

Formal Methods for the Railway Sector

by Maurice ter Beek, Alessandro Fantechi, Alessio Ferrari, Stefania Gnesi (ISTI-CNR, Italy), and Riccardo Scopigno (ISMB, Italy)

Researchers from the Formal Methods and Tools group of ISTI-CNR are working on a review and assessment of the main formal modelling and verification languages and tools used in the railway domain, with the aim of evaluating the actual applicability of the most promising ones to a moving block signalling system model provided by an industrial partner. The research is being conducted in the context of the H2020 Shift2Rail project ASTRail.

Compared with other transport sectors, the railway sector is notoriously cautious about adopting technological innovations. This is commonly attributed to the sector's robust safety requirements. An example is smart route planning: while GNSS-based positioning systems have been in use for quite some time now in the avionics and automotive sectors to provide accurate positioning and smart route planning, the current train separation system is still based on fixed blocks – a block being the section of the track between two fixed points. In signalling systems, blocks start and end at signals, with their lengths designed to allow trains to operate as frequently as necessary (i.e., ranging from many kilometres for secondary tracks to a few hundred metres for a busy commuter line). The block sizes are determined based on parameters such as the line's speed limit, the train's speed, the braking characteristics of trains, sighting and reaction time of drivers, etc. However, the faster trains are allowed to run,

the longer the braking distance and the longer the blocks need to be, thus decreasing the line's capacity. This is because stringent safety requirements impose the length of fixed blocks to be based on the worst-case braking distance, regardless of the actual speed of the train.

With a moving block signalling system, in contrast, a safe zone around the moving train can be computed, thus optimising the line's exploitation (Figure 1). For this solution to work, it requires the precise absolute location, speed and direction of each train, to be determined by a combination of sensors: active and passive markers along the track, as well as trainborne speedometers. One of the current challenges in the railway sector is to make moving block signalling systems as effective and precise as possible, including GNSS and leveraging on an integrated solution for signal outages (think, e.g., of tunnels) and the problem of multipaths [1]. This is one of the main topics addressed by the project ASTRail: SATellite-based Signalling and Automation SysTEms on Railways along with Formal Method and Moving Block Validation (Figure 2). A subsequent aim of the project is to study the possibility of deploying the resulting precise and reliable train localisation to improve automatic driving technologies in the railway sector.

We are currently reviewing and assessing the main formal modelling and verification languages and tools used in the railway domain in order to identify the ones that are most mature for application in the railway industry [2] [3]. This is done by combining a scientific literature review with interviews with stakeholders of the railway sector. This will shortly result in a classification and ranking, from which we will select a set of languages and tools to be adopted in all the relevant development stages of the systems addressed by ASTRail. In particular, we will formalise and validate a

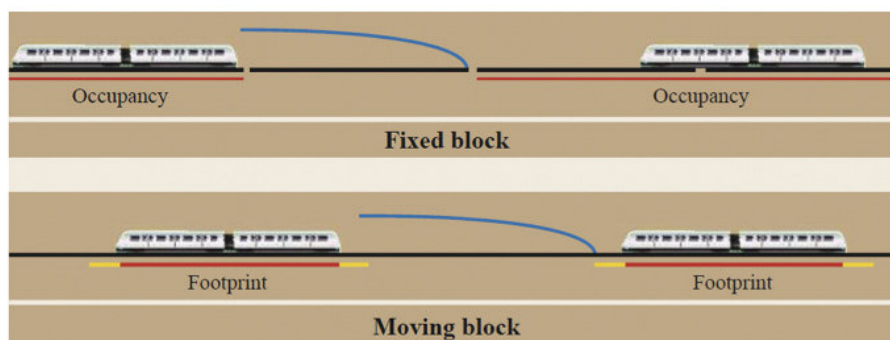


Figure 1: Safe braking distance between trains in fixed block and moving block signalling systems (Image courtesy of Israel.abad/Wikimedia Commons distributed under the CC BY-SA 3.0 license).

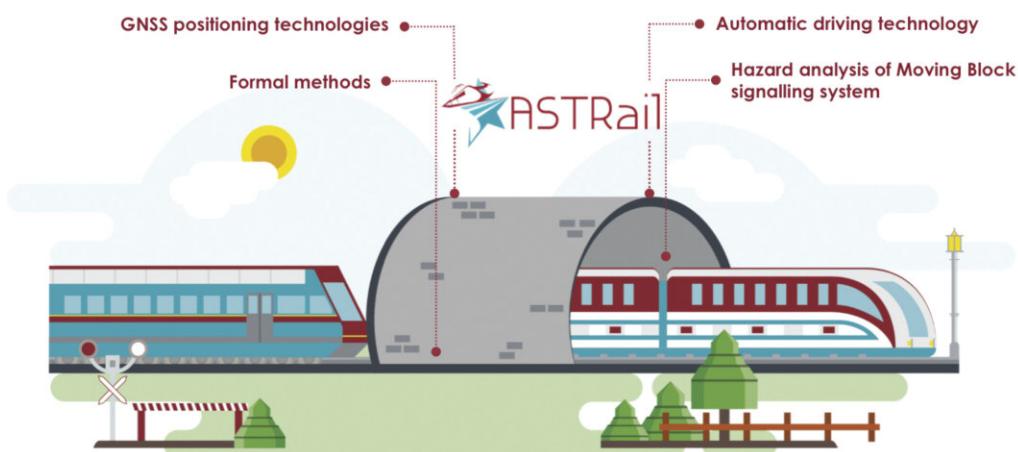


Figure 2: Illustrative summary of ASTRail's objectives.

moving block model developed by SIRTl, which is an industry leader in the design, production, installation and maintenance of railway signalling systems.

ASTRail will run until August 2019 and is coordinated by Riccardo Scopigno from the Istituto Superiore Mario Boella sulle Tecnologie dell'Informazione e delle Telecomunicazioni (ISMB, Italy). Other partners are SIRTl S.p.A. (Italy), Ardanuy Ingeniería S.A. (Spain), École Nationale de l'Aviation Civile (ENAC, France), Union des Industries Ferroviaires Européennes (UNIFE, Belgium) and ISTI-CNR (Italy).

Link:

ASTRail: <http://www.astrail.eu/>

References:

- [1] F. Rispoli, et al.: "Recent Progress in Application of GNSS and Advanced Communications for Railway Signaling", in 23rd Intl. Conf. Radioelektronika, IEEE, 2013, 13-22. DOI: 10.1109/RadioElek.2013.6530882
- [2] A. Fantechi, W. Fokkink, and A. Morzenti: "Some Trends in Formal Methods Applications to Railway Signaling", Chapter 4 in Formal Methods for Industrial Critical Systems: A Survey of Applications (S. Gnesi and T. Margaria, eds.). John Wiley & Sons, 2013, 61-84. DOI: 10.1002/9781118459898.ch4
- [3] A. Fantechi: "Twenty-Five Years of Formal Methods and Railways: What Next?", in Software Engineering and Formal Methods, LNCS, vol. 8368. Springer, 2013, 167-183. DOI: 10.1007/978-3-319-05032-4_13

Please contact:

Stefania Gnesi, ISTI-CNR, Italy
stefania.gnesi@isti.cnr.it

The Impacts of Low-Quality Training Data on Information Extraction from Clinical Reports

by Diego Marcheggiani (University of Amsterdam) and Fabrizio Sebastiani (CNR)

In a joint effort between the University of Amsterdam and ISTI-CNR, researchers have studied the negative impact that low-quality training data (i.e., training data annotated by non-authoritative assessors) has on information extraction (IE) accuracy.

Information Extraction (IE) is the task of designing software artifacts capable of extracting, from informal and unstructured texts, mentions of particular concepts, such as the names of people, organisations and locations – where the task usually goes by the name of "named entity extraction". Domain-specific concepts, such as drug names, or drug dosages, or complex descriptions of prognoses, are also examples.

Many IE systems are based on supervised learning, i.e., rely on training an information extractor with texts where mentions of the concepts of interest have been manually "labelled" (i.e., annotated via a markup language); in other words, the IE system learns to identify mentions of concepts by analysing what manually identified mentions of the same concepts look like.

It seems intuitive that the quality of the manually assigned labels (i.e., whether the manually identified portions of text are indeed mentions of the concept of interest, and whether their starting points and ending points have been identified precisely) has a direct impact on the accuracy of the extractor that results from the training. In other words, one would expect that learning from inaccurate manual labels will lead

Figure 1: Screenshot displaying a radiological report annotated according to the concepts of interest.

