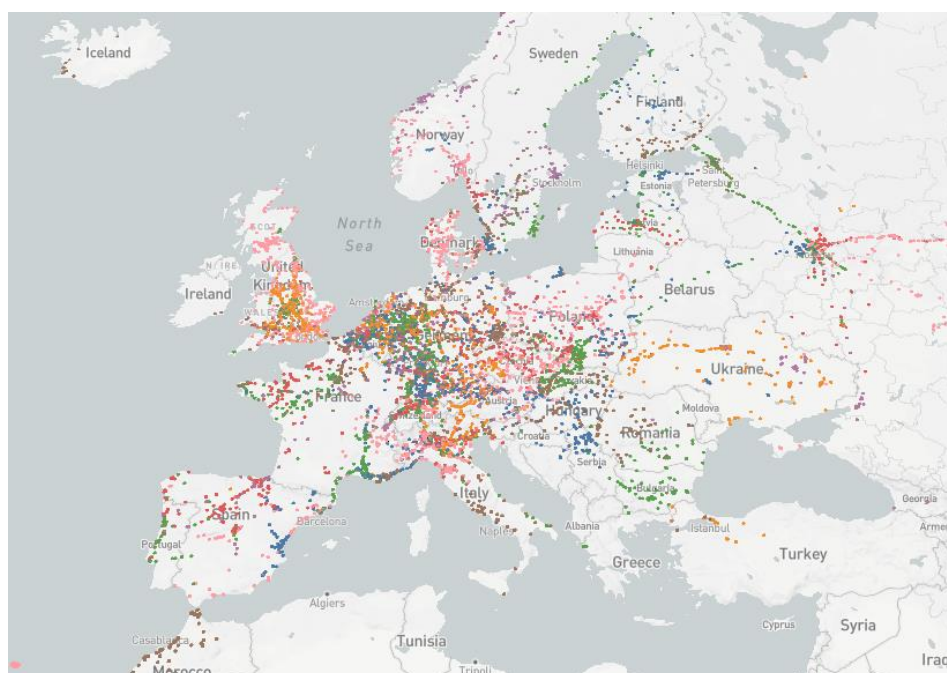


FUMA - Fleet Telematics Big Data Analytics for Vehicle Usage Modeling and Analysis

Publik rapport



Författare: Sara Sylvan och Emil Gustavsson
Datum: 2019-11-28
Projekt inom *BADA – Big Automotive Data Analytics*

FFI Fordonsstrategisk
Forskning och
Innovation

VINNOVA

Energimyndigheten

TRAFIKVERKET

FKG

VOLVO

SCANIA

VOLVO

Innehållsförteckning

1 Sammanfattning	3
2 Executive summary in English.....	4
3 Bakgrund.....	5
4 Syfte, forskningsfrågor och metod	6
4.1 Samarbetsplattform.....	6
4.2 Samarbetsmetod.....	7
5 Mål	7
5.1 Forskningsmål.....	7
5.2 Koppling till FFI och BADA mål.....	7
5.3 Mål för Scania.....	8
5.4 Mål för FCC	8
6 Resultat och måluppfyllelse	8
6.1 Resultat	8
6.2 Måluppfyllelse	16
6.3 Kunskaps- och resultatspridning.....	16
6.4 Publikationer	16
6.5 Övrig kunskapsspridning	17
7 Slutsatser och fortsatt forskning	17
8 Deltagande parter och kontaktpersoner.....	17

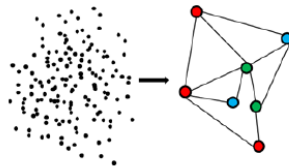
Kort om FFI

FFI är ett samarbete mellan staten och fordonsindustrin om att gemensamt finansiera forsknings- och innovationsaktiviteter med fokus på områdena Klimat & Miljö samt Trafiksäkerhet. Satsningen innebär verksamhet för ca 1 miljard kr per år varav de offentliga medlen utgör drygt 400 Mkr.

För närvarande finns fem delprogram; Energi & Miljö, Trafiksäkerhet och automatiserade fordon, Elektronik, mjukvara och kommunikation, Hållbar produktion och Effektiva och uppkopplade transportsystem. Läs mer på www.vinnova.se/ffi.

1 Sammanfattning

Forskningsprojektet *FUMA – Fleet Telematics Big Data Analytics for Vehicle Usage Modeling and Analysis* är ett samarbetsprojekt mellan Scania AB och Fraunhofer-Chalmers Centre (FCC) där huvudmålet för projektet har varit att utnyttja GPS-information från tunga fordon för att skapa kunskap kring fordons drift and användande. Scania har över 300 000 uppkopplade fordon som kontinuerligt skickar in sina positioner. Från den stora mängden av positioner som genereras av de uppkopplade fordonen har ett transportnätverk extraherats och fordonens beteende på nätverket har studerats.



Algoritmer har utvecklats som tar de råa GPS-positionerna från fordonen och extraherar positionerna där flertalet fordonsstopp har gjorts. För att garantera beräkningsmetoder som är hållbara över tid var målet i projektet att uteslutande använda strömmande beräkningar. Det innebär att nya punkter behandlas allt eftersom de skapas och resultat sparas. I detta forskningsprojekt har målsättningen varit att använda beräkningsmetoder som är moderna och i forskningsfronten samtidigt som det är lika viktigt att vi skapar information som är värdefull för industrin.

De extraherade stopplatserna definierar sedan den transportgraf som studerats i projektet. Fordonens övergångar mellan de olika stopplatserna/noderna i grafen beskriver i sin tur fordonens beteende, och utifrån rörelser på transportgrafan har beskrivningar av dess drift och användning definierats och implementerats. Resultatet var en beskrivning av ett fordons drift där man tydligt kan se skillnad mellan olika typer av fordon som man vet har olika drifter.

I senare delen av projektet analyserades även transportgrafan från ett mer övergripande perspektiv där de aggregerade rörelserna av fordonen studerades. De metoder som framförallt användes var algoritmer där målet är att upptäcka delgrafer som innehåller tätt sammankopplade noder (*eng* community detection).

Samarbetet mellan akademi (FCC) och industri utfördes genom täta avstämningar och gemensamt utvecklande vilket gjorde att kunskapsspridningen mellan projektparterna har varit stor. Samarbetsplattformen som användes under projektet var ett externt datacenter, ICE, utvecklat av RISE SICS där både Scania och FCC kunde implementera och köra kod.

Fyra stycken examensarbetare har examinerats som en del av projektet, där samtliga har varit anställda på Scania. Två artiklar har publicerats och resultaten från projektet har spridits genom publika konferenser samt interna workshops.

2 Executive summary in English

The research project *FUMA – Fleet Telematics Big Data Analytics for Vehicle Usage Modeling and Analysis* is a collaboration between Scania AB and Fraunhofer-Chalmers Centre (FCC). The aim of the project was to develop and implement algorithms that utilize GPS information from heavy vehicles in order to create knowledge regarding their usage mode and operations. Scania has over 200 000 connected vehicles that continuously send their GPS positions to Scania. From the large amount of positions that is generated a transport network was extracted and the vehicles behavior studied and analyzed.

Algorithms and methods were developed that takes the raw GPS positions from the vehicles and then extracts positions where multiple vehicle stops have been performed. In order to guarantee that the utilized algorithms are sustainable and usable for larger data sets, streaming methods have been the main focus.

The extracted stop positions define the transport graph analyzed in the project. The vehicles transitions between the nodes on the graph describe the vehicles behavior, and from the transitions descriptions of their usage modes has been defined and implemented. The results was a summary/description if the usage mode of a vehicle where one can observe clear differences between different type of vehicles.

In the later parts of the project the entire transport graph was analyzed from a more general perspective. Community detection methods was employed on the graph in order to identify communities of nodes/stop positions that are closely connected.

Purpose, research questions, and method

The purpose of the project was to utilize modern computational methods for drawing conclusions regarding the usage mode of vehicles by inly considering their GPS positions. The main research question in the project was how to utilize the vehicle positions in order to obtain knowledge about the transportation system. This knowledge can be utilized for identifying the usage mode of vehicles as well as understanding the entire transport network. Because of increasing amount of data from the connected vehicles, all algorithms and methods analyzed in the project needed to be scalable and applicable on very large data sets.

The project work was performed in close collaborations between the partners and the computational platform used for the project was a data center in Luleå supported by SICS RISE.

Objective

The overall objective of the project was to utilize the latest computational techniques in order to perform computations on the huge sets of GPS positions from Scania vehicles. This overall objective is divided into three more clearly defined aims:

- To analyze, develop, and implement large scale computational methods applicable to big data.
- To understand vehicle movements and to gain knowledge regarding the usage mode of the vehicles. When aggregating the movements of all vehicles knowledge about the entire transport network can also be obtained.
- To develop and implement modern visualization methods for large data sets.

Results

The work performed in the project was divided into six work packages with clear objectives and deliverables. A short summary of the results in each work package follows:

- ***WP1: Definition and extraction of network and features***

In WP1 streaming algorithms was developed that takes in raw GPS positions from the vehicles and returns positions where multiple vehicle stops has occurred. The methods developed are implemented in production at Scania.

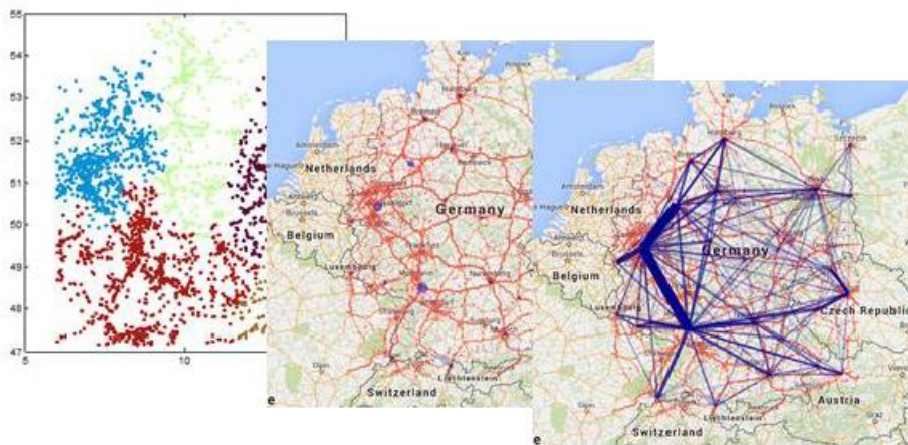
- **WP2: Network component characterization**
In WP2 the nodes/stop positions extracted was enriched with information about the vehicle stops and also with information from external data sources
- **WP3: Definition and characterization of trips**
In WP3 the vehicle transitions between the network nodes was extracted and enriched with more data. Transition cycles and n-grams (consecutive transitions) was also analyzed.
- **WP4: Usage mode classification**
In WP4 a definition of the usage mode of a vehicle was develop and implemented. Validation of the definition was performed on Transportlabets vehicles with good performance results.
- **WP5: Transport chains and networks**
In WP5 a more general analysis of the entire transport network was performed. Community detection methods was implemented which demonstrated that it is possible to extract communities of nodes that are closely connected to each other.
- **WP6: Demonstration software**
In WP6 a demonstrator software was developed by FCC in order to present and demonstrate the project results in interactive web based dashboards.

Four Master thesis projects has been performed as a part of the project, where all have been employed by Scania. Two scientific papers have been published and the results from the project has been disseminated on both public conferences and on internal workshops.

3 Bakgrund

Scanias fordon bygger upp en stor population av uppkopplade fordon som kontinuerligt skickar in sina positioner. I mars 2016 samlade Scania in data från över 200 000 fordon och det är uppskattat att upp till en miljon fordon kommer vara uppkopplade inom en nära framtid. Datan används för tillfället framförallt av individuella kunder för uppgifter relaterade till hantering och ledning av respektive kunds fordonsflotta såsom planering av service och realtidsövervakning av sina fordon. Den historiska datan börjar täcka mer och mer av karaktäristiken av transportsystemet och nya tekniker inom BigData kommer kunna användas för att dra slutsatser angående förarbeteende och transportnätverk. Enbart fordonspositioner visar inte hur fordonen används eller nyttjas utan det behövs en annan abstraktionsnivå där data har bearbetats innan det går att dra den typen av slutsatser.

I en förstudie [Dudas, C., Engström, C., Karlsson, J., Nellros, F., Qi-Gautier, L., Silvestrov, S., & Ulke, J. (2015). *Investigate graph and network algorithms in transport vehicle GPW data to detect and quantify hubs and flow. Swedish study group Mathematics in Industry*] utförd av FCC och Scania där man analyserade rörelserna av en uppsättning fordon lyckades man visa att den information som finns i positionerna som fordonen skickar kan användas för att förstå hur fordonet används. Genom att analysera förflyttningarna fordon gör mellan olika stopplatser (platser där flertalet stopp har gjort av fordon) kan man skapa en förståelse angående fordonets faktiska drift. En förståelse av ett fordonets drift, utifrån de rörelser som fordonet gör, kan till exempelvis användas för att utvärdera om en ägare använder rätt typ av fordon till de uppgifter de har, eller till att skapa sig en mer övergripande förståelse över hur Scanias och deras kunders fordon används.



Figur 1 Transportgraf genererad från fordonsdata i förstudien

Datamängden som Scania samlar in ökar hela tiden vilket gör att ett stort fokus i projektet har varit att alla metoder som analyserats och utvecklats skall vara skalbara och kunna hantera strömmande data. Att hantera stora mängder data kontinuerligt är en utmaning som många företag står inför, det är viktigt för Sverige att tillverkningsindustrin som har många uppkopplade maskiner skaffar sig kompetens att hantera strömmande beräkningar. Det är också viktigt att hantera utmaningen som ligger i att använda data från olika datakällor, att förstå hur olika datamängders kvalitet påverkar det slutgiltiga resultatet.

4 Syfte, forskningsfrågor och metod

Syftet med projektet är att utgående från enbart GPS-positioner med hjälp av moderna beräkningsmetoder göra strömmande beräkningar för att svara på hur fordon nyttjas utifrån både ett individuellt och ett flott-perspektiv.

Den huvudsakliga forskningsfrågan i projektet är hur kan man använda fordonspositioner för att skapa kunskaper kring transportsystem. Den kunskapen kan sedan användas dels för att beskriva driften av ett individuellt fordon, eller för att identifiera när ett fordon ändrat sin drift men informationen kan även användas för att beskriva en stor flotta av fordon för att identifiera transportnätverk.

Det ställs stora tekniska krav på metoderna som utvecklas. För att de ska vara relevanta och skapa värde krävs det att de kan hantera den stora mängd data som kontinuerligt skickas från fordonen. För att åstadkomma detta har stora beräkningskluster använts både för beräkningar och för visualiseringarna. Ytterligare frågeställningar som behandlades i början av projektet var hur man kan utnyttja annan typ av information som man kan få från fordonen (exempelvis varningslampor, bränsleförbrukning, etc.).

4.1 Samarbetsplattform

Tidigt i projektet identifierades utmaningen hur data ska delas mellan industrin och den akademiska världen. Det är viktigt att begränsa tillgång till data så att alla endast har tillgång till ett minimum. Målet med projektet var också att använda moderna metoder vilket kräver stor beräkningskapacitet. Ett sätt att lösa både problemen med datatillgång och tillgång till modern beräkningskapacitet samtidigt som ett av målen som var öka samarbete kring data mellan industrin och akademien var att använda ett externt datacenter utvecklat av RISE SICS. Tack vare finansiering av Vinnova (D-ICE 2016-04318) så kunde projektet nyttja och utvärdera den

kapabilitet som erbjöds via datacentret i Luleå. Datacentret stod för beräkningsresurserna och under större delen av projektet användes den RISE-utvecklade plattformen för beräkningar på stora datamängder, HOPS, som beräkningsplattform.

4.2 Samarbetsmetod

Det dagliga arbetet bedrevs med täta avstämningar mellan akademien och industrin. Arbetssättet har varit nära det agila som med hjälp av korta avstämningar flera gånger i veckan driver arbetet framåt. Som en bieffekt av det arbetssättet har vi sett till att vi har uppnått målet att sprida kunskap mellan industri och akademi.

Samarbetet med utvecklarna av RISE SICS dataplattform har varit inriktat på att vidareutveckling för att förbättra för framtida samarbeten mellan industrin och akademien att använda och dela data. Samarbetet har skett via både fysiska möten och via digitala kanaler. Referensgruppen har varit en förutsättning för att säkerställa att resultatet av projektet ska vara till nytta på Scania. Kvartalsvis har möten av workshopkaraktär hållits med referensgruppen. Målet med dessa möten har varit att se till att informationen som skapas kommer till nytta inom verksamheten samt att säkerställa att vi inom projektet har kontakt med samtliga viktiga intressenter. Referensgruppen har också spelat en aktiv roll i att välja det användarfall som blev det första där informationen nyttjades.

5 Mål

5.1 Forskningsmål

Det övergripande målet med projektet var att använda den senaste tekniken för att göra beräkningar på stora mängder geopositionsdata av Scania tillverkade uppkopplade fordon. Att skapa kod som kontinuerligt tar emot och bearbetar nya positioner och göra det med metoder som är i forskningsfronten för den här typen av analyser.

Målet med projektet kan delas upp i tre delar.

- Den första delen handlar om att utforska och förstå tillämpbarheten av storskaliga beräkningsmetoder. Den kod som utvecklas ska vara modern enligt de senaste forskningsrönen och ska vara produktionsvärdig kod som ska kunna nyttjas och återanvändas.
- Den andra delen av målet är att lära sig förstå rörelsemönster för fordon och därigenom skaffa kunskap kring nyttjandet av ett fordon. Med hjälp av analyserna ska det också vara möjligt att identifiera övergångar mellan olika nyttjandemönster för fordon. Inom projektets mål ryms även att studera hur transportnätverk kan beskrivas.
- Den tredje delen av målet är att utforska och använda moderna metoder för att visualisera dessa stora datamängder på ett sätt som är enkelt att förstå och som går att nyttja i andra projekt som bearbetar stora datamängder.

5.2 Koppling till FFI och BADA mål

Projektet bidrar till följande FFI/BADA-mål:

- Projektet kommer att bidra till Sveriges internationella konkurrensförmåga genom att öka kunskaper kring beräkningar på och hantering av stora datamängder. Både inom industrin och inom den akademiska världen kommer kunskapsnivån kring moderna beräkningsmetoder att öka. Det är viktigt för Sverige att kunna använda moderna tekniker för att tillämpa strömmande beräkningsmetoder.
- Projektet kommer att bidra till att undersöka hur akademien och industrin kan arbeta tillsammans och dela både kod och data på ett sätt som är bibehåller datasäkerheten.

- Det är av stor vikt för samhället att förstå transportsystem och transportnätverk. Denna kunskap är väsentlig när vi ska arbeta med att förändra och förbättra samhällets nyttjande av transporter.
- Projektet kommer att bidra till kunskap kring uppkopplade fordon. Kunskap som är viktig för att stärka industrins konkurrenskraft, genom att förbättra både produkterna och tjänsteutbudet. Kunskap om uppkopplade fordonen är en viktig pusselbit för att skapa förståelse kring transportnätverk vilket behövs när vi går mot allt mer uppkopplade och smarta städer. Därför är det ett av målen med projektet att skapa en beskrivning av transportnätverk.
- För att sprida och nyttja information som skapas krävs det att resultat kan visualiseras och delas, vilket gör att en viktig del av projektet är att skapa en möjlighet för att visa upp och dela resultaten. FCC utvecklar under projektet en demonstratormjukvara för visualisering av beräkningsresultat med målet att i framtiden skapa en kommersiell mjukvara.

5.3 Mål för Scania

Målen för Scania är att möjliggöra för att nyttja positionsinformationen för att påverka produktutvecklingen och för att skapa tjänster till slutkunderna som är baserade på kunskapen. Ett annat tydligt mål för Scania är att öka kompetensen kring analys och strömmade beräkningar på stora datamängder.

5.4 Mål för FCC

Målen för FCC är att förstå hur man beskriver transportsystemet samt att utöka sin förståelse för beräkningar och behandling av stora datamängder. FCC bedriver forskning inom tillämpad matematik, maskininlärning och AI och ett stort mål med projektet var att vidareutveckla metoder och algoritmer så att de anpassas för stora datamängder. Ett annat syfte är att ta utveckla en demonstratormjukvara som innehåller en visualiseringsmiljö som ska kunna återanvändas för att visualisera stora datamängder och de beräkningar som ingår i olika projekt.

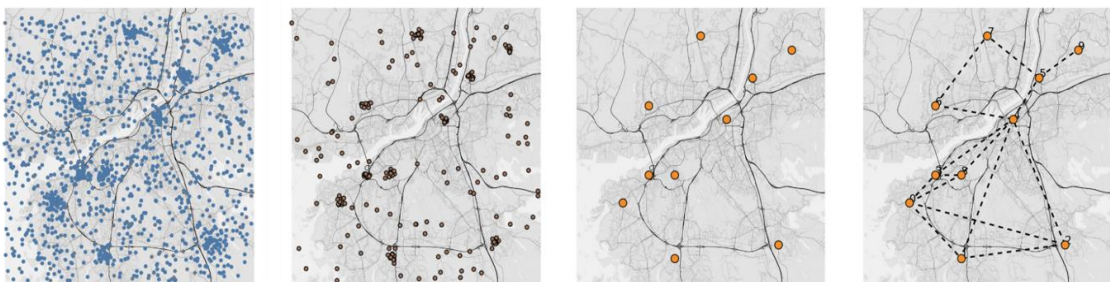
6 Resultat och måluppfyllelse

6.1 Resultat

Projektet har varit uppdelat i 6 arbetspaket som har tydligt avgränsade mål och resultat. Varje arbetspaket bestod av ett antal leverabler. I detta avsnitt presenteras sammanfattade resultat av varje arbetspaket samt en lista på de associerade leverablerna (för mer detaljnivå hänvisas leveransrapporterna som är bifogade).

Arbetspaket 1 – Definition och extrahering av transportnätverk och egenskaper

Målet med Arbetspaket 1 var att transformera den stora mängden råa GPS-positioner från fordonsflottan till ett transportnätverk bestående av noder och länkar. Anledningen var att få en mer konkret beskrivning av Scaniafordonens rörelser.



Figur 2: Schematisk figur över de olika transformationerna som utvecklades i arbetspaket 1.

Transformeringen från GPS-positioner från fordonsflottan till ett transportnätverk gjordes i följande tre steg.

Stoppfiltrering

Grundidén i projektet var att mest information angående ett fordon drift och användning finns inkluderat i de olika stopp som fordonet gör. Det vill säga de GPS-positioner som rapporteras när fordonet kör på en väg inte är lika innehållsrika som de positioner som de faktiskt stannar på för att utföra en uppgift (tanka, rast, lastning, etc.).

Metoder utvecklades inom projektet för att analysera varje fordon tidsserie av GPS-positioner och sedan filtrera ut endast vad som klassificerats som faktiska stopp. Definitionen av ett stopp var när ett fordon hade vistats inom ett begränsat geografiskt område under en viss tidsperiod. Metoderna som utvecklades kunde hantera strömmande data för att kunna implementeras på realtidsdata från Scania's Fleet Management-system. Metoderna är implementerade i Scania's data lake och hanterar alla nya positioner från den uppkopplade flottan kontinuerligt. Scania IT har driftsansvar för att skapa stopp som intressenter på företaget kan använda sig av för att lösa affärsbehov.

Klustring av stopplatser för stopplatsidentifiering

När alla stopp för fordonen filtrerats ut var målet att detektera platser där många antal stopp inträffat (när man analyserar hela flottan med fordon). Flera olika typer av metoder analyserades och utvecklades under projektet för att klustra fordonsstopp till stopplatser där flera stopp inträffat. Kravet på algoritmerna som användes var höga när det handlar om skalbarhet/beräkningskomplexitet eftersom antalet datapunkter var oerhört stort samt att datamängden ökar kraftigt på grund av fler uppkopplade fordon. Exempel på metoder som användes under projektet är:

- *RASTER*: En linjärtids-algoritm utvecklad i projektet för klustring av stora datamängder där dimensionen på datapunkterna är låg (i detta projekt GPS-punkter). En vidareutveckling av algoritmen för strömmande data utvecklades också under projektet (*S-RASTER*). För mer information om algoritmen, se referens i Sektion 7.2.
- *Point of interest (POI) -associering*: En metod där varje stopp associeras till närliggande (inom ett visst avstånd) POI-platser utvecklades och implementerades i projektet.

Skapande av transportnätverk

Det transportnätverk som analyserats i projektet är den graf där stopplatser av stopp definierar noderna och där länkarna skapas beroende på om fordon har åkt mellan stopplatser. Storleken och granulariteten av transportgrafen kan regleras genom att bestämma hur många stopp som måste ha inträffat vid en viss plats för att inkludera den stopplatsen (dvs reglera antalet noder) samt att bestämma hur många fordonsresor som måste ha gjorts mellan två noder för att inkludera en länk i grafen.

Leverabler i arbetspaketet:

- *D1.1 Definition of a network model that dynamically can be extended with arbitrary parameters for network components*
- *D1.2 Big Data framework for computational methods for reducing the scattered raw data to graph representations*
- *D1.3 Map visualizations of the generated models*

Arbetspaket 2 – Karaktärisering av nätverkskomponenter

När stopplatser/hubbar har identifierats från GPS-positionerna från fordonen så var målet i projektet att försöka karaktärisera stopplatserna. Detta gjordes genom att utveckla metoder som extraerade information om varje enskild stopplats från grunddatan.

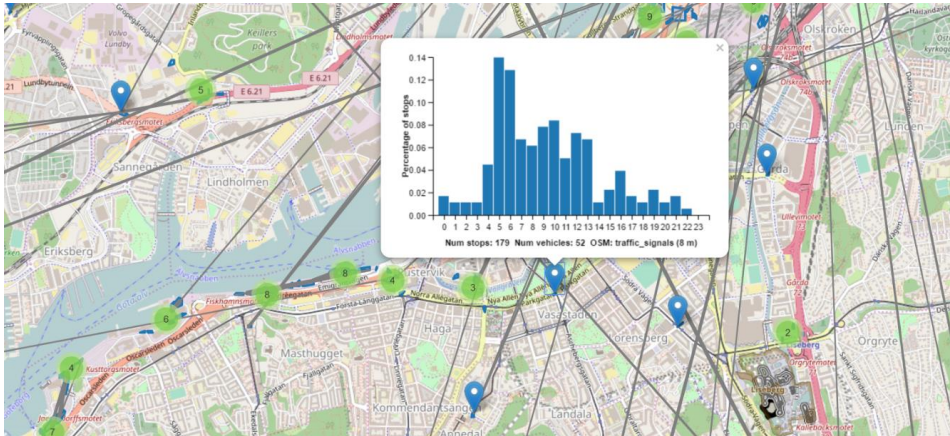
Exempel på egenskaper som noderna och länkarna i grafen:

Noder

Medelposition av stopp
Konvexa höljet av stoppen
Fördelning av tid på dygnet stoppen gjorde
Fördelning av veckodag stoppen gjordes
Duration av stoppen
Antal unika fordon

Länkar

Avstånd mellan stopplatser
Antal resor gjorda av alla fordon
Antal unika fordon som rest mellan stopplatserna
Distribuition av tid på dygnet resorna gjordes



Figur 3: Exempel på en stopplats/hubb där många fordonsstopp skett. Histogrammet visar ett exempel på hur stopplatserna berikades med egenskaper (här fördelningen över timme på dygnet som stoppen inträffade).

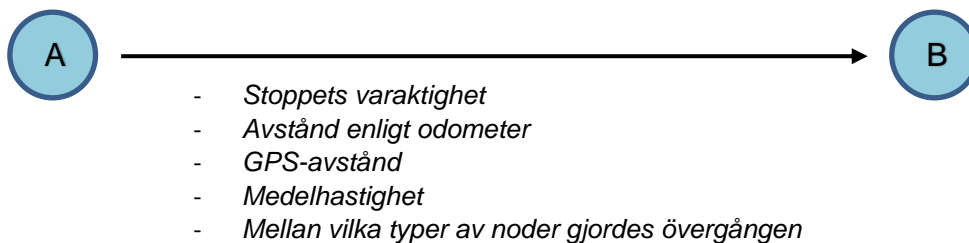
Förutom datadrivna egenskaper som extraherats från grunddatan så berikades noderna med POI-information (om sådan fanns).

Leverabler i arbetspaketet:

- D2.1 Method for segmenting network components into labeled categories
- D2.2 Extension of the network model with found categories for the network components

Arbetspaket 3 – Definition och extrahering av resor

Arbetspaket 3 fokuserade på att analysera hur fordon rör sig i den extraherade transportgraf. Fordonens rörelser översattes till övergångar mellan noder i transportgraf. Varje övergång berikades med egenskaper från grunddatan (exempel på egenskaper som övergångarna berikades med kan ses nedan).



I arbetspaketet analyserades även metoder för att extrahera mer komplexa resor/transitioner som består av fler övergångar mellan noder:

- *Cykler*: Cykliska resor för fordon definierades som ett antal konsekutiva övergångar som återgick till startnoden.
- *N-grams*: Ett *n*-gram definieras som *n* stycken konsekutiva övergångar. Genom att analysera ett fordon's förflyttningar kan man detektera om vissa specifika *n*-grams återkommer frekvent, vilket tyder på att fordonet har en repetitiv drift.

Vissa försök till att definiera samt detektera ett fordon's *hemmabas* gjordes även i arbetspaketet. På grund av svårigheter i att validera metoderna mot direkt observerad data om fordon's hemmabas så användes inte resultaten från dessa metoder i följande arbetspaket.

Leverabler i arbetspaketet:

- *D3.1 Method for detecting and extracting trips.*
- *D3.2 Definition of relevant features for trips and methods to extract these from the network model*
- *D3.3 Method for segmenting trips into labeled categories*
- *D3.4 Map visualization of the generated model(s)*

Arbetspaket 4 – Usage mode-klassificering

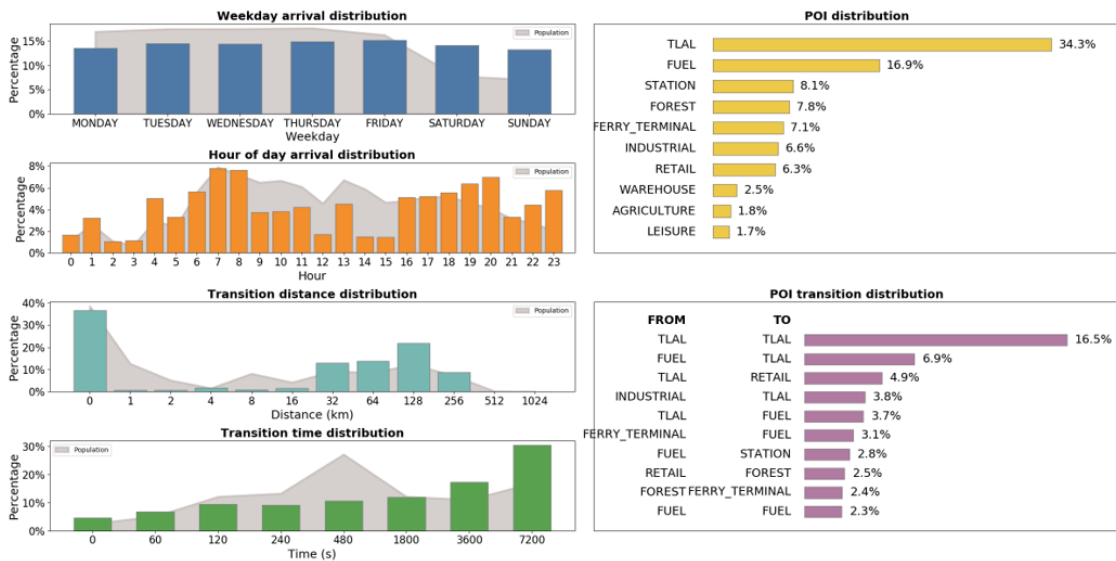
Arbetspaket 4 fokuserade på att analysera fordon's rörelser i transportgrafen för att slutligen kunna segmentera/klassificera fordonens olika drift (*usage mode*). I början av projektet var målet att kunna utifrån ett fordon's rörelser/övergångar bestämma dess faktiska usage mode. På grund av brist på grund sannings-data angående vad fordonen användes för skiftades fokus till att istället försöka sammanfatta och beskriva hur ett fordon använts/dess usage mode. Detta skifte av fokus styrkes av referensgruppen som såg stor nytta i av informationen som skapas av den definitionen av usage mode.

Resultatet blev en sammanfattning av ett fordon's rörelser med hjälp av en lista av egenskaper/features som kan kopplas till dess användning. I projektet definierades och implementerades ungefär 35 stycken olika features. Sex exempel på features som inkluderades i beskrivningen är

- Fördelning av platskategorier (POI) som fordonet besökte
- Fördelning av platskategoriövergångar (POI) som fordonet gjorde
- Fördelning av stopp per veckodag
- Fördelning av stopp per timme på dygnet
- Fördelning av avstånd mellan stoppen
- Fördelning av tid mellan stoppen

I Figurer 4 och 5 kan man se beskrivning av två olika fordon's drift med hjälp av sex olika egenskaper/features. Fordonen ägs av Transportlabbet som är ett Scania internt transportbolag. I bakgrunden syns hela fordon'sflottans medelvärde i grått för jämförelse.

- I Figur 4 beskrivs ett av Transportlabbet's fordon som kör långa resor mellan Södertälje och Zwolle. Man kan tydligt se att fordonet används jämnt fördelat över veckodagar (även helger) samt att övergångsavstånden är längre än genomsnittet. Fordonet besöker framförallt TLAL-platser vilket är Transportlabbet's egendefinierade platser i bland annat Södertälje och Zwolle.
- I Figur 5 beskrivs en av Transportlabbet's bussar. Man kan tydligt se att fordonet är i drift mellan 6 och 18 på vardagar samt att övergångsavstånden är kortare än genomsnittet. De platskategorier som fordonet besöker är framförallt "Stationer" vilket stämmer överens med att fordonet är en buss.



Figur 4: Exempel på beskrivning av usage mode för en av Transportlabets fordon. De egenskaper/features som presenteras är fördelningar av stopp över veckodagar, stopp per timme på dygnet, övergångsavstånd, övergångstid, platskategori samt platskategoritransitioner. Detta specifika fordon är en långtradare som kör långa sträckor mellan Södertälje och Zwolle.



Figur 5: Exempel på beskrivning av usage mode för en av Transportlabets fordon. De egenskaper/features som presenteras är fördelningar av stopp över veckodagar, stopp per timme på dygnet, övergångsavstånd, övergångstid, platskategori samt platskategoritransitioner. Detta specifika fordon är en buss som kör i Södertälje- och Stockholmsregionen

I arbetspaketet analyserades olika metoder för att kunna gruppera in fordonen i olika driftsklasser. Det fanns ingen tillgång till direkt observerad data angående driften på fordonen, så idén var att försöka hitta grupper av fordon där varje grupp består av fordon som har liknande egenskaper/features. Olika klustringsmetoder analyserades och den slutgiltiga implementeringen var en hierarkisk klustringsmetod som bland annat använde Wassersteinmetrik för att kunna jämföra två fordons egenskaper/features. Resultaten från klustringen på Transportlabets fordon där direkt observerad driftdata finns (dvs vad fordonen används för) var positiva och metoden lyckades separera de olika fordonsanvändningarna som Transportlabbet har.

Olika metoder för att detektera när ett fordon ändrar sin usage mode/drift analyserades också i arbetspaketet. För mer information angående dessa se leveransrapport D4.3.

Leverabler:

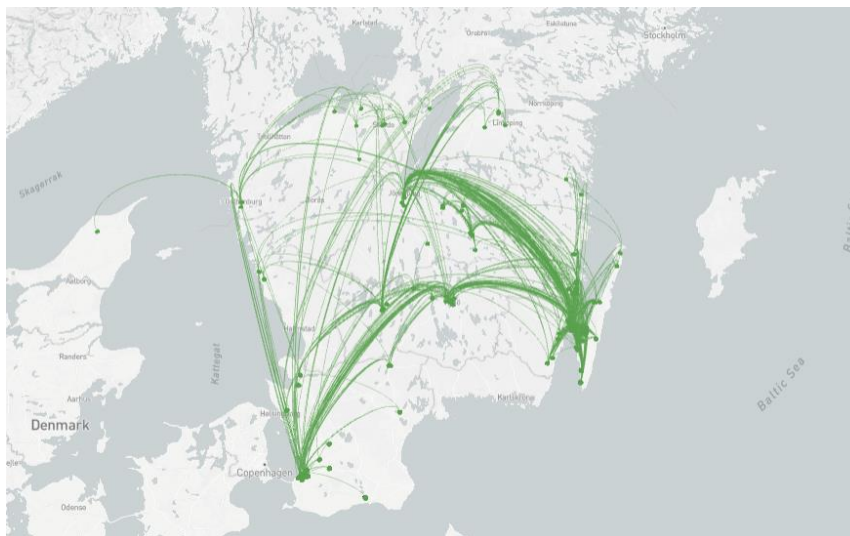
- *D4.1 Algorithm that given a set of trips for a specific vehicle returns usage modes.*
- *D4.2 Definition/model that maps the different categories to verified usage modes*
- *D4.3 Methods for detecting usage mode transitions*

Arbetspaket 5 – Transportkedjor och transportnätverk

I Arbetspaket 5 togs en mer övergripande syn på transportnätverket med mål att detektera transportkedjor och sektorer. Fokus lades på ett koncept som kallas *community detection* som handlar om att detektera sammankopplade delgrupper av noder i en graf. Community detection har traditionellt framförallt använts för att analysera sociala nätverk där varje nod representerar personer och länkarna representerar interaktioner mellan personer. I detta projekt definieras grafen av de övergångar fordon gör mellan olika noder. En community i den grafen skulle representera en uppsättning noder/platser som är tätt sammankopplade genom att många fordonsövergångar har gjorts mellan dem.

I projektet analyserades flera metoder för att göra community detection på stora nätverk och slutresultatet blev en vidareutveckling av en metod som heter BigCLAM. Den utvecklade metoden kan detektera överlappande communities (dvs att varje nod kan tillhöra flera communities) i den stora transportgraf som definieras av Scania och dess kunders fordons rörelser.

Det finns ingen direkt observerad data angående vad som är en faktiskt community i transportgrafan vilket gör validering av metoderna svåra. Genom manuell inspektion av vissa detekterade communities kunde man däremot identifiera flera intressanta communities där de tillhörande noderna visade tydliga likheter. Exempelvis är en community i Sverige där de absolut flesta noder var ICA-butiker, vilket tyder på att fordonen som besöker noderna levererar mat till butiker (se Figure 6). Ett annat exempel var en community i Lettland där alla tillhörande noder var busstationer vilket tyder på att man har detekterat ett lokaltraffiknätverk.



Figur 6: Exempel på detekterad community bestående av 230 noder i södra Sverige. Efter manuell analys av noderna upptäcktes att väldigt många noder var matbutiker eller lagerlokaler vilket tyder på att denna community består av fordon som levererar varor till butiker.

Leverabler:

- D5.1 Report on possible methods for using results from WP1-WP4 to detect transport chains and transport sectors

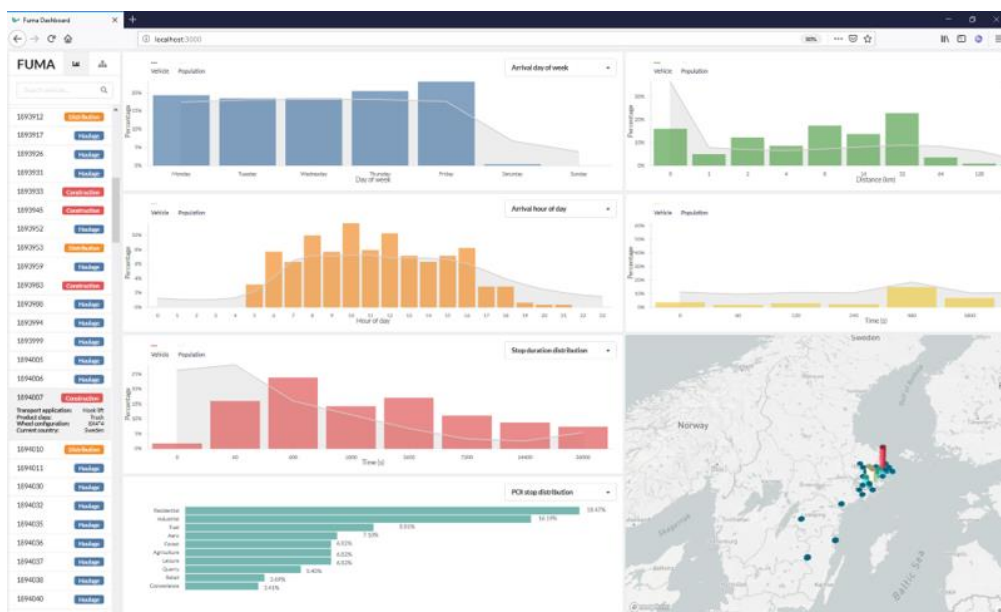
Arbetspaket 6 – Demonstratormjukvara

Målet med arbetspaket 6 var att utveckla en demonstratormjukvara för att kunna presentera och visualisera de projektresultat som tas fram i projektet. Istället för att sprida resultat genom rapporter och presentationer skulle man presentera projektresultat i form av interaktiva dashboards där användarna själv kunde utforska resultaten från algoritmerna.

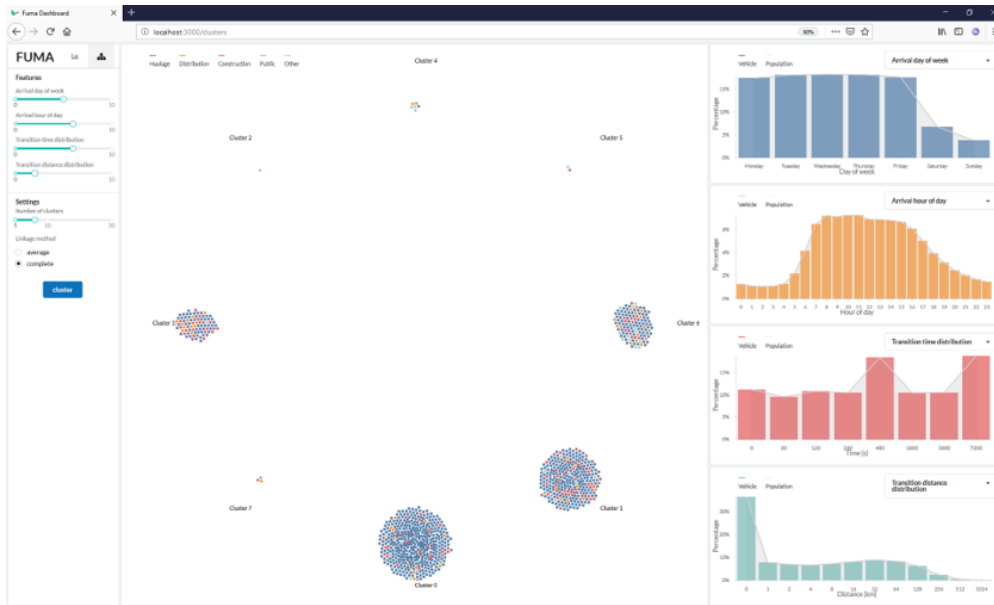
Den utvecklade programvaran är en webbaserad mjukvara där en användare enkelt kan skapa responsiva och interaktiva dashboards.

Några exempel (skärmdumpar) på dashboards utvecklade i projektet för att sprida projektresultaten samt för att visa på nyttan av interaktiva dashboards är:

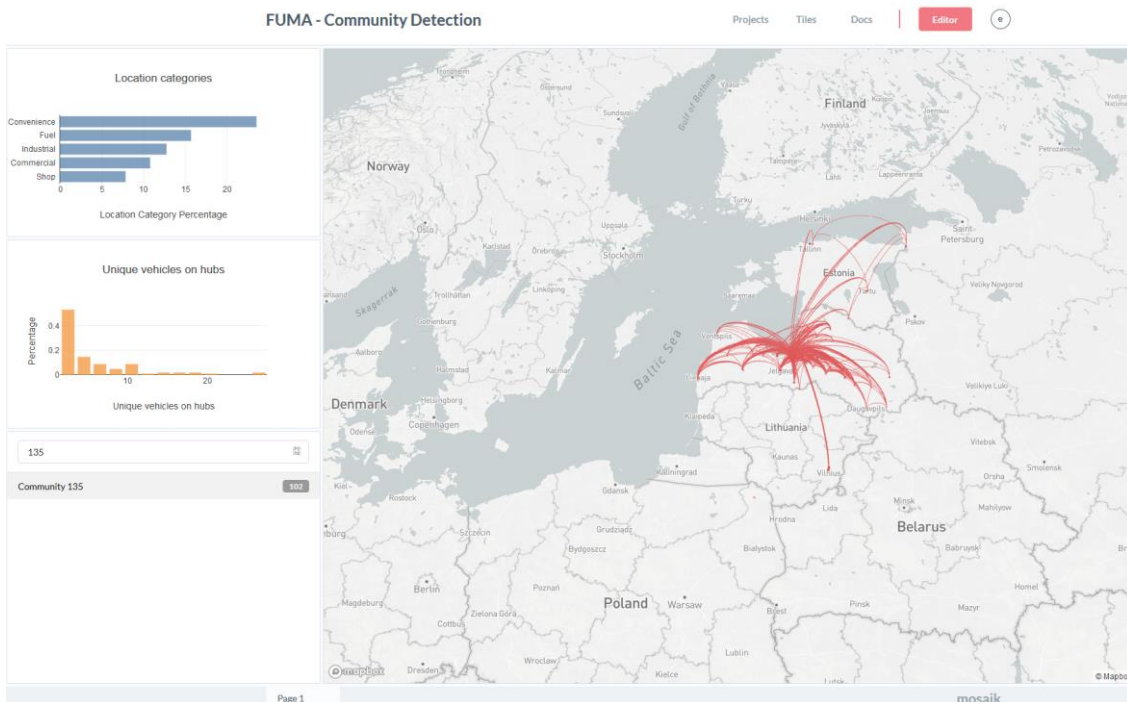
- *Figur 7:* Usage mode-beskrivningen för ett fordon visas. Man kan enkelt söka efter specifika fordon och få en aggregerad och lättförståelig beskrivning av fordonets egenskaper och drift. Om man är intresserad av andra egenskaper/features än de som visas kan man enkelt byta ut en egenskap mot en annan som är av intresse. Även en kartvisualisering där en heatmap visar var fordonet har gjort sina stopp presenteras.
- *Figur 8:* Klustring av fordon utifrån deras egenskaper/features. Till vänster kan man välja hur viktiga varje egenskap är (dvs hur mycket vikt skall läggas på den egenskapen i klustringen). Resultatet från klustringsmetoden visualiseras i grupper av fordon (i mitten) som är representerade av prickar. Till höger ser man medelvärdet av alla fordonens egenskaper för det valda klustret/gruppen.
- *Figur 9:* Resultat från community detection-algoritmer. Till vänster kan man söka och välja vilken community man är intresserad av. Uppe till vänster ses en sammanfattning av den communityns egenskaper och i kartan ser man noderna och länkarna som definierar communityn.



Figur 7: Exempel på dashboard skapad i demonstratormjukvaran som presenterar ett fordon's usage mode. Till vänster kan man söka och välja specifika fordon.



Figur 8: Exempel på dashboard skapad i demonstratormjukvaran som presenterar ett resultat från klustring av fordonen utifrån deras egenskaper/features. Till vänster kan man välja vissa parametrar för klustringsmetoden som exempelvis vilken vikt varje enskild feature skall ha.



Figur 9 Exempel på dashboard skapad i demonstratormjukvaran för att analysera resultaten från community detection-algortimerna.

6.2 Måluppfyllelse

Alla tre forskningsmålen, se redovisningen i föregående kapitel, har uppnåtts. Två av resultaten från forskningsmålen används kontinuerligt på Scania:

- *Storskaliga beräkningsmetoder*: Koden ligger i produktion i Scania och behandlar kontinuerligt fordonspositioner för att skapa kunskap från de uppkopplade fordonen.
- *Rörelsemönstret från fordonen*: Rörelsemönstret undersöks och baserat på resultaten tas affärsbeslut på Scania för att optimera interna processer.

Den tredje målet att utforska visualiseringsmetoder har skapat utforskats och skapat insikter kring hur man kan visualisera geospatial data. Demonstratormjukvaran har använts vid olika presentationer och workshops för att sprida kunskap angående projektresultaten. Mjukvaran vidareutvecklas för tillfället av FCC för en framtida kommersialisering.

Utöver koden som har implementerats har Scania också nått målet av att öka kunskapen kring stora datamängder. Projektet har resulterat i ett stort antal presentationer där resultaten har visats. Flertalet grupper på Scania är intresserade av resultaten och det pågår flera parallella projekt för att nyttja informationen för att stärka olika verksamhetsområden.

6.3 Kunskaps- och resultatspridning

Hur har/planeras projektresultatet att användas och spridas?	Markera med X	Kommentar
Öka kunskapen inom området	X	Resultaten har under projekttiden främst spridits inom projektkonsortiet. Detta genom kontinuerliga arbetsmöten. Kunskapen har också delats med RISE SICS via plattformen för datadelning
Föras vidare till andra avancerade tekniska utvecklingsprojekt	X	Lärdomar från projektet har nyttjas inom forskningsprojektet CODA Resultaten används inom flera tekniska utvecklingsprojekt på Scania
Föras vidare till produktutvecklingsprojekt	X	Scania har satt pipeline i produktion och använder informationen kontinuerligt FCC vidareutvecklar demonstratormjukvaran med mål att skapa en kommersiell produkt.
Introduceras på marknaden		
Användas i utredningar/regelverk/ tillståndsärenden/ politiska beslut		

6.4 Publikationer

- *Contraction Clustering (RASTER): A Very Fast Big Data Algorithm for Sequential and Parallel Density-Based Clustering in Linear Time, Constant Memory, and a Single Pass*
G. Ulm, S. Smith, A. Nilsson, E. Gustavsson, M. Jirstrand,
Submitted to the journal Big Data Research.
- *Contraction Clustering (RASTER): A Big Data Algorithm for Density-Based Clustering in Constant Memory and Linear Time*
G. Ulm, E. Gustavsson, M. Jirstrand.

6.5 Övrig kunskapsspridning

Inom ramen för projektet har det spridits kunskap både internt på FCC och på Scania, samt även utanför forskningsprojektet.

- Fyra stycken examensarbetare har examinerats som en del av projektet, samtliga har varit anställda på Scania.
- FUMA har presenterats på en konferens i Barcelona. Konferensen är en årlig konferens med inriktning på beräkning av stora datamängder. Projektet gavs tillfälle att presentera resultat och metoder under två tillfällen på konferensen
- Projektet har varit en aktiv del i utvärderingen av RISE SICS plattform, vi inom projektet har haft en aktiv roll i att utveckla en plattform för att dela data.
- FUMA har deltagit i tre konferenser ordnade av RISE SICS för att sprida kunskap om hur man kan ta datadrivna beslut inom företag.
- Resultat och metodik angående klustringsmetoden RASTER som utvecklades inom projektet presenterades på The 3rd International Conference on Machine Learning, Optimization and Big Data (MOD 2017) i Volterra, Italien.

7 Slutsatser och fortsatt forskning

Resultaten från projektet anses vara väldigt positiva enligt både Scania och FCC. De algoritmer som utvecklades för att kontinuerligt kunna behandla den strömmande datan är implementerade och produktionsfärdiga på Scania. De beskrivningar av usage mode som definierades och implementerades kunde tydligt visa att man kan skilja på fordon som har olika drift/användning. Resultaten från community detection-algoritmerna visar att man kan identifiera viktiga delar av transportnätverket där fordon opererar.

Alla projektmål uppfylldes och kunskapsspridningen inom projektgruppen har varit väldigt positiv. Scania har ökat sin förståelse för beräkningar av stora datamängder, beskrivning av fordons drift, samt hur man kan identifiera viktiga områden där många Scania-fordon rör sig. FCC har vidareutvecklat sin kompetens inom utveckling och implementering av algoritmer på stora datamängder samt fått en stor förståelse för tunga fordons drift och beteende. Man planerar även att vidareutveckla och i framtiden kommersialisera den demonstratormjukvara som utvecklats som en del av projektet.

Framtida forskningsfrågor som uppkommit under projektets gång men inte kunnat behandlas fullständigt är exempelvis hur man kan detektera när fordon tydligt ändrar sin usage mode. Metoder för sådana usage mode-transitioner analyserades i projektet men implementerades aldrig. Ytterligare forskningsfrågor som diskuterats är hur man kan dra ännu mer nytta av den extraerade transportgraf. Exempelvis om man kan identifiera hur laddstationer bör positioneras utifrån hur fordon verkar röra sig, eller om man kan identifiera marknader/områden där man borde ändra sin marknadsföring.

8 Deltagande parter och kontaktpersoner

Kontaktpersoner för de deltagande parterna Scania AB och Fraunhofer-Chalmers Centre är:

- **Scania AB**
Sara Sylvan och Frida Nellros
- **Fraunhofer-Chalmers Centre**
Emil Gustavsson och Mats Jirstrand



FRAUNHOFER CHALMERS
RESEARCH CENTRE FOR INDUSTRIAL MATHEMATICS



SCANIA