

Reasoning and Learning Services for Coalition Situational Understanding

Dan Harborne^a, Ramya Raghavendra^b, Chris Willis^c, Supriyo Chakraborty^b, Pranita Dewan^b,
Mudhakar Srivatsa^b, Richard Tomsett^d, and Alun Preece^a

^aCrime and Security Research Institute, Cardiff University

^bIBM T. J. Watson Research Center

^cBAE Systems AI Labs

^dIBM Emerging Technology, Hursley, UK

ABSTRACT

Situational understanding requires an ability to assess the current situation and anticipate future situations, requiring both pattern recognition and inference. A coalition involves multiple agencies sharing information and analytics. This paper considers how to harness distributed information sources, including multimodal sensors, together with machine learning and reasoning services, to perform situational understanding in a coalition context. To exemplify the approach we focus on a technology integration experiment in which multimodal data — including video and still imagery, geospatial and weather data — is processed and fused in a service-oriented architecture by heterogeneous pattern recognition and inference components. We show how the architecture: (i) provides awareness of the current situation and prediction of future states, (ii) is robust to individual service failure, (iii) supports the generation of ‘why’ explanations for human analysts (including from components based on ‘black box’ deep neural networks which pose particular challenges to explanation generation), and (iv) allows for the imposition of information sharing constraints in a coalition context where there is varying levels of trust between partner agencies.

Keywords: Situational Understanding, Artificial Intelligence for Situational Understanding, Machine Learning Interpretability

1. INTRODUCTION

This paper considers the situational understanding problem in the context of a coalition operation, where a coalition is taken to mean a temporary alliance of partners formed to achieve some common goal or combined effect, such as a coalition of military forces engaged in a shared mission or multiple relief agencies participating in a response to a major emergency. There is often a time-bounded need to form a coalition, to agree common goals, and to establish policies for command, control, and sharing of information and resources, including data and network services. Considering the problem of coalition situational understanding (CSU), the goals of a coalition are contingent on gaining and maintaining models of an environment and events, drawing on shared information and intelligence sources. Addressing the CSU problem involves being able to exploit a diverse collection of information processing services operating on multimodal data sources and distributed among the coalition partners, to generate and share high-level intelligence products from lower-level data. Additionally, CSU fundamentally involves temporal information processing, as well as requirements for the human to be in the loop.

Situational understanding (SU) is commonly defined as the ‘product of applying analysis and judgment to the unit’s situation awareness to determine the relationships of the factors present and form logical conclusions concerning threats to the force or mission accomplishment, opportunities for mission accomplishment, and gaps in information’.¹ UK military doctrine² defines *understanding* in the following terms:

Further author information: (Send correspondence to Dan Harborne)

Dan Harborne: E-mail: harbored@cardiff.ac.uk

Comprehension (insight) = situational awareness and analysis

Understanding (foresight) = comprehension and judgement

Here, understanding includes *foresight*, i.e., an ability to infer (predict) potential future states, while also being compatible with the common definition above that SU involves being able to draw conclusions concerning threats.

The coalition operations information fusion environment is characterised by its highly dynamic nature, a need for effective information and asset sharing, and operational constraints — including constraints on information sharing — dictated by organisational and mission policies.³

The paper is structured as follows: In Section 3 we outline, as a motivating example, the task of traffic congestion detection that we have been using to explore the CSU problem. Sections 4 and 5 give a summary of two pieces of previous work in the domain of traffic congestion detection that, taken together, offer a coalition system for the detection of current traffic congestion (insight) and models for the prediction of future congestion (foresight) respectively. Section 6 outlines the characteristics and benefits of a coalition systems approach to congestion detection, including the benefits of integrating the models from Section 5 into the system detailed in Section 4. Finally, Section 7 concludes the paper.

2. SITUATIONAL UNDERSTANDING WITHIN THE COALITION CONTEXT

Drawing on previous work,⁴ we identify a set of attributes for the CSU problem as follows:

Level of Understanding A CSU problem will generally span multiple levels of understanding from low to high, in terms of the kinds of semantic entities and relationships considered. For example, at lower levels, CSU may involve the detection, identification and localization of objects such as buildings, people, or vehicles. At higher levels, CSU may involve determining threats, behaviours, or anomalies.

Distributed The CSU problem is distributed in nature, fundamentally due to the structure of a coalition where each coalition partner will typically have data sources and processing services, together with constraints on how those data and resources can be shared with other partners. Moreover, these distributed resources (services and data) will be distributed across the network.

Temporal As noted above, understanding is associated with foresight, requiring the capability to process information temporally. CSU may involve relatively short vs. relatively long timescales, ranging from events in the very near future (hours, minutes, seconds, or less) to events in the more distant future (hours, days, weeks, months).

Multimodal Data Typically, CSU involves the processing and fusion of multimodal data, including data produced by physical sensors (so-called *hard* data) and data originating from humans (usually textual *soft* data).

Human-in-the-Loop The human user needs to interact with a CSU system, generally in terms of setting requirements and preferences, and potentially providing input in terms that change the representation and reasoning of the corresponding CSU system — for example, by providing key information currently unknown to the CSU system. A key requirement is that a CSU system cannot be a ‘black box’ but must have a degree of transparency to users.

3. TRAFFIC CONGESTION AS A SITUATIONAL UNDERSTANDING PROBLEM

To explore CSU, we sought an application where open data from pre-existing services was readily available and where machine learning and reasoning services could plausibly operate at a range of semantic scales (see *Level of understanding* above). We found that the task of monitoring and predicting traffic congestion is such an application. In many cities, multiple organisations offer open sources of information. For example, in London, UK, Transport for London (TfL) offers an application programming interface (API) to view imagery and video from their traffic cameras*, while Open Street Maps (OSM) offers information about roads, e.g., speed limit†. In the US, similar data is made available by various city/state transportation departments‡.

It is worth noting that the task of detecting of traffic congestion could be considered a ‘solved problem’ in the sense that companies such as Google offer live traffic prediction available via their map-related services. However, we are approaching this task purely as a means to explore CSU and, as the data sources mentioned above (and others we have used) come from multiple organisations and offer data in various modalities, we can clearly see that detecting traffic congestion this way simulates the context of operating within a coalition. Furthermore, as we ensured that the system architecture we developed remained decoupled from the specific problem, our final system can be used as a template for approaching other CSU problems by simply exchanging the data sources and services related to the congestion space to ones relevant to alternative CSU problems.



Figure 1. Example Images: uncongested (left), borderline (middle) and congested (right)

4. DETECTING CURRENT CONGESTION USING AN INTERPRETABLE DISTRIBUTED SYSTEM

For our congestion detection work we implemented an instantiation of a conceptual system architecture introduced in.⁵ The structure of the system is shown in Figure 2. The figure shows information flowing from data sources through processing services to decision-support classifications. Dark grey arrows show the path information takes to produce congestion classifications (ratings) and light grey arrows show information flows that produce explanations of those ratings. There are two main service chains producing classifications: one via a CNN (only) and a second via a more complex composition of services that feed in to a reasoning service. We refer to the CNN as the *congestion classifier* and the latter as the *congestion reasoner*. The harnessing of multiple independent machine learning/reasoning-based methods is in line with the portfolio/ensemble approach to achieving system robustness.⁶ In this section we will briefly outline these two main service chains and in Section 6 we will discuss in further details the benefits and implications of the system as a whole with a particular focus on interpretability, robustness, and coalition information flow constraints. For a deeper outline of the methodology within each service and performance results, please refer to our previous work.⁷

4.1 CNN Congestion Classifier

The congestion classifier provides an end-to-end approach to identifying congestion, taking a traffic camera image as input and outputting a class probability of congestion. The model was developed using a transfer learning⁸

*<http://www.trafficdelays.co.uk/london-traffic-cameras/>

†<http://www.openstreetmap.org/>

‡<https://data.austintexas.gov/>

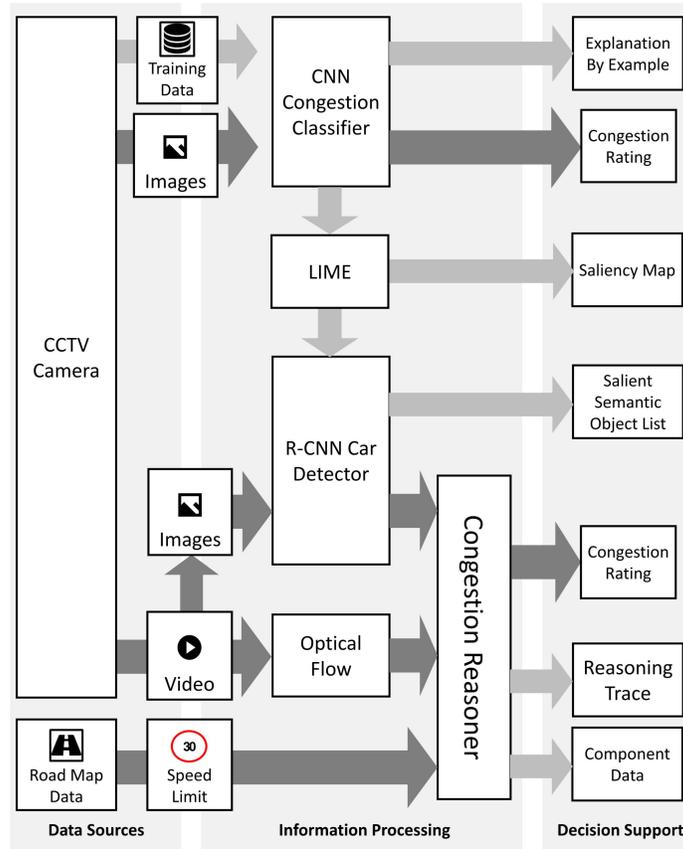


Figure 2. Testbed system architecture: dark grey arrows show information flow leading to congestion ratings; light grey arrows show information flow to generate explanations for ratings

approach, using the GoogLeNet⁹ Inception network, pre-trained on ImageNet data for feature extraction and passes the feature vector to a fully connected layer trained using our labeled traffic image dataset. During our testing of this approach, the recorded precision and recall are 98% and 96% respectively.

4.2 Congestion Reasoner

As shown in Figure 3, the congestion reasoner involves an integration of multiple services and data sources. We use an optical flow (OF) algorithm⁸ on traffic camera video footage to identify movement and then pass the detected blobs of movement to an object detector (a retrained instance of the VGG-16 regional-convolutional neural network, R-CNN, model¹⁰) which filters out any non-vehicle blobs. We can measure the pixel velocities of the vehicles and compare this to the historic trend and speed limit of the road to infer the level of congestion. From testing, the recorded precision and recall are 79% and 87% respectively.

5. FORESIGHT MODELS FOR FUTURE CONGESTION PREDICTION

In our approach to predicting future levels of traffic congestion for a given road, we aimed to predict future traffic speeds and the state of congestion. To do this, we utilise three inputs. The first is the current speed on the road of interest. The second is the current speeds of roads connected to the road of interest. Finally we use the anticipated state of the weather for the prediction time step.

⁸Specifically the implementation of Mixture of Gaussian (MOG2) and Lucas-Kanade found in OpenCV: <http://github.com/itseez/opencv>

For the inputs based on the current speed of a road, we divide the current speed by the historic average speed of the road. This allowed us to use the same model for all roads, even though roads can differ in speed limit and traffic volume trends.

During this work we first trained a regression tree to perform the speed prediction. This had the benefits of performing well (89% accuracy) whilst also being interpretable. We then trained a deep learning model, a LSTM, to make the predictions. This provided slightly improved accuracy but at the cost of losing interpretability. By making both models available within the coalition, this trade-off can be captured by the final system and this will be discussed in Section 6.1.

A final observation from this work was that due to the sparsity of data of traffic speeds in seriously adverse weather conditions, the accuracy of both models when predicting speeds under such conditions was poor. We trained dedicated models for making predictions during such conditions and saw a noticeable improvement (a drop in error of 33% when using the dedicated model over the standard weather conditions model). How the final system can take advantage of these focused models is discussed in section 6.2.

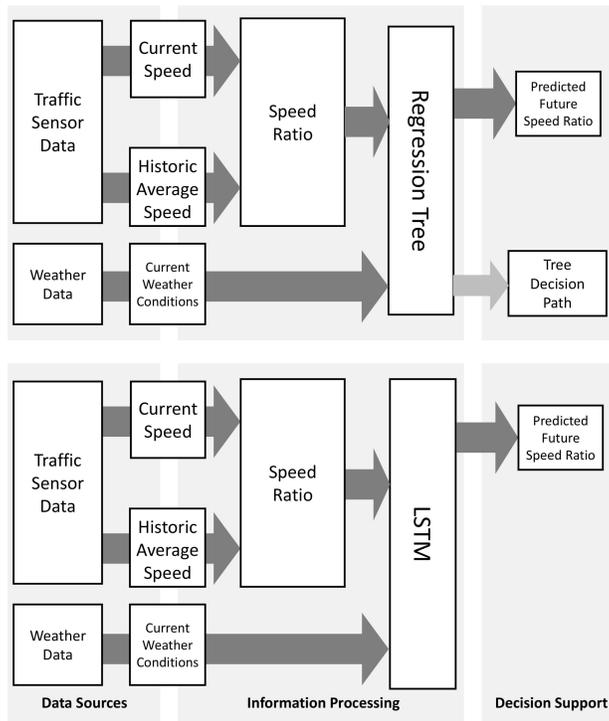


Figure 3. Foresight service chains. Dark grey arrows show information flow leading to predicted road speed; light grey arrows show information flow to generate explanations for predictions

6. INTERPRETABLE DISTRIBUTED SYSTEM FOR CONGESTION FORESIGHT AND DETECTION

In this section we will discuss the benefits and characteristics of utilizing our services as a distributed system including the changes and added benefits of integrating the foresight models to the previously-constructed congestion detection system.

6.1 Interpretability

The interpretability of a service or system refers to the level to which the agents using it (human or machine) can understand why a decision has been made.¹¹ Interpretability can help with troubleshooting identified errors and building user trust. The latter is of particular importance within a coalition setting, where partners can't

carry out their own direct evaluation of the reliability of a system or service. In this case, the reasoning behind a decision often must be supplied by an organisation in order for its partners to act on the intelligence being provided.

In previous work,⁷ we have focused on outlining the interpretability characteristics of the system and its component services. There, we show that by taking advantage of various interpretability techniques, a system can autonomously highlight any decision it has made that may require a human user to provide their judgment. This is particularly important when the volume of decisions and evaluations made by the system is high (e.g., when a system is making an assessment of congestion for every road in a city) and therefore a human user would be unable to review all decisions being made.

In our work, we first generate a saliency map¹² (shown in figure 4) for the decision made by the CNN congestion classifier. This map highlights regions of the input image that were important to the classifier decision. We then pass the highlighted regions to an Regional Convolutional Neural Network (R-CNN) object detector. By checking whether the detector finds semantically relevant objects to congestion (e.g. cars) within the salient regions, we can check the classifier has used features of the input that are sensible for approaching the given task. If the object detector fails to detect any semantically-relevant objects in the salient regions, the decision (along with various generated explanations) can be presented to a human agent for evaluation. In our system, the object detector used in this interpretability technique already existed within our coalition of services as it was created as part of the congestion reasoner service chain. This is a demonstration of how a system can be strengthened by the ability to take advantage of existing services across the coalition.

Finally, to allow for the system’s final decision to be as interpretable as possible, the output of interpretable services can help provide explanations in combination with the output from services which are harder to generate an explanation for. As an example, the system can take advantage of the higher accuracy of the LSTM speed prediction model (outlined in Section 5) and in cases where the regression tree service (outlined in the same section) agrees with the LSTM, it can help provide explanation for the outcome. This relationship is also seen between the CNN congestion detection service (Section 4.1) and the more interpretable congestion reasoner service chain (Section 4.2).



Figure 4. Original input image (left) and saliency map (right: red regions show evidence towards ‘congested’, green regions show evidence towards ‘not congested’)

6.2 Resilience and Robustness

By producing a system that takes advantage of services and data sources across the coalition, we can remove single points of failure and minimise the number of decision being made from information produced by a single service, improving system robustness.⁶ In this section we will highlight some instances where we have taken advantage of such redundancies and resiliences.

Firstly, as mentioned in the previous section (Section 5), we can train multiple models to accomplish the same task but within different conditions (snow conditions vs. no snow). Once these services are made available to the system, the system itself can autonomously make a decision as to which model is most appropriate. This decision can be made by a simple human-provided rule that takes in data about relevant features (e.g., whether it is snowing) or the decision boundary can be machine-learned over time by monitoring the predictions of both models, and comparing them to the eventual ground truth value. This, combined with the input of data about

the relevant feature(s), can allow the system to learn how to autonomously adapt the compositions of services it will use to answer questions.

Further to this, in cases where multiple models that accomplish the same task exist within the coalition but the relevant feature(s) that differentiate their reliability are not known, the system can simply use the outputs of both models to form a (more robust) final decision. In these cases, methods and policies can be placed within the system with regard to how these outputs are aggregated in cases where they don't agree. Some example factors that can be utilised are the associated accuracy of each service, the confidence scores returned with the outputs or simply a given preference for one service over the other. Like with input features, these features of the service or the output can be used within human agent defined rules or within learned rules to then aggregate these outputs into a final piece of information. Alternatively, when disagreement occurs between services, the system can take advantage of a human-in-the-loop at a low level, returning both the outputs and asking the user of the system to provide an assessment. This approach can again be governed by policy. In our instantiated system, this occurs between the congestion classifier model and the congestion reasoner service.

Finally, by bringing together the services available, opportunities to train new services are created that can help build further redundancy in to the system. For example, in order to infer the future levels of congestion, the regression tree service relies on a feed of current average speeds across the road network. If this feed fails, the service can no longer operate. However, available from within the coalition is the optical flow service that is monitoring the flow of traffic. Although this feed outputs different data, the semantics it captures are similar and the modality is also the same. Therefore, the training of a regression tree model to predict future optical flow values would be possible and can quickly develop a new regression tree service that provide an inference of future levels of congestion.

The resilience and redundancy outlined above helps protect the system against attacks tailored towards a particular service (for example, adversarial attacks¹³) on deep learning models). It helps the system to provide accurate information when one services shows signs of a particular weakness (such as the one of our CNN highlighted in the previous section) and finally can allow the system to continue to provide decision support when a particular service goes offline or receives too much demand to handle further requests.

6.3 Operating Under Coalition Information Flow Constraints

When building a system that takes advantage of data sources and services across a coalition, there are a number of information flow constraints that must be factored in. Firstly, there is an importance to adhere to privacy policies set in place by the partner organisations and indeed, have a system that adapts to the privileges of the user both in terms of the organisation they originate from and their position within it.

This can be particularly noticeable when providing explanations for the system's decisions, as many interpretability techniques return an output that reveal the input data (saliency maps are a clear example of this). Without offering an alternative to such interpretability techniques, it is likely that the explanations containing sensitive data would need to be redacted by the system before reaching the end user and thus, removing the ability for the user to trust the decision. In our system we present interpretability techniques that help remove the reliance on sensitive input data. One such method, as outlined in Section 6.1 takes a saliency map and detects semantically relevant objects within the salient regions. This list of detected objects can be returned as a text list removing the reliance on the input data but maintaining the ability for the user to assess whether the system has looked at features of the input image that are appropriate.

In addition to privacy policy constraints, a distributed system can suffer from bandwidth constraints, especially if the user is an agent operating at the tactical edge. Here again, we can use the above salient semantic object technique to change the explanation to a simple text output rather than an image. This can help minimise the bandwidth required to send the output to the agent. As an additional point, this can also help comply to modality constraints (for example if an agent is using a device that can't display imagery).

Finally, we previously outlined ways in which the system can provide redundancy to services across the coalition. This allows the system to deal with coalition factors such as a partner removing access to a service or imposing a usage limit.

7. CONCLUSION

The goal of this paper was to characterise the Coalition Situational Understanding problem — in terms of requiring *insight* and *foresight* with distributed, temporal, human-in-the-loop, and multimodal data requirements — and to outline how this problem can be addressed via a service-oriented approach building on an integration of our previous work. In particular, we have shown how the integrated approach (i) provides awareness of the current situation and prediction of future states (Sections 4 and 5), (ii) is robust to individual service failure (Section 6.2), (iii) supports the generation of ‘why’ explanations for human analysts (including from components based on ‘black box’ deep neural networks which pose particular challenges to explanation generation (Section 6.1), and (iv) allows for the imposition of information sharing constraints in a coalition context where there is varying levels of trust between partner agencies (Section 6.3).

In future work, we look to explore how interpretability techniques can be used to identify the features of misclassified images that are responsible for classification error. This can then help to select highly relevant retraining images to be labeled by humans that target the model’s weakness towards the identified features. Ultimately, the aim is to reduce the amount of retraining images required to remove the weakness from the model and thus decreasing the human resources required for the ground truth labeling effort. Another area of future work is to explore the effectiveness of various forms of interpretability on human-in-the-loop decision making.

ACKNOWLEDGMENTS

This research was sponsored by the U.S. Army Research Laboratory and the UK Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the UK Ministry of Defence or the UK Government. The U.S. and UK Governments are authorised to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

- [1] Dostal, B. C., “Enhancing situational understanding through the employment of unmanned aerial vehicles,” *Army Transformation Taking Shape... Interim Brigade Combat Team Newsletter* (01–18) (2007).
- [2] “Understanding: Joint Doctrine Publication 04 (JDP 04).” Ministry of Defence, UK (2010).
- [3] Pham, T., Cirincione, G., Verma, D., and Pearson, G., “Intelligence, surveillance, and reconnaissance fusion for coalition operations,” in [*Proc 11th International Conference on Information Fusion*], (2008).
- [4] Preece, A., Cerutti, F., Braines, D., Chakraborty, S., and Srivastava, M., “Cognitive computing for coalition situational understanding,” in [*First International Workshop on Distributed Analytics InfraStructure and Algorithms for Multi-Organization Federations*], (2017).
- [5] Nottle, A., Quintana-Amate, S., Harborne, D., Alzantot, M., Braines, D., Tomsett, R., Kaplan, L., Srivastava, M., Chakraborty, S., and Preece, A., “Distributed opportunistic sensing and fusion for traffic congestion detection,” in [*First International Workshop on Distributed Analytics InfraStructure and Algorithms for Multi-Organization Federations*], (2017).
- [6] Dietterich, T. G., “Steps toward robust artificial intelligence,” *AI Magazine* **38**(3), 3–24 (2017).
- [7] Harborne, D., Willis, C., Tomsett, R., and Preece, A., “Integrating learning and reasoning services for explainable information fusion,” in [*First International Conference on Pattern Recognition and Artificial Intelligence*], (2018).
- [8] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M., “Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning,” *IEEE Transactions on Medical Imaging* **35**, 1285–1298 (May 2016).
- [9] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., “Going deeper with convolutions,” in [*Computer Vision and Pattern Recognition (CVPR)*], (2015).
- [10] Simonyan, K. and Zisserman, A., “Very deep convolutional networks for large-scale image recognition,” *CoRR abs/1409.1556* (2014).

- [11] Lipton, Z. C., “The mythos of model interpretability,” in [*2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*], (2017).
- [12] Ribeiro, M. T., Singh, S., and Guestrin, C., “Why should I trust you?: Explaining the predictions of any classifier,” in [*International Conference on Knowledge Discovery and Data Mining (KDD)*], (2016).
- [13] Goodfellow, I. J., Shlens, J., and Szegedy, C., “Explaining and harnessing adversarial examples,” *arXiv:1412.6572* (2014).