greater than most legislators can afford. Another piece of literature on legislative behaviour has explained several reasons why legislators also have incentives to engage in other legislative activities. For example, some members of the US House of Representatives without institutional power to influence policy agenda are more likely to grandstand in hearings to attract voters, particularly when they are in more disadvantaged regions ([Park, 2021](#)). Similarly, in Westminster parliamentary systems, MPs use rebellion speeches to differentiate themselves to attract more electoral support ([Slapin et al., 2018](#); [Proksch and Slapin, 2015](#)).

While nature of parliamentary questions are not subject to the limitations mentioned above, analyses of question content allow us to test the theoretical perspective. As theoretically expected, legislators under SNTV-MMD are expected to target smaller groups of voters by asking questions devoted to mentioning particularistic quantities when facing strong co-partisan. Therefore, legislators tend to oversee the government in the new SMD system by asking general questions related to programmatic and regulatory policies. Thus, it gives rise to the following hypothesis:

**Hypothesis 2.** *Under SMD-MMM, legislators are more likely to ask questions related to programmatic and regulatory policies.*

## Data: Parliamentary Questions

Parliamentary questions recorded in Taiwan Legislative Yuan (立法院) cover a wide range of topics, with nearly 230 unique subjects. To analyse parliamentary questions, I have web scraped parliamentary questions from the official website of Taiwan Legislative Yuan from 1993 to 2020, including relevant information about classified topics, selected keywords and the corresponding question categories.<sup>3</sup>



**Figure 1:** THE NUMBER OF PARLIAMENTARY QUESTIONS FROM 1993 TO 2019

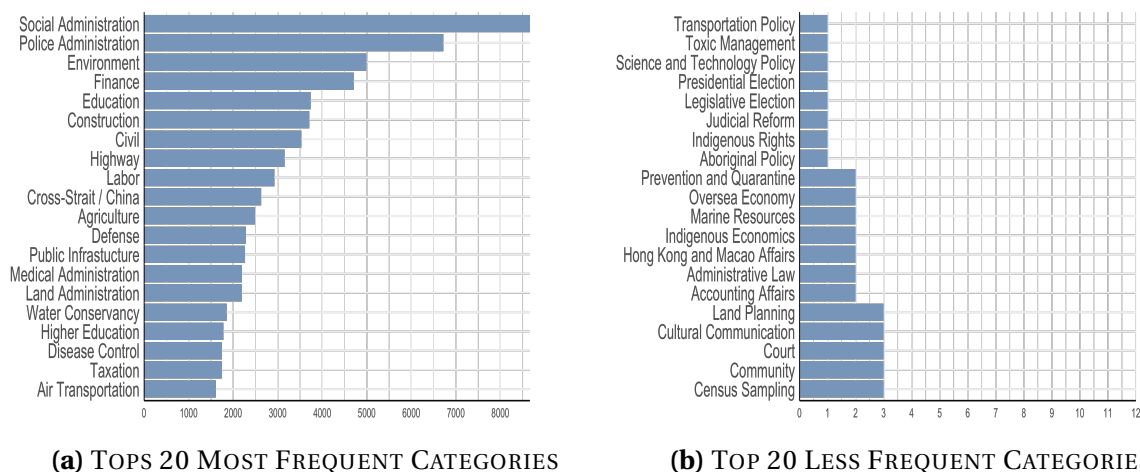
To display the distribution of the questions across years, [Figure 2](#) demonstrates the total number of parliamentary questions documented from 1993 to 2019. During this period, the Taiwan legislature went through the movement of reducing legislative seats, which started in 2000, and the reform occurred in 2008. In the pre-2000 period, the number of questions asked remained roughly stable with multiple fluctuations. Nevertheless, there was a noticeable drop in the number of questions since 2000, right after the start of the movement. This trend persisted, and the total number of questions plummeted.

The [Figure 2a](#) shows the top 20 categories of topics that were frequently asked during the period. Topic categories related to social administration and police administration were the top two topics that appeared in the discussions in the legislature. Coming next are categories

<sup>3</sup>The designed web scrapper programme ([legisCrawler](#)) for retrieving parliamentary questions from Taiwan Legislative Yuan is hosted on author's [GitHub](#).



about environment and finance.



**Figure 2: THE DISTRIBUTION OF PARLIAMENTARY QUESTIONS CATEGORIES ASKED BY LEGISLATORS**

## Pork Barrelling Machine Classifier

We take advantage of the Pork-barrel Legislation Dataset assembled by Prof. Dr. Ching Jyuhn Luor (Luor and Hsieh, 2008; Luor and Liao, 2009; Luor and Chan, 2012) to train the deep learning model. The dataset consists of 7,243 pieces of legislation, which were manually annotated as *Pork* (with label 1) or *Non-Pork* (with label 0) from 2004 to 2008. Additionally, this dataset was cross-coded by three social science researchers to assess its validity, achieving 98

The gold standard for identifying pork barrel legislation is based on the target beneficiaries of the policy (distributed vs. concentrated) and the attributes of policy cost (distributed vs. concentrated), as illustrated in Figure 4 (Wilson and DiIulio, 2001). In particular, typical pork-barrel policies (or legislation) mainly incur distributed costs while generating parochial benefits for specific regions or designated population groups. For instance, the decision to execute an areotropolis project, which involves constructing an airport within a particular area, e.g. Taoyuan City, incurs collective costs for all Taiwanese taxpayers, while the benefit of such a particularistic project is narrowly concentrated within a parochial group of Taoyuan residents (in terms of employment opportunities, economic development, and convenience of the airport) as well as local politicians such as legislators themselves (Luor and Hsieh, 2008; Luor and

立法院議案關係文書 (中華民國 41 年 9 月起編號)  
中華民國 96 年 3 月 28 日印發

院總第 1687 號 委員提案第 7375 號

the purpose of statute

案由：本院委員林重謨、蔡啟芳、侯水盛、潘孟安、盧天麟、高建智、余政道、唐碧娥等 40 人，有鑒於最低基本工資可視為人民維持基本生活水準之意涵，現行每月僅有五千元的老年農民福利津貼相較於一萬五千八百四十元之法定基本工資，實已偏低，不但無法維持老年農民基本生活需求，亦與老年農民長期對國家的貢獻不成正比，更有違政府落實照顧農民的承諾，再加上近年來物價年年高漲，使老年農民生活的艱苦更加雪上加霜。職是之故，基於保障老年農民基本生活水準，肯定老年農民之貢獻並提昇其生活品質，特擬具「老年農民福利津貼暫行條例第四條修正案」，比照國內法定基本工資之標準，將現行老年農民福利津貼發放金額由每月五千元提高至每月一萬五千元。是否有當，敬請公決。

increase farmer's allowance from NTD 5000

improving living standards

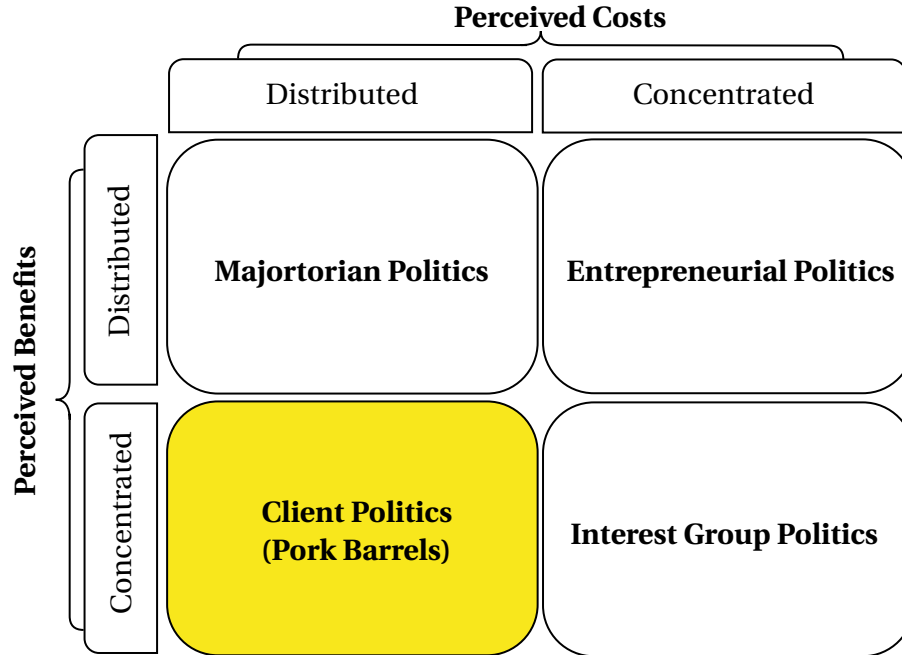
in light of the contribution made by the farmers

to NTD 15000

**Figure 3:** PORK BARREL LEGISLATION REGARDING IMPROVING THE QUALITY OF RETIRED FARMERS' LIFE

Liao, 2009; Luor and Chan, 2012).

Moreover, another example is the subsidy for the targeted populations. As illustrated in **Figure 3**, the purpose of the legislation was to raise the farmers' monthly allowance from NTD 5000 to NTD 15000. In general, the majority of retired farmers are concentrated in agricultural municipalities. Thus, Luor and Hsieh (2008); Luor and Liao (2009) operationalises those concepts commonly found in Taiwan's political context and further categorises the legislation into pork and non-pork.



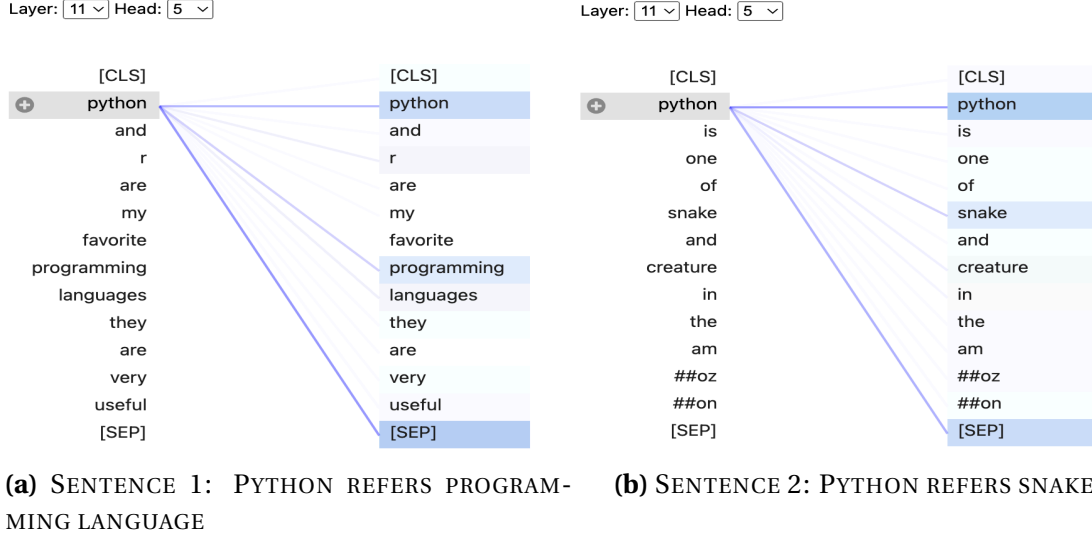
**Figure 4:** CLASSIFYING AND EXPLAINING THE POLITICS OF DIFFERENT POLICY ISSUES, SOURCE: WILSON AND DI IULIO (2001)

## Using BERT Model as Embedding Layers

With increasingly available amounts of political data, the application of the classification task using deep learning methods has received great attention in political science (Chatsiou and Mikhaylov, 2020). In natural language processing, text classification assigns a set of predefined categories to open-ended documents. The approach used in this paper is to combine one of the famous Transformer architectures, BERT developed by Google, with convolutional neural networks.

CNN is one of powerful neural network architectures commonly applied in natural language processing (Zhang, Zhao and Lecun, 2016; Zhang et al., 2017; Kim, 2014; Kim et al., 2016). The architecture can be constructed by a series of convolutions and pooling layers, filtering input vectors and creating a feature map that summarises the input texts. Then, the feature map can be stacked one over another to form a matrix by single-dimensional convolutional filters to extract high-level features. In the context of text classification task, convolution layer essentially learns the condensed features as learning image data.

In practice, the convolutional layers (or recurrent neural net) are recently introduced as en-



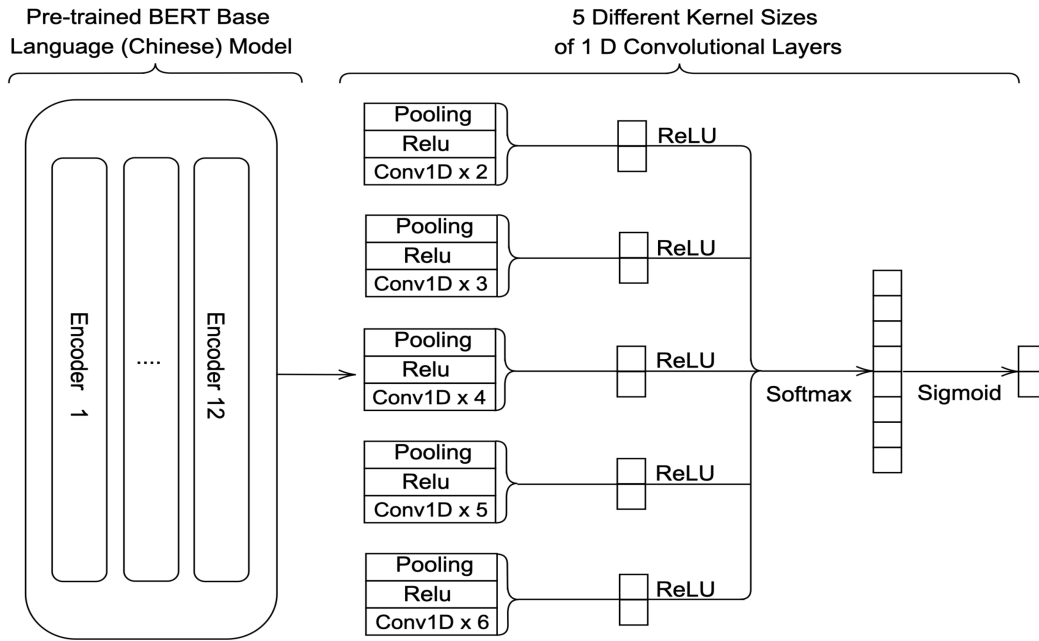
**Figure 5:** EXPLORING THE SELF-ATTENTION MECHANISM IN BERT TRANSFORMERS WITH PRE-TRAINED ENGLISH LANGUAGE MODELS(BERT-BASE-UNCASED)

coders or decoders to deal with semantics problems in modern application of natural language processing. Generally, we can transform input texts with embedding models such as Word2Vec, GloVe or one-hot vector. However, the major challenge of these approaches lies in the fundamental assumption that each input word has fixed representation in different contexts, introducing a potentially severe problem of misrepresentations by referring to inaccurate meaning across sentences. For example, the embedding word “python” as shown in Figure 5 can have different meanings depending on the language context in which it appears. Regardless of polysemous mean in different sentences, the word “python” only renders the same vector. This is because traditional embedding models are context-free, which gives static embedding vectors for the word “python”.

Contrary to earlier approaches, BERT is one of the most powerful Transformer architectures that can detect input tokens in bidirectional semantic context with its prominent feature called self-attention (Devlin et al., 2019; Vaswani et al., 2017). BERT consists of 12 encoder layers, stacked over one another as shown in Figure 6, respectively. For example, giving the single word “python” in two different contexts in Figure 5, “python” in Figure 5a refers to one of the programming languages because contextual embedding is highly associated with “programming”, “languages”, and “r (statistical computing language)”. On the other hand, “python” in Figure 5b refers to an animal in the sentence context as its representation embedding is correlated with

snake and creature.

As discussed above, self-attention is a mechanism that allows neural networks to assign a different amount of attention weight to each element to compute representation embedding in a sequence. This is powerful for performing NLP tasks when dealing with unseen words or tokens not included in the static embedding model like Word2Vec.



**Figure 6:** THE DESIGNED ARCHITECTURE FOR PORK BARREL CLASSIFICATION TASK

## Combining Convolutional Neuralnet with BERT Embedding Layer

Combining CNN with BERT for classification tasks is getting popular in natural language processing in recent years (Safaya, Abdullatif and Yuret, 2020; Lu and Ni, 2019; Lopez and Kalita, 2017). For example, Safaya, Abdullatif and Yuret (2020) use convolutional layers followed by BERT embedding to create a machine learning model that deals with offensive speech identification, while Lu and Ni (2019) deploying the similar approach to patent document classification. Those performance of combination approach is drastically improved from the character-based CNN and the generic BERT model, respectively. The innovative design for machine learning architecture in the literature that motivates the model used is the following. I create the architecture for detecting pork-barrel features in parliamentary questions, as shown in Figure 6.

All the encoders use 12 attention heads and 12 layers. Theoretically, each token can be represented as 768 hidden units (Wolf et al., 2020).<sup>4</sup>

First, we transform each word in the sentence with its word vector using the Chinese BERT pre-trained model.<sup>5</sup> In the BERT layer, I can feed input sentences (pork legislation text) in the encoder layer, where the encoder learns the representation. Afterwards, the decoder generates new output based on the patterns understood by the following encoder. In order to fit embedding vector into CNN layers, I create five kernels of different lengths, aiming to capture different patterns of the n-gram in the original sentence. After the output is passed through ReLU activation, Global Max Pooling is introduced to flatten high-dimensions feature maps into two-dimensional vectors. At the last set of layers, I use the Sigmoid function commonly adapted for binary class with dense layers to get the final outcome, which is the probability of being classified as pork-barrel project.<sup>6</sup>

**Table 1:** A SAMPLED ANALYSIS: LIKELIHOOD OF MENTIONING PARTICULARISTIC GOODS IN PARLIAMENTARY QUESTIONS

Legislators	Probability of Being Pork	Topics	Keywords	Questions
林正峰	0.995515823364258	Health Insurance	Health Insurance Deductions	特別扣除額教育支出...
彭添富	0.992780447006226	Aboriginal Affair	Housing Subsidies	而非采用扣除免稅額...
李復興	0.992780089378357	Old-age Benefits	Elderly Allowance	原住民家庭租屋補助...
盧秀燕	0.992639720439911	Veterans Welfare	Grants for Retired Veterans	補助金發放金額過低...
李顯榮	0.990033149719238	Farmer Welfare	Subsidies; Allowance	政府前後援賽金額高...
丁守中	0.988385319709778	The Handicapped	Living Allowance	身心障礙者生活津貼...
馮定國	0.985531985759735	Elderly Welfare	Unemployment Fund	高齡失業問題嚴重日...
彭添富	0.983698368072510	Agriculture	Crops Subsidies	農作物損失補償問題...
曾華德	0.979519009590149	Military Affair	Increased Pay	救国军补发薪饷问题...
林鴻池	0.978044390678406	Education	University Subsidy	針對諸多已獲得五年...

To validate the classification quality, we sampled 20 pork and non-pork questions respectively in Table 1 and Table 2, automatically classified by the architecture in Figure 6. we merge these questions with corresponding keywords and categories (translated to English) scraped from the website of Taiwan Legislative Yuan. As shown in Table 1, most pork-barrel questions are associated with central government spending for targeted groups and localised infrastructure projects mainly allocated to specific regions. Some legislators raise the question of asking

<sup>4</sup>The traditional Chinese BERT model employed is *bert-base-Chinese*, maintained by CKIP (Chinese Knowledge and Information Processing) at the Institute of Information Science and the Institute of Linguistics, Taiwan Academia Sinica.

<sup>5</sup>This model can be downloaded from <https://ckip-transformers.readthedocs.io/en/stable/> and *HuggingFace* at <https://huggingface.co/bert-base-chinese>.

<sup>6</sup>Regarding the metrics of model performances, see Table B.2 in Supplementary Appendix 2.

the central government to increase rent subsidies for the aboriginal population, while others target policies related to relief and support programmes such as crop subsidy and elderly allowances for specific groups in some municipalities located in western Taiwan. In addition to [Table 2](#) demonstrate many examples of non-particularistic questions. For instance, the questions of keywords and topics are very closely related to regulatory and national policies such as railway management, drug control and criminal investigation.

**Table 2:** SAMPLED PARLIAMENTARY QUESTIONS BEING LESS LIKELY TO MENTION PARTICULARISTIC GOODS

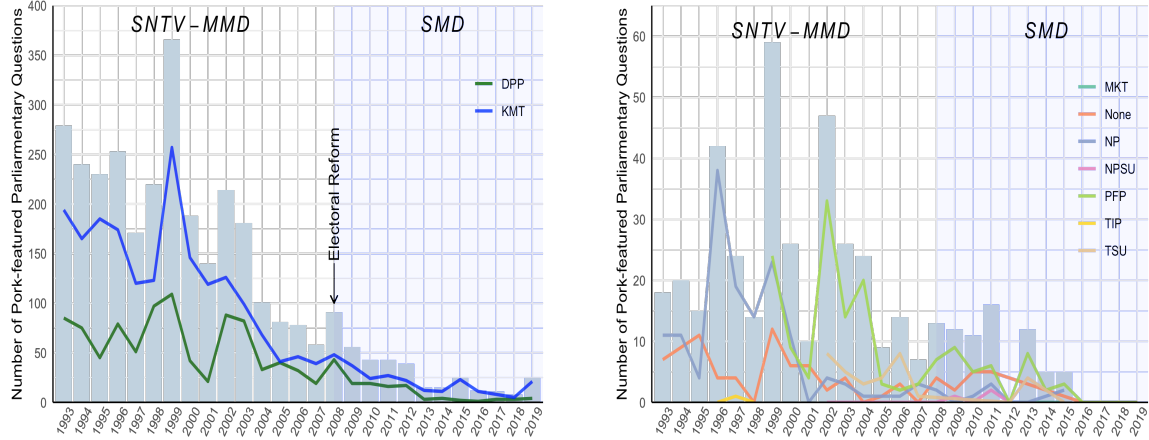
Legislators	Probability of Being Pork	Topics	Keywords	Questions
李復甸	0.000021549063604	Litigation Procedure	Criminal Investigation	鑑於刑事偵察實務上...
林建榮	0.000020212990421	Financial Management	Revolving Interest Rate	明定信用卡、現金卡...
林正峰	0.000019731034627	Energy Policy	Energy Saving	要求各級機關和學校...
林正峰	0.000019187420548	Tobacco Restriction	Departmental Hospital	毒品泛濫，吸毒人數...
王幸男	0.000017634354663	Public Safety	Road Quality	針對道路人孔蓋或管...
管碧玲	0.000013002485503	Railway Management	Taiwan Railway	台灣鐵路管理局發生...
郭榮宗	0.000004869816621	District Court	Drug Abuse	知名提神飲料遭下毒...
陳朝龍	0.000011277100384	Infectious Disease	Avian Influenza	英國政府宣稱台灣出...
林進興	0.000007685628589	Banking Management	Credit Card	行政院金融監督管理...
潘孟安	0.000002590457370	Election	Legislative Elections	單一選區兩票制即將...

## Heterogeneous Effects on Different Sizes of Parties

We next examine the distribution of pork-barrel questions aggregated by party level from 1993 to 2019. [Figure 7](#) displays the distribution of the total number of pork-barrel questions that were requested by legislators from two major parties and the small parties, respectively, as is classified by the deep learning architecture. The left subplot for the two majority parties, KMT (Chinese Nationalist Party) and DPP (Democratic Progress Party), roughly shares a similar pattern as in [Figure 1](#), with a decreasing trend of the numbers overtime after the initiation of the movement that started around 2000. The right subplot [Figure 7b](#) shows the multiple plummets and rises in the total number of questions raised by small parties such as NP (New Party 新黨), PFP (People First Party 親民黨), NPSU (Non-Partisan Solidarity Union, 無黨團結聯盟), TIP (Taiwan Independence Party 台灣獨立黨), TSU (Taiwan Solidarity Union 台灣團結聯盟) and newly established MKT (the Republican Party as known as Minkuotang).

For [Figure 7](#), the changes in the total numbers across times show that parliamentary questions are dwindling steadily, implying that legislators may alternatively use other tools to in-





(a) LARGE PARTIES

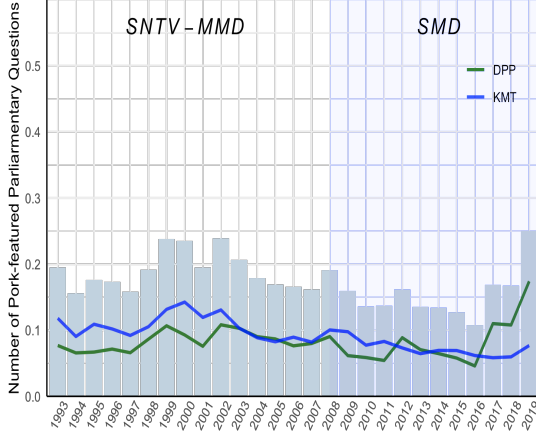
(b) SMALL PARTIES

**Figure 7: THE NUMBER OF PORK BARREL QUESTION BY YEARS**

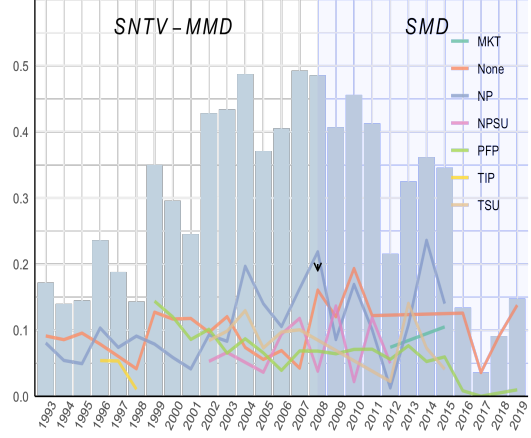
fluence policy for their constituents, such as social media. For example, empirical evidence shows that district members mention more geographical terms on their Facebook and Twitter, whereas party-list members have less tendency to secure pork-barrel projects. In particular, the total number of seats was reduced from 225 to 113 after the reform. In the new system, only 73 seats are elected by district, explaining a fall-off in the number of pork-question since the reform occurred in 2008.

To exclude the effect of volatility in the number of seats across years, we illustrate the average number of pork-barrel questions per seat in the legislature. **Figure 8** plots the average number of questions per seat devoted to pork-barrel project for majority parties (**Figure 8a**) and small parties (**Figure 8b**), respectively. The actual impacts of the reform seem heterogeneous on small parties and large parties (KMT and DPP). In **Figure 8a**, for large parties, the average number persistent declined from around 2001 to 2015, although there was a significant bounce back since 2017. Generally speaking, there are variations in the number of pork questions asked by small-party legislators, compared with legislators from KMT and DPP. From 2003 to 2012, the number of pork questions asked by each legislator from small parties was twice the size of large-party legislators. However, the reform differed from what we expected regarding small-party legislators: their average number increased from 2001 to 2009, yet this trend stopped in 2007. Afterwards, the trend became downward sloping, and the average number kept declining.





(a) LARGE PARTIES



(b) SMALL PARTIES

**Figure 8: THE MEAN NUMBER OF PORK BARREL QUESTION BY YEARS**

## The Presence of the Electoral Reform

How does the reform from SNTV-MMD to SMD-MMM change legislators' representation? To answer the question, we investigate to what extent the reform reduces legislators' incentive to ask pork-barrel questions. In total, legislators asked 116,248 questions throughout the reform transition from 1993 to 2020. While multiple plunges and surges between 2000 and 2015 reveal that the passage of years possibly affects the legislators' intention to propose the questions. To distinguish any irrelevant event to influence our analysis, we first calculated the mean of the pork features by an individual legislator at a quarterly unit and ran the following regression in Equation 0.1,

$$\begin{aligned}
 Pork_{i,t} = & \alpha_0 + \alpha_1 ElectoralReform_t + \alpha_2 + \\
 & \alpha_3(ElectoralReform_t \times t) + \\
 & \theta C_{i,t} + \epsilon_{it},
 \end{aligned} \tag{0.1}$$

where the dependent variable,  $Pork_{i,t}$ , is mean of pork-barrel questions aggregated at quarterly individual level,  $ElectoralReform_t$  is a dummy variable indicating whether the observation is from the period of the post reform,  $C_{i,t}$  includes fixed effects for municipality  $i$  in election  $t$ , and  $\epsilon_{it}$  is the error term. Table 3 shows a negative and significant effect on legislators' motivation to ask pork-barrel questions after reform, both for *Full Model* and *Large Parties*,

suggesting that changing to SMD decreases legislators' incentives to pay attention to parochial interests. Thus, we find generally consistent empirical evidence in favour of **the Hypothesis 1**.

Yet, the heterogeneity issues resulting from natural disasters and political incidents occurred in years that could potentially influence legislators' incentive to ask specific questions or ask more (less) questions. For example, in 2006, a mass movement (*Million Voices against Corruption, President Chen Must Go*) in Taipei Liberty Square led by former DPP Chairman Shih Ming-teh pressured President Chen Shui-bian to resign. This incident was in the media for more than a year, from 2007 to 2008, when the reform was just in practice. To diminish the problem concerning the heterogeneity, we control for the fixed effects from municipalities and legislators across the years. Still, the reform is statistically significant negative, and robust in models (2) and (4).

As such, the empirical evidence shows that increases in time were associated with higher motivation levels to ask pork questions under the old system, SNTV-MMD, and lower levels under SMD. This suggests the institutional change reduces legislators' tendency to ask for the pork-barrel project between KMT and DPP, even if we introduce municipality dummies, legislator dummies, control for the passage of time and other demographics.

To distinguish any possible effects from the passage of time and control for differences between municipality and legislators, we ran the same regression as above by using a subsample only including the large parties (KMT and DPP) and small parties. However, the presence of the reform has an insignificant impact on legislators from *small parties*, revealing that the effects of institutional change do not substantively discourage most small-party members from asking pork questions. Under the SMD, the minority legislators were squeezed for living space in the legislature, which subsequently motivated them to make more efforts to appeal to parochial groups of voters and ask more pork-barrel questions to the government on behalf of their voters in the district.

## Discussion

This chapter contributes to the literature on electoral systems and political representation by demonstrating how the institutional change decrease legislators' incentive to run on their per-

**Table 3:** REGRESSION ANALYSIS FOR PORK-BARREL QUESTIONS

	Full Observation		Large Parties		Small Parties	
	Model (1) Interaction	Model (2) + Controls	Model (3) Interaction	Model (4) (+ Controls)	Model (5) Interaction	Model (6) (+ Controls)
Reform	-14.052*** (3.946)	-13.125*** (5.085)	-14.959*** (4.173)	-16.346*** (5.569)	-11.180 (8.365)	4.954 (19.928)
Year	-.006*** (0.001)	-.006*** (0.001)	-.006*** (0.001)	-.008*** (0.001)	-.002 (0.002)	-.001 (0.003)
Reform $\times$ Year	.006*** (0.001)	.006*** (0.002)	.007*** (0.002)	.008*** (0.002)	.005 (0.004)	-.002 (0.009)
Constant	12.610*** (2.129)	12.713*** (2.735)	13.732*** (2.408)	16.206*** (3.672)	5.091 (5.614)	2.712 (6.546)
District Fixed Effects		✓		✓		✓
Legislator Fixed Effects		✓		✓		✓
The Reform Year	✓	✓	✓	✓	✓	✓
Observations	4,252	4,252	3,539	3,539	713	713
Adjusted $R^2$	0.016	0.218	0.020	0.233	0.003	0.179

Note: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Robust standard errors in parentheses. The district-level pork-barrel parliamentary question at quarterly is regressed on year, electoral reform, and interaction between year and electoral reform, with and without fixed effect controls. The controls include fixed effects for electoral districts, municipalities, and individual legislators and the occurrence of the reform.

sonal reputation by adopting electoral strategies that target the median voter. Combining a dedicated deep learning algorithm with robust regression analysis, we estimate the impacts of the electoral reform through SNTV-MMD to SMD on legislative behaviour. The results show that Taiwan's SMD diminishes legislators' motivation to mention pork-barrel features and increases their awareness of regulatory policies in written parliamentary questions. This finding is in line with a recent study by [Catalinac \(2016\)](#); [Ishima \(2020\)](#) showing that the single-member district system increases elected representatives' attention to national policies ([Catalinac, 2016](#)). Nevertheless, the institutional change has a moderate impact on legislators from smaller parties. This is due to the fact that the new system introduced in Taiwan is particularly disadvantageous to small parties ([Duverger, 1954](#); [Reed, 2001](#); [Huang, 2017](#); [Bawn and Thies, 2003](#)), which deteriorates the effect of the reform on their political behaviours and electoral strategies with the constituencies.

This paper thus comes with some limitations. First, it is noteworthy that there has been a consistent decline in the total number of questions since 2003, despite the reform significantly affecting legislative representation. Specifically, in recent years, social media has emerged as a crucial platform for legislators to establish positions and influence agenda-setting ([Barberá, 2015](#); [Barberá et al., 2019](#); [Schürmann and Stier, 2022](#)).

Another limitation arises from language evolution and its variations over time. The training data was annotated between 2007 and 2009. In other words, the deep learning architecture is trained solely by learning to identify pork barrelling features in legislation within a specific, limited timeframe, which may fail to fully uncover implicit notions developed after 2009. However, BERT's self-attention mechanisms could potentially aid the machine classifier by simulating a skill set similar to that of the human brain, enabling it to identify more complex or previously unseen concepts derived from the labeled data and understand the underlying pork barrel features.

Measuring legislators' preferences towards constituencies under different systems is essential for understanding electoral systems and their impacts on political representation. This paper examines a persistent myth about the impact of the 2008 electoral reforms in Taiwan. While scholars argue about the reform, some evidence suggests that Single-Member Districts (SMD) may not be as effective as initially presumed in addressing the drawbacks of the Single Non-

Transferable Vote (SNTV) system, such as factional politics and intraparty competition (e.g., [Wu, 2003](#); [Batto et al., 2018](#), ). Empirical evidence shows that, over time, there was a decrease in the motivation to run on personal reputation despite intense conflicts between parties when the reform occurred ([Liao, 2020](#)).

## References

- Bäck, Hanna, Markus Baumann and Marc Debus. 2019. "Coordination of Legislative Speech in Times of Crisis: Youth Unemployment and Debates on Redistributive Policies in the Swedish Riksdag, 1994–2014." *International Journal of Social Welfare* 28(4):404–417.
- Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23(1):76–91.
- Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost and Joshua A. Tucker. 2019. "Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data." *American Political Science Review* 113(4):883–901.
- Batto, Nathan, Chi Huang, Alexander Tan and Gary Cox. 2018. *Mixed-Member Electoral Systems in Constitutional Context: Taiwan, Japan, and Beyond*. University of Michigan Press.
- Batto, Nathan F. and Hsin Ta Huang. 2016. Executive Competition, Electoral Rules, and Faction Systems in Taiwan. In *Mixed-Member Electoral Systems in Constitutional Context: Taiwan, Japan, and Beyond*, ed. Nathan F. Batto, Chi Huang, Alexander C. Tan and Gary W. Cox. University of Michigan Press chapter Chapter 4.
- Bawn, Kathleen and Michael F. Thies. 2003. "A Comparative Theory of Electoral Incentives: Representing the Unorganized under PR, Plurality and Mixed-member Electoral Systems." *Journal of Theoretical Politics* 15(1):5–32.
- Cain, Bruce, John Ferejohn and Morris Fiorina. 1987. *The Personal Vote: Constituency Service and Electoral Independence*. Harvard University Press.
- Carey, John M and Matthew Soberg Shugart. 1995. "Incentives to Cultivate a Personal Vote: A Rank Ordering of Electoral Formulas." *Electoral studies* 14(4):417–439.
- Catalinac, Amy. 2016. "From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections." *Journal of Politics* 78(1):1–18.
- Catalinac, Amy. 2017. "Positioning under Alternative Electoral Systems: Evidence from Japanese Candidate Election Manifestos." *American Political Science Review* 112(1):31–48.
- Chatsiou, Kikia and Slava Jankin Mikhaylov. 2020. "Deep Learning for Political Science." *The SAGE Handbook of Research Methods in Political Science and International Relations* pp. 1053–1078.
- Cox, Gary W. 1990. "Centripetal and Centrifugal Incentives in Electoral Systems." *American Journal of Political Science* 34(4):903–935.
- Cox, Gary W. 1996. "Is the Single Nontransferable Vote Superproportional? Evidence from Japan and Taiwan." *American Journal of Political Science* 40(3):740–55.
- Cox, Gary W and Emerson Niu. 1994. "Seat Bonuses under the Single Nontransferable Vote System: Evidence from Japan and Taiwan." *Comparative Politics* 26(2):221–236.

- Devlin, Jacob, Ming Wei Chang, Kenton Lee and Kristina Toutanova. 2019. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." 1.
- Downs, Anthony. 1990. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65(2).
- Duverger, Maurice. 1954. *Political Parties : Their Organization and Activity in the Modern State*. New York: John Wiley & Sons.
- Goplerud, Max and Daniel M. Smith. 2021. "Who Answers for the Government? Bureaucrats, Ministers, and Responsible Parties." *American Journal of Political Science* 00(0):1–16.
- Guinaudeau, Isabelle and Olivier Costa. 2021. "Issue Politicization in the European Parliament. An Analysis of Parliamentary Questions for Oral Answer (2004–19)." *Journal of Common Market Studies* pp. 1–19.
- Hirano, Shigeo. 2006. "Electoral Institutions, Hometowns, and Favored Minorities: Evidence from Japanese Electoral Reforms." *World Politics* 59(1):51–82.
- Høyland, Bjørn and Martin G. Søyland. 2019. "Electoral Reform and Parliamentary Debates." *Legislative Studies Quarterly* 44(4):593–615.
- Huang, Chi. 2017. Electoral System Change and Its Effects on the Party System in Taiwan. In *The Taiwan Voter*, ed. Christopher H. Achen and T. Y. Wang. University of Michigan Press chapter Chapter 10, pp. 223–251.
- Ishima, Hideo. 2020. "How Electoral Reform Alters Legislative Speech: Evidence from the Parliament of Victoria, Australia 1992–2017." *Electoral Studies* 67:102—192.
- Jang, Jinhyeok and Nick C.N. Lin. 2019. "Personal Votes, Electoral Competitiveness of Parties, and Legislative Representation in Taiwan under SNTV." *Japanese Journal of Political Science* 20(1):21–32.
- Kim, Yoon. 2014. "Convolutional Neural Networks for Sentence Classification." *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference* pp. 1746–1751.
- Kim, Yoon, Yacine Jernite, David Sontag and Alexander M. Rush. 2016. "Character-Aware Neural Language Models." *30th AAAI Conference on Artificial Intelligence, AAAI 2016* pp. 2741–2749.
- Kolpinskaya, Ekaterina. 2017. "Substantive Religious Representation in the UK Parliament: Examining Parliamentary Questions for Written Answers, 1997–2012." *Parliamentary Affairs* 70(1):111–131.
- Liao, David Yen-Chieh. 2020. "Electoral Reform and Fragmented Polarization: New Evidence from Taiwan Legislative Roll Calls." *Presentation at 2020 Annual Meeting of The American Political Science Association*.
- Lopez, Marc Moreno and Jugal Kalita. 2017. "Deep Learning applied to NLP."
- Lu, Xiaolei and Bin Ni. 2019. "BERT-CNN: a Hierarchical Patent Classifier Based on a Pre-

Trained Language Model.”

**URL:** <http://arxiv.org/abs/1911.06241>

- Luor, Ching Jyuhn and Chien-liang Liao. 2009. “The Impact of Upcoming Changes in the Electoral System on the Initiation Behavior of Pork Barrel- related Bills from Legislators.” *Taiwanese Political Science Review (in Chinese)* 13(1):3–53.
- Luor, Ching Jyuhn and Fu Yao Chan. 2012. “The Relationship between Particularistic Benefit Bills Initiated by Legislators and the Geographic Distribution of Categorical Grants in Taiwan, FY2005-FY2009.” *Journal of Public Administration* 42:1–32.
- Luor, Ching Jyuhn and Ying-Shih Hsieh. 2008. “The Association of District Size and Pork Barrel Related Bills Initiated by Legislators: An Analysis on the 3rd and 4th Legislative Yuan in Taiwan.” *Public Administration & Policy (in Chinese)* 00(48):1–48.
- Martin, Shane. 2011. “Parliamentary Questions, the Behaviour of Legislators, and the Function of Legislatures: An Introduction.” *The Journal of Legislative Studies* 17(3):259—270.
- Martin, Shane and Olivier Rozenberg. 2017. Roles and Function of Parliamentary Questions. In *Roles and Function of Parliamentary Questions*. Routledge.
- Myerson, Roger B. 1993. “Incentives to Cultivate Favored Minorities Under Alternative Electoral Systems.” *American Political Science Review* 87(4):856–869.
- Park, Ju Yeon. 2021. “When Do Politicians Grandstand? Measuring Message Politics in Committee Hearings.” *Journal of Politics* 83(1):214–228.
- Proksch, Sven-Oliver" and Jonathan B. Slapin. 2009. “How to Avoid Pitfalls in Statistical Analysis of Political Texts: The Case of Germany.” *German Politics* 18(3):323–344.
- Proksch, Sven Oliver and Jonathan B. Slapin. 2015. *The Politics of Parliamentary Debate: Parties, Rebels, and Representation*. Cambridge University Press.
- Reed, Steven R. 1994. “Democracy and the Personal Vote: A Cautionary Tale from Japan.” *Electoral Studies* 12(1):17–28.
- Reed, Steven R. 1995. “The Nomination Process for Japan’s Next General Election: Waiting for the Heiritsu-sei.” *Asian Survey* 35(12):1075–1086.
- Reed, Steven R. 2001. “Duverger’s Law is Working in Italy.” *Comparative Political Studies* 34(3):312–327.
- Richardson, Bradley M. 1988. “Constituency Candidates versus Parties in Japanese Voting Behavior.” *American Political Science Review* 82(3):695–718.
- Russo, Federico. 2021. “Going Local: Parliamentary Questions as a Means of Territorial Representation in the Italian Parliament.” *Political Studies Review* 19(3):410–427.
- Saalfeld, Thomas. 2011. “Parliamentary Questions as Instruments of Substantive Representation: Visible Minorities in the UK House of Commons, 2005–10.” *Journal of Legislative Studies* 17(3):271–289.



- Safaya, Ali, Moutasem Abdullatif and Deniz Yuret. 2020. "KUISAIL at SemEval-2020 Task 12 : BERT-CNN for Offensive Speech Identification in Social Media." pp. 2054–2059.
- Schürmann, Lennart and Sebastian Stier. 2022. "Who Represents the Constituency? Online Political Communication by Members of Parliament in the German Mixed-Member Electoral System." *Legislative Studies Quarterly* pp. 1–16.
- Sheng, Shing-Yuan. 2014*a*. "Change and Continuity in Legislators' Bill-Introduction Before and After the Reform of the Electoral System: An Exploratory Analysis." *Taiwanese Political Science Review* 18(1):73–127.
- Sheng, Shing-Yuan. 2014*b*. "Change and Continuity in Legislators' Bill-Introduction Before and After the Reform of the Electoral System: An Exploratory Analysis." *Taiwanese Political Science Review (in Chinese)* 18(1):73–127.
- Shugart, Matthew Soberg and Martin P. Wattenberg. 2003. *Mixed-Member Electoral Systems The Best of Both Worlds?* Oxford University Press.
- Slapin, Jonathan B., Justin H. Kirkland, Joseph A. Lazzaro, Patrick A. Leslie and Tom O'Grady. 2018. "Ideology, Grandstanding, and Strategic Party Disloyalty in the British Parliament." *American Political Science Review* 112(1):15–30.
- Slapin, Jonathan B. and Sven-Oliver Proksch. 2014. Words as Data: Content Analysis in Legislative Studies. In *The Oxford Handbook of Legislative Studies*, ed. Shane Martin, Thomas Saalfeld and Kaare W. Strøm. Oxford University Press pp. 1–22.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser and Illia Polosukhin. 2017. "Attention is All You Need."
- Wang, Ye-li. 1994. "Promoting Political Reform and Improving Party Politics." *Chinese Kuomintang and National Development* 112:190–197.
- Wilson, James Q and John J Dilulio. 2001. *American Government Institutions and Policies*. In *American Government Institutions and Policies*. MA: Houghton Mifflin Co.
- Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest and Alexander Rush. 2020. "Transformers: State-of-the-Art Natural Language Processing." pp. 38–45.
- Wu, Chung-li. 2003. "Local Factions and the Kuomintang in Taiwan's Electoral Politics." *International Relations of the Asia-Pacific* 3(1):89–111.
- Zhang, Jianwei, Zhenxing Wang, Yuhui Zheng and Guoqing Zhang. 2017. "Cascade Convolutional Neural Network for Image Super-resolution." *Journal of Advanced College of Engineering and Management* 3:1–10.
- Zhang, Xiang, Junbo Zhao and Yann Lecun. 2016. "Character-level Convolutional Networks for Text Classification."

# **Appendices and Supplemental Material:**

## **Electoral Reform and Pork Barrel in Parliamentary Questions**

### **Contents**

<b>1</b>	<b>The Description of Training Data for TextCNN, BERT and CNN-BERT</b>	<b>1</b>
<b>2</b>	<b>Experiments and Results</b>	<b>2</b>

# 1 The Description of Training Data for TextCNN, BERT and CNN-BERT

**Table A.1:** THE DESCRIPTION OF TRAINING DATA FOR TEXTCNN, BERT AND CNN-BERT

	Text CNN			BERT			CNN-BERT		
	Train	Test	Dev	Train	Test	Dev	Train	Test	Dev
<b>Pork</b>	4259	237	237	4259	237	237	4259	237	237
<b>Non Pork</b>	2259	125	126	2259	125	126	2259	125	126
	7243			7243			7243		

Source: Taiwan Pork Barrel Legislation Data ([Luor and Hsieh, 2008](#); [Luor and Liao, 2009](#); [Luor and Chan, 2012](#))

## 2 Experiments and Results

In this section, BERT-CNN will be compared with other models such as TextCNN and generic BERT on their abilities to identify pork-barrel features on the holdout set from 15 % of all pork-barrel legislation. First, **TextCNN** is the generic model without having a BERT embedding layer. TextCNN here uses initialized spaCy's pretrained vectors and the Chinese model used in [Table B.2](#) can be found at <https://spacy.io/models/zh>.

As to the **BERT** model in the second column in [Table B.2](#), I transformed the corpus into word vector representation extracted from BERT with max words 512. At last column, **CNN-BERT** is approach method used to classified pork-barrel. **CNN-BERT** maximizes the utilization of knowledge embedded in pre-trained BERT language models by feeding the contextualized embeddings into several filters and convolution layers of the CNN architecture.

A intuitive approach to evaluate the performance of classifiers is to compare each model's performance on *accuracy* and *recall*, respectively. Precision is constructed by the accuracy of the positive predictions, whereas the ratio of positive instances correctly detected by the classifiers is recall. However, it is much more useful to consider measurement together, like the F1 score. As in Equation [B.2](#), the harmonic mean gives much more weight to low values, whereas the regular mean treats all values equally. As a result, the classifier gets a high *F1 score* the performance of the classifier works better in both recall and precision.

In general, BERT's performance achieves 96% in F1-Score in the weighted average, slightly better than CNN-BERT and then TextCNN. EarlyStop function with patience = 5 from Tensorflow is used to monitor the performance of the training process, which the model stops when there is no improvement within five epochs. Text CNN model was trained until 13 epochs, while the generic BERT model finished at 12 epochs with a learning rate of 1e-08. Last, CNN-BERT costs less time than BERT, completed at 9 epochs.<sup>7</sup>

Last but not least, BERT is a powerful model that has achieved outstanding performance in multiple domains. However, it does indeed require a significant amount of computational resources and training time, making it potentially less ideal in resource-constrained scenarios. In this paper, we present an alternative solution. An effective approach is to extract BERT embeddings, enabling the acquisition of high-quality textual representations without the need for the entire BERT model. This method can significantly reduce computational time and resource requirements while retaining the effectiveness of BERT.

Furthermore, the quality of training data is also a important factor. Even with limited resources, high-quality training data can simplify model requirements, enabling simpler models

---

<sup>7</sup>TensorBoard tracking different epochs of the experiment is available at <https://tensorboard.dev/experiment/2Jm6GKexQiKaLyUz5uKSzg/#scalars>.

to achieve good results. Therefore, when choosing a model, it's essential to balance computational resources, performance, and application requirements. If resources are limited but high-quality training data is available, even simpler models may generate satisfactory results.

In contrast, traditional static embeddings like Word2Vec are also a good option but may not capture the complete contextual information because they are constructed based on statistical information of individual words. On the other hand, extracting embeddings from BERT may help overcome this limitation.

Therefore, when choosing a model, it is important to strike a balance between computational resources, performance, and application requirements. If computational resources are limited, utilizing extracted embeddings from a transformer-based model like BERT might be a ideal compromise in a resource-constrained setting.

$$recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (B.1)$$

$$precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (B.2)$$

$$\begin{aligned} F_1 &= \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = 2 \times \frac{precision \times recall}{precision + recall} \\ &= \frac{TruePositive}{TruePositive + \frac{FalseNegative \times FalsePositive}{2}} \end{aligned} \quad (B.3)$$

**Table B.2:** THE PERFORMANCE OF CNN, BERT AND CNN-BERT

	CNN			BERT			CNN-BERT		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
<b>Non Pork</b>	0.95	0.97	0.96	0.96	0.97	0.97	0.96	0.97	0.97
<b>Pork</b>	0.94	0.91	0.92	0.95	0.93	0.94	0.95	0.93	0.95
<b>Accuracy</b>			0.95	0.96		0.96			0.96
<b>Macro Avg.</b>	0.95	0.94	0.94	0.96	0.95	0.95	0.95	0.94	0.94
<b>Weighted avg.</b>	0.95	0.95	0.95	0.96	0.96	0.96	0.95	0.95	0.95