

# Mapping Jurisprudence of WTO Dispute Settlement Body Using Deep Learning

Suyeol Yun

December 13, 2020

## Abstract

The world trade organization (WTO) legally regulates the world trade system with its dispute settlement body (DSB). There exists a shared understanding among legal experts about how articles of WTO agreements interact with each other. Those understanding has been exclusively shared inside the group of experts. However, the complexity of the WTO legal framework has constrained many developing countries with limited legal knowledge and resources from fully utilizing the WTO DSB. To address this issue, I propose a new method that materializes the relationship between articles of WTO agreements. I collected past 20 years of WTO disputes and trained a neural network that mimics the reasoning process of legal experts that determines which articles to cite for given factual circumstances of legal dispute. Then I collected all the predictions from the trained neural network and fitted a network of articles using Random Forest. I found that this fitted network captures the important interactions between articles closely to the authoritative jurisprudences of WTO DSB. The pattern of interactions are shown as explained in the Panel and Appellate Body reports of the WTO DSB.

## 1 Introduction

The Dispute Settlement Body (DSB) of the World Trade Organization (WTO) deals with trade disputes between WTO members. WTO members file a lawsuit in WTO DSB to claim their impaired benefit related to the WTO agreements as a result of possible illegal action of another member's trade policy. Then a judicial body of WTO DSB, *Panel* or *Appellate Body*, adjudicates the dispute and submits a report in which it expresses its judicial opinion as to whether the challenged trade policy is inconsistent to the rules of the WTO or not (World Trade Organization, 2017).

A lawsuit tends to cite multiple rules of the WTO agreement because one simple rule can't cover the complex characteristics of the trade policy that led to the dispute (Palmer and Mavroidis,

2004). For example, the United States enacted *Continued Dumping and Subsidy Act of 2000* (CDSOA) that distributes the collected anti-dumping duties to its affected domestic producers. This act was challenged by other members with several rules of the WTO agreements such as *rules of anti-dumping* and *rules of subsidy*. This was because this distribution could constitute an illegal subsidy and illegal anti-dumping duty at the same time (*See* Figure 1). Citation of articles of the WTO agreements are evidence of how WTO rules are applied in the real world. For example, principle of *Market Access* which is one of the most important principles in WTO is achieved through a cooperation of multiple articles of the WTO agreements (Figure 23) and those cooperation is realized through the act of citation by members. Therefore, analyzing the network of articles of the WTO agreements provides us a clear view on how WTO DSB constitutes specific norms and regulates specific trade issues. This kind of study has been actively pursued by a group of researchers in numerous literatures (Bown, 2002; Charnovitz, 2004; Trachtman, 2003; Johns and Pelc, 2014), however, those efforts has been limited to study interconnectedness between relatively small numbers (less than 10) of articles. There exists an shared understanding about how rules of the WTO agreements are interconnected as a whole, however, this entire map is exclusively shared among a group of legal experts of the WTO agreements. This exclusiveness becomes more severe as the number of cases requested to WTO DSB increases and this has led to a widening gap of legal capacity between developing and developed countries. This gap now inhibits the effectiveness of the WTO because developing countries are excluded from the WTO DSB to resolve their dissatisfaction over the trade relationship with other members. Therefore, this paper emphasizes the importance of revolutionizing the way of studying network of articles of WTO agreements and proposes a new method that materializes the relationship between the articles of the WTO agreements as a whole. For this purpose, this paper maps the jurisprudence of WTO DSB as a network of legal articles as formally defined in Figure 2 and illustrated in Figure 3. This is because the rules of the WTO agreements explicitly requires *Panel* or *Appellate Body* to address relevant articles together when they construct its jurisprudence related to the meaning, scope and interpretation of any legal text in the WTO agreements as excerpted in Figure 4. Upon this requirement, judicial bodies cite multiple articles together to identify the complex legal identity of a trade policy at issue as clearly opinionated in Figure 1. In addition to it, judicial bodies cite multiple articles together to guide an way of interpretation of the rules of the WTO agreements (*See* Figure 3(b)). To develop a proper method that can find a set of directed edge weights  $W$  defined in Figure 2 as close to a shared understanding of legal experts, this paper points out two main considerations. First, one need to use information inside a textual description of factual

8.1 In the light of our findings, we conclude that **the CDSOA is inconsistent with AD (Anti-dumping) Articles 5.4, 18.1 and 18.4, SCM (Subsidy and Countervailing Measure) Articles 11.4, 32.1 and 32.5**, Articles VI:2 and VI:3 of the GATT 1994, and Article XVI:4 of the WTO Agreement. . . .

8.3 **The CDSOA is a new and complex measure, applied in a complex legal environment.** In concluding that the CDSOA is in violation of the above mentioned provisions, we have been confronted by sensitive issues regarding the use of subsidies as trade remedies. this matter through negotiation.

Figure 1: **Panel’s Judicial Opinion On the *US - Offset (Byrd Amendment; CDSOA)* case:** Panel explicitly expresses the complexity of the trade policy (CDSOA) at issue and cites the rules of anti-dumping (AD) and subsidy (SCM) at the same time to cover its complex characteristics.

circumstances of the legal dispute and the regulatory contents described in the text of each article of the WTO agreements. Second, one need to generalize the members’ strategic citation pattern that is limited to a member’s specific political interest. For example, members strategically cite different rules of the WTO agreements to limit or to encourage the third party participation. Since the third party participation can lead to early settlement of the dispute without continuous legal battle, members cite differently according to their intention to settle the case earlier out of court (Johns and Pelc, 2014). Moreover, members cite articles strategically trying to reshape the legal precedents of WTO DSB in favor of their future interest (Pelc, 2014; Strezhnev, 2014). Upon these two considerations, this paper selects the deep neural network as a technical solution. This is because a deep neural network is empirically known as good at extracting information from text data and generalizing the patterns inside data. Therefore, this paper designs a deep neural network (See Figure 16 and 18) that processes two different types of textual information. One is textual description of the dispute (See an example at Appendix A.1) and the other one is the text of a legal article of the WTO agreements (See an example at Figure 5). This design is improvised to mimic the reasoning process of WTO legal practitioners where the legal practitioners read the textual description of factual circumstances of the dispute and imagine applicable regulatory contents of the legal articles while he/she reads that factual description of the case (See Figure 14, 15 and 16).

To train this neural network, this paper collected textual description of trade policy that led to the dispute and articles of the WTO agreement cited for each dispute case requested to the WTO DSB from 1995 to 2018 (Total 143 cases. Check the list in Appendix A.2). Using this collected

Network of legal articles of WTO agreements is defined as

$$\text{directed weighted graph } G = (V, E, W)$$

where *vertex set*  $V = \{v \mid v \text{ is a legal article of WTO agreement}\}$ ,

*set of directed edges*  $\vec{E} = \{(v_i, v_j) \mid (v_i, v_j) \in V \times V\}$  and

$$w : V \times V \rightarrow \mathbb{R}_+ \text{ s.t. } w(v_j, v_j) = 0 \text{ and } \sum_{v_i \in V} w(v_i, v_j) = 1 \forall v_j \in V$$

Then define *edge weight matrix*  $W = (w_{ij}) \in \mathbb{R}_+^{|V| \times |V|}$  s.t.  $w_{ij} = w(v_i, v_j)$

( $W$  is more formally called *weighted adjacency matrix*)

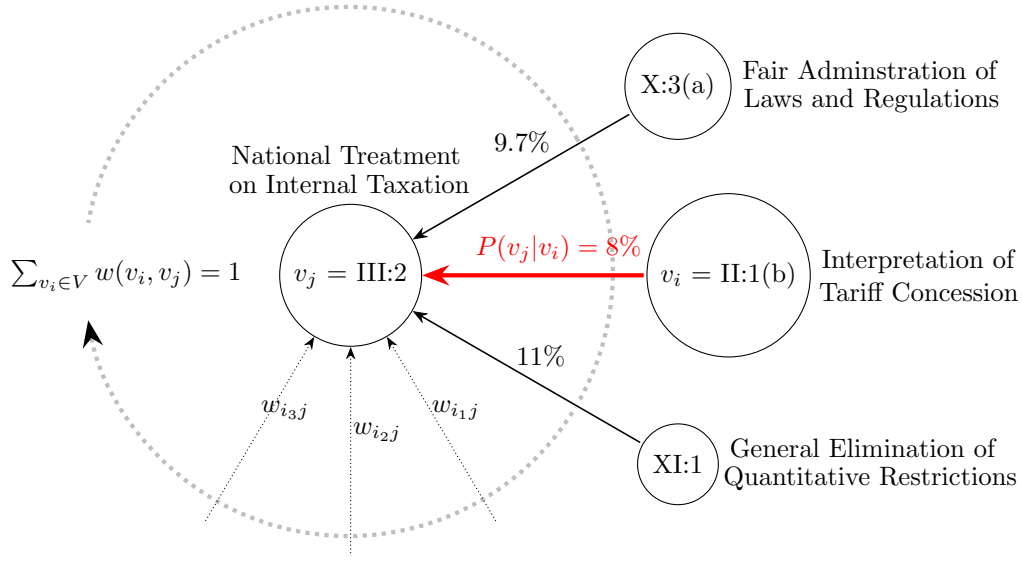
**Figure 2: Formal Definition of Network of Legal Articles of WTO agreements:** I define network of legal articles of WTO agreements as a directed weighted graph where the sum of all weights coming into a node sum up to 1.  $w_{ij}$  is interpreted as conditional probability  $P(v_j|v_i)$  how probably a source node  $v_i$  clarifies the meaning of the target node  $v_j$  compared to other source nodes as illustrated in Figure 3

data, I trained the neural network by enforcing the neural network to answer correctly whether a given article of the WTO agreements can be cited for the given textual description of trade policy. (See Figure 16 and Figure 17).

After training, I fitted a set of directed edge weights  $W^*$  that best explains the variance of each article's citabilities that are predicted by the trained deep neural network using a machine learning technique *Random Forests* (Breiman, 2001; Huynh-Thu et al., 2010).

After fitting  $W^*$ , to check whether this fitted network of articles of the WTO agreements  $G^* = (V, E, W^*)$  maps the jurisprudences of WTO DSB properly, this paper compares the way of the fitted network  $G^*$  explaining how articles of WTO agreements achieve some main principles of WTO, such as *Market Access*, *Reciprocity* and *Non-discrimination* with the jurisprudences of *Panel* and *Appellate Body*. This comparison reveals that the fitted network  $G^*$  captures the interactions between the articles of WTO agreements closely to the jurisprudences of *Panel* and *the Appellate Body*. We can infer from this similarity that the fitted network  $G^*$  closely maps the jurisprudences of WTO DSB. This is because those two judicial bodies authoritatively constitute the jurisprudences over how rules of WTO agreements are working together to achieve those main principles.

Finally, upon this similarity, this paper offers this methodology as an alternative solution to the widening gap of legal capacity between developing and developed countries in WTO DSB. Since this method effectively materializes the shared understanding of legal experts and reveals important interactions between articles inside the system of WTO DSB, it can lower the cost to



(a) Illustrated edge weights of a target node **Article III:2**

“The dictionary definition of the noun ‘excess’ is ‘[t]he amount by which one number or quantity exceeds another’. More specifically, ‘in excess of’ means ‘more than’. Thus, as a textual matter, a particular number or quantity is ‘in excess of’ another number or quantity if it is greater, regardless of the extent to which it is greater. ***Looking at the context of Article II:1(b), first sentence, we note that Article III:2, first sentence, of the GATT 1994 is cast in very similar terms and in fact uses the phrase ‘in excess of’:***

The products of the territory of any contracting party imported into the territory of any other contracting party shall not be subject . . . to internal taxes or other internal charges of any kind in excess of those applied . . . to like domestic products . . .

(b) **Jurisprudence of Panel in *Russia – Tariff Treatment* case:**

Panel clarifies the point that the meaning of the term ‘*in excess of*’ in Article II:1(b) clarifies the meaning of the same phrase in Article III:2.

**Figure 3: Illustration of Network of Legal Articles of WTO agreements:** Every directed edge weight  $w_{ij}$  is interpreted as the conditional probability  $P(v_j|v_i)$  of how probably a source node  $v_i$  constitutes a legal context to clarify the meaning of the target node  $v_j$  among all other source nodes  $v \in V \setminus \{v_i, v_j\}$ . Above subfigure (a) represents how jurisprudence of *Panel* stated in (b) is represented as an edge weight where the source node Article II:1(b) constitutes the legal context of the target node Article III:2 regarding how to interpret its term ‘*in excess of*’ with the 8% of importance compared to other possible source articles.

## Article 7

### Terms of Reference of Panels

1. Panels shall have the following terms of reference unless the parties to the dispute agree otherwise within 20 days from the establishment of the panel:

“To examine, **in the light of the relevant provisions** in (name of the covered agreement(s) cited by the parties to the dispute), the matter referred to the DSB by (name of party) in document . . . and to make such findings as will assist the DSB in making the recommendations or in giving the rulings provided for in that/those agreement(s).”

2. **Panels shall address the relevant provisions** in any covered agreement or agreements cited by the parties to the dispute. . . .

Figure 4: **Explicit Requirement of Interconnecting Articles of WTO agreements:** Dispute Settlement Understanding (DSU) provides a legal guidelines on how judicial bodies of WTO DSB shall adjudicate the requested disputes. It explicitly requires judicial bodies to interweave relevant articles of the WTO agreements to clarify it’s meaning, scope and interpretation.

## Article I

### *General Most-Favoured-Nation Treatment*

1. With respect to customs duties and charges of any kind imposed on or in connection with importation or exportation or imposed on the international transfer of payments for imports or exports, and with respect to the method of levying such duties and charges, and with respect to all rules and formalities in connection with importation and exportation, and with respect to all matters referred to in paragraphs 2 and 4 of Article III, any advantage, favour, privilege or immunity granted by any contracting party to any product originating in or destined for any other country shall be accorded immediately and unconditionally to the like product originating in or destined for the territories of all other contracting parties...

Figure 5: **Example of a legal article of the WTO agreements:** Article I:1 of General Agreement on Tariffs and Trade 1994 that prohibits the discrimination between members of WTO.

build the same amount of legal capacity to understand the WTO DSB. Moreover, rather than keep relying on previous approach that provides legal advice to developing countries that does not create a shared understanding over the system between developing and developed countries, if we shift our focus on how to materialize the current shape of the system, WTO will become more effective as members being able to discuss their trade issues upon the measurable ground of shared understanding about how WTO works.

## 2 Data: Types, Composition and Collection Process

This section explains the composition of data and its collection process in detail.

### 2.1 Overview: How Members Raise Claims in WTO DSB

As explained in the introduction, a trade policy that led to a dispute (preferably called as *Government Measure* in WTO DSB) is pretty much complicated as explicitly expressed by the Panel in Figure 1. To address this complexity, members who raise the claim (preferably called *complainant* in WTO DSB) usually cite multiple articles of the WTO agreements at the same time. For example, in the *US - Offset* case, a group of complainants<sup>1</sup> cited articles as shown in Table 1 from the WTO agreements to claim its inconsistencies of *Continued Dumping and Subsidy Act of 2000* (CDSOA) to those cited articles<sup>2</sup>:

Name of WTO Agreement	Cited Articles
Agreement on Anti-dumping	1, 5.4, 8, 18.1, 18.4
General Agreement on Tariffs and Trade 1994	VI:3, X:3, XXIII:1, VI:2
Agreement on Subsidies and Countervailing Measures	4.10, 7.9, 10, 11.4, 18, 32.1, 32.5
Agreement Establishing the World Trade Organization	XVI:4

Table 1: Cited articles in *US - Offset (Byrd Amendment)* by complainants

Upon this understanding, I collected two different types of data for 143 different dispute cases requested to WTO DSB. (List of cases is available at Appendix A.2). One is textual description of the dispute (*Check* the CDSOA example at Appendix A.1) and the other one is set of articles of the WTO agreements that are cited for each dispute (Appendix A.3). I will explain the data

<sup>1</sup> Australia, Brazil, Chile, European Communities, India, Indonesia, Japan, Korea and Thailand

<sup>2</sup>It is worth noting that the WTO agreements comprises many different agreements covering each specific topic in trade such as *Agreement on Anti-dumping*, *Agreement on Subsidies and Countervailing Measures*, *Agreement on Agriculture* and so on.

source, structure and collection method for two different types of data at the following subsections.

		WT/DS217/R WT/DS234/R Page i
<b>TABLE OF CONTENTS</b>		
		<u>Page</u>
<b>I.</b>	<b>INTRODUCTION .....</b>	<b>1</b>
<b>II.</b>	<b>FACTUAL ASPECTS .....</b>	<b>2</b>
<b>III.</b>	<b>PARTIES' REQUESTS FOR FINDINGS AND RECOMMENDATIONS .....</b>	<b>4</b>
A.	COMPLAINING PARTIES .....	4
B.	UNITED STATES .....	5
<b>IV.</b>	<b>ARGUMENTS OF THE PARTIES .....</b>	<b>5</b>
A.	FIRST WRITTEN SUBMISSION OF THE COMPLAINING PARTIES.....	6
1.	<b>Australia .....</b>	<b>6</b>
2.	<b>Brazil .....</b>	<b>9</b>
3.	<b>Canada .....</b>	<b>15</b>
4.	<b>Chile and Japan .....</b>	<b>24</b>
5.	<b>European Communities, India, Indonesia and Thailand .....</b>	<b>32</b>
6.	<b>Korea .....</b>	<b>40</b>
7.	<b>Mexico .....</b>	<b>45</b>

Figure 6: **Table of Contents of Panel Report:** Panel provides factual aspect in the panel report with its page location.

## 2.2 Factual Aspect: Textual Description of the Dispute

Textual description of the dispute is preferably called as *Factual Aspect* in WTO DSB. Since Panels always provide a factual aspect<sup>3</sup> that summarizes the content of the dispute in the panel report, I wrote a program that can automatically search and collect the panel reports from the WTO official document website<sup>4</sup>. Then I located the factual aspect using the page information inside the table of contents in the panel report as shown in Figure 6. By using this location, I excerpted 143 numbers of different factual aspects from the same number of different panel reports. The case numbers are listed in Figure 7.

### 2.2.1 Joint Adjudication & Early Settlement

The number 143 seems small compared to the total 596<sup>5</sup> number of cases that are requested to WTO DSB. This is due to the following two reasons. First, panel handles different cases together

<sup>3</sup>It's worth noting that Appellate Body doesn't provide any factual aspect because they always use the factual aspect provided by the Panel.

<sup>4</sup><http://docs.wto.org>

<sup>5</sup>As of November 1st, 2020.



if the case is about the same trade policy. For example, in *US - Offset (Byrd Amendment)*, panel merged DS217<sup>6</sup> and DS234 together because they were asking the judicial opinion for the same government measure of the United States as shown in Figure 8. This paper selects the smallest case number as a representative number for this case of joint adjudication. For example, DS217 and DS234 share the same panel report then this paper chooses DS217 as a representative number as shown in Figure 7 where the list includes DS217 but not DS234. Second, members sometimes find *mutually agreeable solution* before the panel expresses its judicial opinion by publishing its panel report. Then Panel stops there and no factual aspect is provided. This paper omitted this kind of *early settled* cases as well.

DS 2, 18, 22, 31, 34, 46, 56, 58, 60, 62, 67, 68, 69, 75, 76, 87, 90, 98, 103, 108, 121, 122, 135, 136, 139, 141, 146, 152, 155, 161, 162, 165, 166, 174, 175, 177, 184, 202, 207, 212, 217, 219, 221, 231, 234, 238, 244, 245, 246, 248, 257, 264, 265, 266, 267, 268, 269, 276, 282, 283, 286, 290, 294, 295, 296, 301, 302, 308, 312, 315, 316, 320, 321, 322, 332, 336, 339, 343, 344, 345, 350, 353, 360, 363, 366, 371, 379, 381, 384, 392, 394, 396, 397, 399, 400, 406, 412, 414, 415, 422, 425, 427, 429, 430, 431, 435, 436, 437, 440, 442, 447, 449, 453, 454, 456, 457, 461, 464, 468, 471, 472, 473, 475, 476, 477, 479, 480, 482, 483, 484, 485, 486, 488, 490, 492, 493, 495, 499, 504, 505, 513, 518, 523

Figure 7: **List of case number of collected panel reports:** “DS + number” uniquely identifies each dispute. For example, DS 523 refers to *US — Pipe and Tube Products (Turkey)* where the United States was challenged by Turkey for its possibly inconsistent anti-dumping measure.

## 2.3 Cited Articles: Set of Articles Cited for the Same Dispute

Every lawsuit in WTO DSB cites multiple set of articles as shown in Table 1. To collect this set of articles claimed for the same dispute, I wrote a program that collects this set of articles cited for the same dispute from the WTO official webpage<sup>7</sup>. The webpage chronologically lists up all dispute cases requested to WTO DSB and the program visits each page of 143 cases and collects the cited articles. Among all the agreements included in the WTO agreements<sup>8</sup> as a component, this paper collected articles from **General Agreement on Tariffs and Trade 1994 (GATT 1994)** only. This is because articles in GATT 1994 constitutes basic set of trade rules of WTO and other

<sup>6</sup>DS refers to *Dispute Settlement*. DS is the official prefix that indicates the case in WTO DSB.

<sup>7</sup>[https://www.wto.org/english/tratop\\_e/dispu\\_e/dispu\\_status\\_e.htm](https://www.wto.org/english/tratop_e/dispu_e/dispu_status_e.htm)

<sup>8</sup>WTO agreements is comprised of multiple agreements, such as General Agreement on Tariffs and Trade 1994, Agreement on Agriculture, Agreement on the Application of Sanitary and Phytosanitary Measures, Agreement on Textiles and Clothing, Agreement on Technical Barriers to Trade, Agreement on Trade-Related Investment Measures, Agreement on Implementation of Article VI of the General Agreement on Tariffs and Trade 1994 (anti-dumping), Agreement on Subsidies and Countervailing Measures, Agreement on Rules of Origin, Agreement on Safeguards and so on.

**UNITED STATES – CONTINUED DUMPING AND SUBSIDY  
OFFSET ACT OF 2000**

***Report of the Panel***

Figure 8: **Cover of a Panel Report Includes Information about Joint Adjudication:** Panel explicitly marks which different cases are handled together in the cover of the panel report. DS217 and DS234 are handled together in this example.

agreements elaborates the articles of GATT 1994 more in detail (World Trade Organization, 1999). For example, the official name of *Agreement on Anti-dumping* is *Agreement on Implementation of Article VI of the GATT 1994* where the name self-explains that it elaborates on the article VI of GATT 1994. The collected result is listed in the Appendix A.2. Figure 9 lists up 80 different articles of GATT 1994 cited in 143 cases without duplication.

I, I:1, II, II:1, II:1(a), II:1(b), II:2, II:3, III, III:1, III:2, III:4, III:5, III:7, IV, IX, IX:2, V, V:1, V:2, V:3, V:3(a), V:4, V:5, V:6, V:7, VI, VI:1, VI:1(a), VI:1(b), VI:2, VI:3, VI:5(a), VI:6, VII, VII:1, VII:2, VII:5, VIII, VIII:1, VIII:3, VIII:4, X, X:1, X:2, X:3, X:3(a), XI, XI:1, XIII, XIII:1, XIII:2, XIII:3(b), XIX, XIX:1, XIX:2, XIX:3, XV, XVI, XVI:1, XVI:4, XVII, XVII:1, XVII:1(c), XVIII, XVIII:10, XVIII:11, XX, XXI, XXII, XXII:1, XXIII, XXIII:1, XXIII:1(a), XXIII:1(b), XXIV, XXIV:12, XXIV:5(b), XXIV:6, XXVIII

Figure 9: **Set of articles of GATT 1994 collected and used in this paper:** These articles comprises the node set  $V$  and their ordered pairs comprise the edge set  $E$  in Figure 2

### 2.3.1 Various Levels of Scope in Cited Articles

As shown in Figure 9, members sometimes cite articles in different levels of scope. For example, For the Article II, member sometimes cites Article II as a whole but sometimes cites Article II:1

Let  $D$  is a set of DS case numbers listed in Figure 7.

Then there exists  $c_d = \{v_d \in V \mid v_d \text{ is an article cited in the case } d \in D\}$

where  $V$  is set of articles listed in Figure 9.

Then define set of cited articles  $C = \{c_d \mid d \in D\}$

Figure 10: **Formal Definition of Set of Cited Articles:** I formally define a set of cited articles  $C$  and the elements of  $C$  are listed in Appendix A.2.

or Article II:1(a). This is because two main judicial bodies of WTO DSB, *Panel and Appellate Body*, both constitute its legal precedents citing articles of the WTO agreements in various levels of scope. Those judicial bodies cite the articles with the level of various scopes, such as *Title, Article, Paragraph, Sentence* or *Term* as shown in Table 2. Following this jurisprudence, members also cite articles in different levels of scope to make their legal claim fit and valid according to the current jurisprudences of WTO DSB.

Table 2: **Various Levels of Scope Adopted to Cite Articles of WTO agreemnts**

Scope	Quote	Source
Title	“As the <b>title of Article 21 makes clear</b> , the task of panels . . . forms part of the process of the ‘Surveillance of Implementation of the Recommendations and Rulings’ of the DSB. . . .”	Appellate Body Report, <i>US – Shrimp (Malaysia)</i> , paras. 86-87.
Article	“The sequence of steps indicated above in the analysis of a claim of justification under <b>Article XX</b> reflects, not inadvertence or random choice, but rather the fundamental structure and logic of Article XX. . . .”	Appellate Body Report, <i>US – Shrimp (Malaysia)</i> , paras. 119-120.

Paragraph	“The verb ‘may’ in <b>Article VI:2</b> of the GATT 1994 is, in our opinion, properly understood as giving Members a choice between imposing an anti-dumping duty or not, as well as a choice between imposing an anti-dumping duty equal to the dumping margin or imposing a lower duty. ...”	Appellate Body Report, <i>US – 1916 Act</i> , paras. 116.
Sentence	“The customary rules of interpretation of public international law as required by <b>the first sentence of Article 17.6(ii) of the Anti-Dumping Agreement</b> , do not admit of another interpretation as far as the issue of zeroing raised in this appeal is concerned.”	Appellate Body Report, <i>US – Zeroing (EC)</i> , paras. 132-133.
Term	“Article II:1(a) provides that a Member shall accord to the ‘commerce’ of other Members treatment no less favourable than that provided for in its Schedule. <b>The term ‘commerce’</b> is defined as referring broadly to the exchange of goods such that, in this provision, the ‘commerce’ of a Member should be understood to refer to all such exchanges of that Member”	Appellate Body Report, <i>Colombia – Textiles</i> , para. 5.34.

### 3 Methodology: Considerations and Development

This section introduces two main considerations that are taken to design the method used in this paper. Those considerations justify the use of the deep neural network. Then I explain the detailed structure of the deep neural network and its training schemes. After finishing training

of the neural network, I explain the process that I conducted to fit the network of articles using Random Forest.

### 3.1 Two Main Considerations: Sufficient Information in Text & Generalization

This paper considered two main points to determine its method to qualitatively fit a set of edge weight  $W$  for the *directed weighted graph*  $G$  defined in Figure 2. One is the importance of using the information inside the text of factual aspect and legal article as exemplified in Appendix A.1 and Figure 5 respectively. The other one is about the way to generalize each member’s strategic citation pattern. Since members of the WTO strategically cite the articles of WTO agreement expecting different outcomes that are fit to each member’s specific interest (Johns and Pelc, 2014; Pelc, 2014; Strezhnev, 2014), this paper selected a method that can generalize these member specific citation patterns. These two considerations and the solution will be explained in the following subsections.

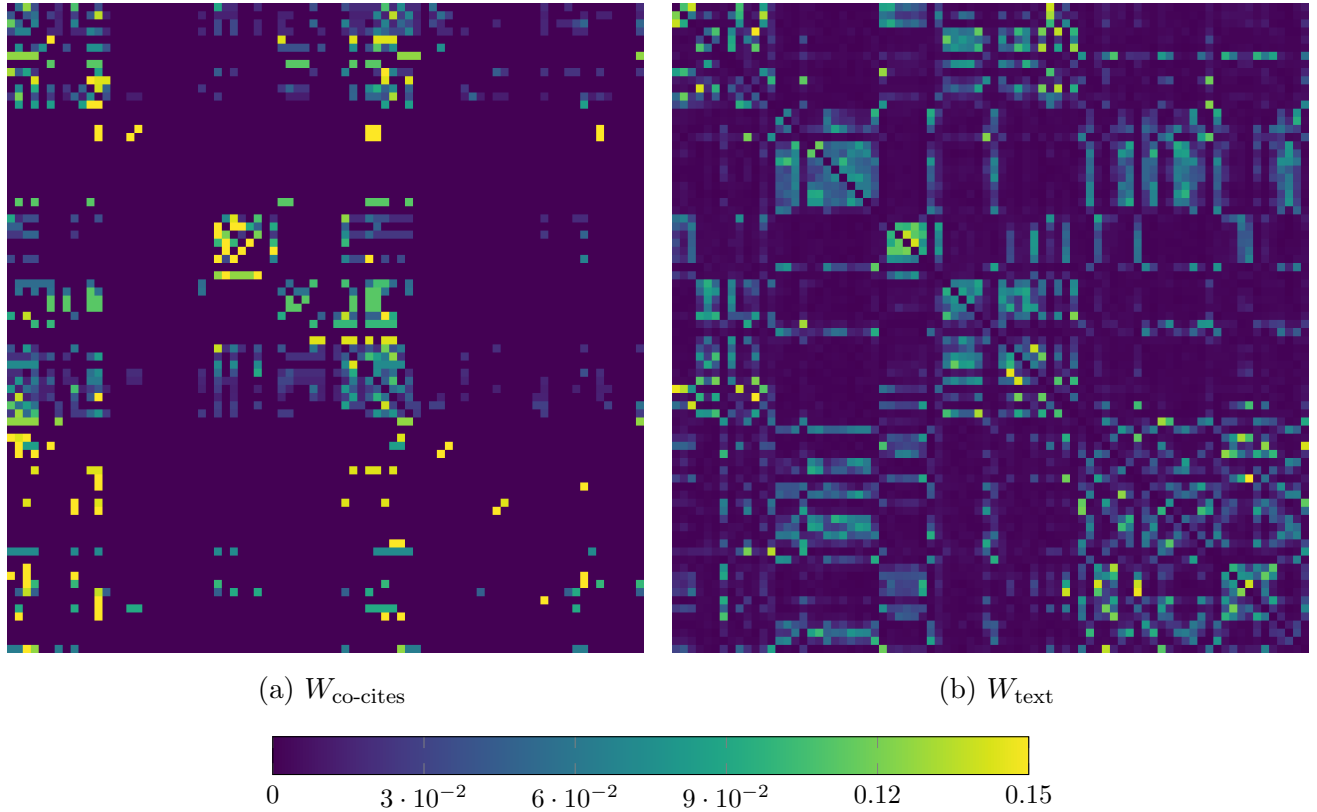


Figure 11: **Heatmap of Two Different Edge Weight Matrices:** Above two subfigures visualizes two different *edge weight matrices*  $W_{\text{co-cites}}$  and  $W_{\text{text}}$ . One can check that  $W_{\text{co-cites}}$  is sparser than  $W_{\text{text}}$ . It means  $W_{\text{co-cites}}$  captures fairly less interactions compared to  $W_{\text{text}}$

### 3.1.1 Importance of Using Textual Information

This paper emphasizes the necessity of using textual information to qualitatively fit a set of edge weight  $W$  defined in Figure 2. One can simply consider a co-citation pattern between the articles of WTO agreements that counts the co-occurrences of each article with other articles, however, it simply allocates a large edge weight for frequently cited articles and fails to explain how articles interact to achieve certain jurisprudence of WTO DSB. This failure is mainly due to the insufficient information in the co-citation matrix. Members cite the articles of the WTO agreements based on the complex characteristics of the trade policy that led to the dispute, however, the co-citation pattern omits this contextual and prior information. To visualize the importance of using textual information, I prepared two different matrices  $W_{\text{co-cites}}$  and  $W_{\text{text}}$  that are following the definition of *edge weight matrix*  $W$  in Figure 2.  $W_{\text{co-cites}}$  is calculated only using the co-citation pattern between the articles of the WTO agreements as formally defined as *Normalized Co-citation Matrix* in Figure 12 and 13.  $W_{\text{text}}$  is the one fitted using the textual information and the way how it's fitted will be explained at the following bodies of this section, in particular in Section 3.3.2. Two heatmaps visualized in Figure 11 shows how sparse the  $W_{\text{co-cites}}$  is compared to the  $W_{\text{text}}$ . This sparsity indirectly refers to the insufficient information to qualitatively map the jurisprudences of WTO DSB. In contrast with it, if we fit the *edge weight matrix*  $W$  using the textual information, we get a more dense and informative matrix as visualized in Figure 11(b). Upon this observation, this paper justifies the use of a deep neural network to process information embedded in the text description of the trade policy and regulatory content of the legal articles. This is because deep neural network is known to extract information from text effectively to perform various tasks such as text classification (Minaee et al., 2020), text summarization (Magdum and Rathi, 2020) and text generation (Guo et al., 2017).

### 3.1.2 Generalization of Each Member's Strategic Citation

This paper aims to map the regulatory system of WTO DSB in a form of *directed weighted graph*  $G$  as defined in Figure 2. To achieve this purpose, we need to fit  $W$  to generalize member specific strategic citation behavior. In terms of generalization, deep neural network is known to generalize well despite its large capacity (Neyshabur et al., 2017), possible instability of training algorithm (Charles and Papailiopoulos, 2017), nonrobustness (Zahavy et al., 2017), and sharp minima (Dinh et al., 2017). Therefore, this paper trains a deep neural network without any member specific information such as geolocation, GDP or specialized industry and so on. By training a deep

Let  $\delta_{ij}^d$  is defined to be 1 if  $\{(v_i, v_j) \mid v_i, v_j \in V \text{ and } i \neq j\} \subset c_{d \in D}$  else 0

where  $V, D$  and  $c_d$  is defined as in Figure 10.

Then let *co-citation matrix*  $M = (m_{ij}) \in \mathbb{N}^{|V| \times |V|}$  s.t.  $m_{ij} = \sum_{d \in D} \sum_{i, j \in V} \delta_{ij}^d$

(a) **Formal Definition of Co-citation Matrix**

	I	I:1	II	II:1	...
I	0	3	7	2	
I:1	3	0	3	4	
II	7	3	0	4	
II:1	2	4	4	0	
⋮					

(b) **Illustration of Co-citation Matrix**

Figure 12: **Formal Definition and Illustration of Co-citation Matrix:** This paper defines co-citation matrix  $M$  as subfigure (a) and it's illustrated as subfigure (b) using the paper's dataset. Note that the co-citation matrix is *symmetric*,  $m_{ij} = m_{ji} \forall i, j \in V$ .

neural network only using a description of the possible inconsistent trade policy and the legal text of the cited articles, this paper expects a fitted  $G^* = (V, E, W^*)$  can show interactions between articles of the WTO agreements without being biased to member specific trade interest and its strategic citation upon it.

## 3.2 Design of Deep Neural Network

Upon the justification of using deep neural network with above two main considerations in Section 3.1, I explain a design process of the deep neural network that can encode the citation patterns in WTO DSB. This paper understands that those citations are performed upon the general understanding over jurisprudences of WTO DSB and encoding of those citation patterns could reveal the jurisprudences of WTO DSB.

For given  $M$  defined in Figure 12(a),

let *normalized co-citation matrix*  $N = (n_{ij}) \in \mathbb{R}^{|V| \times |V|}$  s.t.  $n_{ij} = \frac{m_{ij}}{\sum_{j \in V} m_{ij}}$

(a) **Formal Definition of Normalized Co-citation Matrix**

	I	I:1	II	II:1	...
I	0	0.053	0.125	0.035	..... $\rightarrow \sum_{j \in V} n_{ij} = 1$
I:1	0.040	0	0.04	0.054	
II	0.114	0.049	0	0.065	
II:1	0.032	0.065	0.065	0	
$\vdots$					

(b) **Illustration of Noramlized Co-citation Matrix**

Figure 13: **Formal Definition and Illustration of Normalized Co-citation Matrix:** This paper defines normalized co-citation matrix  $N$  of  $M$  as subfigure (a) and it's illustrated as subfigure (b) using the paper's dataset. Note that normalized co-citation matrix is no more *symmetric*,  $n_{ij} \neq n_{ji} \forall i, j \in V$ . This definition is prepared to fit the definition of co-citation matrix to that of  $W$  in Figure 2.

### 3.2.1 Design Input/Output of Deep Neural Network: by Analogy with How Member Cites in WTO DSB

A rule of thumb to design input and output of deep neural network is to mimic how humans do for a given task. Therefore I present a visualization of how legal experts of WTO agreements determine whether to cite a legal article of WTO agreements with an example of *Korea - Beef* case (Figure 14 and Figure 15). In this case, the United States raised a claim relating to the *Dual-retail system* maintained by South Korea. In *Dual-retail system*, South Korea maintained two distinct retail systems for imported and domestic beef. There existed stores specialized for imported beef and they can sell only imported beef and cannot sell domestic (Korean) beef. U.S. claimed that the *Dual-retail system* is inconsistent to the Article III:4 (National Treatment) of GATT 1994 because *Dual-retail system* discriminates between domestic and imported beef. A measure that discriminates domestic and imported products falls under the scope of the Article III:4 of the GATT 1994, which states the principle of *National Treatment* that prohibits the discrimination between imported and domestic product. However, U.S. didn't cite the Article I:1 of GATT 1994



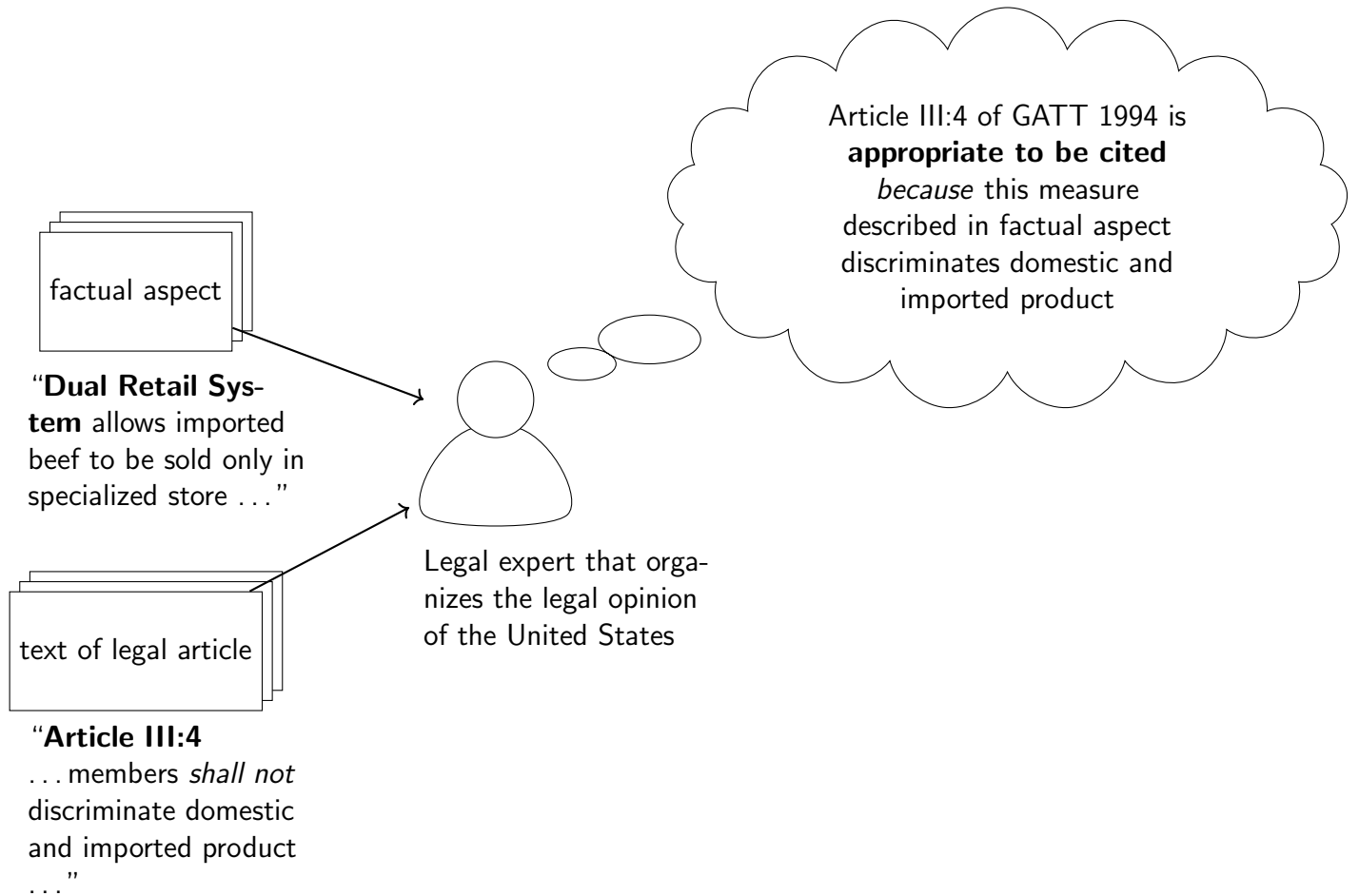


Figure 14: **Visualization of How Member Cites in WTO DSB (Citable Case):** With two different contexts, factual aspect and text of legal articles, member judges whether the given legal article is appropriate to be cited or not.

that prohibits the discrimination between members of WTO because *Dual-retail system* didn't discriminate against the United States from other countries who export beef to South Korea, such as Argentina, Australia, etc.

We can understand that there exists a shared understanding over jurisprudences of WTO DSB among legal experts of the WTO agreements. They follow up new cases and study jurisprudence stated in the *Panel* or *Appellate Body* reports. Then they organize a legal argument by citing certain article(s) upon this shared understanding for given possible inconsistent measures claimed by a member of WTO.

To mimic this reasoning process, I designed the input and output of the deep neural network as illustrated in Figure 16 and formally defined in Figure 17. The neural network is designed to estimate the citability for a given pair of a factual aspect and text of a legal article. By iteratively training the neural network with data explained in Section 2, I expect the neural network can learn a shared understanding over jurisprudences of WTO DSB closely to that of legal experts. The

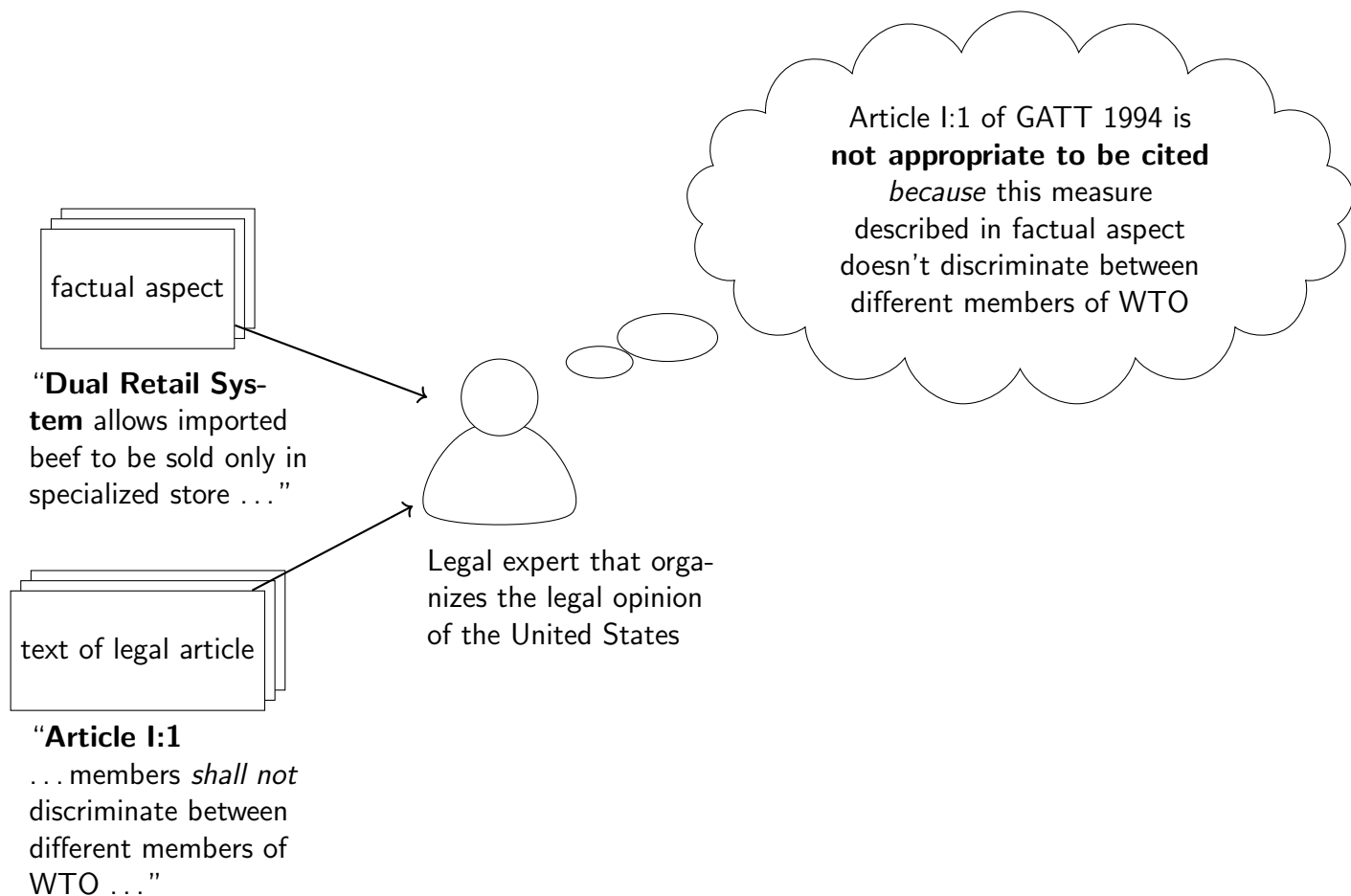


Figure 15: **Visualization of How Member Cites in WTO DSB (Non-Citable Case):** With two different contexts, factual aspect and text of legal articles, member judges whether the given legal article is appropriate to be cited or not.

detailed structure of neural network and training schemes will be explained in the later subsections.

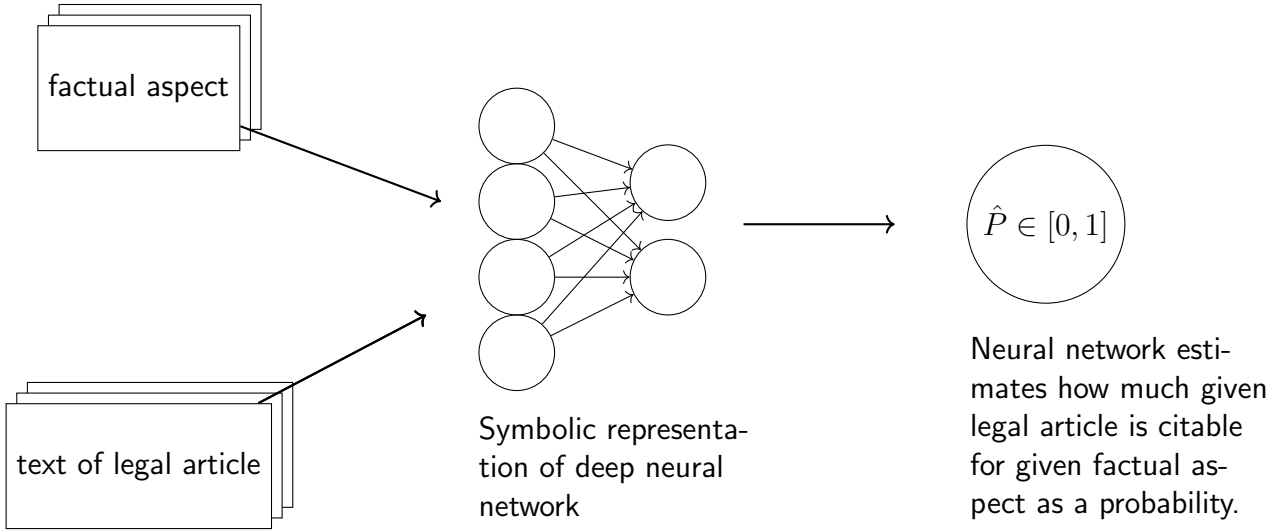


Figure 16: **Design of Training Framework of Deep Neural Network:** I designed a training framework of deep neural network by analogy with how member cites in WTO DSB as visualized in Figure 14 and Figure 15.

For given  $V$ ,  $D$  defined in Figure 9 and Figure10,

let  $E = \{e \mid e \text{ is an english word or special charcter}\}$  and

$n_{\text{factual}}, n_{\text{article}} \in \mathbb{N}$  represents *max token length* of factual aspect and legal article respectively.

Then define  $T = \{t_d \mid t_d = (e_1, e_2, \dots, e_{n_{\text{factual}}}) \text{ s.t. } d \in D \text{ and } e_{i \leq n_{\text{factual}}} \in E\}$

where  $t_d$  represents a factual aspect of one of DS cases listed in Figure 7.

Also define  $A = \{a_v \mid a_v = (e_1, e_2, \dots, e_{n_{\text{article}}}) \text{ s.t. } v \in V \text{ and } e_{i \leq n_{\text{article}}} \in E\}$

where  $a_v$  represents texts of one of a legal article listed in Figure 9.

Now defines a deep neural network  $f$  with a set of parameters  $\theta$

$$f_{\theta} : T \times A \rightarrow [0, 1]$$

Figure 17: **Formal Definition of Input/Output of Deep Neural Network:**  $T$  and  $A$  represents a set of “documents” that are factual aspects and text of legal articles respectively. The term “document” refer to a *tuple* of English words or special characters like  $(e_1, e_2, \dots, e_{n_{\text{max}}})$ . Then a pair of two documents, factual aspect and legal article, is fed into the neural network  $f$  and returns a probability that represents how much the given article is citable for the given factual aspect.

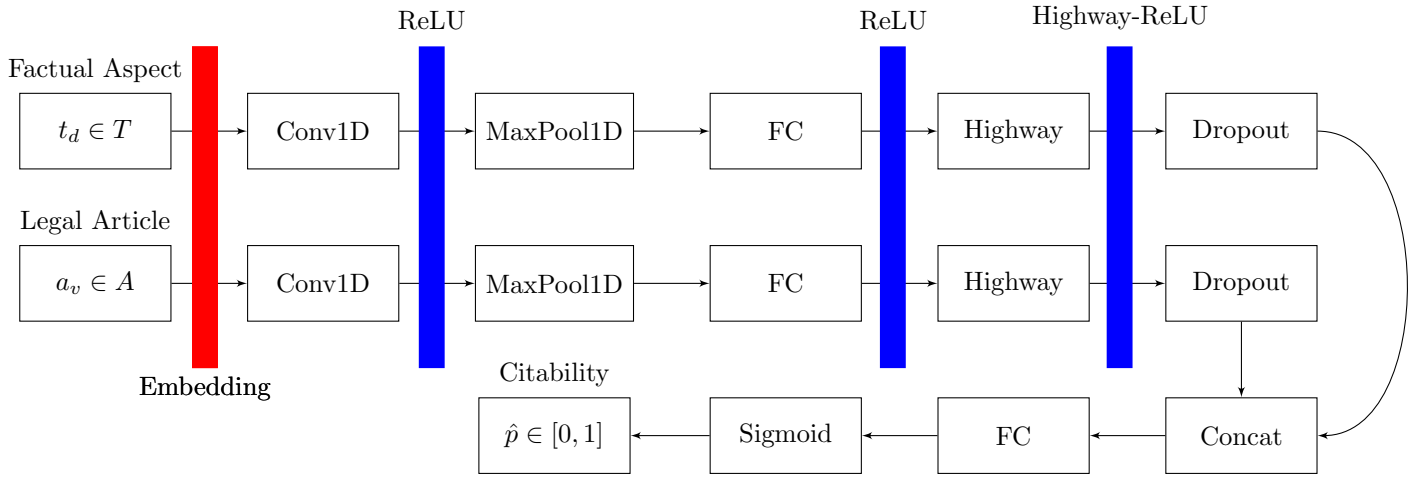


Figure 18: **Layers of Deep Neural Network:** The term “Layer” refers to a mathematical operation. How layers are stacked in which order determines the “structure” of deep neural networks. This figure illustrates the structure of the deep neural network that this paper used. The figure follows the notations defined in Figure 17.

### 3.2.2 Structure of Deep Neural Network

An efficient way to architect the structure of deep neural network is to use the human analogy as I did in subsection 3.2.1. A legal expert reads the text line by line and creates several local understandings. Then he/she merges those local understandings into a global level to summarize an essential information to determine whether the given article is citable or not for the given factual description of the dispute.

To analogy this process, I borrowed *1-Dimensional Text Convolutional Neural Network* (TextCNN) from Kim (2014). This is because of the following two reasons. First, convolutional neural networks are known as good at learning how to integrate local and global features to perform classification tasks (Lawrence et al., 1997; Nahid and Kong, 2017). Second, Kim (2014) has implemented this convolutional neural network into the text domain and has shown its high performance on classification.

Figure 18 demonstrates the entire structure of the deep neural network that is used in this paper. General flow is to return the citability of an legal article that is fed into the neural network with a factual aspect. Each block in Figure 18 represents a set of unique mathematical operations. We prefer to call those blocks as “layer” and each layer has its desired role regarding how to process the information for which purpose. I will explain each block’s role and composition in the following subsections.

### 3.2.2.1 Inputs: a Pair of Documents, Factual Aspect and Legal Article

I defined the neural network in terms of its input and outputs in Figure 17. First we need to *tokenize* the text of factual aspect and legal article. The term “*tokenize*” refer to the process of decomposing the text into the sequence of words or special characters. I used the *off-the-shelf* tokenizer provided by *Spacy* API<sup>9</sup>. Since the neural network is represented as a *function*, it must hold a predefined input and output dimension. I checked the *max token length* over all tokenized results for factual aspects and legal articles. The *max token length* was 35,842 and 20,158 for factual aspects and legal articles respectively. These numbers correspond to  $n_{\text{factual}}$  and  $n_{\text{article}}$  in Figure 17. Then I *padded* a special token [PAD] at the tail in case the token length of a factual aspect or legal article is shorter than those *max token lengths*.

It’s worth noting that the field of deep learning preferably calls this ordered set of tokens as ‘*Document*’. Therefore I prepare documents from the raw data collected in Section 2 at this stage and move on to the next step, *Embedding Layer*.

### 3.2.2.2 Embedding Layer: From Documents to Numerical Vectors

Since the deep neural network is comprised of mathematical operations, we need to transform the word tokens into a form of numerical vectors. This process can be conducted with a single *Embedding Layer*. This layer is defined as  $|Size\ of\ Dictionary| \times k$  matrix where this matrix works as a dictionary for the neural network. Neural network refers to this matrix to find the meaning of a token in a form of  $k$ -dimensional vector and updates its value while it’s trained to reflect domain specific meaning of each token.

For example, WTO DSB prefers to use the word *inconsistent* rather than the word *breach* to refer to the illegality of a member’s trade policy. This kind of domain specific information will be stored in this matrix as a distribution of those numerical vectors in this layer.

I used *Google News Word2Vec*<sup>10</sup> to initialize this embedding layer. This is an open-source pre-trained vector provided by Google. It contains 300 dimensional English word vectors for 3 million unique words. Because of the GPU memory limit, I set 400,000 as a maximum number of word vectors to read from the Google Word2Vec and this number corresponds to the  $|Size\ of\ Dictionary|$ .

Figure 19 represents the “output” of the embedding layer for given document that is tokenized from the example in Appendix A.1. The neural network find correspondent  $k$ -dimensional vectors

---

<sup>9</sup><https://spacy.io/api/tokenizer>

<sup>10</sup><https://code.google.com/archive/p/word2vec/>



Figure 19: **Output of Embedding Layer:** This figure illustrates an output of embedding layer. Embedding layer maps each token in the input document to the specified numerical vector in dimension  $k$ .

for given ordered tokens in the document and returns  $(n_{\text{factual}}, k)$  or  $(n_{\text{article}}, k)$  size of matrix for given two different types of inputs. This matrix is fed into the next layer, *Conv1D*.

### 3.2.2.3 Conv1D: Capturing the Local Features

This subsection explains the how convolution filters runs over the  $(n_{\text{factual}}, k)$  or  $(n_{\text{article}}, k)$  size of matrix that are passed from the previous layer. I define this output matrix from the embedding layer more formally as below following the notation used in Kim (2014)

$$x_{1:n} = x_1 \oplus x_2 \oplus \dots \oplus x_{\max}$$

where  $\oplus$  represents concatenation and

$x_i \in \mathbb{R}^k$  represents the  $i$ -th embedded token from the document.

Let  $x_{i:i+j}$  refer to the concatenation of the embeddings  $x_i, x_{i+1}, \dots, x_{i+j}$

Then a convolution filter  $w \in \mathbb{R}^{h \times k}$  is simply defined as  $(h, k)$  size of matrix where  $h$  represents the filter size and  $k$  represents the embedding dimension which is same to that of the embedding layer. Then the convolution is defined as  $w \cdot x_{i:i+h-1} + b$  where  $b \in \mathbb{R}$  is a bias term. I illustrated

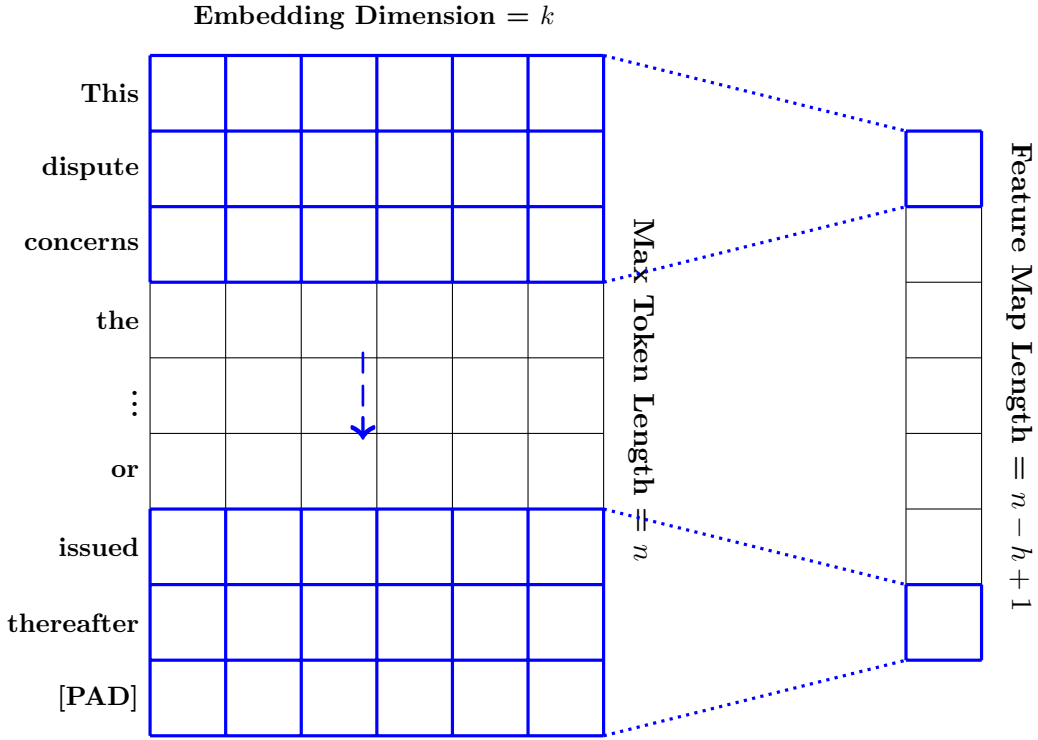


Figure 20: **Conv1D**:  $h$ -sized filter runs over the  $n \times k$  embedding matrix and produces  $(n - h + 1)$  size of feature map.

this convolution operation in Figure 20 where the blue box refers to the convolution filter  $w$  with the filter size 3 and it runs over the each set of 3 embeddings as increasing the  $i$  of  $x_i$  one by one. This will eventually generate a feature vector  $c$  which is a size of  $n - h + 1$ .

Each convolution filter  $w$  learns how to summarize information *locally* for the region of filter size. For example, a convolution filter  $w$  holds its unique way of understanding the meaning of the “***This dispute concerns***” in Figure 20. This convolution filter  $w$  generalizes its own way of understanding 3 token embeddings together over all  $n - h + 1$  number sets of 3 token embeddings.

I prepared 128 different convolution filters  $\{w\}$  so that each of them can hold their unique way of comprising *local understandings* where the size of locality corresponds to its filter size. Moreover, I also prepared three different filter sizes 3, 4 and 5. This means we will eventually get  $(n - 2, 128)$ ,  $(n - 3, 128)$  and  $(n - 4, 128)$  size of vectors by the *Conv1D* operation.

It’s important to note that I introduced *non-linearity* using **ReLU** ( $ReLU(x) = \max(0, x)$ ) after each convolution operation. This means output of convolution operation  $w \cdot x_{i:i+h-1} + b$  becomes 0 in case the value is smaller than 0. This sequential layering of linear (such as convolution operation) and non-linear operations (such as ReLU) lets the model to be able to encode more complex patterns like citation patterns appearing upon the jurisprudence of WTO DSB.

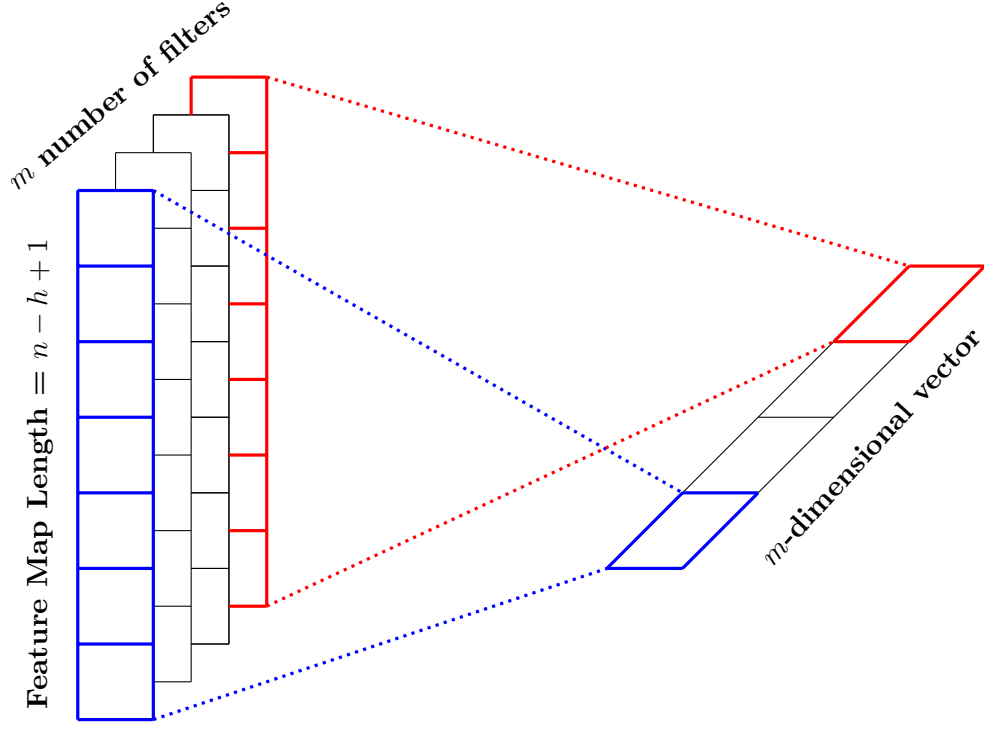


Figure 21: **MaxPool1D**: Filter out max value for all  $m$  number of feature map outputs from  $m$  different convolution filters. MaxPool1D produces a  $m$  dimensional vector as an output for collection of those filtered max values.

#### 3.2.2.4 MaxPool1D: Choose the Most Prominent Feature

Our general goal is to estimate the citability  $\hat{p} \in [0, 1]$  as closely to that of real world citation patterns. Therefore we need an efficient *down-sampling* process to sample the large dimensional features into smaller dimensions. With a convolution, *Max-pooling* is preferred to use for this purpose. *Max-pooling* simply selects the largest values among  $(n - h + 1)$  size of vector where each of dimension is calculated by the same convolution filter  $w$ , i.e.,  $ReLU(w \cdot x + b)$ . Figure 21 illustrates this process. I used 128 ( $= m$  in Figure 21) number of filters for each filter size, thus I got 3 different 128-dimensional vectors as a result of *MaxPool1D*.

#### 3.2.2.5 FC: Enlarge the Capacity

Using the  $(3, 128)$  size of feature map from the *Max-pooling*, I flatten those three (128) dimensional feature maps into a (384) dimensional vector. To increase the model capacity in terms of size of parameters and non-linearity, I introduced a upscaling *Fully Connected Layer* (FC) that is defined as  $W_{fc} \cdot x + b_{fc}$  where  $W_{fc} \in \mathbb{R}^{1024 \times 384}$  is a linear map that increases the feature dimension,  $x$  is the flattened (384) size of feature vector and  $b_{fc} \in \mathbb{R}^{1024}$  is a bias term. Then by applying **ReLU**



again to the output of this *Fully Connected Layer*, I introduced another *non-linearity*. This is to let the model to introduce more complexity in terms of nonlinearity without losing important features from this non-linear operation by introducing upscaled dimension to the feature map. This operation returns a (1024) size of feature vector.

### 3.2.2.6 Highway: Introducing another Non-linearity while Preventing the *Vanishing Gradients*

I implemented *Highway network* (Srivastava, Greff and Schmidhuber, 2015) to add more non-linearity. Adding non-linearities increases the chance that the neural network can approximate more complex patterns in the data, however, as the network gets deeper *vanishing gradients* problem arises. Deep neural network is trained by updating its parameter weight proportional to the partial derivative of the loss function with respect to the current parameter (See Algorithm 1) and computes this gradient by chain rules. Therefore, as the network gets deeper in terms of non-linearity, the value of gradient tends to become *vanishingly small* compared to that of the front layer.

*Highway network* prevents this *vanishing gradient problem* by introducing an additional parameter that learns the *adequate amount of non-linearity*. The highway network is formally defined as following.

$$Y = H(W_H \cdot x + b_H) \cdot T(W_T \cdot x + b_T) + x \cdot (1 - T(W_T \cdot x + b_T))$$

where  $H$  is ReLU,  $T$  is Sigmoid

$$(W_H \in \mathbb{R}^{1024 \times 1024}, W_T \in \mathbb{R}^{1024 \times 1}, b_H \in \mathbb{R}^{1024} \text{ and } b_T \in \mathbb{R} \text{ in this paper's setting})$$

By defining  $W_T$  a linear map that returns 1-dimensional output, the sigmoid of its output  $T(W_T \cdot x + b_T)$  is confined in  $[0, 1]$ . Therefore, the network gets to introduce  $T(W_T \cdot x + b_T)$  amount of ReLU to  $W_H \cdot x + b_H$  where  $x$  is (1024) shape of feature vector from the previous layer.

### 3.2.2.7 Dropout: Ensure more Generalizability

Highway network returns (1024) shape of feature vector and I apply *dropout* operation (Srivastava et al., 2014) with the drop rate 0.5 on this feature vector. Dropout randomly sets the value of each dimension as 0 or scales up by  $1/(1 - \text{drop rate})$  thus dimension-wise sum of the feature vector is unchanged. Dropout is one of the most widely accepted regularization techniques which prevents overfitting and helps the model to achieve its generalizability.

### 3.2.2.8 Concat: Feature Map from Factual Asepct and Legal Content Meets

We have run through the two same processes of Embedding, Conv1D, MaxPool1D, FC, Highway and Dropout for two different types of data, *Factual Aspect* and *Legal Article*. This generates two (1024) sizes of feature vectors. We simply concatenate those two and generate a (2048) size of feature vector.

### 3.2.2.9 FC: Generates Logit

Each feature vector before concatenation in the previous layer corresponds to the understanding of the neural network for each type of data. Now we simply concatenated those two and reduce the size to 1 by applying *Fully Connected Layer* with the liner map  $W_{final} \in \mathbb{R}^{2048 \times 1}$  with bias  $b_{final} \in R$ . This generates a scalar and we consider it as a logit,  $\log \frac{\hat{p}}{1-\hat{p}}$

### 3.2.2.10 Sigmoid: Generates Citability

By applying *Sigmoid* to this size 1 logit, we get  $\hat{p} \in [0, 1]$ .

## 3.2.3 Train of Deep Neural Network

I have total 11,440 number of data instances that is calculated by  $|D| \times |V|$  where  $|D| = 143$  and  $|V| = 80$ . Before training, I randomly split the entire dataset  $T \times A$  into train and test data in 8:2 ratio. The number of split results is 9,153 for train data and 2,287 for test data. I used train data only to train the neural network and used test data to check the trained model's performance. By measuring a performance metric on the inference results of the test data, one can check the generalizability of the trained neural network regarding how well the trained neural network performs over the data that it has never seen before.

The term “training” refers to adjust a set of parameters  $\theta$  that constitutes mathematical operations defined in layers illustrated in Figure 18 and explained in the following subsections. The procedure to fit those set of parameters is described in Algorithm 1.

I fitted the nerual network  $f_\theta$  using *weighted cross entropy loss* as formally defined in Algorithm 1. I used *cross entropy loss* because it measures how much the probability distribution projected by the trained model  $f_\theta$  deviates from the true distribution which represents a shared understanding regarding the citability of certain legal article for given textual description of the case among the group of legal experts of WTO agreements.

---

**Algorithm 1:** Steps to Train Neural Network  $f_\theta$ 

---

**Input:** neural network  $f_\theta$ , train dataset  $T_{train} \times A_{train}$  and set of cited article  $C$  defined in Figure 10

**Output:** fitted set of parameter  $\theta^*$

```
1 Let weight of binary cross entropy  $w = 26.303$ 
2   weighted binary cross entropy loss  $L(y, \hat{p}) = w * y \log \hat{p} + (1 - y) \log (1 - \hat{p})$ 
3   Epochs  $e \in \mathbb{N}$ 
4   Learning Rate  $\alpha \in \mathbb{R}$ 
5 for (  $i = 1, i++, \text{ while } i \leq e$  ) {
6   for (  $(t_d, a_v) \in T_{train} \times A_{train}$  ) {
7      $y \leftarrow 1$  if  $v \in c_d$  else 0
8      $\theta \leftarrow \theta - \alpha * \nabla_\theta L(y, f_\theta(t_d, a_v))$ 
9 return  $f_\theta^*$ 
```

---

It's worth noting that I “weighted” *cross entropy loss*. This is because our dataset is highly imbalanced in terms of citability. Among all 11,440 data instances, I have only 435 data instances where the given article  $a_v$  is actually cited for the given case description  $t_d$ . This is only 3.802%. This is because a case tends to have only 3 - 4 articles cited on average among 80 different articles in  $V$ . Therefore, I adopted a weight of 26.303 which is inverse of the 3.802% to penalize the neural network with higher loss in case the network fails to cite correctly for the positive case where  $y = 1$  in Algorithm 1.

Epochs  $e$  and learning rate  $\alpha$  in Algorithm 1 is a hyperparameter whose value shall be determined before training. First, one epoch refers to one cycle that the neural network trains over the entire training data once and I set  $e$  as 15. Therefore, this neural network sees each training data instance 15 times. Second, the learning rate is about how much we are adjusting the model parameters with respect to the loss gradient. I trained the neural network with learning rate of 0.01 but decays with the rate of 0.95 for seeing every 5000 data instances because gradual decay of learning rate is known as good at preventing the training from being stuck in the local minima.

### 3.2.4 Training Result: AUC-ROC as Performance Metric

Train loss and test loss converged to around 2 and 1.25 after epoch of 10 respectively. However one needs a well-defined performance metric that can measure how well a model performs for a given task. Our task can be understood as a classification problem where the model decides a pair of factual aspects and the legal article falls into the citable case or not. Therefore I used the AUC-ROC metric to measure the model performance.

The AUC-ROC measures model performance at various threshold settings. Since the model

accuracy varies upon the threshold selected, one needs to consider every case of threshold together to measure the model’s classification power conclusively. ROC is measured with a given threshold  $t \in [0, 1]$  and defined as follows.

$$\begin{aligned} \text{ROC}(t) &= \frac{\text{TPR (True Positive Rate)}}{\text{FPR (False Positive Rate)}} \\ \text{where TPR} &= \frac{\text{TP (True Positive)}}{\text{TP (True Positive)} + \text{FN (False Negative)}} \\ \text{FPR} &= \frac{\text{FP (False Positive)}}{\text{FP (False Positive)} + \text{TN (True Negative)}} \end{aligned}$$

Then AUC-ROC measures the “area” under the ROC curve where the curve represents ROC value at every threshold of  $t$ . The maximum value of AUC-ROC is 1 and its baseline is 0.5 where the model randomly estimates the case as citable or not. Our model’s AUC-ROC converged to around 0.85 after the epoch of 5. It means that the model encoded the pattern inside data and predicts better than the naive baseline. However, we need more substantial analysis of what the value 0.85 means. I will generate a network of articles using this trained neural network at the following subsections and will empirically interpret the generated network with my background knowledge about jurisprudences of WTO DSB in Section 4.

### 3.3 Fitting $G^* = (V, E, W^*)$ using Random Forests

This subsection explains how I found the best set of directed edge weights  $W^*$  that closely maps the shared understanding about jurisprudences of WTO DSB among legal experts in a form of *directed weighted graph* as defined in Figure 2 and illustrated in Figure 3 using the trained neural network  $f_{\theta^*}$ .

#### 3.3.1 Definition of $W^*$ : Best Set of Directed Edge Weights

Let  $f_{\theta^*}$  represent the trained neural network that is equipped with an optimized set of parameters  $\theta^*$ . Then we can construct a prediction matrix  $P = (f_{\theta^*}(t_d, a_v)) \in [0, 1]^{|D| \times |V|}$  by collecting all predictions  $f_{\theta^*}(t_d, a_v)$  from the trained neural network  $f_{\theta^*}$  using the all pairs of data  $(t_d, a_v) \in T \times A$  as illustrated in Figure 22.

Upon the assumption that the trained neural network  $f_{\theta^*}$  effectively encodes a shared understanding about jurisprudences of WTO DSB among legal experts of WTO agreements, our task is to find a set of directed edge weights  $W^* = \{w_{ij}^* \mid w_{ij}^* = w^*(v_i, v_j) \text{ s.t. } w^*(v_j, v_j) = 0 \text{ and } \sum_{v_i \in V} w^*(v_i, v_j) = 1 \forall v_j\}$  by exploiting the information encoded in the prediction matrix

$P = (f_{\theta^*}(t_d, a_v)) \in [0, 1]^{|D| \times |V|}$ . This set of directed edge weights  $W^* = \{w_{ij}^*\}$  shall represent a set of conditional probability  $P^*(v_j|v_i)$  how probably a source node  $v_i$  clarifies the meaning of the target node  $v_j$  compared to other source nodes closely to a shared understanding about jurisprudences of WTO DSB among legal experts.

To perform this task, this paper adopts a machine learning technique called *Random Forest* that can rank input features,  $\{v_i \mid v_i \in V \setminus \{v_j\}\}$ , in terms of relative importance to explain the variance of output variables,  $\text{Var}(\{f_{\theta^*}(t_d, a_{v_j}) \mid d \in D\})$  for a given target article  $v_j \in V$ . The step-by-step algorithm of *Random Forest* will be explained in the next subsection.

### 3.3.2 Random Forest on Prediction Matrix $P$ : Finding $W^*$

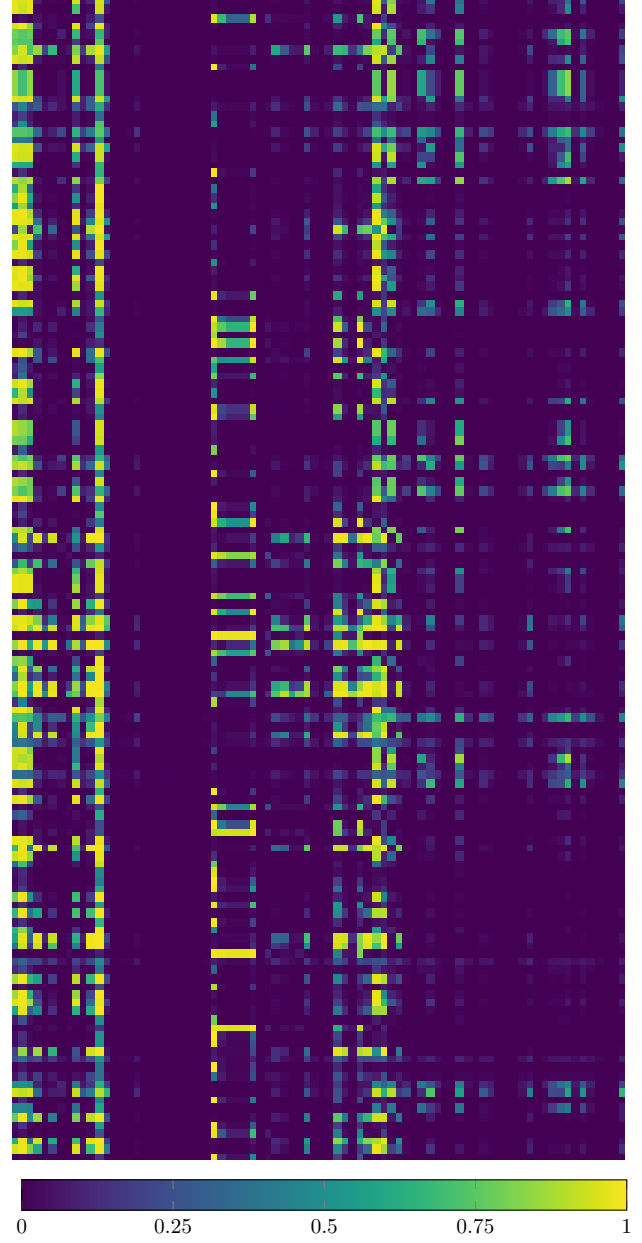
The prediction matrix  $P$  illustrated in Figure 22(b) includes a co-citation pattern between articles of WTO agreements. For example, we can see several highlighted bands in Figure 22(b) and it tells us there exists some co-citation patterns inside the matrix.

To materialize this co-citation pattern in a form of *directed edge weights*  $W$  as defined in Figure 2, I used a machine learning technique called *Random Forest* (RF). RF is an ensemble method that averages out multiple decision trees.

A decision tree consists of several nodes where each node splits the all observations into two with an inequality criterion of an input feature. This split reduces the variance of the output variables gradually as the tree grows. In our case, for a given target article  $v_j \in V$ , the rest of the articles  $v_i \in V \setminus \{v_j\}$  becomes the input features. For example, we have 143 observations in Figure 22(b) where the number equals that of rows. Then we grow a tree where each node of tree splits the observations according to its own criterion whether the value of that input features greater (or lesser) than certain value. By doing so, variance of output variables in 143 observations keep reduced following down the tree. Then we can collect and assign the amount of variance reduced by each input feature to  $w_{ij}$  by interpreting this variance reduction as how much The source article  $v_i$  clarifies the interpretation of the target article  $v_j$ . I normalize the variance of output variables before constructing the tree, thus total variance reduction sums up to 1. Therefore  $W = (w_{ij})$  fits to its definition in Figure 2.

There are three different aspects that distinguish RF from a single decision tree. First, as noted earlier, RF requires to average out multiple decision trees. I ensembled 1,000 decision trees and averaged out all  $w_{ij}$  generated from each decision tree. Second, RF requires *bagging* that random samples observations with replacement before constructing a tree. This is to avoid overfitting by letting each tree being trained on different parts of the same dataset. I sampled 143 observations

	I	I:1	II	II:1	...
DS2	0.950	0.933	0.946	0.068	
DS18	0.950	0.912	0.947	0.013	
DS22	0.070	0.037	0.003	0.042	
DS31	0.933	0.967	0.835	0.135	
⋮					
DS505					
DS513					
DS518					
DS523					



(a) **Illustration of Prediction Matrix:** By defining row as a list of DS case numbers and column as legal articles, we can create  $|D| \times |V|$  matrix that includes  $f_{\theta^*}(t_d, a_v)$  for each cell.

(b) **Prediction Matrix:** This heatmap represents the  $P$  from the actual data and its predictions from  $f_{\theta^*}$

Figure 22: **Illustration of Prediction Matrix:** I collected all the predictions from the trained neural network  $f_{\theta^*}$  and constructed prediction matrix  $P$

with replacement for each construction of decision tree. Third, RF requires to compare the result of random subset of input features to split. I random sampled  $\sqrt{|V| - 1}$  number of input features at each split because Huynh-Thu et al. (2010) reported high performance of use of this parameter for solving regression problem with random forest that is same to our setting.

All the process of finding  $W^*$  from the prediction matrix  $P$  is formally defined in Algorithm 2 and final output of  $W^*$  is visualized as the heatmap  $W_{text}$  in Figure 11(b).

---

**Algorithm 2:** Random Forest to Find  $W^*$

---

**Input:** Prediction Matrix  $P = (p_{dv}) \in [0, 1]^{|D| \times |V|}$  s.t.  $p_{dv} = f_{\theta^*}(t_d, a_v)$

**Output:**  $W^* = (w_{ij}^*) \in [0, 1]^{|V| \times |V|}$

```

1 Let number of features  $n = |V|$  ,
2   number of obseravations  $o = |D|$  and
3   number of trees to ensemble  $m \in \mathbb{N}$ 
4 for (  $v_j \in V$  ) {
5    $X \leftarrow \{x_d \mid x_d = (p_{dv_1}, p_{dv_2}, \dots, p_{dv_n}) \text{ s.t. } v \in V \setminus \{v_j\} \text{ and } d \in D\}$ 
6    $Y \leftarrow \{y_d \mid y_d = p_{dv_j} / \sigma(p_{v_j}) \text{ s.t. } d \in D \text{ and } \sigma(p_{v_j}) \text{ is a standard deviation of}$ 
7      $\{p_{dv_j} \mid d \in D\}\}$ 
8   for (  $k \in \{1, 2, \dots, m\}$  ) {
9     1.  $S = \{(x_d, y_d)\} \leftarrow$  Random sample  $o$  number of  $(x_d, y_d)$  from  $X \times Y$  with
        replacement. Then let  $X_d$  notate set of all sampled  $x_d$ .
10    2. Construct a decision tree  $T_k : X_d \rightarrow \mathbb{R}$  where
11     $T_k = \{N \mid N = (v_i, b, N_p, \hat{y})$  represents a decision node where
12     $b, \hat{y} \in \mathbb{R}, v_i \in V \setminus \{v_j\}$  and  $(v_i, b)$  splits  $S_N \subset S$  that reached the node  $N$ 
13    into  $S_{N_{true}}$  and  $S_{N_{false}}$  with a criterion  $p_{dv_i} \geq b$  with a parent node  $N_p \in T_k$ 
14    if  $N$  is not a root node. Define  $N_p = \emptyset$  if  $N$  is a root node.  $\hat{y}$  represents
15    the node's estimate for given input  $x_d$  and  $\hat{y} = \frac{1}{|S_N|} \sum_{(x_d, y_d) \in S_N} y_d$ .
16     $v_i$  and  $b$  is  $\emptyset$  if  $N$  has no child nodes.  $(v_i, b)$  at each node  $N$  is determined
17    among a random sampled  $\sqrt{|V| - 1}$  number of  $v_i$  from  $V$  that minimizes MSE
         $\frac{1}{|S_{N_p}|} \sum_{(x_d, y_d) \in S_{N_p}} (y_d - T_k(x_d))^2$ . Splitting that generates child node stops when
         $S_N = 1$  at each node  $N$ .}
18    3. for (  $N \in T_i$  ) {
19      if  $v_i$  of  $N \neq \emptyset$  then
20         $I_{v_i \rightarrow v_j}^k(N) \leftarrow I_{v_i \rightarrow v_j}^k(N) + (\text{Var}(S_N) - \text{Var}(S_{N_{true}}) - \text{Var}(S_{N_{false}}))$  where
         $\text{Var}(\cdot)$  is the variance of the output variable  $y_d$  in each subset  $S_N, S_{N_{true}}$ 
        and  $S_{N_{false}}$ 
21      end
22    }
23  }
24 }
25 then  $w_{ij}^* \leftarrow \frac{1}{m} \sum_{k \in \{1, 2, \dots, m\}} I_{v_i \rightarrow v_j}^k(N)$ 
26 return  $W^* = (w_{ij}^*) \in [0, 1]^{|V| \times |V|}$ 

```

---

## 4 Empirical Findings

This section verifies how well the fitted network  $G^* = (V, E, W^*)$  aligns with the jurisprudences of the *Panel* or *Appellate Body* of WTO DSB. Since these two judicial bodies of WTO DSB authoritatively opine on how the regulatory system of WTO DSB is systematically organized, this section will validate the quality of the fitted network  $G^*$  by introducing three different sub-networks of the fitted network  $G^*$  where each sub-network shows how articles of WTO agreements cooperatively achieve important principles of WTO and regulate specific trade issues.

### 4.1 *Market Access*: Ensuring Foreign Goods to Cross Borders and Fairly Compete with Domestic Products

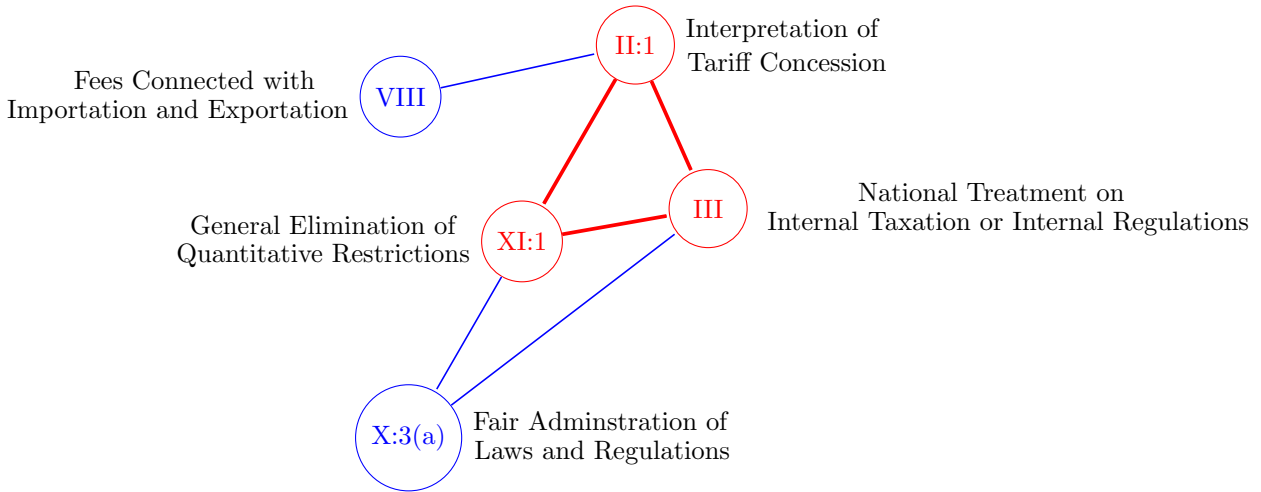


Figure 23: **Network of the Articles that Achieves *Market Access***: This figure demonstrates a network of articles of WTO agreements that cooperatively achieve the principle of *Market Access* in the regulatory system of WTO. Tariff and Non-tariff barriers such as quantitative restriction, internal taxations and extra fees for crossing borders can inhibit the chance of foreign goods to access the foreign market. Therefore, these articles tend to work together to ensure the *Market Access* principle working properly.

*Market Access* refers to the guarantee of the conditions relating to the tariff or non-tariff measures for the entry of goods into the foreign market. For example, if a country set a limit on the quantity of imported goods, no foreign goods can get access to that market above that limit. Also, if a country sets too high tariff rate for the foreign goods that cross the border, it can prevent those foreign goods from accessing that market. Therefore, *Market Access* is the most basic but the most important principle of the WTO because it directly represents the primary goal of the



WTO.

Three different types of measures - *quantitative restriction*, *tariff* and *internal regulations* - are mostly preferred by the countries to protect their domestic market from the importation of foreign goods. For example, a country can discriminate imported and domestic products with its internal regulation as shown in the *Korea - Beef* case in Section 3.2. These three types of measures that are mostly relevant to the *Market Access* are illustrated in Figure 23. Article XI:1, II:1 and III correspond to the obligations that regulate quantitative restriction, tariff and internal regulations respectively. Edges are colored in red if the mean of the two directed edge weights are greater than 10%. For example,  $w^*(\text{II:1}, \text{III}) = 0.117$  and  $w^*(\text{III}, \text{II:1}) = 0.122$  thus their simple mean is 11.95% which is greater than 10%. This red triangle that is comprised of Article XI:1, II:1 and III shows that the fitted network  $G^* = (V, E, W^*)$  captures the principle of *Market Access* as a cluster of these three articles with relatively large edge weights ( $> 10\%$ )<sup>11</sup>. These three articles are *cohesively* interconnected because they need each other to clearly distinguish the role of each one from others to achieve the *Market Access*. For example, *Panel* explicitly opinionated that Article XI:1 and Article III shall take in charge of *quantitative restriction* and *internal requirements* respectively without duplication of their responsibilities as following:

“the General Agreement distinguishes between measures affecting the ‘importation’ of products, which are regulated in Article XI:1, and those affecting ‘imported products’, which are dealt with in Article III. **If Article XI:1 were interpreted broadly to cover also internal requirements, Article III would be partly superfluous.**”

In addition to it, the fitted network  $G^*$  also captures sub-articles that support the *Market Access* triangle. For example, I colored edges in blue if the mean of the two directed edge weights are greater than 5% and less than 10%. The Article X:3(a) which prohibits *unfair administration of laws and regulations* are connected to the Article XI:1 and III because those two are mostly deteriorates the principle of *Market Access* by administering their relevant laws and regulations unfairly. Also, in a similar notion, Article VIII that prohibits *the collection of abnormal amount of fees connected with importation or exportation* is connected to the Article II:1 because often those fees are indistinguishable from the tariff schedule whose Interpretation is regulated under Article II:1. For example, in *Argentina - Textile* case, Argentina insisted that they didn’t collect fees connected to importation inconsistently with the obligation under Article VIII because they

---

<sup>11</sup>Compare 10% to the simple mean of total edge weight 100% over all 79 possible source nodes, 1.26% ( $\sim 100\%/79$ )

included the fees into the tariff schedule. However, *Panel* refuted this logic as following<sup>12</sup>:

“The provisions of the WTO Understanding on the Interpretation of **Article II:1(b)** of GATT 1994, dealing with ‘other duties and charges’, make clear that **including a charge in a schedule of concessions in no way immunizes that charge from challenge as a violation of an applicable GATT rule** . . . Therefore, we consider that the fact that Argentina’s statistical tax is included in its Schedule is not a defence to its inconsistency with the provisions of **Article VIII** of GATT.”

## 4.2 *Reciprocity* in Non-Tariff Barriers: Compensate or Retaliate as Much as You have Protected

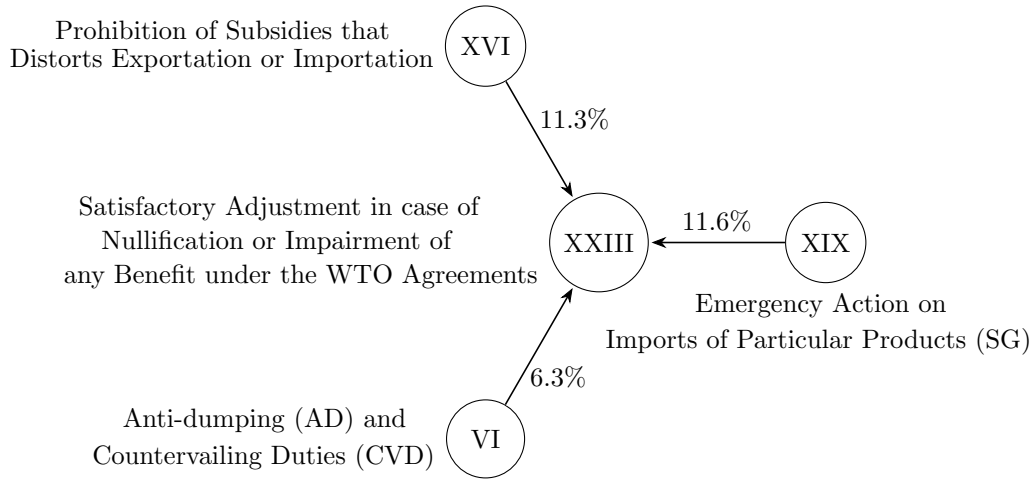


Figure 24: **Network of Articles that Achieves *Reciprocity* for Non-Tariff Barriers:** This figure demonstrates a network of articles of WTO agreements that cooperatively regulates *Non-Tariff Barriers* (NTBs) in WTO DSB. Three major NTBs such as *Anti-dumping (AD)*, *Countervailing Duties (CVD)* and *Safeguard (SG)* are relying on Article XXIII to resolve its inconsistency with the principle of *Reciprocity*.

Ensuring the *Market Access* principle is a basic approach to smooth the world trade flow as explained in the previous section. However, members are sometimes allowed to apply higher duties or to maintain a quantitative restrictions in certain circumstances, such as to act against low-price dumping of foreign producers, to offset the distorted competitive relationship resulted by the illegal subsidies or to take an emergency measure in case irreversible injury to domestic industry is expected. Each of these examples are called *Anti-dumping (AD)*, *Countervailing Duties (CVD)* and *Safeguard (SG)* respectively. These measures are recognized as three major types of

<sup>12</sup>Panel Report, Argentina – Textiles and Apparel, paras. 6.81-6.82.

*Non-Tariff Barriers* (NTBs) since these three measures often fail to satisfy the legal requirements of the WTO agreements and evolve into trade disputes.

The principle of *reciprocity* is a general notion that regulates these three major NTBs. This principle requires change of value of imports affected by a country's trade policy shall be balanced with the equal value of exports across trading partners affected (Bagwell and Staiger, 1999). The rules of the WTO agreements, in particular the Article XXIII of GATT 1994, confers a right to insist its nullified or impaired benefit that is expected under the WTO agreements and require satisfactory adjustment to the member who results in such nullification or impairment. Moreover, this article confers a right to retaliate if no satisfactory adjustment is being fulfilled.

Therefore, Article XXIII regulates this action-reaction characteristics of *AD*, *CVD* and *SG* to achieve the principle of *reciprocity*. For example, *Panel* stated that a member can resort to Article XXIII and raise a legal claim in case another member levies unjustified antidumping duties that fails to fully explain the causal relationship between the dumping and the material injury to the related industry in its report on *New Zealand - Imports of Electrical Transformers from Finland*<sup>13</sup>:

“Panel believed that if a contracting party affected by the determination (of levying **anti-dumping duties**) could make a case that the importation could not in itself have the effect of causing material injury to the industry in question, that contracting party was entitled, under the relevant GATT provisions, **in particular Article XXIII**, that its representations be given sympathetic consideration and that eventually, **if no satisfactory adjustment was effected, it might refer the matter to the CONTRACTING PARTIES**<sup>14</sup>, as had been done by Finland in the present case.”

This relationship is captured by the 6.3% weight of directed edge from Article VI to Article XXIII. This is because citing Article VI (Antidumping duties and CVD) naturally leads to the citation of Article XXIII to insist its impaired benefit and solicit the satisfactory adjustment relying on the principle of *reciprocity* in WTO DSB. Also, other two types of measures, *Illegal Subsidy* and *SG* whose legal requirements are regulated under the Article XVI and XIX respectively also heavily rely (> 10%) on the Article XXIII to solicit its impaired benefit and satisfactory adjustment based on the principle of *reciprocity* in case the breach of those requirements occurs.

---

<sup>13</sup> *WTO official document* L/5814, adopted on 18 July 1985, pp.67-68, para.4.4.

<sup>14</sup> It means raising a legal claim through WTO DSB

### 4.3 *Non-discrimination* Principle for Formulating Regional Trade Agreements (RTA)

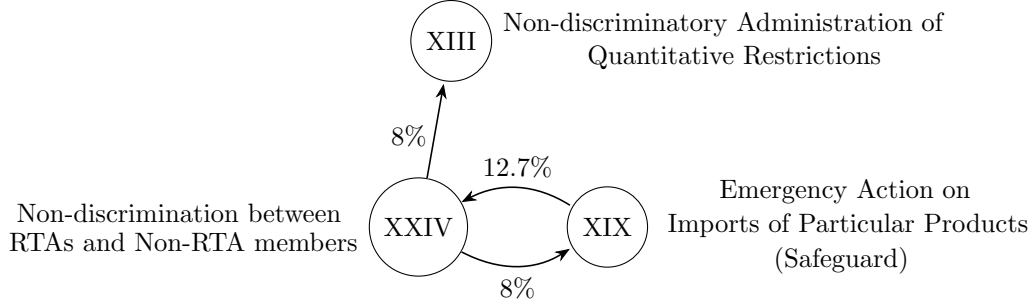


Figure 25: **Network of Articles that Achieves *Non-discrimination* for Formulating *RTAs*:** A rudimentary obligation of Article XXIV that prohibits discrimination between non-RTA and RTA members become realized in a relationship with different types of trade measures, such as *Safeguard* or *Quantitative Restriction*.

The term *Regional Trade Agreements* (RTAs) refer to any reciprocal trade agreements between two or more trade partners. WTO permits its members to enter into RTAs under specific conditions written in the Article XXIV. However, Article XXIV roughly draws a basic condition to formulate RTAs and its detailed regulatory system has evolved with a number of disputes arises from numerous discriminatory uses of RTA. For example, in *US - Line Pipe* case, the United States applied its safeguard measures to its trade partners excluding Canada and Mexico who are members of the NAFTA although the United States had included Canada and Mexico in the analysis of whether increased imports caused serious injury to domestic production. Therefore, *Appellate Body* concluded this discriminatory application of safeguard measures as inconsistent with the rules of the WTO agreements. This case, where the breach of the Article XIX (Safeguard) leads to the breach of the Article XXIV (Non-discrimination between RTA and Non-RTA members) is captured in the 12.7% weight of directed edge from Article XIX to Article XXIV in Figure 25. This relatively high edge weight results from the numerous real world cases where the members frequently discriminate non-RTA members from the RTA members when they maintain its *Safeguard* measures to protect their domestic industry. Also, we can understand the opposite direction of 8% weighted edge from Article XXIV to Article XIX as the *non-discrimination* principle in Article XXIV is extended to the Article XIX as members of WTO have repeatedly used RTAs to justify their discriminatory application of *Safeguard measures* between RTA and non-RTA members.

In addition to the case of *Safeguard*, members frequently try to justify their discriminatory

use of quantitative restriction by insisting that Article XXIV has conferred its right to unequally distribute the quantitative restriction between RTA and Non-RTA members. For example, in the *Turkey – Textiles* case, Turkey defended himself that Article XXIV authorizes members forming a customs union to deviate from the obligations of Article XIII that requires members to apply quantitative restriction without discrimination between members. However, *Appellate Body* explicitly refuted this logic as saying<sup>15</sup>:

“Article XXIV does not allow Turkey to adopt, upon the formation of a customs union with the European Communities, quantitative restrictions . . . **which were found inconsistent with Articles XI and XIII of GATT 1994** and Article 2.4 of the ATC. However, the Appellate Body stressed that it was only finding that **Turkey’s quantitative restrictions at issue were not justified by Article XXIV . . .**”

This legal context that Article XXIV does not exempt the obligation under Article XIII is captured by the 8% weighted directed edge from Article XXIV to Article XIII in Figure 25.

## 5 Conclusion

Upon the understanding that articles are interactively working together to constitute the jurisprudences of WTO DSB, this paper proposed a new method that can materialize the relationship between articles in a form of weighted and directed network using deep and machine learning.

As shown in the section 4, the result network captures the important interactions between articles of WTO agreements that are covered by the *Panel* and *Appellate Body* with relatively large edge weights. As legal experts understand those important cohesions between articles that constitute the important jurisprudences of WTO DSB, this method can also provide a general guideline about where to focus and how to mingle the articles of WTO agreements to justly claim the illegality of a trade policy at issue and satisfactory adjustment for one’s impaired benefit. Moreover, though jurisprudence of WTO DSB constantly evolves with new cases, this method can adapt to the newly created jurisprudences by training new cases.

Therefore, this method could be considered as a technical solution to narrow the gap between developing and developed countries in terms of legal capacity in WTO DSB. Without spending lots of time to study and elaborate the jurisprudences explained in the *Panel* and *Appellate Body* report, this method provide a legal guidelines where to focus and how to compose the

---

<sup>15</sup>Appellate Body Report, *Turkey – Textiles*, para. 65.

legal claim with which articles that fits to the case at issue. Moreover, the previous approach that provided legal advice to developing countries has been ineffective because it cannot create a shared understanding over the system between developing and developed countries. If we shift our focus to these technical solutions that can materialize the current shape of the system, WTO DSB will become more effective as members can discuss their trade issues upon the measurable ground of shared understanding about how WTO DSB works.

## References

- Bagwell, Kyle and Robert W. Staiger. 1999. “An Economic Theory of GATT.” *The American Economic Review* 89(1):215–248.
- Bown, Chad P. 2002. “The Economics of Trade Disputes, the GATT’s Article XXIII, and the WTO’s Dispute Settlement Understanding.” *Economics and Politics* 14(3):283–323.
- Breiman, Leo. 2001. *Random Forests*.
- Charles, Zachary and Dimitris Papailiopoulos. 2017. “Stability and Generalization of Learning Algorithms that Converge to Global Optima.”.
- Charnovitz, Steve. 2004. “THE WTO AND COSMOPOLITICS.” *Journal of International Economic Law* 7(3):675–682.
- Dinh, Laurent, Razvan Pascanu, Samy Bengio and Yoshua Bengio. 2017. “Sharp Minima Can Generalize For Deep Nets.”.
- Guo, Jiaxian, Sidi Lu, Han Cai, Weinan Zhang, Yong Yu and Jun Wang. 2017. “Long Text Generation via Adversarial Training with Leaked Information.”.
- Huynh-Thu, Vân Anh, Alexandre Irrthum, Louis Wehenkel and Pierre Geurts. 2010. “Inferring Regulatory Networks from Expression Data Using Tree-Based Methods.” *PLOS ONE* 5(9):1–10.
- Johns, Leslie and Krzysztof J. Pelc. 2014. “Who Gets to Be In the Room? Manipulating Participation in WTO Disputes.” *International Organization* 68(3):663–699.
- Kim, Yoon. 2014. “Convolutional Neural Networks for Sentence Classification.” *CoRR* abs/1408.5882.

- Lawrence, S., C. L. Giles, Ah Chung Tsoi and A. D. Back. 1997. "Face recognition: a convolutional neural-network approach." *IEEE Transactions on Neural Networks* 8(1):98–113.
- Magdum, P. G. and Sheetal Rathi. 2020. A Survey on Deep Learning-Based Automatic Text Summarization Models. In *Advances in Artificial Intelligence and Data Engineering*, ed. Niranjan N. Chiplunkar and Takanori Fukao. Singapore: Springer Singapore pp. 377–392.
- Minaee, Shervin, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu and Jianfeng Gao. 2020. "Deep Learning Based Text Classification: A Comprehensive Review."
- Nahid, A. and Y. Kong. 2017. Local and Global Feature Utilization for Breast Image Classification by Convolutional Neural Network. In *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. pp. 1–6.
- Neyshabur, Behnam, Srinadh Bhojanapalli, David McAllester and Nathan Srebro. 2017. "Exploring Generalization in Deep Learning."
- Palmeter, D. and P.C. Mavroidis. 2004. *Dispute Settlement in the World Trade Organization: Practice and Procedure*. Dispute Settlement in the World Trade Organization: Practice and Procedure Cambridge University Press.
- Pelc, Krzysztof J. 2014. "The Politics of Precedent in International Law: A Social Network Application." *The American Political Science Review* 108(3):547–564.
- Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *Journal of Machine Learning Research* 15(56):1929–1958.
- Srivastava, Rupesh Kumar, Klaus Greff and Jürgen Schmidhuber. 2015. "Highway Networks." *CoRR* abs/1505.00387.
- Strezhnev, BuenoAnton. 2014. "Using Latent Space Models to Study International Legal Precedent: An Application to the WTO Dispute Settlement Body." *American Political Science Association 2014 Annual Meeting*.
- Trachtman, Joel. 2003. "Toward open recognition? Standardization and regional integration under Article XXIV of GATT." *Journal of International Economic Law* 6:459–492.

World Trade Organization. 1999. *WTO Agreements Series*. Number no. 2 in “WTO Agreements Series” World Trade Organization.

World Trade Organization. 2017. *A Handbook on the WTO Dispute Settlement System*. A WTO Secretariat publication Cambridge University Press.

Zahavy, Tom, Bingyi Kang, Alex Sivak, Jiashi Feng, Huan Xu and Shie Mannor. 2017. “Ensemble Robustness and Generalization of Stochastic Deep Learning Algorithms.”.

## Appendix A

### A.1 Factual Aspect Example

Excerpt below is from the panel report for the *US - Offset Act (Byrd Amendment)*<sup>16</sup> case.

#### II. FACTUAL ASPECTS

2.1 This dispute concerns the Continued Dumping and Subsidy Offset Act of 2000 (the “CDSOA” or the “Offset Act”), which was enacted on 28 October 2000 as part of the Agriculture, Rural Development, Food and Drug Administration and Related Agencies Appropriations Act, 2001.<sup>1</sup> The CDSOA amends Title VII of the Tariff Act of 1930 by adding a new section 754 entitled Continued Dumping and Subsidy Offset. Regulations prescribing administrative procedures under the Act were brought into effect on September 21, 2001.

2.2 The CDSOA provides that :

“ Duties assessed pursuant to a countervailing duty order, an anti-dumping duty order, or a finding under the Antidumping Act of 1921 shall be distributed on an annual basis under this section to the affected domestic producers for qualifying expenditures. Such distribution shall be known as “the continued dumping and subsidy offset”. ”

2.3 The term “affected domestic producers” means :

---

<sup>16</sup>Panel Report, United States — Continued Dumping and Subsidy Offset Act of 2000, WTO Doc. WT/DS217/R (adopted Jan. 27, 2003).



“ a manufacturer, producer, farmer, rancher, or worker representative (including associations of such persons) that –

(A) was a petitioner or interested party in support of the petition with respect to which an anti-dumping duty order, a finding under the Antidumping Act of 1921, or a countervailing duty order has been entered, and

(B) remains in operation.

Companies, business, or persons that have ceased the production of the product covered by the order or finding or who have been acquired by a company or business that is related to a company that opposed the investigation shall not be an affected domestic producer. ”

2.4 In turn, the term “qualifying expenditure” is defined by the CDSOA as “expenditure[s] incurred after the issuance of the anti-dumping duty finding or order or countervailing duty order in any of the following categories: “

(A) Manufacturing facilities.

(B) Equipment.

(C) Research and development.

(D) Personnel training.

(E) Acquisition of technology.

(F) Health care benefits to employees paid for by the employer.

(G) Pension benefits to employees paid for by the employer.

(H) Environmental equipment, training or technology.

(I) Acquisition of raw materials and other inputs.

(J) Working capital or other funds needed to maintain production.” ”

2.5 The CDSOA provides that the Commissioner of Customs shall establish in the Treasury of the United States a special account with respect to each order or finding<sup>8</sup> and deposit into such account all the duties assessed under that Order.<sup>9</sup> The Commissioner of Customs shall distribute all funds (including all interest earned on the funds) from

the assessed duties received in the preceding fiscal year to affected domestic producers based on a certification by the affected domestic producer that he is eligible to receive the distribution and desires to receive a distribution for qualifying expenditures incurred since the issuance of the order or finding.<sup>10</sup> Funds deposited in each special account during each fiscal year are to be distributed no later than 60 days after the beginning of the following fiscal year.<sup>11</sup> The CDSOA and regulations prescribe that (1) if the total amount of the certified net claims filed by affected domestic producers does not exceed the amount of the offset available, the certified net claim for each affected domestic producer will be paid in full, and (2) if the certified net claims exceed the amount available, the offset will be made on a pro rata basis based on each affected domestic producer's total certified claim.

2.6 Special accounts are to be terminated after “(A) the order or finding with respect to which the account was established has terminated; (B) all entries relating to the order or finding are liquidated and duties assessed collected; (C) the Commissioner has provided notice and a final opportunity to obtain distribution pursuant to subsection (c); and (D) 90 days has elapsed from the date of the notice described in subparagraph (C).” All amounts that remain unclaimed in the Account are to be permanently deposited into the general fund in the US Treasury.<sup>12</sup>

2.7 The CDSOA applies with respect to all anti-dumping and countervailing duty assessments made on or after 1 October 2000<sup>13</sup> pursuant to an anti-dumping order or a countervailing order or a finding under the Antidumping Act of 1921 in effect on 1 January 1999 or issued thereafter. [END]

## A.2 Collected Cited Articles for 143 WTO DSB Cases

DS refers to *Dispute Settlement* and this notation is officially adopted by WTO DSB.

WTO DSB identifies each dispute with a unique number for each case such as DS2 and DS18.

Case Number	Cited Articles (GATT 1994)
DS 2	I, III, XXII:1
DS 18	XI, XIII

DS 22	VI:3, VI:6
DS 31	III, XI
DS 34	XI, XIII, XXIV
DS 46	XVI
DS 56	II, VII, VIII, X
DS 58	I, XI, XIII, XX
DS 60	VI
DS 62	II
DS 67	II, XXIII, XXIII:1
DS 68	II, XXII:1, XXIII:1
DS 69	II, III, X, XIII, XXVIII
DS 75	III:2
DS 76	XI
DS 87	III:2
DS 90	XI:1, XIII, XVIII:11
DS 98	XIX
DS 103	X, XI, XIII
DS 108	III:4, XVI
DS 121	XIX
DS 122	VI
DS 135	III, XI, XXIII, XXIII:1(b)
DS 136	III:4, VI
DS 139	I:1, III:4, XXIV
DS 141	I, VI
DS 146	III, XI
DS 152	I, II, III, VIII, XI
DS 155	III:2, X:3(a), XI:1
DS 161	II, III, X, XI, XVII
DS 162	III, III:4, VI, XI
DS 165	I, II, VIII, XI

DS 166	I, XIX
DS 174	I, III:4
DS 175	III, III:4, XI, XI:1
DS 177	I, II, XIX
DS 184	VI, X
DS 202	I, XIII, XIX
DS 207	II, XIX:1
DS 212	VI:3
DS 217	VI:2, VI:3, X:3, XXIII:1
DS 219	I, VI
DS 221	VI, VI:2, VI:3, VI:6
DS 231	I, III, XI:1
DS 234	VI, VI:2, VI:3, X, X:3, XXIII:1
DS 238	XIX:1
DS 244	VI, X
DS 245	XI
DS 246	I:1
DS 248	I:1, XIII, XIX:1
DS 257	VI, VI:3, X:3
DS 264	VI, X:3
DS 265	III:4, XVI
DS 266	III:4, XVI
DS 267	III:4, XVI
DS 268	VI, X
DS 269	II, II:1, XXIII, XXIII:1, XXVIII
DS 276	III, III:4, XVII, XVII:1
DS 282	VI, X
DS 283	III:4
DS 286	II, XXII
DS 290	I, I:1, III, III:4

DS 294	VI
DS 295	VI, VI:2
DS 296	VI:3, X:3
DS 301	I:1, III:4, XXIII:1
DS 302	II:1, III:2, III:4, X:1, X:3, X:3(a), XI:1, XV
DS 308	III
DS 312	VI:1, VI:2(a), VI:2(b), VI:6
DS 315	X:1, X:3
DS 316	III:4, XVI:1, XXIII:1
DS 320	I, II
DS 321	I, II
DS 322	VI, VI:1, VI:2(a)
DS 332	I:1, III:4, XI:1, XIII:1
DS 336	VI:3, X:3
DS 339	II:1, III:1, III:2, III:4, III:5, XI, XIII:1
DS 343	I:1, II, II:1, III, VI, VI:2, X:3(a), XI:1, XIII:1, XX
DS 344	VI, VI:1, VI:2
DS 345	I, II, II:1, VI, VI:2, VI:3, X, X:1, X:2, XI, XIII
DS 350	VI:1, VI:2
DS 353	III:4
DS 360	II:1, III:2, III:4
DS 363	III:4, XI:1
DS 366	I:1, II:1, III:2, V:6, VII, VII:1, X:3, X:3(a), XI, XIII:1
DS 371	II:1(b), II:3, III:2, III:4, VII:1, VII:2, VII:5, X:1, X:3, X:3(a)
DS 379	I, VI
DS 381	I, III
DS 384	III:4, IX, IX:2, X:3, X:3(a), XXIII:1(b)
DS 392	I:1, XI:1
DS 394	VIII, VIII:1, VIII:4, X, X:1, X:3, XI, XI:1
DS 396	III:1, III:2

<b>DS 397</b>	I, I:1, VI:1, X:3(a)
<b>DS 399</b>	I:1, II, XIX
<b>DS 400</b>	I:1, III:4, XI:1, XXIII:1(b)
<b>DS 406</b>	III:4, XX, XXIII:1(a)
<b>DS 412</b>	III:4, III:5, XXIII:1
<b>DS 414</b>	VI
<b>DS 415</b>	I:1, II:1, XIX:1, XIX:2
<b>DS 422</b>	VI:1, VI:2(a), VI:2(b)
<b>DS 425</b>	VI:1, VI:6
<b>DS 427</b>	VI, VI:3
<b>DS 429</b>	VI:1, VI:2, VI:2(a), X
<b>DS 430</b>	I, XI
<b>DS 431</b>	VII, VIII, X, X:3(a), XI, XI:1
<b>DS 435</b>	III:4
<b>DS 436</b>	I, VI
<b>DS 437</b>	VI, XXIII
<b>DS 440</b>	VI
<b>DS 442</b>	VI, X:3(a)
<b>DS 447</b>	I:1, III:4, XI:1
<b>DS 449</b>	VI, X
<b>DS 453</b>	I:1, III:2, III:4, XI:1
<b>DS 454</b>	VI
<b>DS 456</b>	III:4
<b>DS 457</b>	II:1(a), II:1(b), X:1, X:3(a), XI, XI:1
<b>DS 461</b>	II:1, II:1(b), VIII:1, X:3(a)
<b>DS 464</b>	VI, VI:1, VI:2, VI:3
<b>DS 468</b>	II:1(b), XIX:1
<b>DS 471</b>	VI:2
<b>DS 472</b>	I:1, II:1(b), III:2, III:4, III:5
<b>DS 473</b>	VI:2

<b>DS 475</b>	I:1, III:4, XI:1
<b>DS 476</b>	I, III, X, XI
<b>DS 477</b>	III:4, X:1, XI:1
<b>DS 479</b>	VI
<b>DS 480</b>	VI, VI:1, VI:2
<b>DS 482</b>	VI
<b>DS 483</b>	VI
<b>DS 484</b>	III:4, X:1, X:3, XI:1
<b>DS 485</b>	II:1(a), II:1(b), VII
<b>DS 486</b>	VI
<b>DS 488</b>	I, X:3
<b>DS 490</b>	I:1, XIX:1, XIX:2
<b>DS 492</b>	I, I:1, II, II:1, II:2, XIII, XIII:1, XIII:2, XXVIII
<b>DS 493</b>	VI
<b>DS 495</b>	XXIII:1
<b>DS 499</b>	I:1, III:4, X:3(a), XI:1, XIII:1
<b>DS 504</b>	VI
<b>DS 505</b>	VI:3
<b>DS 513</b>	I:1, X:1, X:2, X:3(a), XI:1
<b>DS 518</b>	I:1, II:1(b), XI:1, XIX:1
<b>DS 523</b>	VI:3