logistics

MDPI

Article

Social Media and Logistics: Uncovering Challenges and Solutions Through YouTube Data

Hisatoshi Naganawa [1] and Enna Hirata [2,*]

[1] Faculty of Ocean Science and Technology, Kobe University, Kobe 658-0022, Japan; 2117053z@gsuite.kobe-u.ac.jp
[2] Graduate School of Maritime Sciences, Kobe University, Kobe 658-0022, Japan
* Correspondence: enna.hirata@platinum.kobe-u.ac.jp; Tel.: +81-78-431-6200

Abstract: *Background*: Logistics challenges, such as driver shortages, are a major global issue, with many countries struggling to find effective solutions. YouTube, as a social networking platform, has a growing user base and is increasingly used not only for entertainment but also for social interaction, such as commenting, searching, and browsing, and it can thus potentially be used as an indicator of the topic under discussion. *Methods*: This study collects YouTube data containing keywords related to logistics issues—particularly the 2024 problem—and applies natural language processing (NLP) techniques to explore potential solutions. It is the first study to analyze both subtitle and comment data extracted from YouTube audio as large-scale text data in the field of logistics. *Results*: The analysis identified four primary areas of concern in logistics: time management, driver welfare, technological investment, and policy transparency. Sentiment analysis revealed a predominant negative sentiment in user discussions, highlighting dissatisfaction with current logistics policies and operations. *Conclusions*: The findings provide new insights that could inform the development of effective logistics policies and improve services for logistics companies while also proposing innovative research methods using NLP.

Keywords: logistics challenges; YouTube; NLP; video data mining; 2024 problem

check for updates

Academic Editor: Booi Kam

Received: 26 February 2025
Revised: 14 April 2025
Accepted: 21 April 2025
Published: 23 April 2025

Citation: Naganawa, H.; Hirata, E. Social Media and Logistics: Uncovering Challenges and Solutions Through YouTube Data. *Logistics* **2025**, *9*, 56. https://doi.org/10.3390/logistics9020056

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The shortage of professional drivers in the logistics industry is a global issue [1]. The reasons for the shortage are mainly due to remunerations, well-being, and the amount of working hours [2]. The long working hours of workers in the logistics industry is a widely known issue, being one of the biggest obstacles to attracting and retaining a sustainable labor force of drivers. In this study, we study the case of Japan in particular, using social media data retrieved from YouTube, in order to identify effective solutions to resolve logistics challenges, particularly on the issue of driver shortage.

While logistics challenges in Japan are multifaceted, this study specifically focuses on the issue of driver shortages, one of the most pressing factors in the "2024 problem", which is commonly used to refer to logistics challenges in Japan. The 2024 problem was triggered by the April 2024 driver overtime cap, which was already a serious problem before being enforced. In 2018, the Japanese government enacted the work style reform law to restrict working hours and bridge the disparity between regular and irregular workers. As a result, the labor standards were revised, limiting overtime work to 45 h per month and 360 h per year in principle. However, in the logistics and transportation industry, due to the nature of its operations, as a special case, overtime working hours have been much longer than the standard in the past. To regulate long working hours, a law stipulating the upper limit of

960 h per year for truck drivers was enacted as of April 2024. This enforcement worsened the shortage of truck drivers and stagnation of logistics functions, which is the so-called "2024 problem".

With the spread of the internet, social issues such as "the 2024 problem" have gained substantial attention on social media. In addition to facilitating new forms of human connection, social media has also contributed to the rise of intermediary media that bridge personal information sharing with mass communication, reshaping the way information is disseminated. Video online platforms (VOPs) stand out from traditional media, which primarily rely on text and images, due to their strong visual appeal and vast amount of content. The growth of on-demand content distribution and specialized internet channels has further expanded media diversity, creating multiple avenues for information consumption. These evolutions have played a critical role in spreading information, especially with respect to the 2024 problem.

In traditional content services, users are positioned as consumers who receive content delivered unilaterally. However, with the deepening and diversification of social media functions implemented in VOP, users have come to function not only as bearers of interactive communication (receivers and senders) but also as co-creative participants in content services. As a result, the speed of information diffusion has dramatically increased, and the social influence of certain topics and trends has been strengthened to spread instantly.

While there are a variety of VOPs, such as NicoNico video, TikTok, and bilibili, YouTube has popularized a new video viewing and sharing behavior worldwide, and in Japan, it is the VOP with the largest amount of users, and it is a social media platform that is attracting attention with a usage rate of over 87.9% among all generations [3]. YouTube is very effective in analyzing user psychology and market trends, as users' opinions and impressions are reflected as comments in real time. There is also an abundance of videos and comments related to the 2024 problem, making it a valuable data source for highlighting the challenges and issues facing the domestic logistics industry. By collecting these comments and utilizing natural language processing techniques, it is possible to extract useful information from the large volume of comment data and analyze it efficiently.

While previous research has primarily focused on analyzing YouTube video comments, this study introduces an innovative approach by converting video audio to text for further analysis.

The key contributions of this research are fourfold. First, it highlights the importance of optimizing time management, often overlooked in current strategies, to improve cost efficiency and service delivery. Second, it emphasizes driver welfare, advocating better working conditions, fair pay, and work–life balance. Third, it encourages investment in technology and workforce development to increase operational efficiency while reducing environmental impact. Fourth, it promotes transparency in pricing, cost control, and sustainability to create a more ethical and forward-looking logistics system that goes beyond traditional cost-cutting approaches.

This research explores how social networking service (SNS) data can be used to derive policy insights. Specifically, we present a novel method that integrates both YouTube comments and subtitle data, providing a unique perspective that has not been applied in logistics-related studies.

The following section reviews previous studies. Section 3 proposes the methodology of the study. Section 4 presents the data used. Section 5 presents the experiments and results. Section 6 discusses the results. Finally, Section 7 concludes the work.

2. Literature Review

2.1. Truck Driver Shortage

The problem of truck driver shortages is a global problem. According to a study by the International Road Transport Union [4], there could be a shortage of more than seven million drivers by 2028 in the 36 countries surveyed if no action is taken to recruit more drivers. This includes 4.9 million in China, 745,000 in Europe, and 200,000 in Turkey [4]. The common problems cited in each country are (1) an aging workforce, (2) long working hours, and (3) low wages. Hyland et al. [2] examined the factors contributing to driver shortages in Ireland and Europe as a whole. Mittal et al. investigated the global nature of the truck driver problem by examining separate sets of demographics and socio-economic, legal, and psychological characteristics in two different regions of the world, India and the USA [5]. As a result, they found that in order to devise new strategies to address the problem of truck driver shortages, it is important that the objectives of three main stakeholders are met: the drivers themselves, the industry as a whole, and the government as the policy maker [5]. Ju et al. found that wages were a much stronger predictor of crashes, both in terms of importance and magnitude of effect, compared to hazardous driving and driver fitness [6].

In Japan, since the economic recession of the 1990s, companies have started to reduce operating costs, including logistics-related costs, a trend that continues to this day. Understanding Japanese attitudes towards logistics issues is of great significance in raising awareness of policy and practical responses, particularly in relation to the supply-and-demand situation for logistics services [7].

Yano attributes the shortage of truck drivers to an aging population and low wages in the industry [8]. Possible solutions include ensuring fair wages and reducing working hours, which would include minimizing waiting times for loading cargo. As a matter of fact, trucks in Japan do not use detachable trailers, unlike those in many other developed countries. In addition, there is no standardization in the size of cardboard boxes. This lack of uniformity means that cargo has to be loaded and unloaded by hand, a task typically performed by forklifts in other countries but performed by truck drivers in Japan, often without compensation for the extra work.

The Japanese government has proposed several phased changes to ease the challenges facing the trucking industry, such as encouraging shippers to pay truck drivers fair wages. Prime Minister Kishida has also pledged to work on increasing driver pay [9]. However, the impact of these measures remains to be seen.

2.2. YouTube

Social media is characterized by a variety of mechanisms that encourage users to connect with each other and to visually understand their relationships with each other. According to Khan, YouTube offers various functions besides watching videos to encourage user involvement [10]. Compared to typical human relationship-focused social media platforms, YouTube has interactive features such as a “low rating” button, in addition to emphasizing video viewing [10]. Therefore, YouTube is considered to be used not only as entertainment but also as a place for social interaction, such as through commenting, searching, and browsing [11]. In addition, according to Shao, YouTube can be seen as an integration of traditional entertainment, such as television, music, and games [12]. Furthermore, YouTube can be seen as a platform for learning [13]. A study by Orús et al. found that active and experiential learning through YouTube helps to improve user learning and satisfaction [14]. Chan also revealed the usefulness of YouTube as a learning platform for digital natives [15].

The text data that can be obtained from YouTube are the subtitles of the videos and comments on the videos. In previous studies, comments on YouTube videos on various topics were collected and analyzed. Liyih et al. conducted a sentiment analysis of the war between Hamas and Israel in YouTube comments [16]. Contreras analyzed feminist discussions on YouTube in Spain [17]. Gao et al. classified whether YouTube comments indicate suicide risk [18]. On the topic of logistics and transportation, Das et al. and Li et al. collected comments on YouTube videos about self-driving cars and performed natural language processing analyses, such as sentiment analysis and polarity mapping, in order to understand consumers of self-driving cars [19,20]. Nikolaidou et al. conducted a literature review on research using social media in transport.

On the other hand, no studies analyzing subtitles in videos were found on any of the topics. Moreover, no studies have conducted an analysis directly related to logistics. From the above, to the best of our knowledge, this is the first study that applies natural language processing analysis to investigate logistics-related issues through YouTube text data.

3. Method

The research methodology applied in this study can be summarized as five phases (Figure 1): (1) preparing data using API to download data from YouTube, (2) preprocessing data using google translate and NLTK, (3) analyzing data involving sentiment and topic modeling, (4) obtaining the output, and (5) obtaining implications from the output.

<table>
<tr><th colspan="2">1. Data preparation</th><th>2. Data Preprocessing</th><th colspan="2">3. Data analyzing</th></tr>
<tr><td>a. Identify target period through interest levels</td><td>b. Select target video and download</td><td>c. Translate Japanese into English and preprocessing</td><td>d. Sentiment analysis</td><td>e. Topic modeling</td></tr>
<tr><td>Google Trends</td><td>YouTube Data API v3</td><td>Google Translate</td><td>cardiffnlp /twitter-roberta-base-sentiment-latest</td><td></td></tr>
<tr><th colspan="3">4. Output</th><th colspan="2">5. Implications</th></tr>
<tr><td>a. Word cloud</td><td>b. Sentiment</td><td>c. Topics</td><td>a. Implications from subtitle data</td><td>b. Implications from comment data</td></tr>
</table>

Figure 1. Outline of research process.

In the data preparation phase, YouTube Data API v3 [21] was used to select target videos from YouTube and collect comment data. The video audio was captured using yt-dlp [22] and then converted to text using whisper v3 large [23].

In the data preprocessing phase, the extracted Japanese text data were converted into English text using Google Translate. Google Translate was chosen for two reasons. First, it offers higher-quality translation compared to other tools such as DeepL. DeepL outperforms Google Translate in some cases [24], but sometimes, Google Translate turns out to be more effective in capturing contextual and expressive nuances in cases such as translating tourist websites [25]. Second, we were able to use the Google Translate API for free.

Thereafter, we first applied the Natural Language Toolkit (NLTK) to special marks, including emojis, symbols, and punctuations. We removed stop words, which are common words such as "the" and "a", to focus on the more meaningful and informative words in the database. Additional content-specific stop words such as URLs, hashtags, and user mentions (@username) were also removed. For better accuracy, words were further lemmatized, which is a process used to convert words in past tense or plural forms into their stem, for example, "commented" and "comments" are converted to "comment". The cleansed data were then split to the smallest unit, which is called tokenization, for analysis.

In the data analysis phase, sentiment analysis and topic modeling were performed. For sentiment analysis, the "cardiffnlp/twitter-roberta-base-sentiment-latest" [26] model was employed. Each chunk of text was split into 512 tokens and then run through the model. The output of the model is a sentiment prediction (positive, neutral, or negative). After all chunks were processed, a weighted average was calculated based on the individual sentiment scores of each chunk. This weighted average gives an overall sentiment score for the entire text. The purpose of this process is to account for variations in sentiment between different parts of the text. For example, one part of the text may be very positive, while another part may be neutral. The weighted average ensures that the overall sentiment reflects the distribution of sentiment throughout the text. The highest value out of three sentiments, positive (POS), neutral (NEU), and negative (NEG), was taken as the sentiment of the text.

For topic modeling, BERTopic [27] was deployed to extract topics for each emotion. Class-term frequency–inverse document frequency (Class-TF-IDF) is used for feature word identification, and uniform manifold approximation and projection (UMAP) is used for dimensionality reduction, while hierarchical density-based spatial clustering of applications with noise (HDBSCAN) is used for clustering.

4. Data

YouTube was chosen for two reasons. First, it is Japan's largest online video platform, with a usage rate of 87.9% across all age groups [3]. Second, it provides multiple perspectives by capturing both video creators' views and audience feedback, making it a valuable two-sided source of information.

Using YouTube Data API v3, we collected the IDs of videos related to the 2024 problem from January 2020 to July 2024. The search keywords were "the logistics 2024 problem", "The 2024 problem", and "logistics crisis" in Japanese, and duplicate videos and videos with low relevance were removed from the database. The database consists of a total of 640 videos. Subtitle data were extracted from the IDs using a downloader called yt-dlp [22] and converted into text using whisper-small [28].

The number of videos peaked in April 2023 and April 2024 (Figure 2). The classification of contributors is dominated by individuals, followed by old media such as TV stations. However, the number of views was found to be the highest for old media (Figure 3). Comments were collected from video IDs using YouTube Data API v3. The database of video comments consists of a total of 96,749 comments. Old media had the highest number of comments, followed by individuals. The number of comments per video is 218 for old media, 399 for net media, 75 for individuals, and 14 for organizations, indicating a relatively high amount of net media.

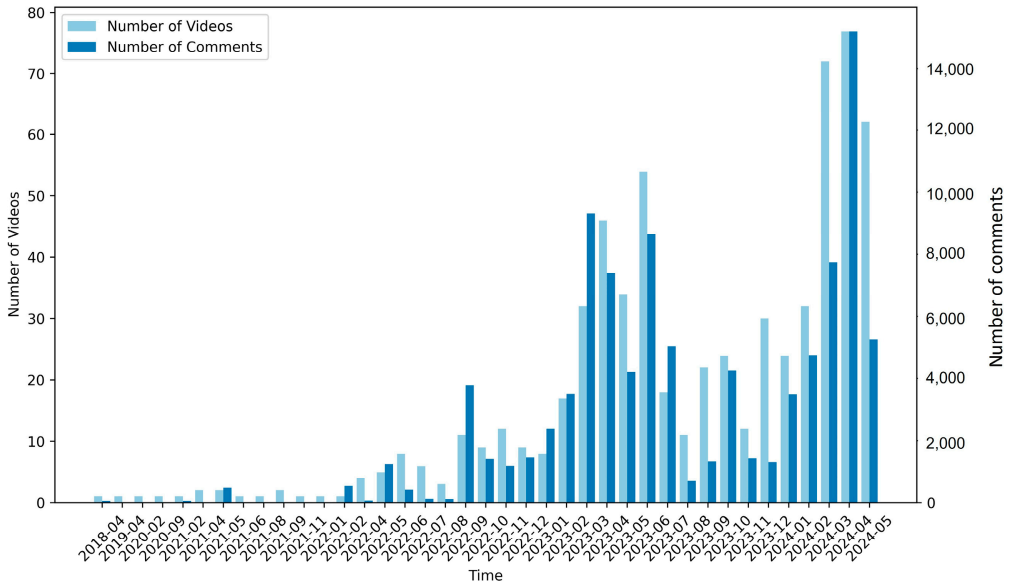


Figure 2. Time trends in videos and comments.

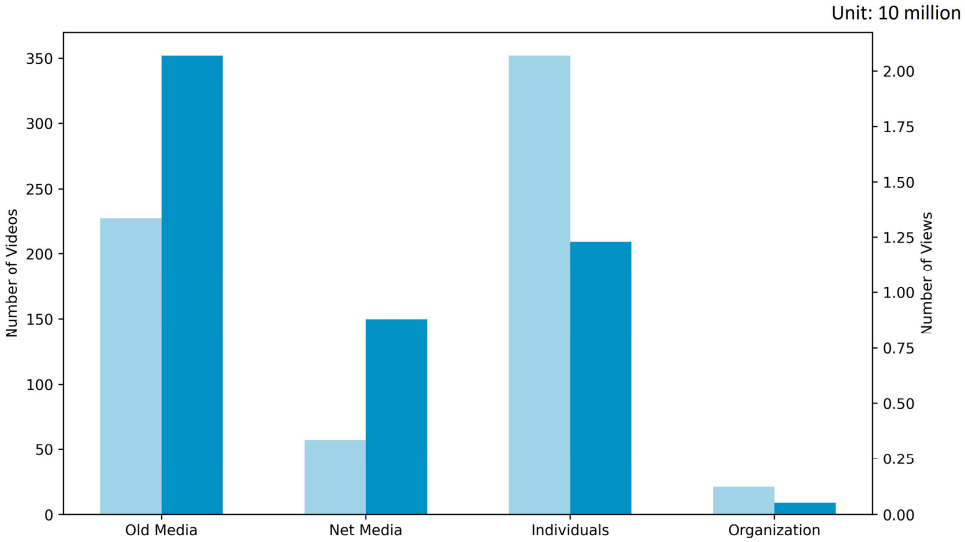
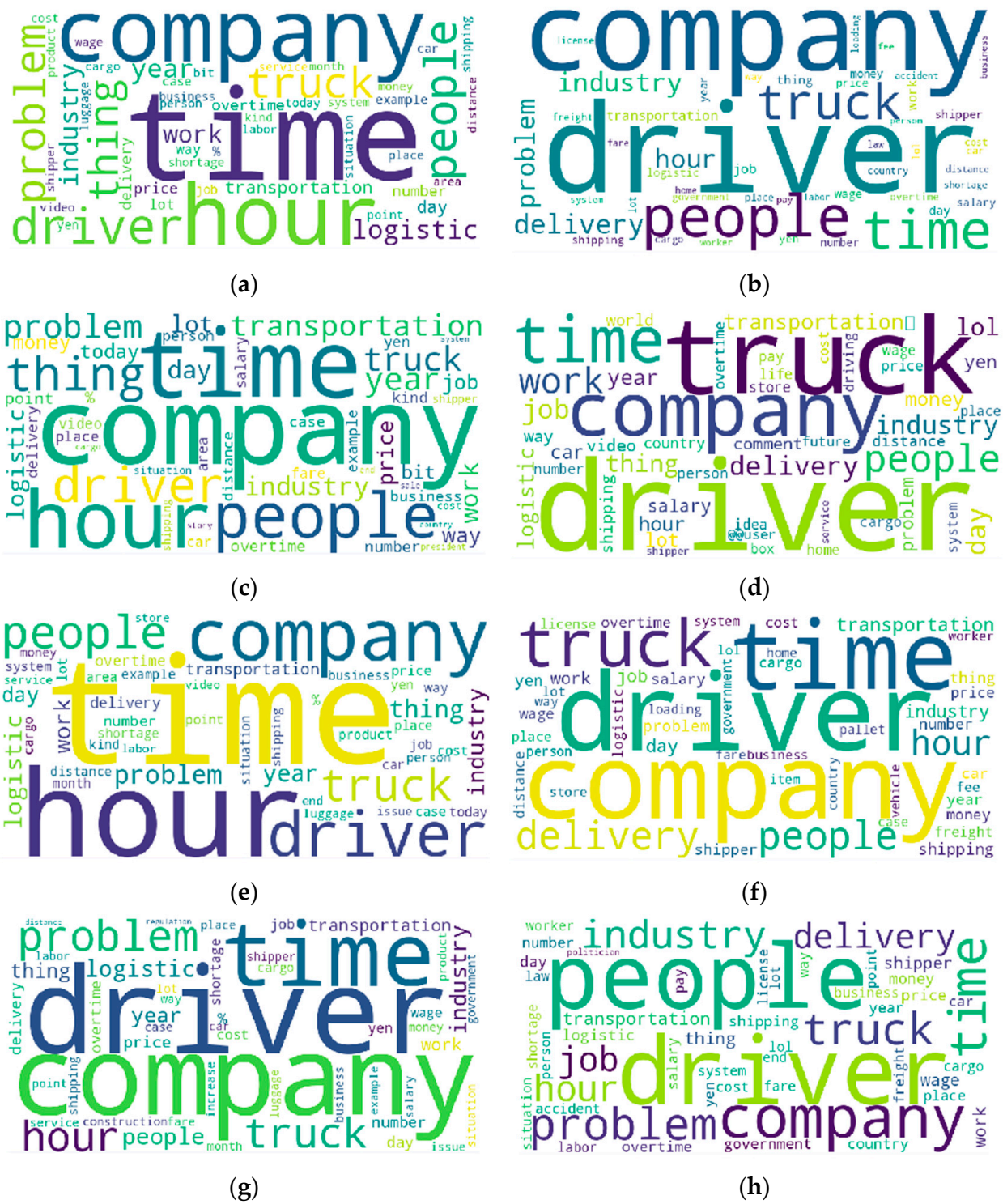


Figure 3. Video category. Old media refers to news media such as newspapers, magazines, television, and radio, such as NHK; net media refers to media that disseminate and contain information shared via the internet, such as Yahoo; individuals include users who are individuals; organizations include users who are an organization and are neither old media nor net media.

5. Results

5.1. Word Cloud

A word cloud was created from the target text data. Separate word clouds were created for subtitles and comments, and for each emotion as well. Words such as "company", "time", and "person" are commonly used in subtitles and comments. Other words such as "logistics", "truck", and "driver" are found in subtitles, and "manner", "occupation", and "money" are found in comments. There are no significant differences in words by emotion (Figure 4a–h).



(**a**) (**b**)

(**c**) (**d**)

(**e**) (**f**)

(**g**) (**h**)

Figure 4. (**a**) subtitles; (**b**) comments; (**c**) subtitles (POS); (**d**) comments (POS); (**e**) subtitles (NEU); (**f**) comments (NEU); (**g**) subtitles (NEG); (**h**) comments (NEG). Bigger and bolder words indicate higher frequency.

5.2. Sentiment Analysis

Subtitles and comments were analyzed for sentiment using cardiffnlp/twitter-roberta-base-sentiment-latest [29]. Subtitles were given an emotion score for each of the 512 tokens, and the weighted average of the scores was used as the emotion of the video. The highest score among the three, namely positive (POS), neutral (NEU), and negative (NEG), was employed.

We first compared the sentiment metric by category (Table 1). Overall, for subtitle data, about 14% of the YouTube videos related to the "2024 problem" are positive, 64% are neutral, and 22% are negative. For comment data, about 10% are positive, 39% are neutral, and 51% are negative. Subtitles often reflect the actual spoken content in the video, so they can be more neutral in tone. YouTube creators may provide factual or informative content that tends to avoid strong emotional language or bias. As a result, the sentiment distribution in subtitles might lean toward neutral or factual language. Viewer comments, however, are typically more opinionated and emotionally charged. People may express stronger reactions, either positive or negative, based on their personal feelings about the video, the topic, or the video creator. The higher percentage of negative comments (51%) and lower percentage of positive comments (10%) may suggest frustration, skepticism, or disagreement among viewers regarding the video content.

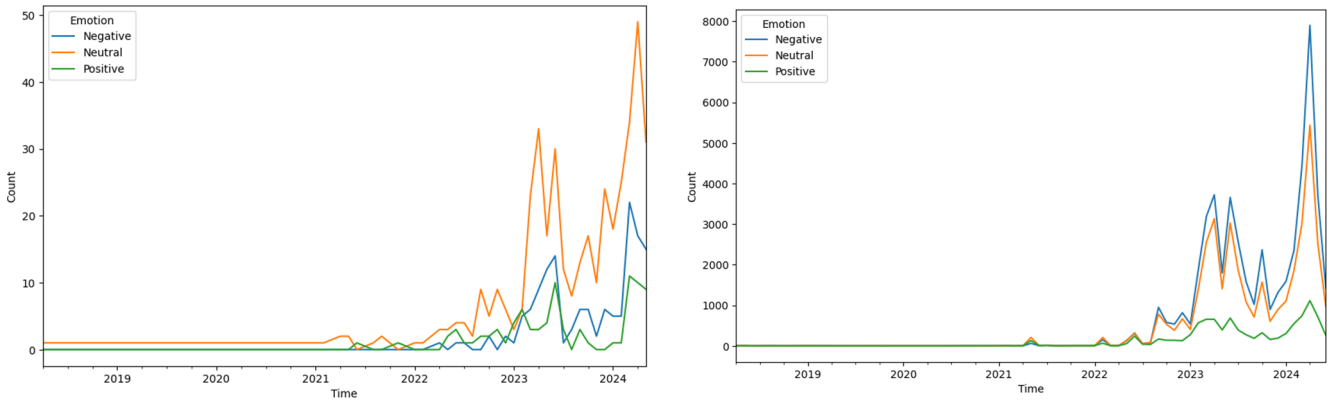
Table 1. Sentiment metrics by category.

Category	Subtitle				Comments			
	POS	**NEU**	**NEG**	**Total**	**POS**	**NEU**	**NEG**	**Total**
Old media	6%	77%	16%	100%	10%	40%	51%	100%
Net media	12%	46%	42%	100%	8%	38%	55%	100%
Individuals	18%	59%	23%	100%	13%	37%	50%	100%
Organization	14%	62%	24%	100%	15%	37%	48%	100%
All categories	14%	64%	22%	100%	10%	39%	51%	100%

In summary, the sentiment results reflect a combination of the emotional engagement of the audience, the nature of the content (informational vs. opinionated), platform dynamics (such as anonymity and audience bias), and technical factors (limitations of sentiment analysis). Subtitles' more neutral tone reflects the objective or informational nature of the videos, while the negativity of comments reflects the more passionate or critical reactions of viewers.

From a category perspective, old media are generally neutral, while most of the videos posted via online media are negative.

Figure 5a,b visualize the number of comments by sentiment for each time series. The "neutral" and "negative" emotions are the most common for subtitles and comments, respectively. There is a correlation between the percentage of each emotion in the number of videos and the number of comments.



(**a**) (**b**)

Figure 5. (**a**) Emotion trends in subtitles. (**b**) Emotion trends in comments.

Sentiment analysis focuses on determining the emotional tone behind a text—whether the sentiment is positive, negative, or neutral. It seeks to understand how people feel about a particular topic. To complement sentiment analysis, we also conducted topic modeling, which focuses on identifying the underlying themes or topics in a large corpus of text. By combining the two techniques, we can obtain a clearer picture of the discussions happening within the text.

5.3. Topic Modeling

Topic modeling was performed according to sentiment for each of the scripts and comments using BERTopic. Dimensionality reduction was performed using UMAP. Emotion categories were divided into three categories, positive (POS), neutral (NEU), and negative (NEG), taking into account the themes and social contexts they represent. The parameters are briefed in Table 2.

Table 2. Parameter value.

Parameter	Description	Value
n_neighbors	This parameter controls the balance UMAP to retain between local and global structure	30
n_components	This parameter determines the reduced	2
min_dist	This parameter controls how closely UMAP can cluster data points	0.3
metric	This parameter determines the distance to be applied for similarity calculation	cosine

5.3.1. Topics (Subtitles)

For subtitles, we were able to divide them into 3, 10, and 3 topics, respectively, by emotion (Table 3). The reason for the small number of positive and negative topics could be attributed to the fact that about 64 percent of the subtitles were neutral, which means that the numbers of positive and negative subtitles were small. The cumulative contribution ratio (CCR) was obtained to indicate the total percentage of a data set that is represented by a specific subset of data values.

Table 3. Number of topics and CCR (subtitles).

	POS	NEU	NEG
Topics	3	10	3
CCR	100%	100%	100%

The three topics belonging to positive emotions are related to time and company. The topics belonging to neutral emotions mainly refer to objective events and situations. The main topics are those related to drivers and logistics companies, topics related to products carried, topics related to disasters, and topics related to time. The topics belonging to the negative sentiment focus mainly on challenges and problems: the first and second topics relate to drivers and delivery times, while the third is related to rates (Table 4).

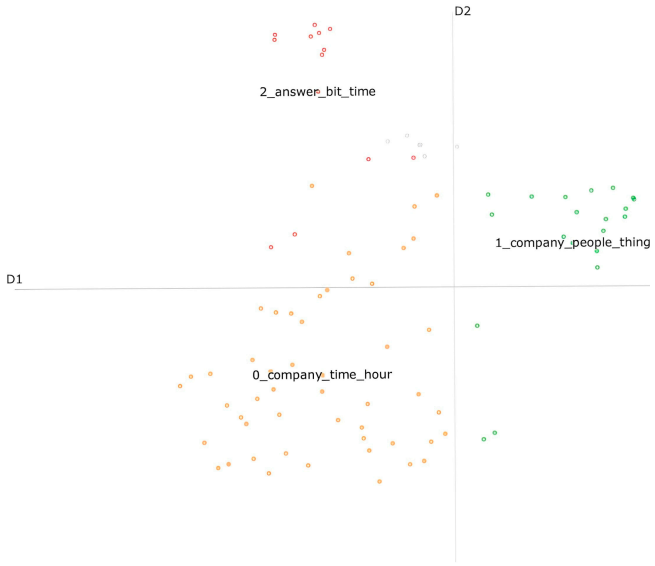
Table 4. Top topic clusters by sentiment (subtitles).

Cluster	POS	NEU	NEG
0	company, time, hour, driver	driver, truck people	driver, company, logistic, truck
1	company, people, thing, time	people, problem, time	driver, time, hour, delivery
2	answer, bit, time, question	driver, company, time	construction, company, price, cost
3		bookstore, book, time	
4		logistic, thing, company	

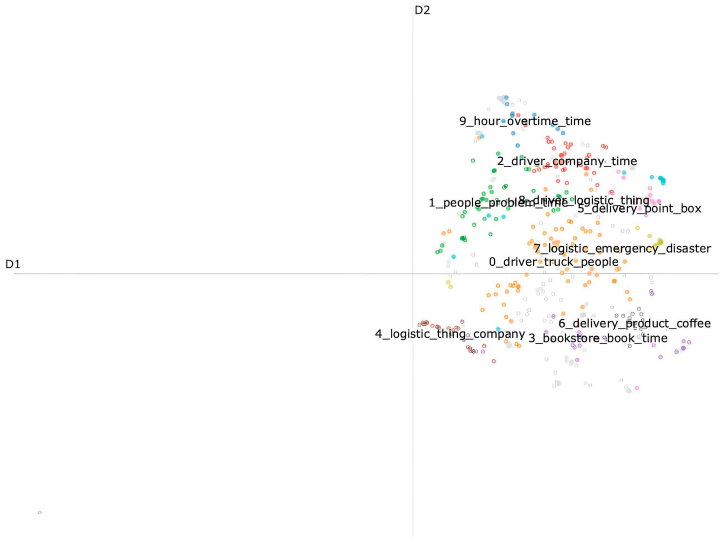
Table 4. *Cont.*

Cluster	POS	NEU	NEG
5		delivery, point, box	
6		delivery, product, coffee	
7		logistic, emergency, disaster	
8		driver, logistic, thing	
9		hour, overtime, time	

Figure 6a–c are two-dimensional coordinate diagrams depicting the relationship between each topic for each emotion.

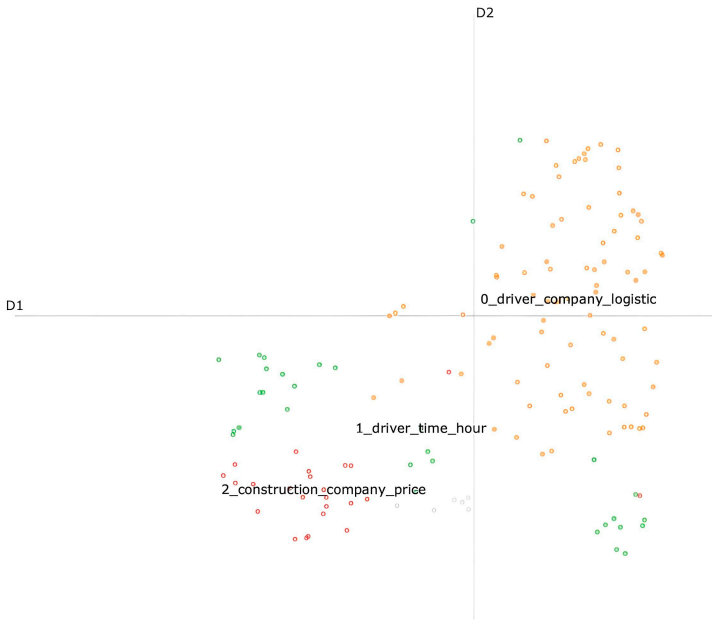


(**a**)



(**b**)

Figure 6. *Cont.*



(**c**)

Figure 6. (**a**) Two-dimensional coordinate diagram (subtitles (POS)); (**b**) two-dimensional coordinate diagram (subtitles (NEU)); (**c**) two-dimensional coordinate diagram (subtitles (NEG)).

5.3.2. Topics (Comments)

For comments, the topics could be divided into 94, 306, and 409 topics by sentiment, respectively. The CCR for the top 10 topics was as follows (Tables 5 and 6).

Table 5. Number of topics and CCR (comments).

	POS	NEU	NEG
Topics	94	306	409
CCR	55.59%	51.13%	42.19%

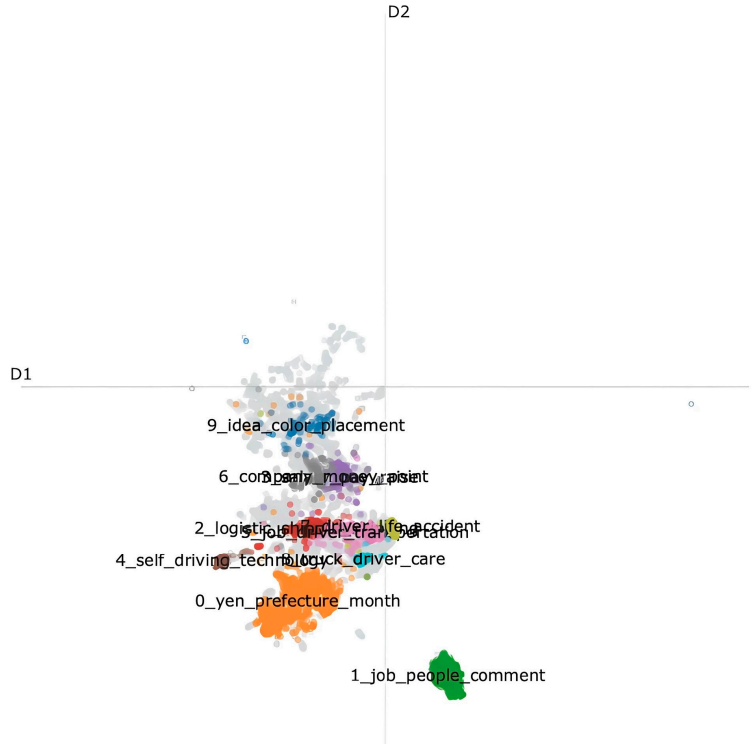
Table 6. Top topic clusters by sentiment (comments).

Cluster	POS	NEU	NEG
0	yen, prefecture, month	trucking, job, car	japan, yen. japanese, kishida
1	job, people, comment	yen, income, prefecture	delivery, specify, date, package
2	logistic, shipper, shipping	driver, truck, accident	politicians, bureaucrats, government, country
3	salary, pay, raise	driving, self, robot	transportation, deregulation, industry, fares
4	self, driving, technology	overtime, hour, month	pallets, pallet, bulk, loading
5	job, driver, transportation	train, rail, railway	logistics, stop, collapse, blood
6	company, money, point	job, nursing, environment	sorry, saskiyoshiaki, comment, icomikisan
7	driver, life, accident	wage, salary, raise	overtime, unpaid, work, pay
8	truck, driver, care	fare, black, increase	speed, limiter, 80 km, limit
9	idea, color, placement	shipping, display, seller	reporting, solve, change, infidelity

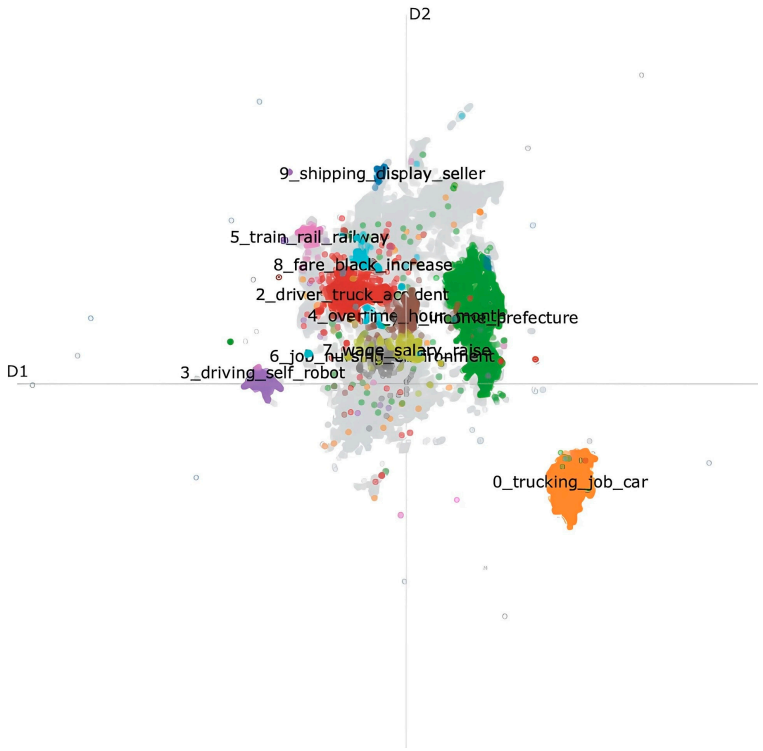
As for comments, topics belonging to positive sentiments mainly include those related to salary, such as salary and pay; those related to the driver's job, such as job and driver; and those related to accidents and disasters, such as life and accident. Neutral sentiment topics include those related to non-truck transportation modes such as train and railway, working hours such as overtime, and new technology such as robot, in addition to those similar to positive sentiment topics. Negative topics include those related to government

and public entities, such as Japan and Kishida; those related to speed, such as speed and 80 km; and those related to loading and unloading, such as pallet and bulk.

Figure 7a–c are two-dimensional coordinate diagrams depicting the relationship between each topic for each emotion for comment data.

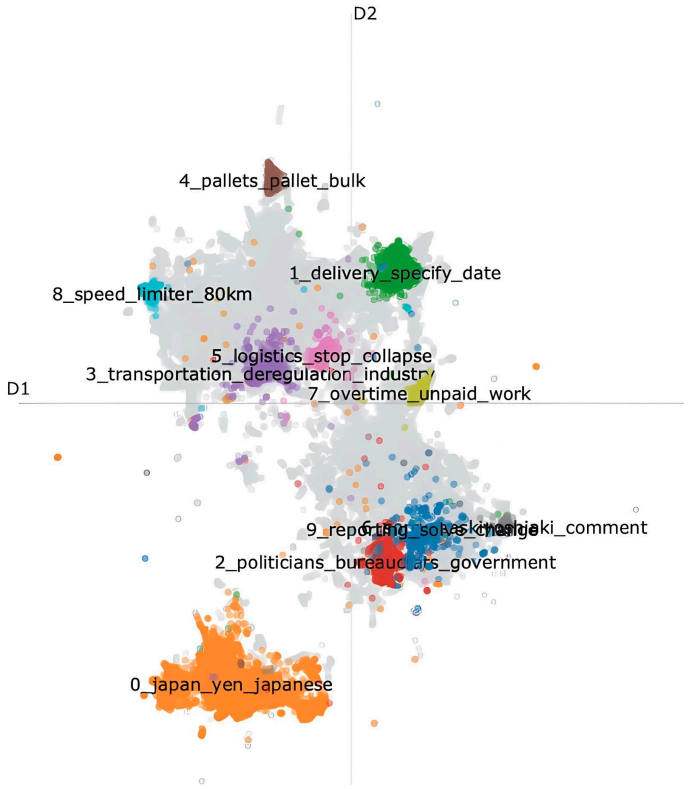


(**a**)



(**b**)

Figure 7. *Cont.*



(**c**)

Figure 7. (**a**) Two-dimensional coordinate diagram (comments (POS)); (**b**) two-dimensional coordinate diagram (comments (NEU)); (**c**) two-dimensional coordinate diagram (comments (NEG)).

6. Discussion

Since the purpose of this study is to identify solutions to logistics challenges, we focus primarily on the implications from positive and negative sentiments. In this section, the implications from subtitle and comment data are discussed separately.

6.1. Implications from Subtitle Data Analysis

6.1.1. Positive Topic Clusters

Cluster 1 indicates the importance of "company, time, hour, driver". The implications include the following: (1) Policies should focus on reducing inefficiencies related to time management within logistics operations. This can include encouraging the adoption of technologies like route optimization software, which helps drivers take the most efficient routes, reducing both time and fuel consumption. Additionally, regulations could promote work hour flexibility, allowing drivers to operate during optimal hours for traffic and weather conditions to improve efficiency and reduce delays. (2) Relating to drivers' well-being, there should be strict policies to ensure that drivers' work hours are regulated to avoid fatigue, a major factor in road accidents. Policies could mandate reasonable driving hours, enforce mandatory rest breaks, and set minimum wage standards to improve driver retention and morale. Additionally, drivers should be offered regular training programs for both safety and productivity, such as learning the latest navigation technologies or techniques to reduce wear on vehicles. (3) Company-supported flexibility is crucial. Logistics companies could be incentivized to offer more flexible working hours or staggered shifts for drivers, ensuring that the workforce is well rested and that delivery deadlines are met without overworking drivers. These policies should also foster a culture of efficiency within companies, helping logistics firms optimize schedules and fleet usage.

Cluster 2 focuses on "company, people, thing, time". Policy implications include the following: (1) The promotion of people-centric logistics. Policies should encourage logistics companies to invest in their employees' well-being through benefits, career development, and workforce training. A focus on people ensures better service delivery and reduces turnover rates in a highly competitive industry. For instance, policies could mandate training programs that teach new technologies, customer service, or sustainability practices. (2) Encouraging innovation in logistics. Policies could offer tax breaks or grants to logistics companies that invest in innovative tools, such as automation or IoT systems, which can improve the speed and accuracy of deliveries while optimizing inventory management. Incentives for adopting smart logistics tools could also help reduce costs and increase productivity. (3) Optimizing time with technology. Governments could provide incentives for logistics companies to implement time-saving technologies like real-time tracking, predictive analytics, and AI-based planning tools. These technologies help companies anticipate and solve problems in real-time, improving overall efficiency. Furthermore, policies could encourage the adoption of green technologies (such as electric vehicles) that help reduce environmental impact while saving operational costs over time. (4) Sustainability and product responsibility. Policies could mandate companies to implement eco-friendly practices (e.g., using sustainable packaging, such as modular containers, and reducing emissions), aligning with environmental goals. These policies could also promote product safety, ensuring that logistics companies adhere to rigorous quality standards that protect both consumers and the environment.

Cluster 3 contains "answer, bit, time, question". Policy implications include the following: (1) Promoting transparency. Policies should promote a culture of transparency and accountability in logistics. This includes ensuring that companies respond quickly to customer inquiries about delivery status, pricing, and delays. Real-time tracking systems should be widely adopted so that customers can track shipments and receive immediate responses, reducing dissatisfaction. (2) Continuous improvement. Policies should encourage logistics companies to implement mechanisms that allow employees and customers to raise questions and provide feedback on inefficiencies or service problems. Incentives could be offered to logistics companies that demonstrate an active commitment to addressing these "bits" of feedback and making small but meaningful improvements, whether in process optimization or customer service. (3) Innovation-driven solutions. A "questioning" mindset should be encouraged at all levels of the logistics industry, from drivers to management. Policies could include grants or recognition for logistics companies that innovate to solve common logistics challenges—whether it is better inventory management, faster response times, or improved customer experience. (4) Leveraging data for answers. Policies could mandate the use of big data and analytics to gather actionable insights, enabling logistics companies to provide accurate and timely answers to customers and improve operational efficiency. Policies could also support partnerships in sharing data across companies to reduce redundant processes and improve decision making.

Overall, in addressing logistics issues through these positive issue clusters, policies should focus on (1) optimizing time management, (2) ensuring driver welfare, (3) promoting innovation, and (4) encouraging transparency. Companies should be incentivized to invest in both technology and people to improve operational efficiency, reduce environmental impact, and increase customer satisfaction. A culture of inquiry and continuous improvement should be encouraged, with a focus on using data and feedback to solve logistical challenges.

6.1.2. Negative Topic Clusters

Cluster 1 includes "driver, company, logistic, truck". Policy implications include the following: (1) Driver safety and welfare. To address the negative effect of overworked or poorly compensated drivers, policies should enforce regulations that ensure fair wages, adequate rest periods, and reasonable working hours. Strict guidelines on driving hours (e.g., limiting daily driving hours and mandating rest breaks) can reduce driver fatigue and accidents. (2) Truck maintenance and safety. Regulations should stipulate regular inspections and maintenance of trucks to ensure safety standards are met. Policies should enforce safety features, such as mandatory use of electronic logging devices (ELDs) to track hours of service and compliance with safety regulations, which would reduce mechanical failures and accidents. (3) Logistics company accountability. Policies should hold logistics companies accountable for their role in ensuring driver safety and proper vehicle maintenance. Companies could be incentivized to improve driver support systems, provide training on safe driving practices, and ensure fair working conditions. This could be achieved through government incentives for companies that implement robust safety protocols and programs. (4) Truck emissions and sustainability. Policies should promote greener logistics operations by encouraging logistics companies to adopt electric or hybrid trucks. Regulations could incentivize the gradual replacement of older, more polluting trucks with cleaner alternatives, helping to reduce the environmental footprint of logistics operations.

Cluster 2 contains "driver, time, hour, delivery". Policy implications include the following: (1) Driver time management. Policies should regulate driving hours and ensure adequate rest to avoid overworking drivers, which can lead to accidents and inefficiency. Limiting delivery hours to prevent exhaustion, as well as mandating scheduled breaks, would improve driver safety and well-being. (2) Efficient delivery scheduling. Policies should incentivize the use of digital tools that optimize time management in delivery operations, such as real-time traffic updates, route optimization algorithms, and automated scheduling systems. These tools can help reduce delays and inefficiencies in the delivery process, benefiting both companies and customers. (3) Delivery times and realistic expectations. Policies should help set realistic expectations for delivery times, ensuring that companies do not put undue pressure on drivers to meet impractical deadlines. Addressing issues such as traffic congestion or delivery window expectations helps ensure that drivers have sufficient time to meet delivery requirements without compromising safety. (4) Pay for time. Policies should consider fair compensation for the hours worked by drivers, including overtime pay for extended delivery times or working during peak hours. In addition, performance-based pay models could be implemented to reward drivers for timely, safe, and efficient deliveries.

Cluster 3 highlights "construction, company, price, cost". Policy implications include the following: (1) Cost control and price transparency. Policies should promote price transparency in the logistics sector, ensuring that companies clearly communicate delivery costs, potential fees, and price fluctuations. This would help customers make informed decisions and prevent exploitative pricing practices. Policies could focus on standardized pricing models for common logistics services, which would create more predictability and fairness in pricing. (2) Supply chain cost efficiency. Policies could encourage logistics and construction companies to adopt cost-saving strategies, such as bulk purchasing, joint procurement, or shared infrastructure for transporting goods. Providing incentives for companies to optimize routes and warehouse logistics would reduce operational costs and improve delivery times. (3) Subsidizing costs for small and medium-sized enterprises (SMEs). To address the disproportionate impact of rising logistics costs on smaller businesses, policies could provide targeted subsidies, grants, or tax breaks to SMEs in logistics or construction. This would enable smaller players to stay competitive despite the rising costs in the industry.

(4) The cost of construction materials. With rising costs in construction, policies should focus on improving the efficiency of construction logistics, reducing delays, and cutting down on excess material costs. This could include incentivizing the use of local suppliers or promoting digital tools that track and manage inventory to avoid overordering and reduce wastage. (5) Environmental cost regulations. Regulations can focus on the long-term costs of environmental impact, encouraging logistics companies and construction firms to adopt sustainable practices. Policies can introduce carbon pricing, where companies are taxed based on their carbon emissions, or provide incentives for adopting greener technologies (e.g., electric trucks and sustainable construction methods) that help reduce environmental costs in the long run.

Overall, in addressing logistics issues related to these negative issue clusters from subtitle data, policies should focus on (1) improving driver welfare, (2) regulating working hours, (3) ensuring safety, and (4) incentivizing the use of technology to improve operational efficiency. Regulations should (5) promote transparency in pricing, cost control, and sustainable practices in logistics and construction. By addressing these concerns, policymakers can reduce inefficiencies, improve safety, lower operating costs, and create a more sustainable and equitable logistics system.

6.2. Implications from Comment Data Analysis

6.2.1. Positive Topic Clusters

Cluster 1 highlights "yen, prefecture, month". Policies may focus on implementing regional logistics support programs that offer subsidies or grants to logistics companies in different prefectures, helping them manage shipping costs and boosting local economies.

Cluster 2 involves "job, people, comment". It highlights the need to encourage open communication between companies and employees through feedback mechanisms to improve working conditions, job satisfaction, and service quality in logistics.

Cluster 3 discusses "logistic, shipper, shipping". This implicates the importance of creating standards for efficient and sustainable shipping practices, incentivizing companies to adopt greener shipping technologies and optimize logistics routes to reduce costs and emissions.

Cluster 4 focuses on "salary, pay, raise". This indicates a need to enforce fair pay regulations across the logistics sector, ensuring competitive wages for drivers and warehouse staff to improve retention, reduce turnover, and enhance job satisfaction.

Cluster 5 involves "self, driving, technology". This indicates the importance of promoting the adoption of autonomous vehicle technologies in logistics, offering funding for pilot projects, and ensuring a regulatory framework for safe and efficient integration into existing transportation systems.

Cluster 6 highlights "job, driver, transportation". This indicates the need for policies to support driver training programs and certifications, ensuring better safety standards, improved driver skills, and job security while addressing driver shortages in the logistics sector.

Cluster 7 focuses on "company, money, point". This indicates the need to offer tax incentives or rebates to companies investing in innovation, sustainability, and technology to optimize logistics operations and improve financial performance.

Cluster 8 discusses "driver, life, accident". The implications include mandating safety training, better insurance, and accident prevention programs for drivers to reduce accidents and improve their life quality.

Cluster 9 highlights "truck, driver, care". This highlights the need to introduce regulations to ensure that drivers have access to proper rest areas, health programs, and mental health support, as well as regular truck maintenance schedules to ensure safety and well-being.

Cluster 10 features "idea, color, placement". This implies that policies should encourage logistics companies to innovate in packaging design, optimizing space and reducing material usage. Incentives for environmentally friendly packaging that is efficient and reduces overall logistics costs should be offered.

Overall, the positive topics highlight the need for (1) regional support for shipping costs. It proposes (2) the promotion of sustainable shipping and greener technologies, such as autonomous vehicle adoption with funding support and the introduction of eco-friendly packaging innovations. It also suggests (3) the importance of enhancing communication for better job satisfaction and (4) calls for fair pay to improve driver retention. It further emphasizes (5) providing support for driver training to enhance safety and accident prevention, including the provision of tax incentives for innovation. Finally, it highlights (6) improvements in driver welfare through rest areas and support.

6.2.2. Negative Topic Clusters

Cluster 1 involves "japan, yen, japanese, kishida". Kishida is the former president of Japan. This indicates the need to promote government policies that support logistics innovation in Japan, including funding for digital transformation in the logistics sector to enhance efficiency and cost-effectiveness.

Cluster 2 discusses "delivery, specify, date, package". This highlights the need to enforce clearer regulations for delivery timeframes and tracking to improve transparency, reduce delays, and ensure customer satisfaction in package deliveries.

Cluster 3 includes "politicians, bureaucrats, government, country". This indicates the need to implement stronger coordination between government agencies and logistics companies to streamline regulations, reduce bureaucratic hurdles, and support industry growth.

Cluster 4 features "transportation, deregulation, industry, fares". There is a need to consider carefully balanced deregulation of transportation fares to increase competition while ensuring safety and quality standards in the logistics sector.

Cluster 5 discusses "pallets, pallet, bulk, loading". Policies may consider encouraging the adoption of standardized loading and palletization systems to improve warehouse efficiency and reduce handling times in logistics operations.

Cluster 6 involves "logistics, stop, collapse, blood". Policies may consider establishing contingency plans and infrastructure investments to prevent logistics disruptions (e.g., strikes or system collapses) that could severely impact supply chains, particularly in critical sectors.

Cluster 7 includes "sorry, saskiyoshiaki, comment, icomikisan". The two distinct names are the names of YouTubers. This indicates that individual users are playing increasing roles in tackling logistics issues. Policies may consider encouraging transparency in addressing logistics issues by promoting clear communication and accountability from both companies and government officials when delays or errors occur.

Cluster 8 highlights "overtime, unpaid, work, pay". Policies may consider enforcing regulations that prevent unpaid overtime work in the logistics industry, ensuring that workers are fairly compensated for extra hours worked and improving labor conditions.

Cluster 9 indicates "speed, limiter, 80 km, limit". This implies the need to introduce regulations that require speed limiters in logistics vehicles to ensure safety, improve fuel efficiency, and reduce the risk of accidents.

Cluster 10 includes "reporting, solve, change, infidelity". Policies may consider strengthening reporting mechanisms and oversight to address unethical practices in logistics, ensuring accountability for any mismanagement or dishonest practices.

Overall, the top topic clusters from comments data suggest (1) the importance of government support in logistics innovation and digital transformation. They indicate (2) the need for clearer delivery time regulations and tracking. They highlight (3) the importance of coordination between government agencies and logistics companies and (4) call for balanced deregulation of transportation fares to enhance competition. They further propose (5) providing support to standardized loading systems to improve efficiency and (6) urge contingency planning for logistics disruptions. They also encourage (7) transparency in addressing logistics issues and (8) call for regulations against unpaid overtime, strengthening regulations for speed limiters in vehicles for safety.

6.3. Summary

By comparing the insights derived from subtitle and comment data, we found that subtitle data provide a more focused, policy-driven approach aimed at mitigating negative challenges in logistics, while comment data provide a more holistic framework by integrating both positive and negative insights, innovation, workforce development, and continuous improvement (Table 7).

Table 7. Overview of policy recommendations.

From Subtitle Analysis		**From Comment Analysis**	
Optimize time management	Improve driver welfare	Support for shipping costs	Government support in logistics innovation
Ensure driver welfare	Regulate working hours	Promote sustainable shipping	Clearer delivery time regulations and tracking
Promote innovation	Ensure safety	Enhance communication	Coordination between government agencies and logistics companies
Encourage transparency	Incentivize use of technology	Fair pay to improve driver retention	Balanced deregulation of transportation fares
	Promote transparency in pricing	Support to driver training	Support of standardized loading systems
		Improve driver welfare	urges contingency planning for logistics disruptions
			Transparency in addressing logistics issues
			Regulations against unpaid overtime
			Strengthening regulation for speed limiters

As a result, insights extracted from comment data provide a more forward-looking and comprehensive set of policy recommendations.

Taking advantage of both, and combining insights extracted from both subtitle and comment data, the following can be concluded: (1) To address logistics issues, policies should prioritize optimizing time management, ensuring driver welfare, promoting innovation, and fostering transparency. (2) Companies should be encouraged to invest in technology and workforce development to improve efficiency, reduce environmental impact, and increase customer satisfaction. (3) A culture of continuous improvement should be promoted, focusing on data and feedback to solve logistical challenges. (4) Policies should focus on improving driver welfare, regulating working hours, and ensuring safety. Incentives for technology adoption should be promoted to improve efficiency. (5) Regulations should promote transparency in pricing, cost control, and sustainability in logistics and construction, ultimately reducing inefficiencies, improving safety, lowering costs, and creating a more sustainable logistics system.

7. Conclusions

The application of NLP technology to the field of logistics offers several advantages over traditional research methods. First, traditional research methods in logistics often involve manual analyses of large data sets, which can be time-consuming and are prone to human error. NLP enables automated processing and analysis of large amounts of unstructured data, such as social media posts, speeding up the research process. Second, traditional methods may focus on structured data (e.g., numerical data from logistics systems), but much of the valuable information in logistics comes from unstructured

sources, such as emails, SNS, or documents. NLP can analyze these sources and extract actionable insights that traditional methods may miss. Third, NLP can provide deeper insights by identifying patterns in large data sets that would be difficult for traditional methods to detect. For example, analyzing stakeholder sentiment in the SNS channel can help improve product offerings and/or policy decisions. In conclusion, NLP provides a more scalable, automated, and insightful way to analyze data in logistics, improving efficiency, accuracy, and decision making compared to traditional research methods.

In this study, we applied NLP techniques to analyze subtitles and comments from YouTube videos related to logistics issues. By converting audio data into text and analyzing it along with comment data, we introduced a novel approach to social media analysis. Our findings suggest that subtitle data provide a more focused, policy-oriented perspective on addressing logistics challenges, while comment data provide a broader framework that includes both positive and negative insights. These two data sources can be used complementarily to gain a more complete understanding of logistics issues and identify solutions.

This research presents a more holistic, forward-looking framework for logistics policy that prioritizes technology, worker welfare, transparency, and sustainability to address both operational and societal challenges. These contributions go beyond existing policies, which often fail to effectively integrate these elements or fully exploit the potential of data-driven and innovative solutions.

Firstly, the research findings' emphasis on optimizing time management is a significant addition, as many existing logistics strategies often overlook the time management aspect, which has a direct impact on both cost efficiency and service delivery. This suggests a shift toward more systematic scheduling, route optimization, and time tracking technologies, areas that are still developing in some regions.

Secondly, the research findings suggest that an increased focus on ensuring driver welfare and regulating working hours is a critical contribution, particularly as existing policies in some places are insufficient to address the physical and mental well-being of drivers. By emphasizing fair pay, rest breaks, and work–life balance, this research is in line with emerging global calls to improve labor standards in logistics and provide a more comprehensive framework for worker protection.

Thirdly, the research suggests that the focus on encouraging companies to invest in both technology and workforce development to improve operational efficiency and reduce environmental impact goes beyond many current policies that focus primarily on infrastructure or regulatory compliance. The research highlights how embracing innovation and fostering a culture of continuous improvement can enable logistics companies to remain competitive, efficient, and sustainable, and it points to the long-term benefits of these investments.

Fourthly, the research suggests that promoting transparency in pricing, cost control, and sustainability is a notable addition to traditional logistics policies, which tend to focus on cost reduction or efficiency without addressing environmental and ethical concerns. This research positions transparency and sustainability as central principles for creating a more equitable and forward-looking logistics system.

It is important to recognize that our study has limitations. First, we focused on videos uploaded to YouTube related to the "2024 problem" posted through July 2024. We did not include similar information published in languages other than Japanese or information published after the study period, leaving room for future research in these areas. In particular, countries such as the U.S. and Europe are also actively working to improve the problem of declining truck driver numbers. The views and municipal interests of these countries should also be included in future research.

Second, while topic modeling and sentiment analysis methods are powerful, they are not perfect. For example, a model might classify a sarcastic comment as positive when it is intended to express frustration or dissatisfaction. Also, sentiment analysis tools often struggle to detect sentiment in ambiguous or context-dependent statements. For example, a sentence like "This problem is really troubling, but I think we can handle it" could be interpreted as negative because of the word "troubling", even though the overall sentiment is positive because of the phrase "we can handle it". Future research is encouraged to explore solutions for improving the sentiment model. It is also recommendable to gain a deeper understanding by utilizing state-of-the-art methods, such as GraphRAG, for the data collected in this study.

Third, since YouTube users can create an account with a pseudonym, the data collected lack reliable detailed demographic information—such as age, profession, or social group—which could impact the interpretation of the result.

In conclusion, our study demonstrates the effectiveness of NLP techniques in extracting the top priorities in various issues related to truck drivers. It is difficult for humans to manually review and summarize all of the content. Not only does this study provide valuable insights, but it shows that NLP techniques have proven useful for both academic researchers and industry professionals.

Author Contributions: Conceptualization, H.N. and E.H.; methodology, H.N. and E.H.; software, H.N.; validation, E.H.; data curation, H.N.; writing—original draft preparation, H.N. and E.H.; writing—review and editing, H.N. and E.H.; visualization, H.N.; supervision, E.H.; funding acquisition, E.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI, Grant Number JP 23K04076.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chandiran, P.; Ramasubramaniam, M.; Venkatesh, V.G.; Mani, V.; Shi, Y. Can Driver Supply Disruption Alleviate Driver Shortages? *A Systems Approach. Transp. Policy* **2023**, *130*, 116–129. [CrossRef]
2. Ji-Hyland, C.; Allen, D. What Do Professional Drivers Think about Their Profession? An Examination of Factors Contributing to the Driver Shortage. *Int. J. Logist. Res. Appl.* **2022**, *25*, 231–246. [CrossRef]
3. Hottolink 2025 Ranking of SNS Utilization in Japan and Worldwide. Available online: https://www.hottolink.co.jp/column/20250106_114872/ (accessed on 14 February 2025).
4. IRU Global Truck Driver Shortage Report 2023 | IRU | World Road Transport Organisation. Available online: https://www.iru.org/resources/iru-library/global-truck-driver-shortage-report-2023 (accessed on 15 January 2025).
5. Mittal, N.; Udayakumar, P.D.; Raghuram, G.; Bajaj, N. The Endemic Issue of Truck Driver Shortage-A Comparative Study between India and the United States. *Res. Transp. Econ.* **2018**, *71*, 76–84. [CrossRef]
6. Ju, S.; Belzer, M.H. Follow the Money: Trucker Pay Incentives, Working Time, and Safety. *Econ. Labour Relat. Rev.* **2024**, *35*, 7–26. [CrossRef]
7. Hirata, E.; Matsuda, T. Examining Logistics Developments in Post-Pandemic Japan through Sentiment Analysis of Twitter Data. *Asian Transp. Stud.* **2023**, *9*, 100110. [CrossRef]
8. Yano, Y. The Status of the Driver Shortage Problem in the Trucking Industry and Future Actions. *Jpn. Inst. Labour Policy Train.* **2024**, *66*, 51–65.
9. Ueno, H.; Yoon, J.; Masuike, H. How a Crisis in Truck Driving Could Change Life in Japan. *The New York Times (Digital Edition)*, 23 December 2023.
10. Khan, M.L. Social Media Engagement: What Motivates User Participation and Consumption on YouTube? *Comput. Hum. Behav.* **2017**, *66*, 236–247. [CrossRef]
11. Liu, J.; Fukuzawa, K. A Study on Popular Videos in Video Distribution Services: From a Large-Scale Data Analysis of YouTube Channels. *Trans. Jpn. Soc. Bus. Diagn.* **2023**, *23*, 7–13. [CrossRef]

12. Shao, G. Understanding the Appeal of User-generated Media: A Uses and Gratification Perspective. *Internet Res.* **2009**, *19*, 7–25. [CrossRef]
13. Duffy, P. Using Youtube: Strategies for Using New Media in Teaching and Learning. In *Enhancing Learning Through Technology*; World Scientific: Singapore, 2008; pp. 31–43. ISBN 978-981-279-944-9.
14. Orús, C.; Barlés, M.J.; Belanche, D.; Casaló, L.; Fraj, E.; Gurrea, R. The Effects of Learner-Generated Videos for YouTube on Learning Outcomes and Satisfaction. *Comput. Educ.* **2016**, *95*, 254–269. [CrossRef]
15. Chan, Y.M. Video Instructions as Support for beyond Classroom Learning. *Procedia-Soc. Behav. Sci.* **2010**, *9*, 1313–1318. [CrossRef]
16. Liyih, A.; Anagaw, S.; Yibeyin, M.; Tehone, Y. Sentiment Analysis of the Hamas-Israel War on YouTube Comments Using Deep Learning. *Sci. Rep.* **2024**, *14*, 13647. [CrossRef] [PubMed]
17. Serrano-Contreras, I.-J. NLP on YouTube: A Look on Feminism. In *Digitalization of Democratic Processes in Europe: Southern and Central Europe in Comparative Perspective*; Musiał-Karg, M., Luengo, Ó.G., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 127–138. ISBN 978-3-030-71815-2.
18. Gao, J.; Cheng, Q.; Yu, P.L.H. Detecting Comments Showing Risk for Suicide in YouTube. In *Proceedings of the Future Technologies Conference (FTC), Vancouver, BC, Canada, 18 October 2018*; Arai, K., Bhatia, R., Kapoor, S., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 385–400.
19. Das, S.; Dutta, A.; Lindheimer, T.; Jalayer, M.; Elgart, Z. YouTube as a Source of Information in Understanding Autonomous Vehicle Consumers: Natural Language Processing Study. *Transp. Res. Rec.* **2019**, *2673*, 242–253. [CrossRef]
20. Li, T.; Lin, L.; Choi, M.; Fu, K.; Gong, S.; Wang, J. YouTube AV 50K: An Annotated Corpus for Comments in Autonomous Vehicles. In Proceedings of the 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 15–17 November 2018; pp. 1–5.
21. Google YouTube Data API. Available online: https://developers.google.com/youtube/v3 (accessed on 15 January 2025).
22. Yt-Dlp/Yt-Dlp 2025. Available online: https://github.com/yt-dlp/yt-dlp (accessed on 15 January 2025).
23. Openai/Whisper-Small · Hugging Face. Available online: https://huggingface.co/openai/whisper-small (accessed on 15 January 2025).
24. Moreno, A.I.; Mora, M.E.D. Google Translate versus DeepL in Spanish to English Translation of Don Quixote. *Transl. Translanguaging Multiling. Contexts* **2025**, *11*, 65–87. [CrossRef]
25. Freskila, E.P.; Jayantini, I.G.A.S.R. Translation Equivalence of Tourism Website Content: A Comparison Between Google Translate and DeepL Translate. *Indones. J. EFL Linguist.* **2025**, *10*, 1–20. [CrossRef]
26. Camacho-collados, J.; Rezaee, K.; Riahi, T.; Ushio, A.; Loureiro, D.; Antypas, D.; Boisson, J.; Espinosa Anke, L.; Liu, F.; Martínez Cámara, E. TweetNLP: Cutting-Edge Natural Language Processing for Social Media. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 38–49.
27. Grootendorst, M. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure 2022. Available online: https://arxiv.org/abs/2203.05794 (accessed on 15 January 2025).
28. Radford, A.; Kim, J.W.; Xu, T.; Brockman, G.; McLeavey, C.; Sutskever, I. Robust Speech Recognition via Large-Scale Weak Supervision. 2022. Available online: https://arxiv.org/abs/2212.04356 (accessed on 15 January 2025).
29. Loureiro, D.; Barbieri, F.; Neves, L.; Espinosa Anke, L.; Camacho-collados, J. TimeLMs: Diachronic Language Models from Twitter. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Dublin, Ireland, 22–27 May 2022; pp. 251–260.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.