Future Generation Computer Systems 126 (2022) 110–122

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A reinforcement learning-based framework for disruption risk identification in supply chains

Hamed Aboutorab [a,*], Omar K. Hussain [a], Morteza Saberi [b], Farookh Khadeer Hussain [b]

[a] School of Business, UNSW Canberra, Canberra, ACT, Australia
[b] School of Computer Science, University of Technology, Sydney, NSW, Australia

ARTICLE INFO

Article history:
Received 12 January 2021
Received in revised form 4 August 2021
Accepted 5 August 2021
Available online 9 August 2021

Keywords:
Supply chains
Disruption risks
Proactive risk identification
Reinforcement learning

ABSTRACT

Risk management is one of the critical activities which needs to be done well to ensure supply chain activities operate smoothly. The first step in risk management is risk identification, in which the risk manager identifies the risk events of interest for further analysis. The timely identification of risk events in the risk identification step is crucial for the risk manager to be proactive in managing the supply chain risks in its operations. Undertaking this step manually, however, is tedious and time-consuming. With the increased sophistication and capability of advanced computing algorithms, various eminent supply chain researchers have called for the use of artificial intelligence techniques to increase efficiency and efficacy when performing their tasks. In this paper, we demonstrate how reinforcement learning, which is one of the recent artificial intelligence techniques, can assist risk managers to proactively identify the risks to their operations. We explain the working of our proposed Reinforcement Learning-based approach for Proactive Risk Identification (RL-PRI) and its various steps. We then show the performance accuracy of RL-PRI in identifying the risk events of interest by comparing its output with the risk events which are manually identified by professional risk managers.

1. The motivation of this paper

The growth of globalisation has increased the dependence of supply chain (SC) companies on others situated across geographic borders. While this has increased the efficiency of SC companies, it has also exposed them to additional uncertainties and risks that can negatively impact their operations. To manage these risks and ensure the success of the operations in such a networked business environment, proactive risk identification has become an important step. However, as mentioned in our previous work [1], this step is also extremely complex as it requires each SC company to achieve the following three requirements, namely (i) consider the external risks impacting its operations apart from only the internal risk events (termed R1), (ii) identify new risk events negatively impacting its operations apart from the known ones (termed R2), and (iii) process a high volume of operational data in a timely manner and flag it to the risk manager's attention if it is of further interest (termed R3). In the current era of big data, undertaking these tasks manually is extremely challenging and impractical. Thus, various algorithms have been used in the literature to assist the risk manager in the task of risk identification, as shown in Table 1. As our focus in this paper is on the task of risk identification and not on the other tasks of risk management, such as risk assessment, risk evaluation etc., Table 1 only shows those techniques that assist in this task.

However, as seen in Table 1, while approaches do exist that attempt to achieve R1, R2 or R3 individually while identifying risk, none of them achieve this collectively [1]. This is despite researchers acknowledging the need for such approaches [21] that will assist a risk manager to proactively identify risks to its operations. As we highlighted in [1], achieving R1-R3 collectively requires an interdisciplinary effort between SC risk managers and data scientists and the use of advanced artificial intelligence-based algorithms that can process a massive amount of data captured from not only the internal but also the external operating environment and determine if it needs to be brought to the attention of the risk manager or not. Reinforcement learning (RL), a type of machine learning, aims to automatically learn from the environment and act accordingly. It has been used in different applications such as Atari games [22–27], robotics [28–30], medicine [31,32]. RL algorithms need several essential components to work, namely, agent, environment, state, action, and reward or penalty. At an abstract level, an agent captures the current state of the environment and selects the best possible action to take in the given scenario to achieve its goal. Based on the action taken, the agent receives either a reward or a penalty. These returns are constantly used by the agent to improve on the

* Corresponding author.
E-mail addresses: h.aboutorab@unsw.edu.au (H. Aboutorab), o.hussain@adfa.edu.au (O.K. Hussain), morteza.saberi@uts.edu.au (M. Saberi), farookh.hussain@uts.edu.au (F.K. Hussain).

https://doi.org/10.1016/j.future.2021.08.004
0167-739X/

Table 1
Techniques used for risk identification in the literature.
Source: Reproduced from [1].

Paper	Technique used for identifying risks	R1	R2	R3
[2]	Integration of literature review and social system theory	✓	✓	×
[3]	Scenario analysis method	✓	×	×
[4]	Value-focused process engineering (VFPE)	✓	✓	×
[5]	Hazard and operability (HAZOP).	×	×	×
[6]	Historical data analysis and interview	×	×	×
[7]	Risk identification hierarchy structure and risk control model and relative measures	✓	×	×
[8]	State–space	×	✓	×
[9]	Engineering judgements and historical records	✓	✓	×
[10]	Semi-structured interviews and fuzzy-failure mode and effects analysis (FMEA)	×	×	×
[11]	Three-dimensional risk matrix	×	✓	×
[12]	Association rule mining (ARM)	✓	×	✓
[13]	Focus groups and subsequent Delphi study	✓	×	×
[14]	Fuzzy-failure mode and effects analysis (FMEA)	×	×	×
[15]	Pareto chart and time series analysis	✓	×	×
[16–19]	Supply-chain operations reference (SCOR) model	×	×	×
[20]	Integration of the supply-chain operations reference (SCOR) model and value-focused process engineering (VFPE)	×	✓	×
This paper	Reinforcement Learning-based framework for Risk Identification in Supply Chains	✓	✓	✓

next action it takes in the next state of the environment it is in. By adopting such a trial-error method of sense and action, the agent constantly learns and aims to achieve its goal in the best possible way [33,34]. Researchers have proposed RL algorithms which differ in terms of the policy they use and the space from which they can choose the possible actions based on the possible states in which they can be. For example, [35] develops Asynchronous Advantage Actor–Critic (A3C) which is an on-policy algorithm with continuous action–space and state–space. [36] proposes the Deep Deterministic Policy Gradient (DDPG), which is an off-policy algorithm in a continuous space. In this paper, we develop an RL-based model that learns from the environment and brings to the risk manager's attention any new event of disruption risk (in the form of news articles) that has the potential to impact the organisation's operations. Disruption risks are defined as those risks that have the potential to negatively impact the operations between any two SC companies. The developed model uses the Q-learning algorithm in which the learning process is independent of the agent's action (off-policy), and the action–space and state–space are discrete [33].

While existing research has developed RL-based news recommender systems in different domains, none of them has focused on the domain of SC for the identification of operational risks. For example, [37] formulates an RL music recommender based on the exploration and exploitation strategy, which improves the process of recommending by not just focusing on the user ratings but also recommending new songs which may be attractive to the user. [38] integrates Markov chain processes (MDPs) and RL to develop a recommender system that models the sequential interactions with users. [39] proposes a recommender system that suggests stores to users on the Internet using Deep RL. [40] presents a news recommender by utilising Deep RL that considers not only both current and future reward, but also both click/no click labels and user return patterns for feedback. [41] develops an interactive RL recommender system by using a knowledge graph to learn from prior knowledge instead of learning from scratch to avoid the discomfort that may be caused to the user by RL explorative policies. In this paper, we utilise the recommendation ability of the RL-based model to bring to the SC risk manager's attention any news article that it should consider as it has the potential to impact its operations. Due to the ability of RL-based models to recommend appropriate actions, researchers have applied it in different SC activities. For example, [42–44] apply RL in inventory management to determine what quantity of inventory to order. [45] applies it to manage customer demand, whereas [46] investigates its use in supplier selection. In the presence of uncertain operating conditions, [47,48] attempt to optimise a close-loop supply chain's performance, whereas [49] focuses on optimising a company's behaviour for recovery from a disaster. For the sake of brevity, we do not discuss the working of these approaches in this paper, as our aim is to develop an RL-based recommendation model for the identification of disruption risks in supply chain operations. However, no research in the literature has attempted to develop such an approach in an open-loop supply chain. In this paper, we address this gap by developing a Reinforcement Learning-based approach for the Proactive Risk Identification (RL-PRI) of supply chain disruption risks. The structure of the paper is as follows. In Section 2, we explain our proposed RL-PRI approach for the proactive identification of supply chain disruption risks. In Section 3, we explain the working of each module of the proposed framework. In Section 4, we evaluate the performance of the proposed framework, followed by Section 5, which concludes the paper with a discussion on future work.

2. Reinforcement Learning-based approach for the Proactive Risk Identification (RL-PRI) of supply chain disruption risks

As shown in Fig. 1, our proposed RL-PRI approach has four modules, namely, data preparation, data collection, entity recognition, and RL-based recommender system. In the data preparation module, the risk manager identifies the risk events in which they are interested. In the data collection module, relevant news on the risk events is collated from different news sources. In the entity recognition module, the details of each news item are identified. Finally, in the RL-based recommender system module, a score for each news item is calculated and the news which is the most relevant to the selected risk events is shown to the user. The combined output of these modules helps achieve the goal of automatedly and intelligently detecting information related to potential disruption risks and brings this to the attention of the risk manager with minimal human intervention. The working of each module is explained in the subsequent sub-sections.

2.1. Data preparation module

This is the first module of RL-PRI that kick-starts the process of proactive risk identification in supply chain operations. It does this by defining the *disruption risk events* that may have a chance to negatively impact the supply chain company's operations. In other words, the working of this module defines the scope and direction which the risk identification process should take. As

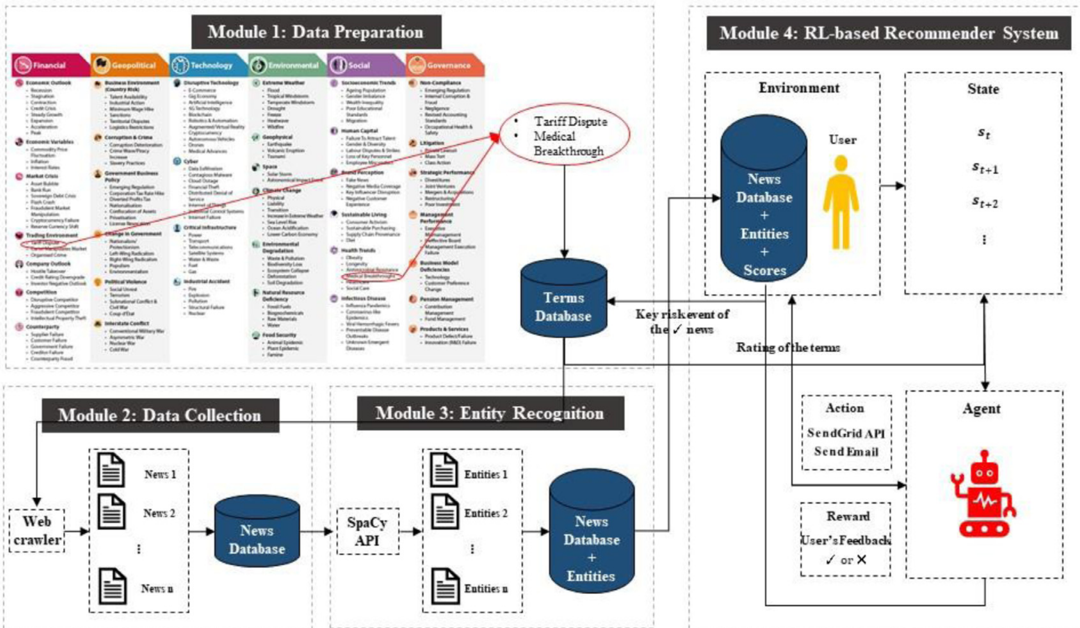


Fig. 1. Conceptual model of the proposed RL-PRI approach for SC disruption risk identification.

the scope and direction depend on the objectives that the supply chain company wants to achieve, the steps of this module are done manually by the risk manager according to the service level agreements (SLA) or the objectives to be achieved. To avoid the ambiguities and complexities that may arise from the risk managers defining the disruption risk events in their own terms, RL-PRI uses the Cambridge Taxonomy of Business Risks [50]. As shown in Fig. 2, this standardised taxonomy lists the different possible disruption risk events, from which the risk manager can identify disruption risk events that may negatively impact its operations. The identified disruption risk events identified by the risk manager are stored in the *Terms database* for further processing in the next modules.
Input: Cambridge Taxonomy of Business Risks, supply chain company's operations or SLAs.
Output: Terms database, which stores disruption risk events that have a chance of negatively impacting the SLAs.
Process: Manually done by the risk manager.

2.2. Data collection module

The second module of RL-PRI collects real-time information related to the disruption risk events defined in the first module. This is done by a web crawler that searches all the trusted sources of news such as BBC, CNN, Bloomberg, Google News, for disruption risk events. News articles which match the disruption risk events will be stored in the *News database* along with their details such as source, title, description, URL, and the disruption risk event term (defined in module 1) to which it links. These details will be used in the next modules.
Input: Terms database.
Output: News database.
Process: Automatedly done by the web crawler.

2.3. Entity recognition module

The third module of RL-PRI augments each collected piece of news with details to increase its contextual description. The input of this module is the *News database*, and the output is the *Augmented News database*.

Table 2
Entity types with which each news article is augmented.

Type	Description
KRE	Most relevant risk event from the Cambridge Taxonomy of Business risks (Key Risk Event)
PERSON	People, including fictional
NORP	Nationalities or religious or political groups
FAC	Buildings, airports, highways, bridges, etc
ORG	Companies, agencies, institutions, etc
GPE	Countries, cities, states.
LOC	Non-GPE locations, mountain ranges, bodies of water
EVENT	Named hurricanes, battles, wars, sports events, etc

As shown in Table 2, the contextual description added to each news article consists of the most relevant risk event from the Cambridge Taxonomy of Business Risks [50], the person entity to whom the news article is related, the nationalities of the persons being discussed in the news article, the physical buildings and companies which are mentioned in the news article, the event mentioned in the news article along with the geographic location concerning the event mentioned. These augmented details for each news article will be used in the next module to decide if that piece of news is relevant to the risk manager's operation for it to be shown to them for further risk management analysis.
Input: News database.
Output: Augmented News database with the entity types for each news article.
Process: Automatedly using the spaCy API [51].

2.4. RL-based recommender module

The fourth and last module of RL-PRI utilises the *Augmented News database* as input and decides which of them should be shown to the risk manager for their proactive follow-up for better risk management. This decision process is done using an RL-based approach that determines the chances of a news article of being of interest to the risk manager or not based on the SLAs objectives to be achieved. In other words, if the RL-PRI decides that a particular news article is relevant to the disruption risk event terms and the SLA to be achieved, then it is shown to the risk manager.

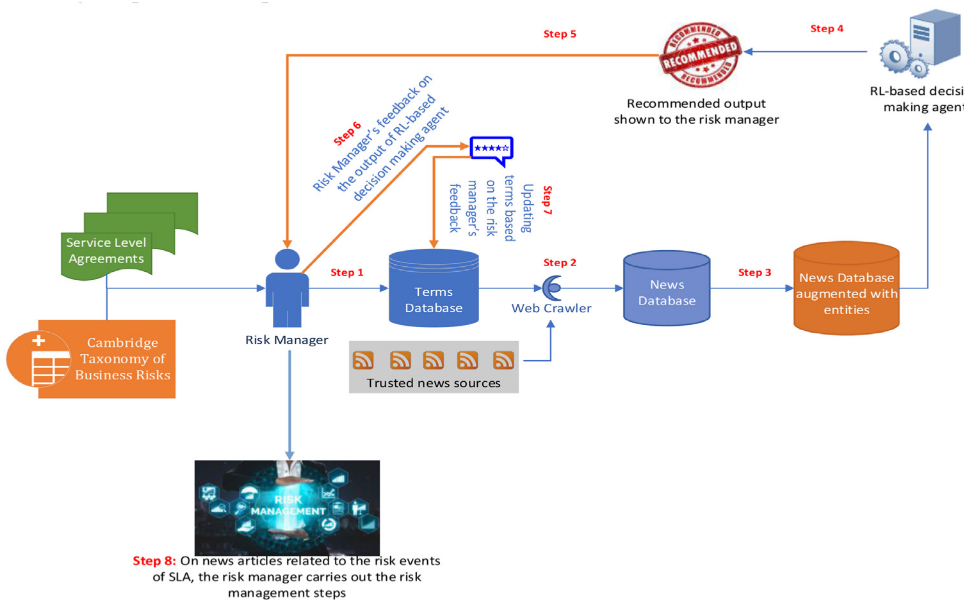


Fig. 2. Working of the RL-PRI's modules classified into different working steps.

Otherwise, it is not. To rate the performance output of the RL-PRI in bringing relevant news articles to the attention of the risk manager, the RL-PRI decision-making agent is given feedback by the risk manager, either as a reward or penalty. This feedback process ensures that RL-PRI brings the human in the loop concept to adapts to the risk manager's preferences in terms of what is considered relevant to the SLAs. The risk manager's feedback is based on their response to the recommended news article, which can either be a ✓ or ×. ✓ represents that the risk manager is interested in the recommended news articles by the RL-PRI and wishes to gather further information about the risk event it relates to. In this case, this response is considered as a *reward* by the RL-PRI because it showed the risk manager a news article which was of interest as it impacts the objectives that the risk manager wants to achieve. On the other hand, × represents that the risk manager is not interested in the recommended news article. In this case, this action by the risk manager is considered as a *penalty* by the RL-PRI, because it showed a news article in which the risk manager is not interested as it does not impact the objectives that they want to achieve. The RL-PRI considers the reward or penalty that it is given and adapts its future workings accordingly to incorporate the exploration and exploitation characteristics of RL-based learning. It does the exploration by utilising the ϵ-greedy strategy, where $0 \leq \epsilon \leq 1$ is the probability of how often the exploration should happen compared to exploitation [52]. During exploration, RL-PRI randomly selects a risk event from the Cambridge Taxonomy of Business Risks, regardless of the SLA, and investigates the risk manager's interest in it.

Input: Augmented News database with the entity types for each news article, Terms database.

Output: News articles recommended by the RL-PRI.

Process: Recommendation process is automated by RL-PRI. The feedback process is manual, based on the risk manager's response to each recommended news article.

Fig. 2 shows the overall working of the different RL-PRI's modules classified into various working steps. Step 1 of Fig. 2 involves the risk manager in defining the Terms database according to the SLAs to be achieved and the Cambridge Taxonomy of Business Risks. This step is undertaken by the *Data Preparation Module*. Step 2 deals in crawling the trusted news sources and collecting news articles that relate to the risk events represented in the Terms database. This is undertaken by the *Data Collection Module*. Step 3 involves augmenting the collected news articles with different types of entities to assist in better decision-making. This step is undertaken by the *Entity Recognition Module*. Steps 4–7 are undertaken by the *RL-based Recommender Module*. Step 4 determines which news items collected in Step 3 are more relevant to the risk manager. Step 5 shows the selected news articles from Step 4. Step 6 collates the risk manager's feedback in terms of how relevant or irrelevant the news articles were to its operations. Step 7 updates the search process for the next iteration based on the risk manager's feedback. Fig. 2 also shows Step 8, in which the risk manager further investigates the risks to its operations arising from the recommended news article in Step 5. The process of assessing and managing the risks is not in the scope of the working of the proposed RL-PRI and hence is not discussed in this paper. The details of the workings steps 1–7 are explained in the next section.

3. Working details of each module of RL-PRI

In this section, we explain the working details of each RL-PRI module of supply chain disruption risks. We do this by taking a hypothetical real-world supply chain scenario and explaining the working of steps 1–7 in this scenario. We consider ABC as a global logistics company based in Australia. The company is dependent on its global suppliers that are based in South Africa and the United Kingdom (UK) to supply it with raw materials for its operations. Let us also consider that ABC exports the goods that it makes to China. As the lead time of getting the goods from these suppliers is high, the risk manager of ABC wants to be proactive in managing the company's operations to avoid any disruption risks that may impact it and its suppliers in an uncertain environment in addition to considering the Australian operational environment. The proposed RL-PRI assists ABC's risk manager to achieve this aim as follows:

Step 1: Selecting the terms and creating the *Terms database* in the Data Preparation Module

This step is done by the Data Preparation Module. As discussed in Section 2.1, this is a manual step in which the ABC's risk manager, based on the SLAs to be met by its suppliers, identifies the risk events from the Cambridge Taxonomy of Business Risks that may impact the organisation and its suppliers. Let us assume that due to the current situation in the world as a result of COVID-19 and China's tariff on Australian trade, the risk manager considers the *Tariff Dispute* and *Medical Breakthrough* from the taxonomy as possible disruption risks impacting its local operations, along with its UK and South Africa-based suppliers and its exporters in China. At the end of this step, the RL-PRI has a list of terms shown in Table 3 in the *Terms database*. In other words, these terms are those risk events which the risk manager identifies as having a possible negative impact on its operations. These terms can either be *confirmed* or *interim*. *Confirmed* represents the risk event term's status as being approved by the risk manager to be added in the Terms database, whereas *interim* indicates that the term has been added temporarily and its status will be confirmed based on the feedback received by the RL-PRI from the risk manager in the next steps. This is explained further in Steps 6 and 7. Step 1 also assigns a rating value (tr) of zero to each term in the Terms database. The rating value of each term will be updated after the analysis of the next steps once their first iteration has occurred. Algorithm 1 shows the working of this step.

```
Algorithm 1: Data Preparation
Taxonomy ← list(Cambridge Taxonomy of Business Risks)
SLA' risk management objectives'
for t in Taxonomy do
    if t in SLA then
        terms ← t
    end if
end for
for i in range(len(terms)) do
    tr ← 0
    type ← 'confirmed
end for
Terms database ← {'Term': terms, 'Rating': tr, 'Type': type}
return (Terms database)
```

Step 2: Collecting news articles related to the risk event terms in the Data Collection Module.

This step is done by the Data Collection Module. As discussed in Section 2.2, this is an automated step wherein a web crawler scans the sources of trusted news and collects articles that match the risk event terms defined in Table 3 to make the *News database*. Table 4 shows a snapshot of the collected web articles that match the risk event terms of *Tariff Dispute* and *Medical Breakthrough* from the Terms database. The web articles in the News database are listed according to the specifics of source, title, description, URL, and term. Algorithm 2 shows the working of this step.

```
Algorithm 2: Data Collection
function extract news (term)
    for news in web do
        if 'term' in news text then
            source ← news source
            title ← news title
            description ← news description
            url ← news url
        end if
    end for
    News database ← {'Source': source, 'Title': title, 'Description': description,
'URL': url, 'Term': term}
    return (News database)
end function
    for t in Terms database do
        extract news (t)
    end for
```

Step 3: Augmenting the News database with different types of entities in the Entity Recognition Module

As mentioned in Section 2.3, this is an automated step undertaken by the Entity Recognition Module. This step is performed on the articles collected in the News database and analyses the

Table 3
Terms database.

Term	Rating (tr)	Type
Tariff Dispute	0	Confirmed
Medical Breakthrough	0	Confirmed

URLs of each news item to obtain the values of the entity types defined in Table 2.

As an example, Table 5 shows the entities of KRE and GPE augmented to the News database of Table 4 by using the spaCy API [51]. For the sake of brevity, the details for the columns of source, title, description, URL, and terms in Table 5 are shown by labels as they are similar to the ones shown in Table 4. Algorithm 3 shows the working of this step.

Steps 4 and 5: Determining which news to be shown to the risk manager in the RL-based Recommender Module

As discussed in Section 2.4, steps 4–7 are undertaken by the *RL-based Recommender Module*. Step 4 determines which news items from those collected in Step 3 are more relevant to show to the risk manager. This is done by the RL-based agent by matching the relevance of each news article to the objectives defined in the SLA. Step 5 shows the articles which were shortlisted for the risk manager. The workings of these steps are as follows:

Task 1: Calculate the initial *news score* (ns) for each article in the Augmented News database

In the first task of Step 4, a *news score* (ns) for each news article in the Augmented News database of Table 5 is determined. This score is based on the relevance of the entities of each news article to the risk manager's objectives or the SLA in question. The relevance of *KRE* of each news article is the similarity (si) between the KRE and the *risk event term*. This similarity is calculated by comparing the word embeddings which represent the multi-dimensional meaning of words by using the spaCy API [51]. For the other entities, their relevance, as entity score (es), to the SLAs is determined automatedly by the RL-based Recommender Module, and a value of 1 or 0 is assigned to them. A relevance value of 1 as the es is assigned if the specifics of each augmented entity for a news article is in the interests of the risk manager. Otherwise, a value of 0 is assigned. Once the values for all the augmented entities for a news article have been assigned, its *news score* (ns) is determined by calculating the mean of these scores, as shown in Eq. (1).

$$ns = \frac{si + \sum_{j=1}^{ne-1} es_j}{ne} \tag{1}$$

where ne represents the number of augmented entities for each news article.

Continuing with the example of the ABC company earlier this section, Table 5 shows the augmented entities for the first two news articles. Based on the entity values of KRE and GPE, an si value of 0.28 is assigned to KRE of the first news article, which is the similarity between Emerging Regulation and Tariff Dispute, and an es value of 0 is assigned to GPE, which is not relevant to the SLA. For the second news article, an si value of 0.18 is calculated for KRE, which is the similarity between Logistics Restrictions and Tariff Dispute, and an es value of 1 is determined for GPE. Therefore, using Eq. (1), the ns values for the news articles are computed, as shown in Table 6.

Task 2: Estimate Q-value for each news article

The second task of Step 4 focuses on estimating the Q-value for each news article, which is used by the RL-PRI to decide which news article is the most beneficial to show to the risk manager. The news article with the highest Q-value has the highest chance of receiving a reward as feedback from the risk manager as

Algorithm 3: Entity Recognition

```
function spaCy NER (news)
    if entity (kre, person, norp, fac, org, gpe, loc, event)  in news text then
        entities ← {'KRE': kre, 'PERSON': person, 'NORP': norp, 'FAC': fac,'ORG':
org, 'GPE': gpe, 'LOC': loc, 'EVENT': event}
    end if
    return (news + entities)
end function
for news in News database do
    spaCy API (news)
end for
```

Table 4
Snapshot of News database.

No	Source	Title	Description	URL	Term
1	Sputnik International	Why US-Led 'Overtly Anti-China Coalition' Probably Will not Appear Under Biden	Joe Biden's concept of "coalitions of democracies" to counterbalance rising China could be part of a broader plan at forming a new world order, where the US-aligned nations will flock to Washington leaving those who do not play by the rules out in the cold.	https://sputniknews.com/...	Tariff Dispute
2	Reuters	Air freight prices 'outrageous' as COVID-19 shots rolled out, says WHO expert	Some carriers are seeking "outrageous" prices to fly dry ice and other medical equipment in the pre-holiday rush, but a capacity squeeze should ease in 2021 when the roll-out of COVID-19 vaccines is expected, the World Health Organisation's logistics chief said.	https://reuters.com/...	Medical Breakthrough

Table 5
News database augmented with entity extensions.

No	Source	Title	Description	URL	Term	KRE	GPE
1	Source of news No.1	Title of news No.1	Description of news No.1	URL of news No.1	Tariff Dispute	Emerging Regulation	US
2	Source of news No.2	Title of news No.2	Description of news No.2	URL of news No.2	Medical Breakthrough	Logistics Restrictions	UK

Table 6
News database with entities and score extensions.

No	Source	Title	Description	URL	Term	KRE	GPE	***ns***
1	Source of news No.1	Title of news No.1	Description of news No.1	URL of news No.1	Tariff Dispute	Emerging Regulation	US	0.14
2	Source of news No.2	Title of news No.2	Description of news No.2	URL of news No.2	Medical Breakthrough	Logistics Restrictions	UK	0.59

feedback vice versa. The objective of RL-PRI is to maximise the rewards, thereby assisting the risk manager in bringing to their attention the news articles related to risk events of interest to its operations. The Q-value of each news article is determined by the optimal value functions [33]. Markov decision processes (MDPs) structure the process of learning from an interaction to attain a goal. An *agent* by using an off-policy Q-learning process, learns and makes decisions based on the *environment* it is in. More specifically, at each time step t and environment *state*, $s_t \in S$, the agent takes an *action*, $a_t \in A$, and receives a *reward*, $r(s_t, a_t) \in R$, based on that action and moves to the next environment state, s_{t+1}. The agent tries to maximise the expected returns to achieve its goal (G_t), which is the sum of the discounted rewards, as shown in Eq. (2):

$$G_t = r(s_t, a_t) + \gamma r(s_{t+1}, a_{t+1}) + \gamma^2 r(s_{t+2}, a_{t+2}) + \cdots \tag{2}$$

where $0 \le \gamma \le 1$ and represents the discount rate.

The discount rate defines the present value of future rewards and assists the agent to decide to what extent to consider those rewards. Eq. (3) shows a modified version of Eq. (2) to show the relation of successive time steps in a way that is critical for the RL theory:

$$\begin{aligned} G_t &= r(s_t, a_t) + \gamma\left(r(s_{t+1}, a_{t+1}) + \gamma r(s_{t+2}, a_{t+2}) + \cdots\right) \\ &= r(s_t, a_t) + G_{t+1} \end{aligned} \tag{3}$$

Eq. (2) is also used to estimate the value functions of the RL algorithm, which measures the quality for the agent to be in a state, and the quality to select an action in a given state. The notion of quality here is defined in terms of expected returns.

$$V_\pi(s_t) = E_\pi\left[G_t | s_t\right] = E_\pi\left[\sum_{k=0}^{\infty} \gamma^k\, r(s_{t+k}) | s_t\right] \tag{4}$$

$$Q_\pi(s_t, a_t) = E_\pi\left[G_t | s_t, a_t\right] = E_\pi\left[\sum_{k=0}^{\infty} \gamma^k\, r(s_{t+k}, a_{t+k}) | s_t, a_t\right] \tag{5}$$

Eq. (3) defines the value function of state s_t under policy π, and Eq. (4) represents the value of taking action a_t in state s_t under policy π. The optimal value functions are then determined as follows:

$$V_*(s_t) = \max_\pi V_\pi(s_t) \tag{6}$$

$$Q_*(s_t, a_t) = \max_\pi Q_\pi(s_t, a_t) \tag{7}$$

Based on the Bellman optimality equation, the value of a state under an optimal policy must be equal to the expected return for the best action from that state

$$V_*(s_t) = \max_{a_t} Q_{\pi_*}(s_t, a_t) = \max_{a_t} E_{\pi_*}\left[G_t | s_t, a_t\right] \tag{8}$$

Using Eq. (3) and function $p(s_{t+1}, r | s_t, a_t)$ which defines the dynamics of the MDP and specifies a probability distribution for each choice of s and t, Eq. (8) can be written as follows:

$$V_*(s_t) = \max_{a_t} \sum_{s_{t+1}, r} p(s_{t+1}, r | s_t, a_t)\left[r + \gamma V_*(s_{t+1})\right] \tag{9}$$

The Bellman optimality equation for $Q_*(s_t, a_t)$ is as follows:

$$Q_*(s_t, a_t) = E\left[r(s_{t+1}, a_{t+1}) + \gamma \max_{a_{t+1}} Q_*(s_{t+1}, a_{t+1}) \middle| s_t, a_t \right] = \sum_{s_{t+1}, r} p(s_{t+1}, r | s_t, a_t) \left[r + \gamma \max_{a_{t+1}} Q_*(s_{t+1}, a_{t+1}) \right] \quad (10)$$

In state s_t, we use Eq. (10) to determine the maximum cumulative reward for showing each news article from the augmented News database. To compute the Q-value for each news article, first, we calculate the rating score (rs) using Eq. (11), which is the probable reward for state s_t.

$$rs = \frac{ns + tr}{2} \quad (11)$$

As shown in Eq. (11), rs for a news article is the average of the *term rating* (tr) in the Terms database and the ns value of the news article. The more augmented entities of a news article match the SLA of the logistics company, the higher the rs value of the news article. In other words, the rs value of a news article is the reward which the RL-based agent will receive for showing it to the risk manager in the current state s_t.

Continuing with the example of the ABC company from earlier in this section, in the first iteration of the RL-PRI, the tr value of each term in the Terms database is 0, as shown in Table 3. The ns values of the news article after the first round of RL-PRI are shown in Table 6. Thus, the rs values for the first two news articles are respectively 0.07 and 0.29, as computed by Eq. (11). In other words, the probable reward for showing the second article to the risk manager is higher than showing the first one as the second article speaks about a risk event in a location which is of interest to the risk manager's operations. Even though the first article is more relevant to the SLA, based on the si value, its geographic location is not Australia, United Kingdom, South Africa, or China. Thus, the probable reward for RL-PRI in the current state s_t to show that piece of news to the risk manager is less compared to the second one. However, as shown in Eq. (10), Bellman optimality also needs to find the future state s_{t+1} in which the risk manager will be, when a news article in state s_t is shown to them, and whether that news article of state s_{t+1} will be beneficial to their operations or not. In the context of dynamic SC risk identification, the future state s_{t+1} relates to finding out what other news articles the risk manager can be shown as a result of showing an article in the current state s_t, and whether these news articles of state s_{t+1} will be of interest to them according to their operations or SLAs. To achieve this, RL-PRI considers:

- the KRE of each news article that can possibly be shown to the user in state s_t. This KRE is from an article that relates to the risk event term of the Terms database in Table 6.
- adding the KRE of the news article in Table 6 (news article that can possibly be shown to the user in state s_t) as an *interim term* to the Terms database. News articles related to this interim term are then searched to determine the risk manager's state in s_{t+1} by repeating the process of Steps 2 and 3.

The rationale behind this approach is that as the KRE of an article that can possibly be shown to the user in state s_t relates to the risk event term of the Terms database in Table 6, then in state s_{t+1}, RL-PRI which other news articles related to the news article in state s_t should be shown to the risk manager. In other words, as RL-PRI uses an off-policy Q-learning process, it estimates the ns_{t+1} score of all the probable news articles in the partially observed state s_{t+1}, given state s_t. It does this searching for news articles that can be shown in s_{t+1} related to the KREs of the news article in s_t of Table 6 by repeating the process of Steps 2 and 3. The probable reward (ns) of showing these news articles to the risk manager in state s_{t+1} is computed using Task 1 of Step 4. Using these rewards of showing news articles in state s_t and s_{t+1}, the Q-value of showing each news article to the risk manager at the state $\mathrm{s_t}$ can be determined by using Eq. (12).

Table 7
Updated terms database.

Term	Rating (tr)	Type
Tariff Dispute	0	Confirmed
Medical Breakthrough	0	Confirmed
Emerging Regulation	0	Interim
Logistics Restrictions	0	Interim

$$Q_*(s_t, a_t) = rs_t + \gamma \max_{a_{t+1}}(ns_{t+1}) \quad (12)$$

Continuing with the example of the ABC company, Table 6 shows the KREs of the news articles along with the probable reward of showing them to the risk manager in state s_t. To compute the future state s_{t+1} and probable reward which RL-PRI will obtain in showing these news articles, the KREs of Table 6 are added as *Interim Terms* to the Terms database, as shown in Table 7. Steps 2 and 3 discussed above are repeated for the newly added terms, and the result is shown in Table 8 for the two terms of Emerging Regulation and Logistics Restrictions. The ns value for each news article determined is shown in Table 8.

Therefore, the Q-values of showing (action — a_1) news article 1 and news article 2 (action — a_2) of Table 6 to the risk manager in state (s_t) with the assumption of $\gamma = 0.5$ is calculated using Eq. (12) as follows:

$$Q_*(s_t, a_1) = 0.07 + 0.5 \times 0.2 = 0.17$$
$$Q_*(s_t, a_2) = 0.29 + 0.5 \times 0.69 = 0.63$$

Task 3: Selecting the optimal action

The third task of Step 4 is to select the optimal action. In other words, this action selects the news article to be shown to the risk manager that maximises the Q-value. This is done using Eq. (13) as follows:

$$a_{*t} = \underset{a_t}{\operatorname{argmax}}\, Q_*(s_t, a_t) \quad (13)$$

Step 5 sends the selected news articles of interest to the risk manager by email using SendGrid API [53]. Continuing with the example of the ABC company, we obtain the optimal action using Eq. (13). From the above analysis, the probable reward of showing news article 2 in the current state (s_t) is more than that of news article 1. This makes sense as the KREs and GPEs of news articles in state s_{t+1} arising from news article 2 of Table 6 are more related to the SLAs of the risk manager than news article 1. So, the RL-PRI notifies these news articles to the risk manager by email and waits for the feedback from the risk manager on its recommendation.

Steps 6 and 7: Receiving the feedback of the risk manager and calculating the rating values in the RL-based Recommender Module

Step 6 determines the feedback from the risk manager in terms of either ✓ or ✕. Depending on this feedback, Step 7 updates the searching process in the next iterations. If the risk manager is interested in the recommended news article (n_i), the rating (tr_{n_i}) related to the corresponding risk event term to that article will be updated using Eq. (14). Otherwise, it is updated using Eq. (15).

$$tr_{n_i} = \frac{tr_{n_i} + 1}{2} \quad (14)$$

Table 8
News database with entities and score extensions for 'Emerging Regulation' and 'Logistics Restrictions'.

	Source	Title	Description	URL	Term	KRE	GPE	*ns*
1	Forbes	There are Plenty of Concerns for Aerospace and Defence in 2021 but Deloitte's Study Finds Some Positives	In its 2021 Aerospace and defence Outlook, Deloitte lays out a variety of familiar pandemic and politically related downside risks for the sectors this coming year.	https://www.forbes.com/...	Emerging Regulation	Innovation (R&D) Failure	China	0.2
2	Telegraph	Why ministers must ramp up COVID-19 vaccination to 2M a week or face a devastating third wave	New modelling suggests even a full national lockdown would not be enough to prevent a third wave more deadly than the first	https://www.telegraph.co.uk/...	Logistics Restrictions	Coronavirus-like Epidemics	UK	0.69
3	CBC News	Confusion reigns over holiday COVID-19 restrictions	Without a detailed list of what is considered an essential good, the definition varies from one business to another.	https://www.cbc.ca/...	Logistics Restrictions	Coronavirus-like Epidemics	Canada	0.19

Table 9
Updated terms database.

Term	Rating (tr)	Type
Tariff Dispute	0	Confirmed
Medical Breakthrough	0.5	Confirmed
Logistics Restrictions	0	Confirmed

$$tr_{n_i} = \frac{tr_{n_i} + 0}{2} \tag{15}$$

In the next iterations, such an updated rating value of a term will result in determining an accurate *rs* score of a news article which has the same risk term. The higher the (tr_{n_i}) value of a term, the higher the *rs* score of the news article, showing its importance to the risk manager's operations and vice versa. In Step 7, the type of interim risk events used to determine the risk manager's state in s_{t+1} will be updated based on the risk manager's feedback on the recommended news article in state s_t. If the risk manager is interested in the recommended news article (n_i) in s_t, then the type of the corresponding term which was used to determine the risk manager's state in s_{t+1} is updated from *interim* to *confirmed*. This means that the next time RL-PRI is run, this term is also considered to be a risk event and news articles on these risk event terms are searched to identify the disruption risks related to it. On the other hand, if the risk manager is not interested in the recommended news article, and/or the news article related to the *interim* term is not recommended to be shown to the risk manager, then the *interim* term/s are removed from the Terms database. This is because it is considered that such terms are not important and hence should not be searched to identify disruption risks in the future interactions of the framework.

Continuing with the example of the ABC company, if the risk manager responds with a ✓ on the recommended news article 2 in the current state (s_t), the rating for the term, *Medical Breakthrough*, will be calculated using Eq. (14), and the type of the term *Logistics Restrictions* will be changed from interim to confirmed in the Terms database, as shown in Table 9. Algorithm 4 shows the working of steps 4–7. Fig. 3 shows the working of the proposed framework which is based on the continuous implementation of steps 1–7. The frequency of the implementation can vary from daily or hourly, depending on how frequently the risk manager wants to scan the environment.

Algorithm 4: RL-based recommender system

function spaCy similarity (a, b)
  $si \leftarrow$ calculate similarity of a and b
  return (si)
end function
function entity score (type)
  SLA' risk management objectives'
  if type in SLA then
    $es \leftarrow 1$
  else
    $es \leftarrow 0$
  end if
  return (es)
end function
function Sendgrid API (x)
  send email (x)
end function
for news in *News database* do
  spaCy similarity (Term, KRE)
  for x in entities do
    entity score (x)
  end for
  $ne \leftarrow$ number of entity
  calculate news score $\leftarrow \frac{si + \sum_{j=1}^{n-1} es_j}{ne}$
  return (news + entities + news score)
end for
calculate $Q_*(s_t, a_t) \leftarrow rs_t + \gamma \max_{a_{t+1}} (ns_{t+1})$
calculate $a_{*t} \leftarrow \underset{a_t}{\operatorname{argmax}}\, Q_*(s_t, a_t)$
return (a_{*t} (the most relevant news (news*))
Sendgrid API (news*)
if news* is ✓ then
  calculate $tr_{\text{news}*} \leftarrow \frac{tr_{\text{news}*}+1}{2}$
  terms.append(KRE*)
else
  calculate $tr_{\text{news}*} \leftarrow \frac{tr_{\text{news}*}+0}{2}$
end if

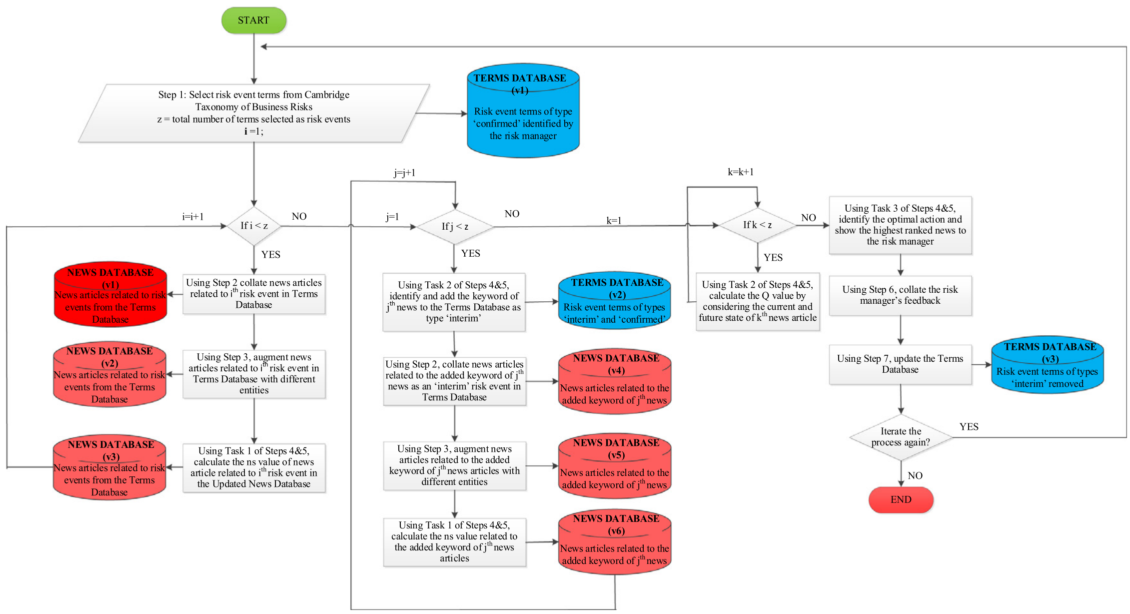


Fig. 3. Working of the RL-based approach for the proactive identification of supply chain disruption risks.

Table 10
Number of news articles collected from Google News for each risk event.

Tarif dispute	Recession	Sea level rise	Flood	Total
205	315	280	348	1148

4. Evaluation of RL-PRI in proactively assisting the risk manager in identifying the disruption risk events

In this section, we show the working of RL-PRI in assisting the risk manager to proactively identify supply chain disruption risks to its operation. For an ideal evaluation, a like-by-like comparison should be performed in which the results of RL-PRI are compared with other similar automated risk identification approaches. However, as discussed in Section 1, while different types of risk identification approaches are proposed in the literature, none are based on the working of an RL-based approach, whose aims are to (a) proactively identify disruption risk events types that may negatively impact on their operations and (b) collate real-time information related to the occurrence of such disruption risk events and bring them to the attention of the risk manager for them to decide if it is sufficiently important to be considered in the further analysis of risk management or not. So, in our evaluation, we compare the performance of RL-PRI against that of a human risk manager. Our evaluation approach is as follows:

- In Stage 1, we modify the example of the ABC company from Section 3. We consider ABC as a company based in Australia with a market in Australia and China. According to the SLAs of ABC, we, from the perspective of the risk manager, identify the disruption risk events from the Cambridge Taxonomy of Business Risks that will negatively impact its operations. This is further explained in Section 4.1.
- In Stage 2, we collect news articles related to the identified disruption risk events in Stage 1. We do this in two ways. First is in an automated way, where the proposed RL-PRI proactively identifies and collects news articles related to the disruption risk events of interest. Second is from the perspective of a risk manager who uses Google to manually identify and collect the news articles related to the disruption risk events of interest. To reach agreement on the collected datasets of the news articles using both approaches, two or more experts, as risk managers, determine if a news article is relevant to the SLAs. Then Cohen's kappa metric is used to measure the agreement between the experts' analysis. This is further explained in Section 4.2.
- In Stage 3, we measure the *relevancy* and evaluate the *correctness* of the recommended articles by (a) RL-PRI compared to the manually collected ones, and (b) manually collected ones compared to RL-PRI. By doing so, we measure and determine how accurately RL-PRI performs compared to what a human would do manually and vice versa. The definition of the *relevancy* and *correctness* metrics and the process of determining their values is as follows:
- *Relevancy* : This metric is used only on the articles recommended by RL-PRI and it measures *how relevant each recommended news article is to the risk manager's operations*. As explained in Task 2 of Steps 4 and 5, RL-PRI computes the $Q_*(s_t, a_t)$ value for each news article that represents the *relevance* score of that article to the SLA operations. The higher the score, the higher the relevance for this news article to the SLAs, and vice versa.
- *Correctness* : This metric measures *how correct the recommended outputs are of RL-PRI compared to the manually identified ones* and vice-versa. This is done by determining the accuracy (a), precision (p), recall (r), specificity (s), and $F1$ score of the RL-PRI and manual outputs. Before determining such scores, each output of RL-PRI and the manual process is classified into one of the following categories:
    - True Positive (TP) → also known as *relevant* and *showing*. This metric shows the number of news articles that are relevant to the SLAs and are recommended to the risk manager by the manually collated dataset and the one recommended by RL-PRI
    - False Positive (FP) → also known as *irrelevant* and *showing*. This metric shows the number of news articles that are not relevant to the SLAs and are recommended to the risk manager by the manually collated dataset and the one recommended by RL-PRI.
    - False Negative (FN) → also known as *relevant* and *not showing*. For the news dataset manually collated by the

Table 11
Scoring statistics by the experts for the manual news collection.

Scoring metrics	Tarif dispute	Recession	Sea level rise	Flood
Both experts score 1 for the news articles	46	91	26	250
Both experts score 0 for the news articles	159	224	252	96
Total number of news articles ranked as 1 by expert 1 and 0 by expert 2	0	0	0	0
Total number of news articles ranked as 1 by expert 2 and 0 by expert 1	0	0	2	2

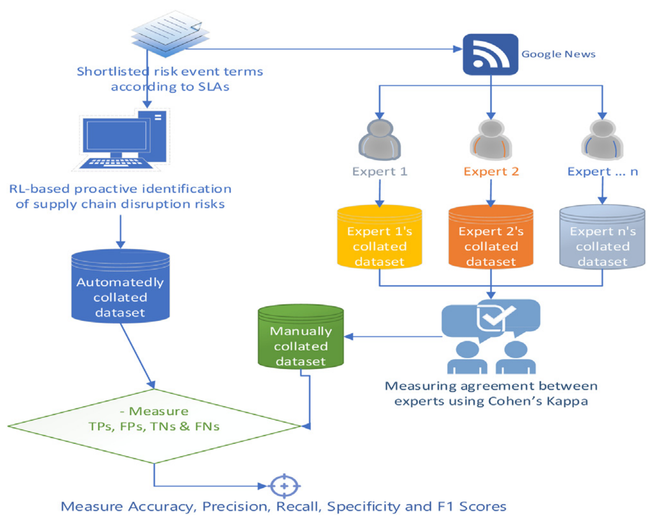


Fig. 4. Evaluation process to test the accuracy of the RL-PRI with the manual approach.

experts, the value of this metric will be 0, as Google shows all the news related to the risk factor of interest. For RL-PRI, this shows the number of news articles that were identified as relevant by the experts in the manual process but not recommended by it.

- True Negative (TN) → also known as *irrelevant* and *not showing*. Again, for the news dataset manually collated by the experts, the value of this metric will be 0, as Google shows all the news related to the risk factor of interest. For RL-PRI, this shows the number of news articles which it did not show to the risk manager compared to the manual process as they were irrelevant to their SLAs.

Next, the a, p, r, s, and $F1$ scores are determined using Eqs. (16) to (20).

$$a = \frac{tp + tn}{tp + fp + fn + tn} \quad (16)$$

$$p = \frac{tp}{tp + fp} \quad (17)$$

$$r = \frac{tp}{tp + fn} \quad (18)$$

$$s = \frac{tn}{tn + fp} \quad (19)$$

$$F1 = \frac{2pr}{p + r} \quad (20)$$

The working of this step is further explained in Section 4.3. Fig. 4 shows the schematic representation of the evaluation process.

4.1. Defining the scope of the analysis process

The defined scope for the analysis process is as follows:

Table 12
Number of news articles recommended by RL-PRI for each risk event.

Tarif dispute	Recession	Sea level rise	Flood	Total
52	116	30	246	444

ABC is a company based in Australia and in addition to its national market, it exports goods to China. From Q1 of 2020, the trade relations between Australia and China deteriorated. Furthermore, due to the economic crisis during the COVID-19 situation, there is a probability of a recession in the Australian economy. In addition, as a result of La Nina developing across the Pacific Ocean, there is a very high chance of floods occurring in Australia resulting in climate change.

The risk manager of the ABC company wants to proactively manage the risks arising from such disruptions and uncertainties. So, they identified the terms 'tariff dispute', 'recession', 'sea level rise' and 'flood' as the risk event terms from the Cambridge Taxonomy of Business Risks that will negatively impact its operations.

4.2. Collecting the news articles related to the risk events manually and automatedly

In this step, we collect the news articles related to the risk terms of 'tariff dispute', 'recession', 'sea level rise' and 'flood' manually and automatedly. For the manual collection, we use Google to search for the terms. The search process returned a total number of 1148 news articles. Table 10 shows the news articles returned for each risk term.[1] Each of these news articles was then analysed by two experts to confirm if it is relevant to the considered disruption risk events or not. Each expert was asked to read each of the news articles and give them a score of either 0 or 1, along with an explanation. A score of 0 signifies that the expert does not consider the news article as relevant to ABC's operations, and 1 signifies otherwise. This process took 240 min for expert 1 and 300 min for expert 2. Table 11 shows the scoring statistics by the experts.

To determine the level of agreement among the experts, Cohen's kappa coefficient is calculated using Eq. (21) as discussed in [54].

$$k = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e} \quad (21)$$

where p_o is the observed agreement among experts, and p_e is the hypothetical probability of random agreement. Based on Eq. (21), the kappa co-efficient k was determined as 1, which shows almost perfect agreement among the experts. For the automated data collection, we used RL-PRI, which took less than one minute to retrieve a total number of 444 articles. Table 12 shows the number of news articles returned for each risk term by RL-PRI.

[1] It is essential to note the disparity in terms of the sample size shown by Google for a search term. On page 1 of the search results, Google reports an inflated number of news articles matching the search term. But when we go to the last page of the search results, the number of news articles reduces drastically and differs from what is shown on page 1 of the search results. Furthermore, the search results for a term vary according to the date when it was searched.

Table 13
Scoring statistics by the experts for RL-PRI output.

Scoring metrics	Tarif dispute	Recession	Sea level rise	Flood
Both experts score 1 for the news articles	43	89	23	223
Both experts score 0 for the news articles	9	27	5	21
Total number of news articles ranked as 1 by expert 1 and 0 by expert 2	0	0	0	0
Total number of news articles ranked as 1 by expert 2 and 0 by expert 1	0	0	2	2

Table 14
Number of TP, FP, FN, and TN for the risk events in the manual search.

Tarif dispute		Recession		Sea level rise		Flood	
Category	Total number	Category	Total number	Category	Total number	Category	Total number
TP	46	TP	91	TP	26	TP	250
FP	159	FP	224	FP	254	FP	98
FN	0	FN	0	FN	0	FN	0
TN	0	TN	0	TN	0	TN	0

Table 15
Number of TP, FP, FN, and TN for the risk events in RL-PRI.

Tarif dispute		Recession		Sea level rise		Flood	
Category	Total number	Category	Total number	Category	Total number	Category	Total number
TP	43	TP	89	TP	23	TP	223
FP	9	FP	27	FP	7	FP	23
FN	3	FN	2	FN	3	FN	27
TN	150	TN	197	TN	247	TN	75

Table 16
a, p, r, s, and $F1$ scores from the manual search.

Tarif dispute		Recession		Sea level rise		Flood	
Metric	Score	Metric	Score	Metric	Score	Metric	Score
a	0.22439	a	0.288889	a	0.092857	a	0.718391
p	0.22439	p	0.288889	p	0.092857	p	0.718391
r	1	r	1	r	1	r	1
s	0	s	0	s	0	s	0
$F1$	0.366534	$F1$	0.448276	$F1$	0.169935	$F1$	0.83612

Table 17
a, p, r, s, and $F1$ scores from RL-PRI.

Tarif dispute		Recession		Sea level rise		Flood	
Metric	Score	Metric	Score	Metric	Score	Metric	Score
a	0.941463	a	0.907937	a	0.964286	a	0.856322
p	0.826923	p	0.767241	p	0.766667	p	0.906504
r	0.934783	r	0.978022	r	0.884615	r	0.892
s	0.943396	s	0.879464	s	0.972441	s	0.765306
$F1$	0.877551	$F1$	0.859903	$F1$	0.821429	$F1$	0.899194

Table 18
a, p, r, s, and $F1$ average scores for the manual search and RL-PRI.

Manual		RL-PRI	
Metric	Score	Metric	Score
a	0.331	a	0.917
p	0.331	p	0.817
r	1	r	0.922
s	0	s	0.890
$F1$	0.455	$F1$	0.864

Similar to the manual process, these news articles were then analysed by the two experts to confirm if they are relevant to the considered disruption risk events or not. Table 13 shows the scoring analysis of the experts on which Cohen's kappa agreement was determined as equal to 1 using Eq. (21).

In the next step, we compare the *relevancy* of RL-PRI's output and evaluate the *correctness* of the articles recommended by RL-PRI and the manually collected ones.

4.3. Measuring the relevancy and evaluating the correctness of the articles recommended by each approach

The Q-value of each article determined by RL-PRI shows the *relevancy* of each recommended news article to the risk manager's operations. To measure *Correctness*, Tables 14 and 15 show the number of TP, FP, FN, and TN for the news articles for both the manual and automated collections from RL-PRI, respectively.

The accuracy, precision, recall, specificity, and $F1$ scores for each risk event obtained in the manual search compared to the RL-PRI is calculated using Eqs. (16) to (20), as shown in Tables 16 and 17, respectively.

The average accuracy, precision, recall, specificity, and $F1$ scores of the analysis for all the four risk events are shown in Table 18. From the analysis, it can be seen that RL-PRI achieves very high accuracy, precision, specificity, and F1 scores for recommending important news articles to the risk manager for proactive risk identification. The recall value of the manual process is high because it recommends all the news related to risk events, but of these, there are a lot of irrelevant news which RL-PRI does not show to the risk manager, thereby it has a high specificity score compared to the manual process. Furthermore, the time it takes for RL-PRI to recommend these articles is less than one minute compared to 540 min taken by the two experts in the manual process. This shows the benefits and improvements of our proposed approach and the advantage it gives to the risk manager to proactively and quickly identify the risks to their operations compared to undertaking this manually. Based on these outputs, the risk manager performs Step 8 in Fig. 2, i.e., to further analyse the severity of these risk events impacting their operations. As previously stated, this step is not within the paper's scope and hence is not discussed.

5. Conclusion and future work

In this paper, we proposed RL-PRI, which is an RL-based approach that assists risk managers in the proactive identification of disruption risk events in their operations. We demonstrated the accuracy of RL-PRI against the manual approach currently undertaken by risk managers to identify disruption risk events. The experiments show the effectiveness of the RL-PRI approach in risk identification using artificial intelligence techniques to facilitate the efficacy of supply chain operations. The applicability of the proposed approach is not just limited to either disruption risks or the domain of supply chains but can also be applied in areas such as the stock market to identify the relevant events that may affect stock prices. While the results from RL-PRI are significant, there are many improvements that can be made in each of its modules to further increase its performance. These areas of further improvements, which are our future work, are as follows:

- In Module 1, we utilised the Cambridge Taxonomy of Business Risks to identify the disruption risk events, however a broader taxonomy can be investigated to detect other types of risk events.
- Module 2 can be improved by expanding the number of news sources, especially local news sources from the location of interest. Another source of information that can be considered for more extensive exploration is social media. In addition, this module can be further developed by applying semantic analysis to remove ineffective information such as multiple articles with the same meaning but different wording.
- Module 3 can be improved by utilising appropriate NLP APIs for the more precise extraction of entities. This will increase the performance of the RL module.
- In Module 4, different RL algorithms can be used to select the best possible action in each state. For example, to speed up the learning process and to perform in a more complex environment, methods such as the Deep Deterministic Policy Gradient (DDPG) [36] and Asynchronous Advantage Actor–Critic (A3C) [35] can be implemented.

CRediT authorship contribution statement

Hamed Aboutorab: Conceptualization of the idea, Formal analysis, Investigation, Data curation, Development, Completion of all the technical parts of the paper, Initial draft version and Finalisation of the manuscript. **Omar K. Hussain:** Co-conceptualization of the idea, Led the project in setting its scope. He also assisted in supervision, completing the experiments part and in the overall writeup of the manuscript. **Morteza Saberi:** Assisted in the RL-based aspects of the paper along with the experiment setup stage. Review of the paper write up and feedback towards finalization. **Farookh Khadeer Hussain:** Overall project guidance, Final draft review and feedback, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

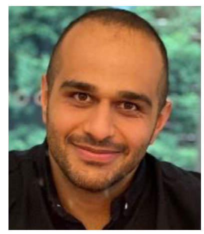
Acknowledgements

The first author acknowledges the financial support received from The University of New South Wales, Australia for this work. This research was supported partially by the Australian Government through the Australian Research Council's Linkage Projects funding scheme (project LP160100080).

References

[1] H. Aboutorab, O.K. Hussain, M. Saberi, F. Hussain, E. Chang, A survey on the suitability of risk identification techniques in the current networked environment, J. Netw. Comput. Appl. (2021) 102984.
[2] J. Nan, Z.H. Jia, Supply risk identification and analysis of iron and steel enterprises supply chain based on social system theory, 2008, pp. 1–4, http://dx.doi.org/10.1109/WiCom.2008.2470.
[3] J. Gao, Z. Wang, Research on real estate supply chain risk identification and precaution using scenario analysis method, in: IE and EM 2009 - Proceedings 2009 IEEE 16th International Conference on Industrial Engineering and Engineering Management, 2009, pp. 1279–1284, http://dx.doi.org/10.1109/ICIEEM.2009.5344449.
[4] D. Neiger, K. Rotaru, L. Churilov, Supply chain risk identification with value-focused process engineering, J. Oper. Manage. 27 (2009) 154–168, http://dx.doi.org/10.1016/j.jom.2007.11.003.
[5] A. Adhitya, R. Srinivasan, I.A. Karimi, Supply chain risk identification using a HAZOP-based approach, AIChE J. 55 (6) (2009) 1447–1463, http://dx.doi.org/10.1002/aic.11764.
[6] A. Mody, Improving the Risk Identification Process for a Global Supply Chain, Massachusetts Institute of Technology, 2012.
[7] H. Wang, X. Gao, S. Wang, Research on risk identification and control of medicine supply chain, in: Proceedings of 2012 3rd International Asia Conference on Industrial Engineering and Management Innovation (IEMI2012), 2013, pp. 367–377.
[8] X.H. Zhao, P.P. Li, X. Duan, X.Y. Li, Supply chain risk identification based on state space, in: Advanced Materials Research, Vol. 915, 2014, pp. 1495–1499.
[9] N.E. Yeboah, Y. Feng, O.-S. Daniel, N.B. Joseph, Agricultural supply chain risk identification-a case finding from ghana, J. Manage. Strateg. 5 (2) (2014) 31.
[10] C. Yu, L.H. Huatuco, Supply chain risk management identification and mitigation: A case study in a Chinese dairy company, Smart Innov. Syst. Technol. 52 (2016) 475–486, http://dx.doi.org/10.1007/978-3-319-32098-4_41.
[11] Y. Juan, H. Liu, F. Xiao, Supply chain hazards identification and a three-dimensional risk matrix model for food safety, Rev. Ibérica Sist. Tecnol. Inf. (E11) (2016) 420–432, [Online]. Available: https://login.wwwproxy1.library.unsw.edu.au/login?qurl=https%3A%2F%2Fsearch.proquest.com%2Fdocview%2F1861825360%3Faccountid%3D12763.
[12] A. Salamai, M. Saberi, O. Hussain, E. Chang, Risk identification-based association rule mining for supply chain big data, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), in: LNCS, vol. 11342, 2018, pp. 219–228, http://dx.doi.org/10.1007/978-3-030-05345-1_18.
[13] D.-W. Kwak, V.S. Rodrigues, R. Mason, S. Pettit, A. Beresford, Risk interaction identification in international supply chain logistics, Int. J. Oper. Prod. Manag. 38 (2) (2018) 372–389, http://dx.doi.org/10.1108/IJOPM-03-2016-0121.
[14] L. Lu, R. Zhou, R. De Souza, Enhanced FMEA for supply chain risk identification, road to a digit, Supply Chain Manag. 25 (2018) 311–330, http://dx.doi.org/10.15480/882.1783.
[15] D. Sugiarto, A. Ariwibowo, I. Mardianto, D. Surjasa, Identification of rice supply chain risk to DKI Jakarta through Cipinang primary rice market, E ES 106 (1) (2018) 12013.
[16] Z. Luan, C. Xie, L. Duan, Z. Wang, Y. Xiong, Risk identification based on OEM supply chain, 2009, pp. 1–4, http://dx.doi.org/10.1109/ICMSS.2009.5303508.
[17] H. Huo, J. Zhang, Research on retail enterprise supply chain risk identification based on SCOR, in: MSIE 2011, 2011, pp. 1302–1305.
[18] J.R. Ríos, D.F.M. Duque, J.C.O. Gómez, Operational supply chain risk identification and prioritization using the SCOR model, Ing. Univ. 23 (1) (2019) http://dx.doi.org/10.11144/Javeriana.iyu23-1.oscr.
[19] I.P. Tama, R. Yuniarti, A. Eunike, I. Hamdala, W. Azlia, Risk identification in cassava chip supply chain using SCOR (Supply Chain Operation Reference), IOP Conf. Ser.: Mater. Sci. Eng. 494 (1) (2019) http://dx.doi.org/10.1088/1757-899X/494/1/012050.
[20] Q. Zhu, L. Liu, Y. He, Application of process analysis based on value objective improvement in risk identification of supply chain, in: Proceedings - 2019 Chinese Automation Congress, CAC 2019, 2019, pp. 4213–4218, http://dx.doi.org/10.1109/CAC48633.2019.8996166.
[21] T. Britton, Difference between reactive, predictive and proactive risk management in aviation SMS, 2019.
[22] V. Mnih, et al., Playing atari with deep reinforcement learning, 2013, arXiv Prepr. arXiv:1312.5602.
[23] M.G. Bellemare, Y. Naddaf, J. Veness, M. Bowling, The arcade learning environment: An evaluation platform for general agents, J. Artificial Intelligence Res. 47 (2013) 253–279.
[24] J. Oh, X. Guo, H. Lee, R.L. Lewis, S. Singh, Action-conditional video prediction using deep networks in atari games, in: Advances in Neural Information Processing Systems, 2015, pp. 2863–2871.

[25] J. Schulman, S. Levine, P. Abbeel, M. Jordan, P. Moritz, Trust region policy optimization, in: International Conference on Machine Learning, 2015, pp. 1889–1897.
[26] V. Mnih, et al., Human-level control through deep reinforcement learning, Nature 518 (7540) (2015) 529–533.
[27] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, N. Freitas, Dueling network architectures for deep reinforcement learning, in: International Conference on Machine Learning, 2016, pp. 1995–2003.
[28] A.A. Rusu, M. Večerík, T. Rothörl, N. Heess, R. Pascanu, R. Hadsell, Sim-to-real robot learning from pixels with progressive nets, in: Conference on Robot Learning, 2017, pp. 262–270.
[29] S. Levine, C. Finn, T. Darrell, P. Abbeel, End-to-end training of deep visuomotor policies, J. Mach. Learn. Res. 17 (1) (2016) 1334–1373.
[30] B. Depraetere, M. Liu, G. Pinte, I. Grondman, R. Babuška, Comparison of model-free and model-based methods for time optimal hit control of a badminton robot, Mechatronics 24 (8) (2014) 1021–1030.
[31] S. Wang, P. Ren, Z. Chen, Z. Ren, J. Ma, M. de Rijke, Order-free medicine combination prediction with graph convolutional reinforcement learning, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1623–1632.
[32] A. Jonsson, Deep reinforcement learning in medicine, Kidney Dis. 5 (1) (2019) 18–22, http://dx.doi.org/10.1159/000492670.
[33] R.S. Sutton, A.G. Barto, Introduction to Reinforcement Learning, Vol. 135, MIT press, Cambridge, 1998.
[34] L.P. Kaelbling, M.L. Littman, A.W. Moore, Reinforcement learning: A survey, J. Artificial Intelligence Res. 4 (1996) 237–285.
[35] V. Mnih, et al., Asynchronous methods for deep reinforcement learning, in: International Conference on Machine Learning, 2016, pp. 1928–1937.
[36] T.P. Lillicrap, et al., Continuous control with deep reinforcement learning, 2015, arXiv Prepr. arXiv:1509.02971.
[37] X. Wang, Y. Wang, D. Hsu, Y. Wang, Exploration in interactive personalized music recommendation: a reinforcement learning approach, ACM Trans. Multimed. Comput. Commun. Appl. 11 (1) (2014) 1–22.
[38] X. Zhao, L. Zhang, L. Xia, Z. Ding, D. Yin, J. Tang, Deep reinforcement learning for list-wise recommendations, 2017, arXiv Prepr. arXiv:1801.00209.
[39] I. Munemasa, Y. Tomomatsu, K. Hayashi, T. Takagi, Deep reinforcement learning for recommender systems, in: 2018 International Conference on Information and Communications Technology (Icoiact), 2018, pp. 226–233.
[40] S. Zhang, L. Yao, A. Sun, Y. Tay, Deep learning based recommender system: A survey and new perspectives, ACM Comput. Surv. 52 (1) (2019) 1–38.
[41] S. Zhou, et al., Interactive recommender system via knowledge graph-enhanced reinforcement learning, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 179–188.
[42] A. Oroojlooyjadid, M. Nazari, L.V. Snyder, M. Takáč, A deep Q-network for the beer game: Deep reinforcement learning for inventory optimization, Manuf. Serv. Oper. Manag. (2021).
[43] A. Kara, I. Dogan, Reinforcement learning approaches for specifying ordering policies of perishable inventory systems, Expert Syst. Appl. 91 (2018) 150–158.
[44] M.H.F. Zarandi, S.V. Moosavi, M. Zarinbal, A fuzzy reinforcement learning algorithm for inventory control in supply chains, Int. J. Adv. Manuf. Technol. 65 (1–4) (2013) 557–569.
[45] C.-F. Chien, Y.-S. Lin, S.-K. Lin, Deep reinforcement learning for selecting demand forecast models to empower Industry 3.5 and an empirical study for a semiconductor component distributor, Int. J. Prod. Res. 58 (9) (2020) 2784–2804.
[46] R.U. Bilsel, S.R.T. Kumara, A reinforcement learning approach for dynamic supplier selection, in: 2007 IEEE International Conference on Service Operations and Logistics, and Informatics, 2007, pp. 1–6.
[47] S. Barat, et al., Reinforcement learning of supply chain control policy using closed loop multi-agent simulation, in: International Workshop on Multi-Agent Systems and Agent-Based Simulation, 2019, pp. 26–38.
[48] S. Barat, et al., Actor based simulation for closed loop control of supply chain using reinforcement learning, in: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, 2019, pp. 1802–1804.
[49] S. Yang, Y. Ogawa, K. Ikeuchi, Y. Akiyama, R. Shibasaki, Firm-level behavior control after large-scale urban flooding using multi-agent deep reinforcement learning, in: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation, 2019, pp. 24–27.
[50] A. Coburn, D. Ralph, M. Tuveson, S. Ruffle, G. Bowman, A taxonomy of threats for macro-catastrophe risk management, in: Cent. Risk Stud. Cambridge Univ. Cambridge, Work. Pap. July, 2013, pp. 20–24.
[51] SpaCy · industrial-strength natural language processing in Python, 2021, https://spacy.io/ (accessed Jan. 06, 2021).
[52] A.D. Tijsma, M.M. Drugan, M.A. Wiering, Comparing exploration strategies for Q-learning in random stochastic mazes, in: 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 2016, pp. 1–8, http://dx.doi.org/10.1109/SSCI.2016.7849366.
[53] Email delivery service | SendGrid, 2021, https://sendgrid.com/ (accessed Jan. 07, 2021).
[54] M.L. McHugh, Interrater reliability: the kappa statistic, Biochem. Med. Biochem. Med. 22 (3) (2012) 276–282.



Hamed Aboutorab is currently a Ph.D. candidate at the University of New South Wales, Canberra. His research is in data science and currently focusing on the application of Reinforcement learning in supply chain risk management. His research work has been published in reputable international journals such as Expert Systems with Applications and Journal of Network and Computer Applications.

Omar Hussain is an Associate Professor at the University of New South Wales, Canberra. His research interests are in business intelligence, cloud computing and logistics informatics. In these areas, his research work focuses on utilising decision making techniques for facilitating smart achievement of business outcomes. His research work has been published in various top international journals such as Information Systems, The Computer Journal, Knowledge Based Systems, Future Generation Computer Systems, etc. He has won awards and funding from competitive bodies such as the Australian Research Council for his research.

Morteza Saberi is a Lecturer with School of Information, Systems and Modelling, University of Technology Sydney, Australia and has an outstanding research records and significant capabilities in the area of Business Intelligence, Data Mining and Applied Machine Learning. He has published more than 150 papers in reputable academic journals and conference proceedings. His Google Scholar citations and h-index are 2210 and 22 respectively. He was a Lecturer at the Department of Industrial Engineering at University of Tafresh. He is also the recipient of the 2006–2011 Best Researcher of Young Researchers Club, Islamic Azad University (Tafresh Branch). He is also the recipient of National Eminent Researcher Award among Young Researchers Club, Islamic Azad University members.

Dr. Farookh Khadeer Hussain is an Associate Professor in School of Computer Science, University of Technology Sydney. He is an Associate Member of the Advanced Analytics Institute and a Core Member of the Centre for Artificial Intelligence. His key research interests are in trust-based computing, cloud of things, blockchains and machine learning. He has published widely in these areas in top journals such as FGCS, The Computer Journal, JCSS, IEEE Transactions on Industrial Informatics, IEEE Transactions on Industrial Electronics etc.