# Mapping Regulatory System of WTO Dispute Settlement Body Using Deep Learning

Suyeol Yun

November 18, 2020

**Abstract**

## 1 Introduction

The Dispute Settlement Body (DSB) of the World Trade Organization (WTO) deals with trade disputes between WTO members. WTO members can file a lawsuit in WTO DSB to claim their impaired benifit related to the WTO agreements as a result of possible illegal action of a other 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 (Palmeter 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 multiple rules of the WTO agreements such as *rules of anti-dumping* and *rules of subsidy* because this distribution could constitute an illegal subsidy and 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 Acceess* which is one of the most important principles in WTO, is achieved through the cooperation of multile articles of the WTO agreements (*See* Figure 20) where the cooperation is realized thorugh the act of citation. Therefore, analyzing the

network of articles of the WTO agreements provides us a clearer view on how WTO DSB constitutes specific norm or 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 number (less than 10) of articles of the WTO agreements and has not been conducted in a level of entire WTO agreements. There exists an 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 and researchers of the WTO agreements. This exclusiveness becomes more severe as the number of requested cases increases to WTO DSB and this has led to a widening gap of legal capcity between developing and developed countries in WTO. 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 addresses importane of revolutionizing the way of studying network of articles of WTO agreements and propose a new methodology to materialize the relationship between the articles of the WTO agreements as a whole. For this purpose, this paper maps the regulatory system of WTO DSB as a network of legal articles of the WTO agreements 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 the trade policy as exemplified 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 circumnstances of the dispute and the reuglatory contents described in 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 or vice versa (Johns and Pelc, 2014). Moeroever, members cite articles strategically trying to reshape the legal precedents of

> 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-dumpig (AD) and subsidy (SCM) at the same time to cover its complex characteristics.

WTO DSB in favor of their future interest (Pelc, 2014; Strezhnev, 2014).

Upon these two considerations, this paper adopts the deep neural network as a technical solution. This is because a deep neural network is generally known as good at effectively extracting information from text data and generalizing the patterns inside data. Therefore, this paper designs a deep neural network (Figure **??**) 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 the factual description (*See* 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 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 that led to the dispute (*See* Figure 16 and Figure 17). After training, I fitted a set of directed edge weight $W^*$ that best explains the variance of each article's citability predicted by the trained deep neural network using a ensemble of random forests (Breiman, 2001; Huynh-Thu et al., 2010).

To check whether this fitted network of articles of the WTO agreements $G^* = (V, E, W^*)$ maps the regulatory system of WTO DSB properly, this paper compares the way of the fitted network $G^*$ achieving the main pricples of WTO, *Market Access*, *Reciprocity* and *Non-discrimination*, with the jurisprudence of WTO DSB made by *Panel* and *Appellate Body*. The comparison reveals that

Network of legal articles of WTO agreements is defined as

*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

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

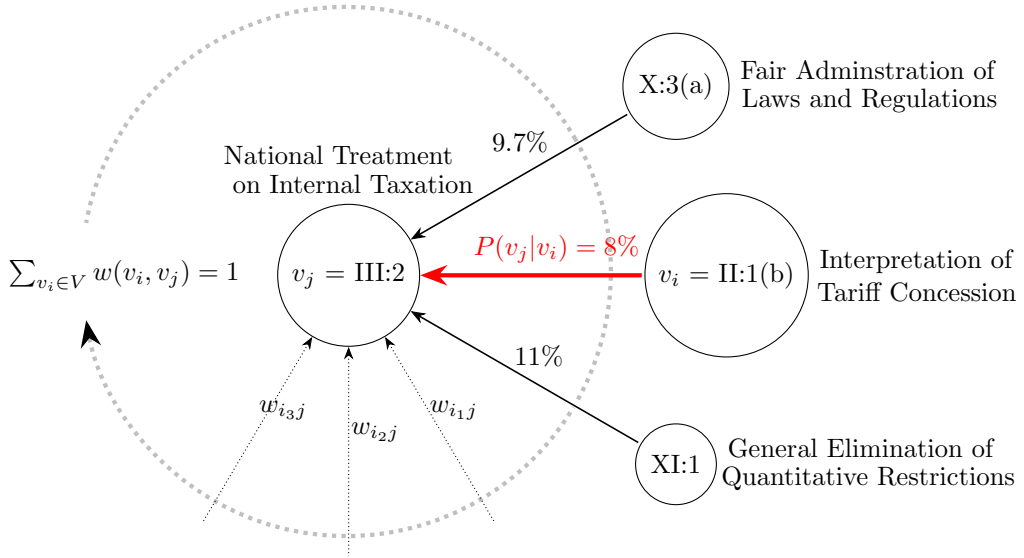
Then define *edge weight matrix* $W = (w_{ij}) \in {\mathbb{R}_+}^{|V| \times |V|} \; s.t. \; w_{ij} = w(v_i, v_j)$

($W$ is more formally called *weighted adjancency 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

the fitted network $G^*$ captures the interaction between the articles of WTO agreements similarly with the jurisprudence of *Panel* and *the Appellate Body*. This similarity guarantees that the fitted network $G^*$ closely maps the regulatory system of WTO DSB since these two judicial bodies can authoritatively consitute the jurisprudence over how rules of WTO agreements are working together to achieve those main principles of WTO.

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 between legal experts and reveals important interactions between articles inside the system, it can induce a lower cost to build the same amount of legal capacity to understand the WTO DSB.




(a) **Illustrated edge weights of a source node Article II:1(b)**

"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 (8%) where the source node Article II:1(b) constitutes the legal context of the target node Article III:2 with the probability of 8% among all 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: **Article 7 of the Dispute Settlement Understanding (DSU):** DSU provides a legal guidelines on how judicial boides of WTO shall adjudicate the requested disputes. It explictly 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.

# 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 explictly 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 cited articles as shown in Table 1 from the WTO agreements to claim its inconsitencies of *Continued Dumping and Subsidy Act of 2000* (CDSOA) to those cited articles[1]:

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

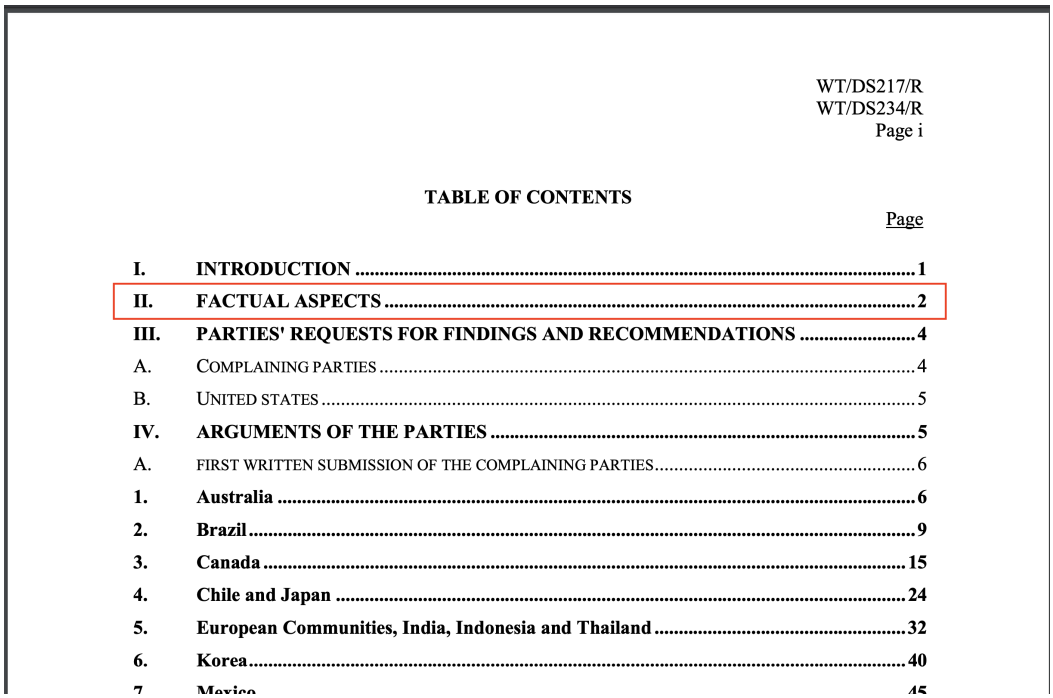
## 2.2 Factual Aspect: Textual Description of the Dispute

Textual description of the dispute is preferably called as *Factual Aspect* in WTO DSB. Since Panels always provide a factual aspect[2] that summarizes the content of the dispute in the panel report,

[0] Australia, Brazil, Chile, European Communities, India, Indonesia, Japan, Korea and Thailand

[1] 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.

[2] It's worth noting that Appellate Body doesn't provide any factual aspect because they use the factual aspect provided by the Panel.



WT/DS217/R
WT/DS234/R
Page i

**TABLE OF CONTENTS**

Page



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

I wrote a program that can automatically search and collect the panel reports from the WTO official doucment website[3]. Then I located the factual asepct 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 number of different factual aspects from the sanme 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 number (596[4]) of cases requested to WTO DSB. This is due to the two reasons.

First, panel handles different cases together if the case is about the same trade policy. For example, in *US - Offset (Byrd Amendment)*, panel merged DS217[5] 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 representatitve number for this kind of joint Adjudication. For example, since DS217 and DS234 shares the same panel report, this paper choose DS217 as a representatitve number as shown in Figure 7 where the list includes DS217 but not DS234.

[3] `http://docs.wto.org`
[4] As of November 1st, 2020.
[5] DS refers to Dispute Settelement. DS is official prefix that indicates the case in WTO DSB.

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 available. 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 United States was challenged by Turkey for its possibly inconsistent anti-dumping measure.

**WORLD TRADE ORGANIZATION**

**WT/DS217/R**
**WT/DS234/R**
16 September 2002
(02-4742)

Original: English

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

***Report of the Panel***

Figure 8: **Cover of Panel Report:** Panel explicitly marks which different cases are handled together in the cover of the panel report. DS217 and DS234 are handled togehter in this example.

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

Every lawsuit in WTO DSB cites multiple set of aritcles 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[6]. The webpage chronologically lists up all dispute cases requested to WTO DSB and the prgoram visits each page of 143 cases and collects the cited articles. Among all the articles from different agreements of the WTO agreements[7], 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 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:2, VI:2(a), VI:2(b), 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

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

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

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.

[6] `https://www.wto.org/english/tratop_e/dispu_e/dispu_status_e.htm`

[7] WTO agreeemnts is comprised of mutiple 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 (antidumping), Agreement on Subsidies and Countervailing Measures, Agreement on Rules of Origin, Agreement on Safeguards and so on.

### 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 VI, member sometimes cites Article VI as a whole but sometimes cites Article VI:2 or Article VI:2(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. Both judicial bodies cite the legal articles with the level of *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 jurisprudence of WTO DSB.

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

| **Scope** | **Quote** | **Source** |
|---|---|---|
| Title | "As the ***title*** **of Article 21 makes clear**, 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, *US – Shrimp (Malaysia)*, paras. 86-87. |
| Article | "The sequence of steps indicated above in the analysis of a claim of justification under **Article XX** reflects, not inadvertence or random choice, but rather the fundamental structure and logic of Article XX. . . ." | Appellate Body Report, *US – Shrimp (Malaysia)*, paras. 119-120. |
| Paragraph | "The verb 'may' in **Article VI:2** 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, *US – 1916 Act*, paras. 116. |

| | | |
|---|---|---|
| Sentence | "The customary rules of interpretation of public international law as required by **the first sentence of Article 17.6(ii) of the Anti-Dumping Agreement**, do not admit of another interpretation as far as the issue of zeroing raised in this appeal is concerned." | Appellate Body Report, *US – Zeroing (EC)*, 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. **The term 'commerce'** 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, *Colombia – Textiles*, para. 5.34. |

# 3 Methodology: Considerations and Development

This section introduces two main considerations to design the method used in this paper. Then it explains the method that is used to fit the network of articles of WTO agreement under those considerations.

## 3.1 Two Main Considerations For Design of Method

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 importance of using the information represented in a form of textual description inside the content of dispute 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 startegic citation pattern. Since members of the WTO strategically cite the articles of WTO agreement expecting different outcomes thats serves member-specific national

interest (Johns and Pelc, 2014; Pelc, 2014; Strezhnev, 2014), this paper selected a method that can generalize this member speicific citation pattern. These two considerations and the solution will be explained in the following subsections.

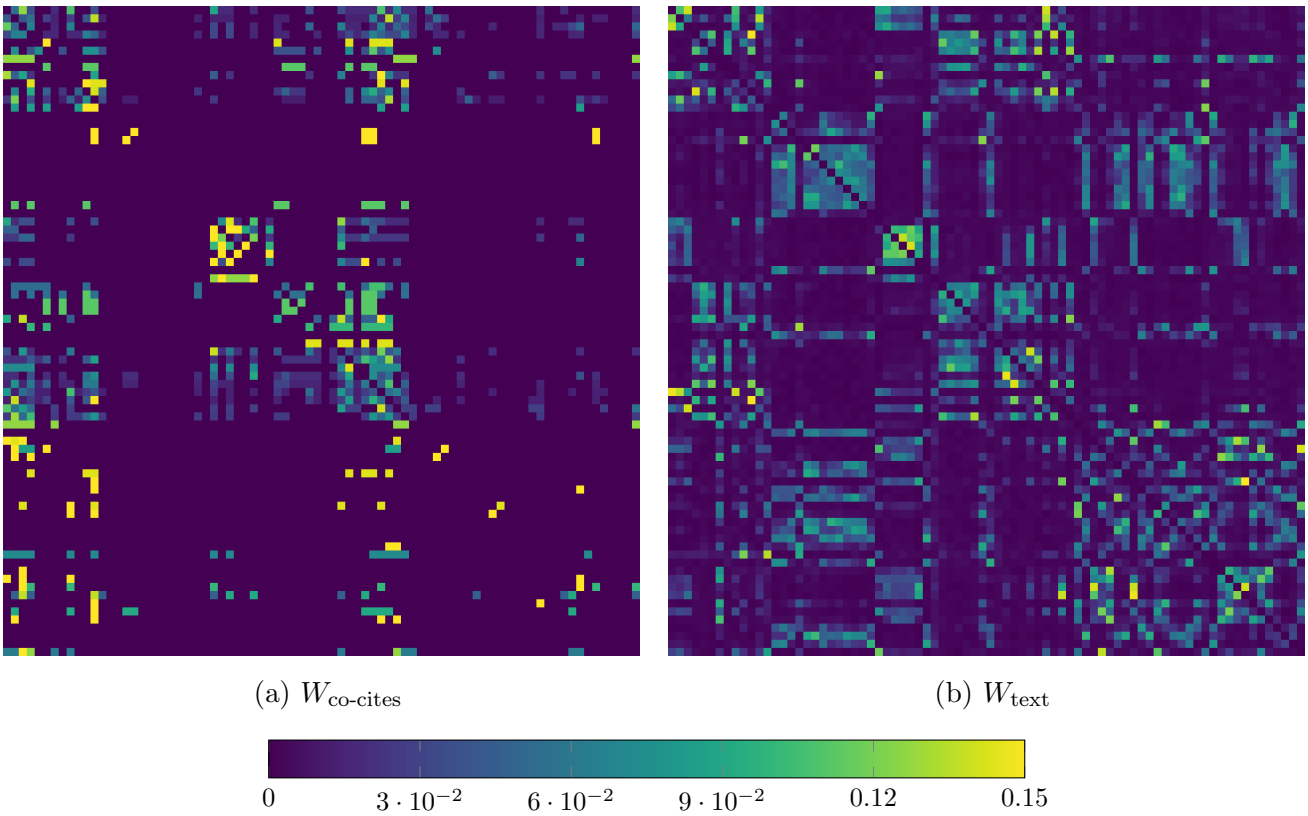



Figure 11: **Heatmap of Two Different Edge Weight Matrices:** Above two subfigures visualizes two different *edge weight matrices* $W_{\text{co-cites}}$ and $W_{\text{text}}$. One can check that $W_{\text{co-cites}}$ is sparser than $W_{\text{text}}$.

### 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 main principles of WTO.

This failure is mainly due to the insufficient information in the co-citation matrix. Members tend to cite the articles of the WTO agreements based on the complex characteristics of the trade policy that led to the dispute, however, co-citation pattern omits this contextual information. To visualize the importance of using textual information, I prepared two differeent 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 13. $W_{\text{text}}$ is the one that is fitted using the textual information and the way how it's fitted will be introduced at the following bodies of this section. Two Heatmaps visualized in Figure 11 shows how sparse the $W_{\text{co-cites}}$ is compared to the $W_{\text{text}}$. This sparisty indirectly refers to the insufficient information to qualitatively map the regulatory system of WTO DSB. In contrast with it, if we fit the *edge weight matrix* $W$ using the textual information, we get more dense, informative matrix as visualized in Figure11(b).

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** | $\cdots$ |
|---|---|---|---|---|---|
| **I** | **0** | **3** | **7** | **2** | |
| **I:1** | **3** | **0** | **3** | **4** | |
| **II** | **7** | **3** | **0** | **4** | |
| **II:1** | **2** | **4** | **4** | **0** | |
| $\vdots$ | | | | | |

(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 co-citation matrix is *symmetric*, $m_{ij} = m_{ji}\ \forall i, j \in V$.

Upon this observation, this paper justifies the use 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 the 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).

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

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

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

| | **I** | **I:1** | **II** | **II:1** | $\cdots$ | |
|---|---|---|---|---|---|---|
| **I** | **0** | **0.053** | **0.125** | **0.035** | | $\sum_{j\in V} n_{ij} = 1$ |
| **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.

### 3.1.2 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 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 strategic citation.

## 3.2 Design of Deep Neural Network

Upon the justification of using deep neural network with two main reasons explained in Section 3.1, I introduce a design of deep neural network that is expected to efficiently encode members' legal citation pattern.

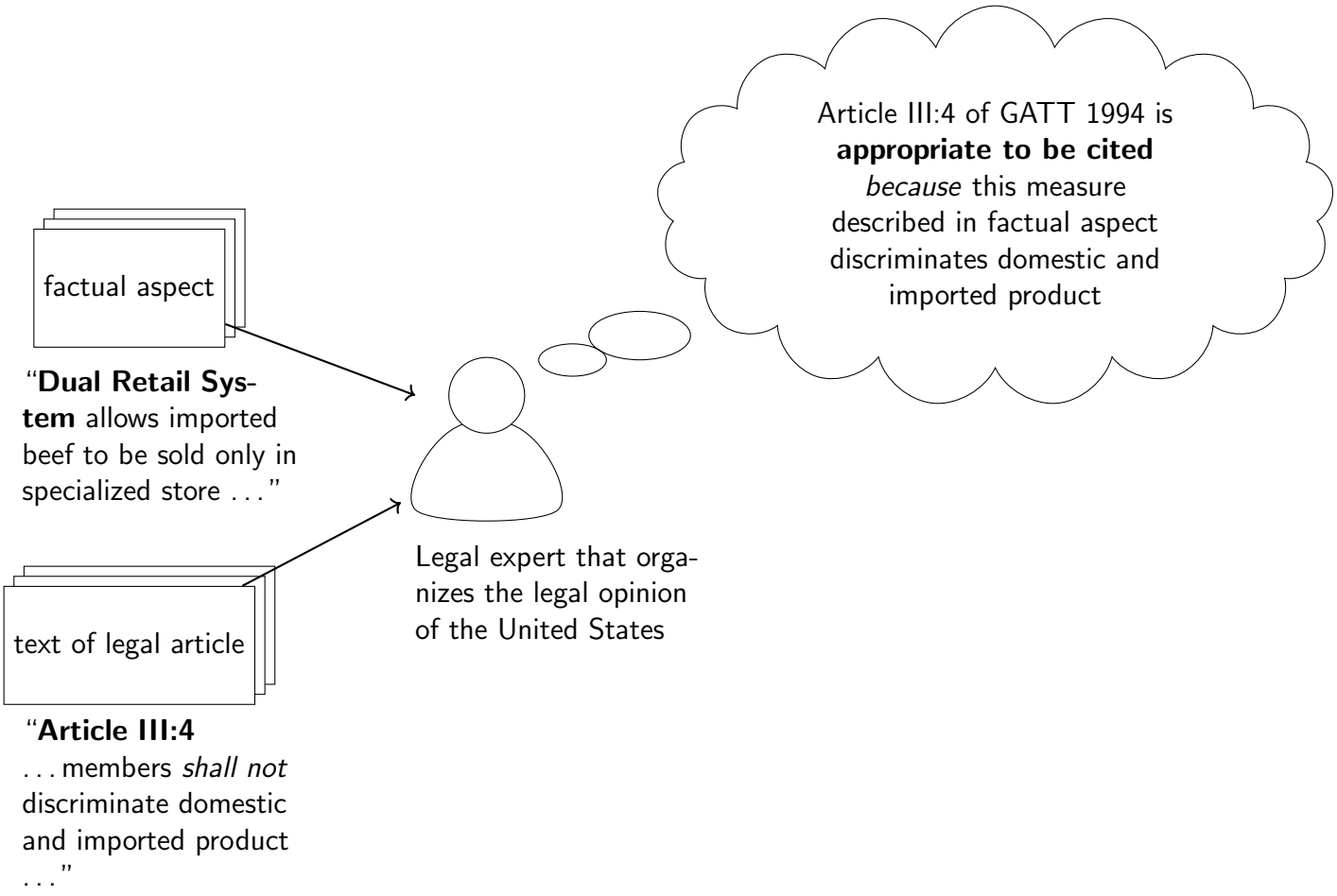



Figure 14: **Visualization of How Member Cites in WTO DSB (Citable Case):** With two different contexts, factual aspect and text of legal articles, member judges whether the given legal article is appropriate to be cited or not.

### 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 expert of rules of WTO determines 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

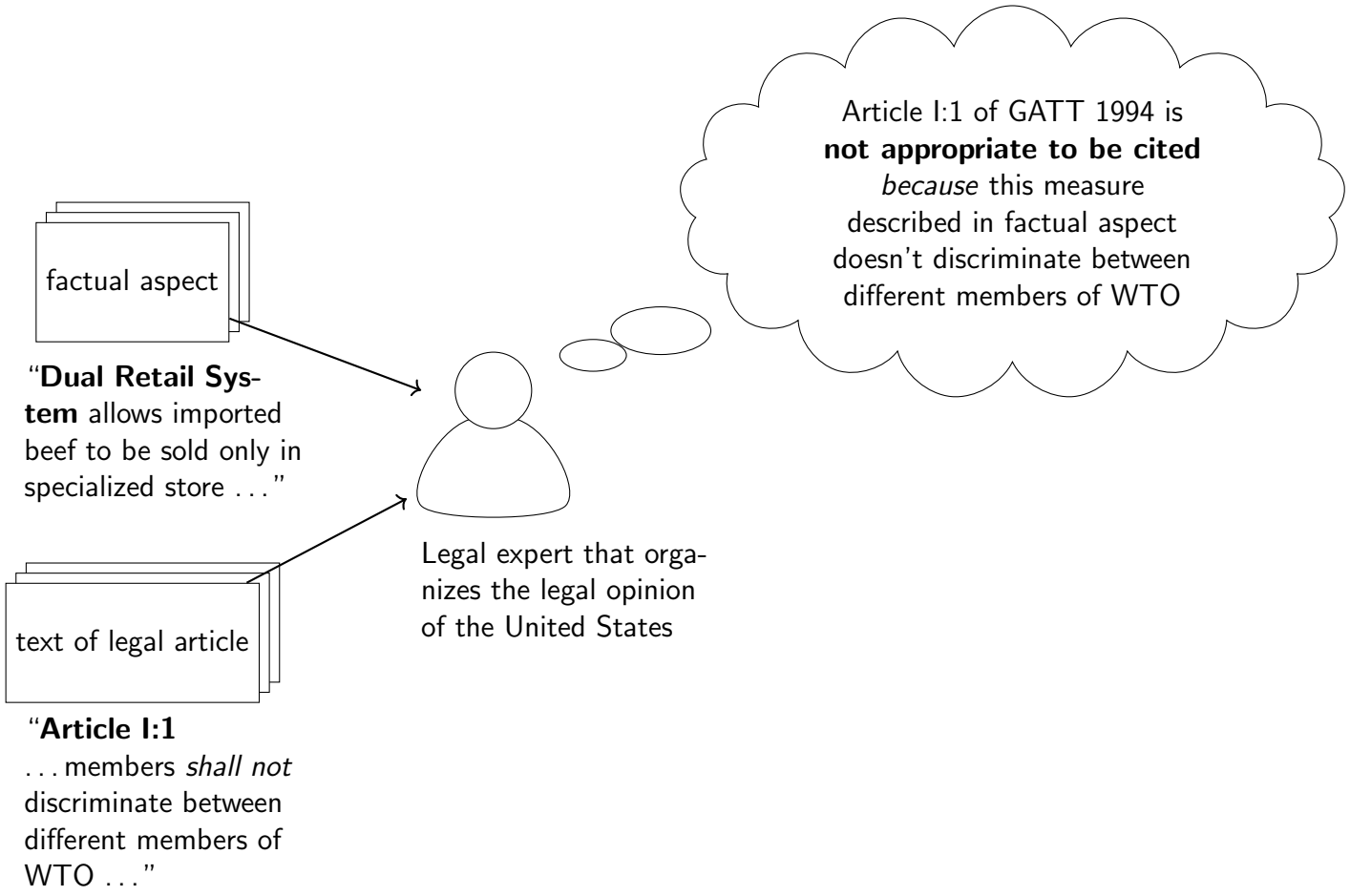



Figure 15: **Visualization of How Member Cites in WTO DSB (Non-Citable Case):** With two different contexts, factual aspect and text of legal articles, member judges whether the given legal article is appropriate to be cited or not.

two distinct retail systems for imported and domestic beef, i.e., there exist stores specialized for imported beef and they can sell only imported beef and cannot sell domestic 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* has a chance to discriminate 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 *like product*. However, U.S. didn't cite the Article I:1 of GATT 1994 that prohibits the discrimination between members of WTO agreements because *Dual-retail system* didn't discriminate the United States with other countries that exports beef to South Korea, such as Argentina, Austrailia, etc.

We can understand that there exists a shared understanding over regulatory system of WTO DSB among legal experts of the rules of WTO. They follow up new cases and study jurisprudence stated in the *Panel* or *Appellate Body* reports. Then they organize a legal argument by citing

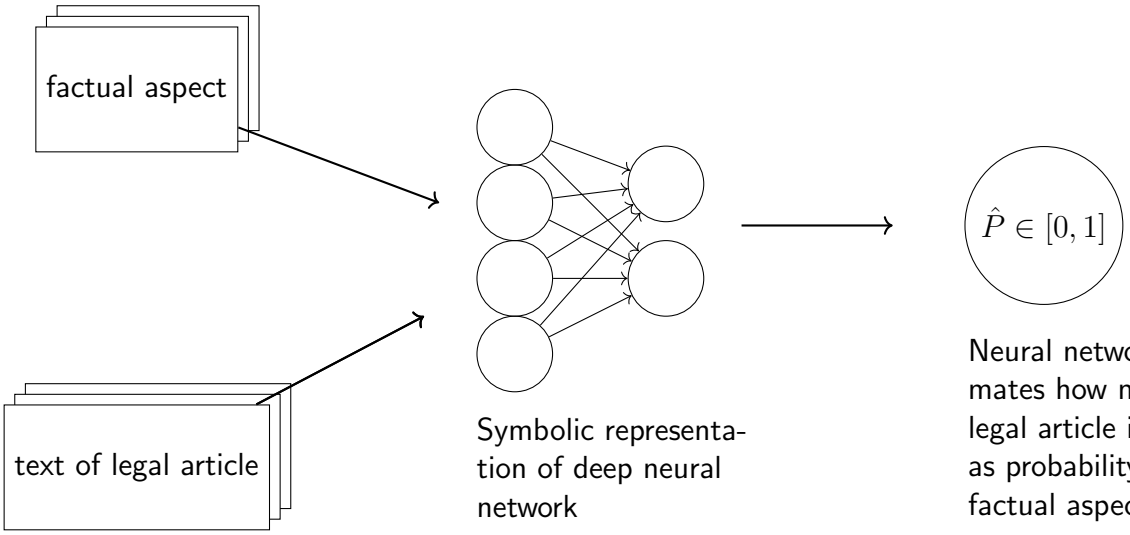



Figure 16: **Design of Training Framework of Deep Neural Network:** I desgined a training framework of deep neural network by analogy with how member cites in WTO DSB as visualized in Figure 14 and Figure 15.

certain article(s) upon this shared understanding for given possible inconsistent measure held 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. The neural network is designed to estimate the citable probability for given pair of a factual aspect and text of a legal article. By iteratively trains the neural network with data explained in Section 2, I expect the neural network can learn a shared understanding over reglatory system of WTO DSB as closely as to that of legal experts. The detailed structure of neural network and training technique will be explained in the later subsections.

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

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

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

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 $\theta$

$$f_\theta : T \times A \to [0, 1]$$

Figure 17: **Formal Definition of Input/Ouput of Deep Neural Network**: $T$ and $A$ represents a set of "documents" that are factual aspect and text of legal article respectively. The term "document" refer to a *tuple* of English words or special characters like $(e_1, e_2, \ldots, e_{n_{\max}})$. Then a pair of document, tuple of a 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.

### 3.2.2 Structure of Deep Neural Network

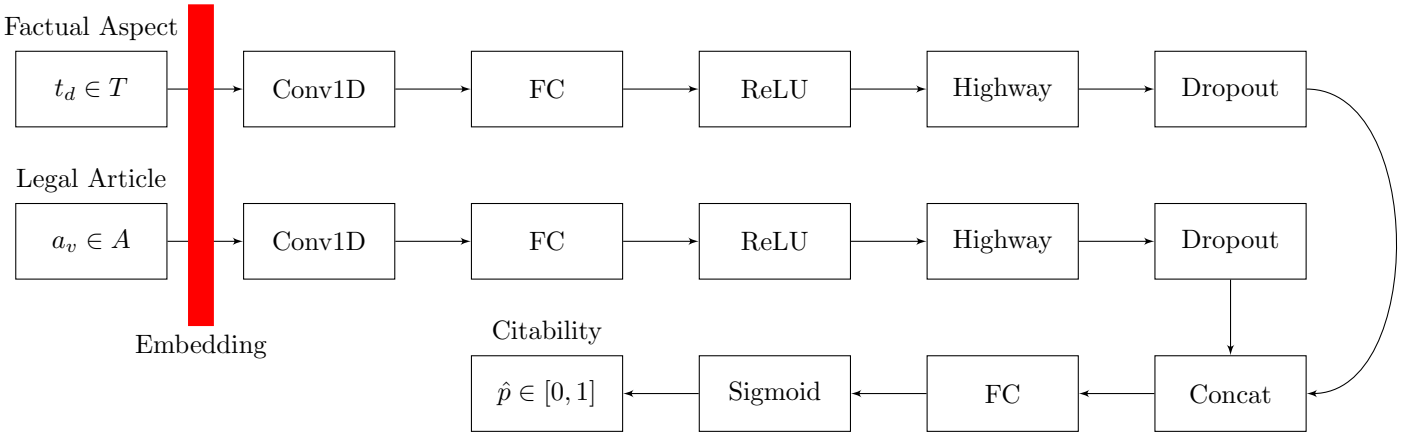



Figure 18: **Layers of Deep Neural Network:** The term "Layer" refers to an mathematical operation. How layers are stacked in which order determines the "structure" of deep neural network. This figure illustrates the structure of the deep neural network tha this paper used. The figure follows the notations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

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 deterimine 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 network is known as good at learning how to integrate local and global features to perform classification task (Lawrence et al., 1997; Nahid and Kong, 2017). Second, Kim (2014) has implemented this convolutional neural network into the text domain (TextCNN) and has shown its high performance on classification.

I explained technical Details of each layer illustrated in Figure 18 in Appendix.

### 3.2.3 Train of Deep Neural Network

I have total 11,440 number of data instances that is calculated by $|D| * |V|$ 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 result 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 selected 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 can perform over the data that it has never seen before.

The term "training" refers to fitting a set of parameters $\theta$, that constitutes mathematical operations defined in layers illustrated in Figure 18. The procedure to fit those set of parameters is as described in Algorithm 1.

**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 $\theta^*$

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

I fitted the nerual network $f_\theta$ using *weighted cross entropy loss* as explained in Algorithm 1. I used *cross entropy loss* because it measures how much the probability distribution $\hat{Y}$ projected by the trained model $f_\theta$ deviates from the true distribution $Y$ which symbolically 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.

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 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 $y \leftarrow 1$ in Algorithm 1.

More techinical details regarding epochs and learning rates will be covered in Appendix. and the training result (model performance) will be explained in Appendix.

## 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 of 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 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 19.

Upon the assumption that the trained neural network $f_{\theta^*}$ effectively encodes a shared understanding among legal experts of WTO agreements, our task is to find a set of directed edge weights $W^* = \{w^*_{ij} \mid w^*_{ij} = w^*(v_i, v_j)\ s.t.\ w^*(v_j, v_j) = 0 \text{ and } \sum_{v_i \in V} w^*(v_i, v_j) = 1\ \forall v_j\}$ 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 praobability $P^*(v_j|v_i)$ how probably a source node $v_i$ clarifies the meaning of the target node $v_j$ compared to other source nodes as closely as to a shared understanding of legal experts.

To perform this task, this paper adopts a machine learning technique called *Random Forest* that can rank input features, $\{f_{\theta^*}(t_d, a_{v_i}) \mid d \in D \text{ and } v_i \in V \setminus \{v_j\}\}$, in terms of relative importance to explain the variance of output variables, $\{f_{\theta^*}(t_d, a_{v_j}) \mid d \in D\}$ for 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 $W^*$

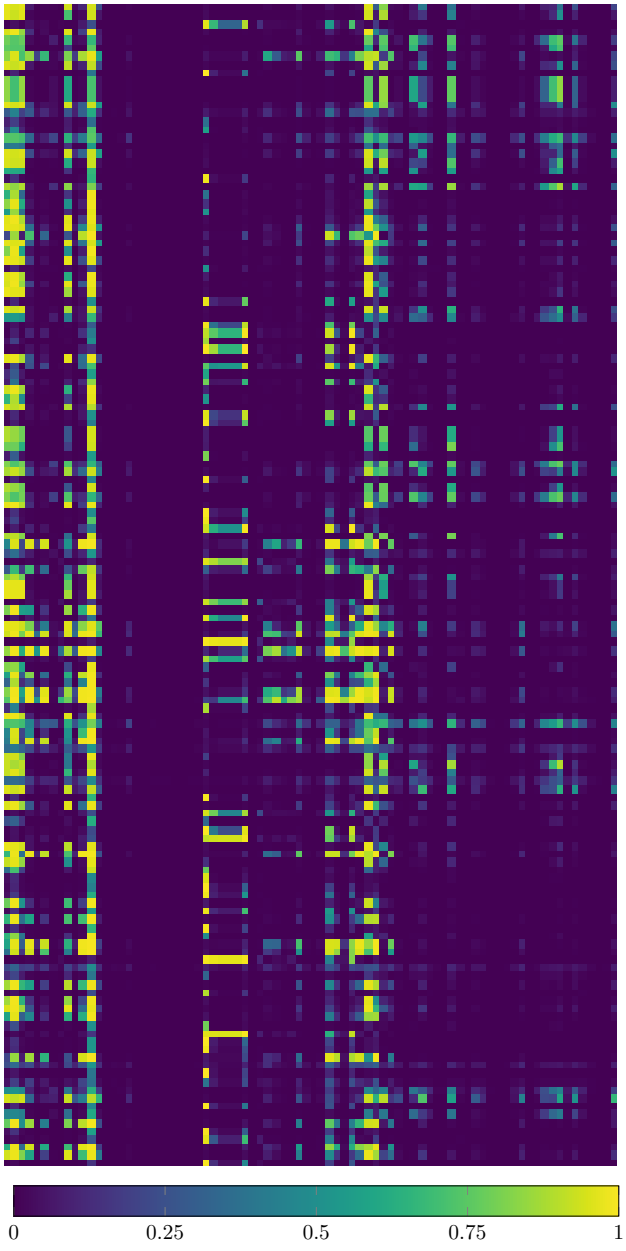
The prediction matrix $P$ illustrated in Figure 26(b) includes co-citation pattern between articles of WTO agreements. For example, we can see several highlighted bands in Figure 26(b) and it tells us there exists some co-citation patterns inside the matrix.

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 in data into two with a inequality criterion of an input feature. This split reduces the varaince of the output variables gradually as tree grows. In our case, for given target article $v_j \in V$, rest of the articles

| | 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 19: **Illustration of Prediction Matrix:** I collected all the predictions from the trained nerual network $f_{\theta^*}$ and constructed prediction matrix $P$

$v_i \in V \setminus \{v_j\}$ becomes the input features. For example, we have 143 observations in Figure 26(b) where the number equals to 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 feature greater (or lesser) than certain value. By doing so, varaince of output variables in 143 observations keep reduced follwing 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 noramlize 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 distinguishes RF from single decision tree. First, as noted earlier, RF requires to averages out multiple decision trees. I ensembled 1,000 decision trees and averages 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 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. Final output of $W^*$ is visualized as an 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:** $W^* = (w^*_{ij}) \in [0,1]^{|V|\times|V|}$

Let number of features $n = |V|$ ,
    number of obseravations $o = |D|$ and
    number of trees to ensemble $m \in \mathbb{N}$
**for** ( $v_j \in V$ ) {
    X $\leftarrow \{x_d \mid x_d = (p_{dv_1}, p_{dv_2}, \ldots, p_{dv_n})\ s.t.\ v \in V \setminus \{v_j\}$ and $d \in D\}$
    Y $\leftarrow \{y_d \mid y_d = p_{dv_j} / \sigma(p_{v_j})\ s.t.\ d \in D$ and $\sigma(p_{v_j})$ is a standard deviation of $\{p_{dv_j} \mid d \in D\}\}$
    **for** ( $k \in \{1, 2, \ldots, m\}$ ) {
        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 $x_d$.
        2. Construct a decision tree $T_k : X_d \to \mathbb{R}$ where
        $T_k = \{N \mid N = (v_i, b, N_p, \hat{y})$ represents a decision node where
        $b, \hat{y} \in \mathbb{R}, v_i \in V \setminus \{v_j\}$ and $(v_i, b)$ splits $S_N \subset S$ that reached the node $N$
        into $S_{N_{true}}$ and $S_{N_{false}}$ with a criterion $p_{dv_i} \geq b$ with a parent node $N_p \in T_k$
        if $N$ is not a root node. Define $N_p = \emptyset$ if N is a root node. $\hat{y}$ represents
        the node's estimate for given input $x_d$ and $\hat{y} = \frac{1}{|S_N|} \sum_{(x_d,y_d)\in S_N} y_d$.
        $v_i$ and $b$ is $\emptyset$ if $N$ has no child nodes. $(v_i, b)$ at each node N is determined
        among a random sampled $\sqrt{|V| - 1}$ number of $v_i$ from $V$ that minimizes MSE $\frac{1}{|S_{N_p}|} \sum_{(x_d,y_d)\in S_{N_p}} (y_d - T_k(x_d))^2$ }
        3. **for** ( $N \in T_i$ ) {
            **if** $v_i$ *of* $N \neq \emptyset$ **then**
                $I^k_{v_i \to v_j}(N) \leftarrow I^k_{v_i \to v_j}(N) + (\mathrm{Var}(S_N) - \mathrm{Var}(S_{N_{true}}) - \mathrm{Var}(S_{N_{false}}))$ where $\mathrm{Var}(\cdot)$ is the variance of the output variable $y_d$ in each subset $S_N, S_{N_{true}}$ and $S_{N_{false}}$
            **end**
        }
    }
}
**then** $w^*_{ij} \leftarrow \frac{1}{m} \sum_{k\in\{1,2,\ldots m\}} I^k_{v_i \to v_j}(N)$
**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 jurisprudence of the *Panel* or *Appellate Body* of WTO DSB. Since these two judicial bodies of WTO DSB authoritatively opinionate how the regulatory system of WTO DSB systemetically 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 different articles of WTO agreements cooperatively achieves important principles of WTO and regulates specific trade issues.

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

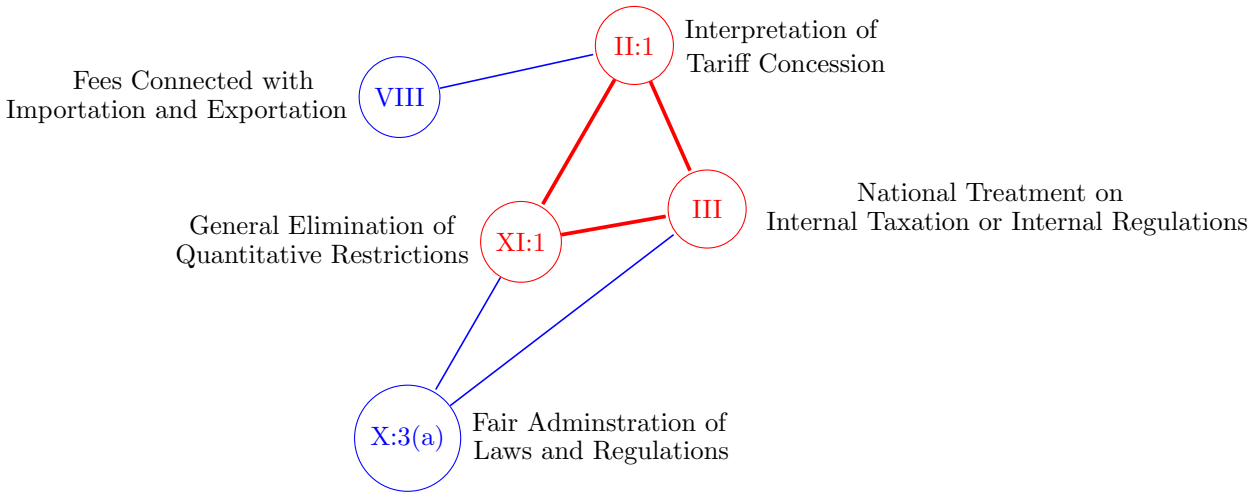



Figure 20: **Network of the Articles that Achieves *Market Access*:** This figure demonstrates a network of articles of WTO agreements that cooperatively achieves the principle of *Market Access* in the regulatory system of WTO. Tariff and Non-tariff barriers such as quantitative restriction, internal taxations and extra fees for crossing border 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 above that limit can get access to that market. Also, if a country set a high tariff rate for a imported goods to cross the border, it also prevents foreign goods from accessing to that market. Therefore, *Market Access* is the most basic but important principle of the WTO because it directly represent the primary goal of the WTO, smoothing the world trade flow.

Three different types of measures - quantitative restriction, tariff and internal taxation or regulations - are recocgnized as three major types of measures that protectionistic country prefer to use. For example, a country can discriminate imported and domestic product with internal regulation such as enforcing the specialized retail channel for the imported product as *Korea - Beef* case as introduced in section 3.2. This discrimination provide a unfair ground of competition between domestic and imported products and it can eventually exclude imported beef from the Korean beef market. These three main types of measures that are relevant to the *Market Access* is illustrated in Figure 20. Article XI:1, II:1 and III correponds to the obligations that regulates quantitative restriction, tariff and internal taxation 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 articles with large edge weights. These three articles are cohesively interconnected because they need each other to clealry determine the role of each one to achieve the *Market Access* principle. For example, *Panel* explained the relationship between Article XI:1 and III in *India - Autos* case as following[8]:

> "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 supports this *Market Access* triangle. For example, I colored edges in blue if the mean of the two directed edge weights are greather than 5% and less than 10%. The Article X:3(a) which prohibits a unfair adminstration of laws and regulations are connected to the Article XI:1 and III that are mostly deteriorates the principle of *Market Access* by unfair adminstration of laws and regulations relating to the quantitative restriction and competitive relationship between imported and domestic products. Also, in a similar notion, Article VIII get connected to the Article II:1 because often abnormal amount of fees connected with the importation are charged and it deteriorates the principle of *Market Access* as well. For example, in *Argentina - Textile* case, Argentina insisted that they didn't collect fees connected to importation inconsistently with the obligation under Article VIII

[8]Panel Report, India – Autos, para. 7.220.

because they included the fees into the tariff schedule which is regulated by the Article II. However, *Panel* refuted this logic as following[9]:

> "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

Ensuring the *Market Access* principle by enforcing the binding tariff to imported products, preventing the internal regulations and taxtions from discrimintating the domestic and imported products and eliminating the quantitative restrictions on the border is a basic ideas to smooth the world trade flow.

However, members are exceptionally 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 subsidies and to take an emergency measure in case irrevisible injury to domestic industry is expected. Each of these examples are called *Anti-dumping (AD)*, *Countervailing Duties (CVD)* and *Safeguard (SG)* respectively and recognized as three major types of *Non-Tariff Barriers* that can constitue illegal trade barriers (NTBs) since these three measures often fail to satisfy the legal requirements of the WTO agreements and evolve into trade dispute in WTO DSB.

The principle of *reciporocity* is a general notion that regulates these three major NTBs. This principle requires change of value of imports introduced 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*, confers a right to insist its nullfied or imparied benefit that is expected under the WTO agreements and require satisfactory adjustement to the member who results in such nullification or impariment. Moreover, the article confers a right to retaliate if no satisfatory adjustment being fulfilled.

[9]Panel Report, Argentina – Textiles and Apparel, paras. 6.81-6.82.

Therefore, *Article XXIII* regulates this action-reaction characteristics of *AD*, *CVD* and SG to achieve the priniple of *reciprocity* in WTO DSB as shown in Figure 21. For example, *Panel* stated that a member can resort to *Artile XXIII* to raise legal claim in case of unjustified leving of antidumping duties of other member that is not fully explained causual realtionship between the dumping and material injury to relevant industry in its report on *New Zealand - Imports of Eletrical Transformers from Finland*[10] as following:

> "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**[11], as had been done by Finland in the present case."

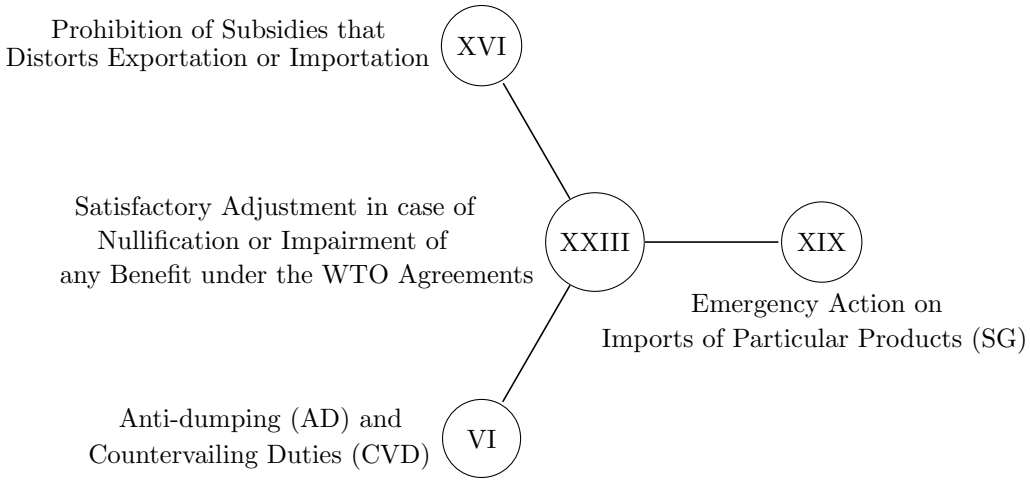



Figure 21: **Network of Articles that Achieves *Reciprocity* for Non-Tariff Barriers:** This figure demonstrates a network of articles of WTO agreements that cooperatively regulates *Non-Tariff Barriers* (NTBs) in the regulatory system of WTO. Three major NTBs such as *Anti-dumping (AD)*, *Countervailing Duties (CVD)* and *Safeguard (SG)* .

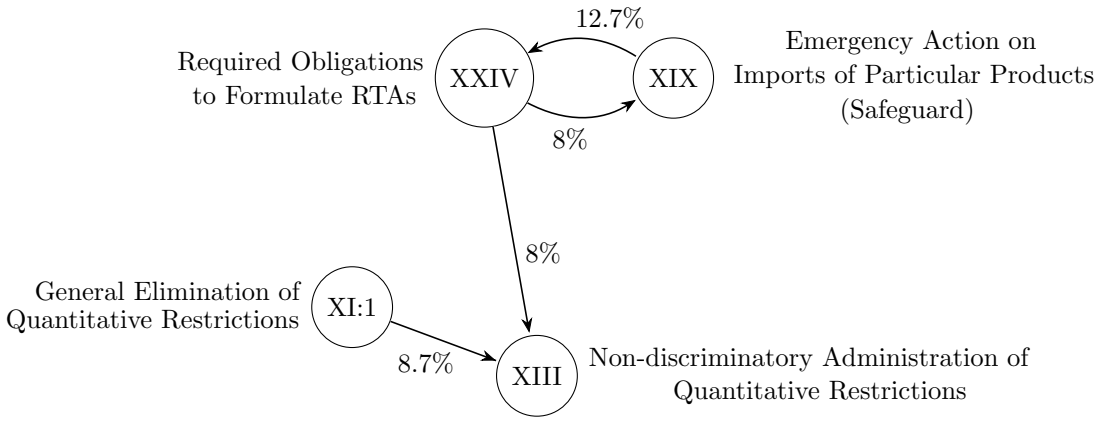



Figure 22: **Network of Articles that Achieves *Non-discrimination* for Formulating *RTAs*:** A rudimentary obligation of Article XXIV that prohibits discrimination between non-RTA and RTA members become realized in a relationship with different types of trade measures, such as *Safeguard* or *Quantitative Restriction*.

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

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* as illustrated in Figure 22. 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 the RTAs. 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. Figure 22 captures this directive relationship where the breach of the *Article XIX (Safeguard)* leads to the breach of the *Article XXIV* where it requires members not to discriminate between non-RTA and RTA members while formulate and maintain RTAs with an 12.7% weighted directive edge from *XIX* to *XXIV*. Also, we can understand the directive weighted edge from *XXIV* to *XIX* as *non-discriminatory* principle in the *Article XXIV* become extended to the *Article XIX* as members of WTO have used RTAs to justify their discriminatory application of *Safeguard measures* between RTA and non-RTA trade

[10] *WTO official document* L/5814, adopted on 18 July 1985, pp.67-68, para.4.4.
[11] It means raising a legal claim through WTO DSB

partners.[12]

In addition to the case of *Safeguard*, members sometimes try to justify their discriminatory use of quantitaitive restriction on imported products. For example, in the *Turkey – Textiles* case, Turkey defenced himself that *Article XXIV* authorizes members forming a customs union to deviate from the obligations of *Article XIII* that requires members to apply quantitaitive restrction without discrimination between members. However, *Appellate Body* explicitly refuted this logic as saying[13]:

> "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 tha *Article XXIV* does not exempt the obligation of the *Article XIII* is captured by the directive weighted edge from the *XXIV* to *XIII* in Figure 22. Moreover, *non-discrimination* obligation under *Article XIII* is interpreted within the meaning of the *Article XI* according to a number of panel reports[14] which prohibits quantitative restriction in general without any special occassions of reciprocal understanding like RTAs, this context is captured in the directive weighted edge from *XI* to *XIII* in the same Figure.

# 5 Conclusion

[12]It's worth noting that we have total 79 source articles per target article. Therefore, naive baseline of the edge weight can be calculated by simple mean, $100\%/79 \sim 01.26\%$

[13]Appellate Body Report, Turkey – Textiles, para. 65.

[14]Panel Report, *US – Shrimp*, para.7.22; Panel Report, *India – Quantitative Restrictions*, para. 5.17, Panel Report in *Colombia – Ports of Entry* para. 7.291.

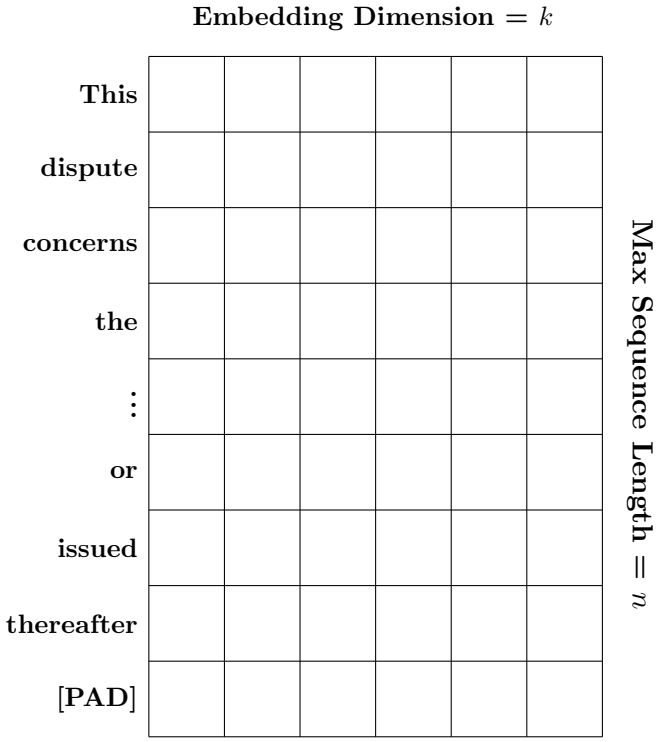



Figure 23: **Embedding Layer:** $n \times k$ matrix representation of factual aspect

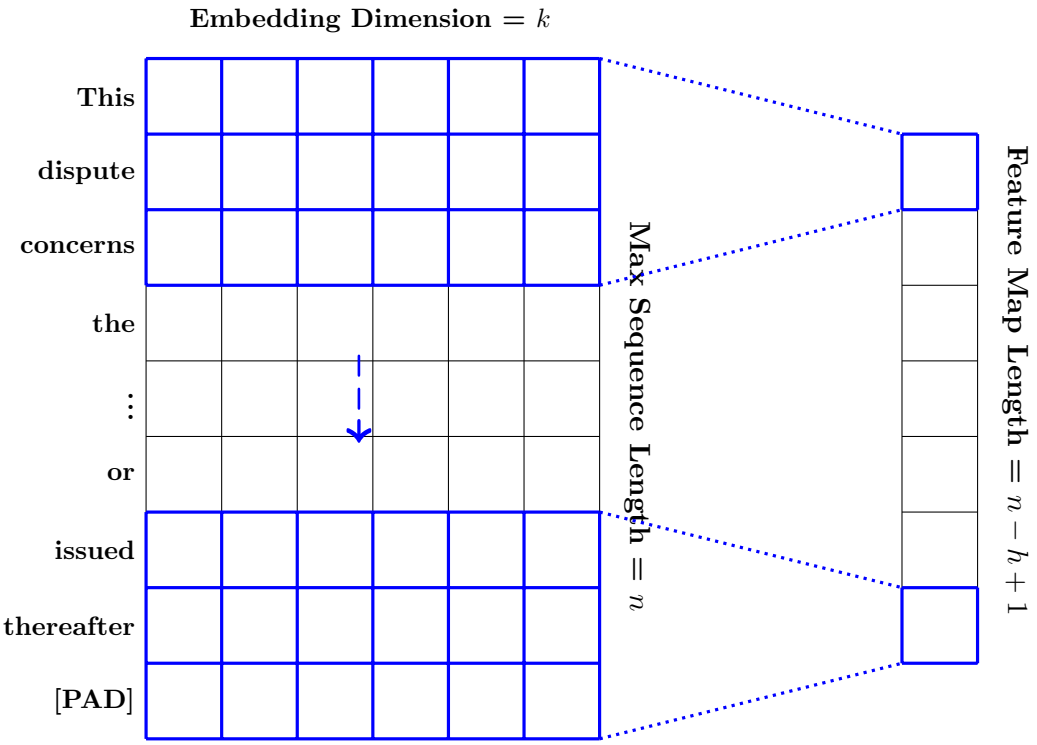



Figure 24: **Conv1D:** $h$-sized filter runs over the $n \times k$ embedding matrix and produces $(n-h+1)$ size of feature map.

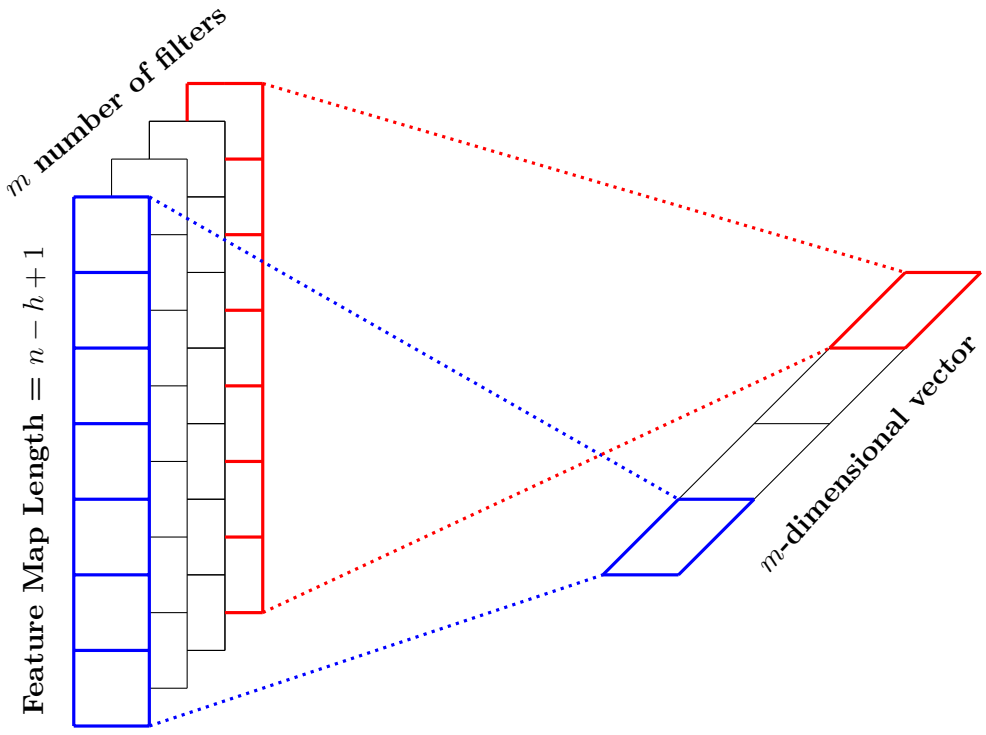



Figure 25: **MaxPool1D:** Filter out max value for all $m$ number of feature map outputs from $m$ different convolution filters. MaxPool1D produces $m$ dimensional vector as an output for collection of those filtered max values.

# 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.

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)* [15] 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, 2001.1 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 :
" 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.

[15] Panel Report, United States — Continued Dumping and Subsidy Offset Act of 2000, WTO Doc. WT/DS217/R (adopted Jan. 27, 2003).

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 finding8 and deposit into such account all the duties assessed under that Order.9 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.10 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.11 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.12

2.7 The CDSOA applies with respect to all anti-dumping and countervailing duty assessments made on or after 1 October 200013 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 Settelement* 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 |

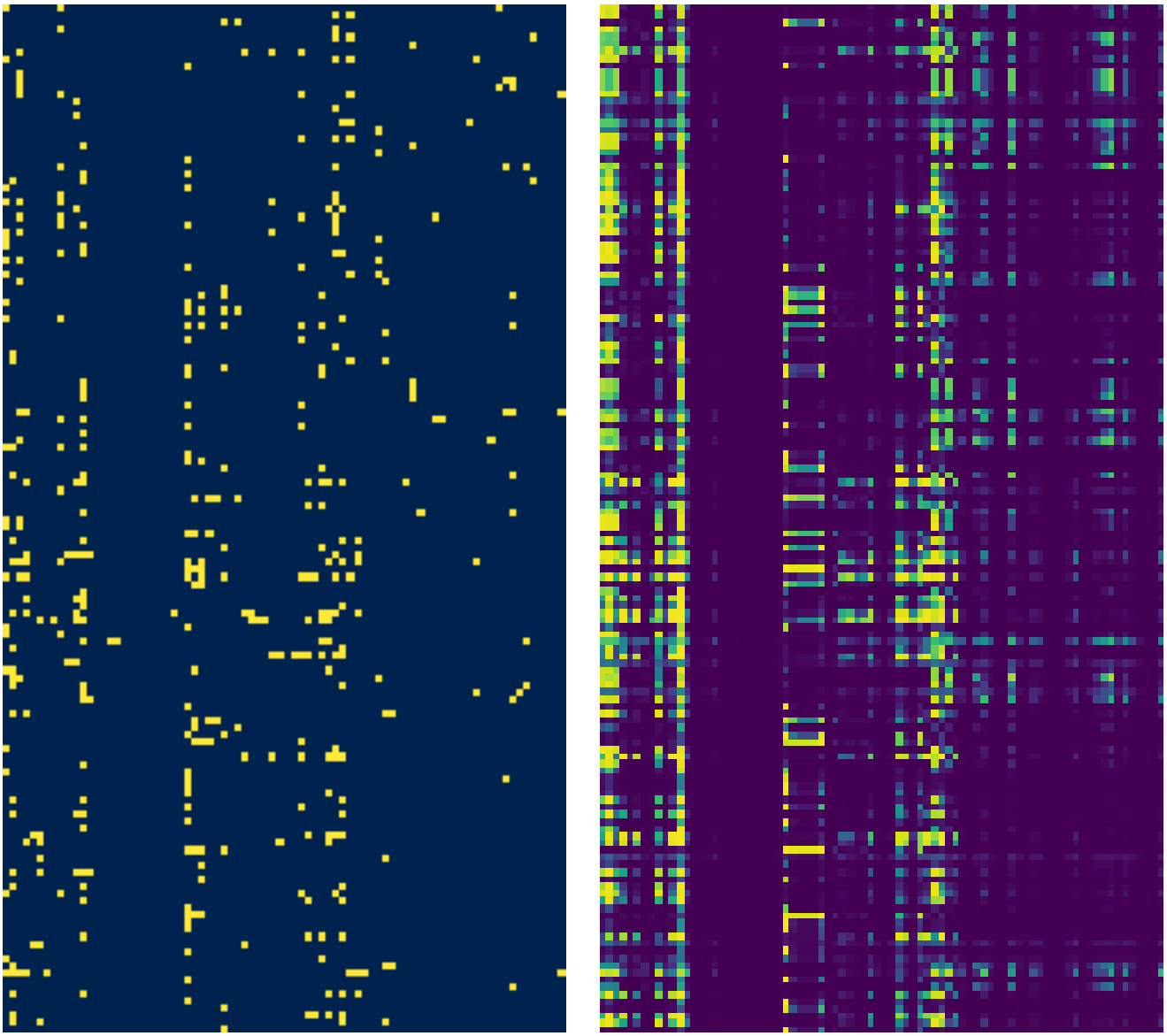
| | |
|---|---|
| **DS 219** | I, VI |
| **DS 221** | VI, VI:2, VI:3, VI:6 |
| **DS 231** | I, III, XI:1 |
| **DS 234** | VI, VI:2, VI:3, X, X:3, XXIII:1 |
| **DS 238** | XIX:1 |
| **DS 244** | VI, X |
| **DS 245** | XI |
| **DS 246** | I:1 |
| **DS 248** | I:1, XIII, XIX:1 |
| **DS 257** | VI, VI:3, 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 |
| **DS 397** | I, I:1, VI:1, X:3(a) |
| **DS 399** | I:1, II, XIX |
| **DS 400** | I:1, III:4, XI:1, XXIII:1(b) |
| **DS 406** | III:4, XX, XXIII:1(a) |
| **DS 412** | III:4, III:5, XXIII:1 |
| **DS 414** | VI |
| **DS 415** | I:1, II:1, XIX:1, XIX:2 |
| **DS 422** | VI:1, VI:2(a), VI:2(b) |
| **DS 425** | VI:1, VI:6 |

| | |
|---|---|
| **DS 427** | VI, VI:3 |
| **DS 429** | VI:1, VI:2, VI:2(a), X |
| **DS 430** | I, XI |
| **DS 431** | VII, VIII, X, X:3(a), XI, XI:1 |
| **DS 435** | III:4 |
| **DS 436** | I, VI |
| **DS 437** | VI, XXIII |
| **DS 440** | VI |
| **DS 442** | VI, X:3(a) |
| **DS 447** | I:1, III:4, XI:1 |
| **DS 449** | VI, X |
| **DS 453** | I:1, III:2, III:4, XI:1 |
| **DS 454** | VI |
| **DS 456** | III:4 |
| **DS 457** | II:1(a), II:1(b), X:1, X:3(a), XI, XI:1 |
| **DS 461** | II:1, II:1(b), VIII:1, X:3(a) |
| **DS 464** | VI, VI:1, VI:2, VI:3 |
| **DS 468** | II:1(b), XIX:1 |
| **DS 471** | VI:2 |
| **DS 472** | I:1, II:1(b), III:2, III:4, III:5 |
| **DS 473** | VI:2 |
| **DS 475** | I:1, III:4, XI:1 |
| **DS 476** | I, III, X, XI |
| **DS 477** | III:4, X:1, XI:1 |
| **DS 479** | VI |
| **DS 480** | VI, VI:1, VI:2 |
| **DS 482** | VI |
| **DS 483** | VI |
| **DS 484** | III:4, X:1, X:3, XI:1 |
| **DS 485** | II:1(a), II:1(b), VII |

| | |
|---|---|
| **DS 486** | VI |
| **DS 488** | I, X:3 |
| **DS 490** | I:1, XIX:1, XIX:2 |
| **DS 492** | I, I:1, II, II:1, II:2, XIII, XIII:1, XIII:2, XXVIII |
| **DS 493** | VI |
| **DS 495** | XXIII:1 |
| **DS 499** | I:1, III:4, X:3(a), XI:1, XIII:1 |
| **DS 504** | VI |
| **DS 505** | VI:3 |
| **DS 513** | I:1, X:1, X:2, X:3(a), XI:1 |
| **DS 518** | I:1, II:1(b), XI:1, XIX:1 |
| **DS 523** | VI:3 |

## A.3 Technical Details



(a) Co-citation (b) Prediction

Figure 26: **Co-citation & Prediction**