

УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ



Момир Аџемовић

ПРЕДВИЂАЊЕ ТРАЈЕКТОРИЈА ВИШЕ
ОБЈЕКТА НА СЦЕНИ

мастер рад

Београд, 2022.

Ментор:

др Младен Николић, ванредни професор
Универзитет у Београду, Математички факултет

Чланови комисије:

др Јована Ковачевић, доцент
Универзитет у Београду, Математички факултет

др Александар Картељ, доцент
Универзитет у Београду, Математички факултет

Датум одбране: 15. септембар 2022.

посвета... у изради...

Наслов мастер рада: Предвиђање трајекторија више објеката на сцени

Резиме: У изради...

Кључне речи: машинско учење, аутономна вожња, растеризација, графовске неуронске мреже

Садржај

| | | |
|----------|--|-----------|
| 1 | Увод | 1 |
| 1.1 | Дефиниција аутономне вожње | 1 |
| 1.2 | Поставка проблема | 2 |
| 1.3 | HD мапе | 3 |
| 1.4 | Циљ рада | 4 |
| 2 | Преглед основних градивних елемената | 5 |
| 2.1 | Неуронске мреже | 5 |
| 2.2 | Конволутивне неуронске мреже | 5 |
| 2.3 | Графовске неуронске мреже | 5 |
| 2.4 | Механизам пажње | 5 |
| 3 | Преглед постојећих приступа | 6 |
| 3.1 | Технике прилагођене репрезентацијама временских серија | 7 |
| 3.2 | Технике засноване на растеризацији | 9 |
| 3.3 | Технике засноване на графовским репрезентацијама | 10 |
| 3.4 | Хибридне технике | 11 |
| 3.5 | Технике засноване на облацима тачака | 11 |
| 4 | Метрике за евалуацију модела | 13 |
| 5 | Припрема података | 16 |
| 5.1 | Argoverse интерфејс | 16 |
| 5.2 | Припрема трајекторије агента | 17 |
| 5.3 | Припрема трајекторија суседа | 18 |
| 5.4 | Припрема путних сегмената | 19 |
| 5.5 | Припрема кандидата путних сегмената | 20 |
| 5.6 | Преглед дефинисаних параметара | 25 |
| 5.7 | Визуализација | 25 |

| | | |
|----------|---|-----------|
| 6 | Техника заснована на разумевању контекста обрадом сцене представљене графом | 28 |
| 6.1 | Модел <i>VectorNet</i> | 29 |
| 6.2 | Модификована верзија <i>VectorNet</i> са узоркованим крајњим тачкама трајекторија | 32 |
| 6.3 | Припрема података | 37 |
| 6.4 | Експерименти и резултати | 38 |
| 7 | Техника заснована на разумевању контекста обрадом растеризоване сцене | 43 |
| 7.1 | Растеризација <i>HD</i> мапа | 44 |
| 7.2 | Модел <i>HOMER</i> | 47 |
| 8 | Закључак | 53 |
| | Библиографија | 54 |

Глава 1

Увод

Аутономна возња подразумева потпуну или парцијалну аутоматизацију процеса контроле возила коришћењем рачунарских и осталих технологија. Агент (возило) у овом случају мора да има способност опажања окружења и самосталног кретања кроз то окружење ради остваривања задатог циља (краткорочно: паркирање, скретање, ... дугорочно: транспорт, **обрада земљишта на фармама**, ...). Кључне компоненте које се интегришу у овај систем су:

- Детекција и праћење објеката у околини (опажање окружења са акцентом на покретне објекте)
- Разумевање њихових циљева како би се сам циљ агента ускладио са њима, **што је** први корак за омогућавање самосталног кретања кроз окружење.

Прва компонента се реализује помоћу специјалних сензора као што су *LiDAR* сензори за конструкцију 3D мапе окружења, **радар сензори** за детекцију удаљености и брзине објеката (погодан и у случају лошег времена) и камере који прикупљају 2D слике окружења. Саме камере помоћу метода машинског учења могу да извршавају исте послове као и радари (као што то човек ради користећи само вид). Циљ овог рада се односи на други део тј. анализа постојећих радова чија је **тема** разумевање циљева објеката на **сцени** где се агент налази. Будуће трајекторије тих објеката се посматрају као циљеви, а сама трајекторија агента се усклађује у односу на њих.

1.1 Дефиниција аутономне возње

Министарство саобраћаја САД и NHTSA (*National Highway Traffic Safety Administration*) је усвојила *Society of Automotive Engineers (SAE)* стандард који прописује 6 нивоа аутоматизације од нултог, где човек има потпуну контролу до петог, где рачунар самостално управља возилом [1]:

- **Ниво 0:** Човек има потпуну контролу над возилом. Возило може да има једноставније аутоматизације као што је на пример аутоматизовано кочење у хитним случајевима.
- **Ниво 1:** Човек и рачунар сарађују у процесу управљања возилом (сва одговорност је на возачу).
- **Ниво 2:** Рачунар има потпуну контролу, али човек мора да буде присутан и да реагује у било ком тренутку ако је то неопходно (сва одговорност је на возачу).
- **Ниво 3:** Рачунар има потпуну контролу ако су испуњени одређени услови, а човек је у том случају потпуно ослобођен свих дужности вожње. Ако услови више не важе (нпр. пада снег) онда човек мора да преузме одговорност над возилом. Када су испуњени услови, одговорност је на произвођачу софтвера за аутономну вожњу.
- **Ниво 4:** Слично као ниво 3, али возило је испрограмирано тако да се само заустави у случају да услови нису испуњени. Када сви услови опет важе, возило-рачунар може да настави да врши свој задатак.
- **Ниво 5:** Рачунар има потпуну контролу над возилом у сваком тренутку.

1.2 Поставка проблема

Предвиђање трајекторија или предвиђање кретања једног агента на сцени са више покретних објеката (суседи аналогни агенту) се формално дефинише као предикција скупа вредности $\{(x_a^t, y_a^t) \mid t \in [0, T]\}$, где је T дужина трајекторије агента која се предвиђа, под претпоставком да су дате следеће информације:

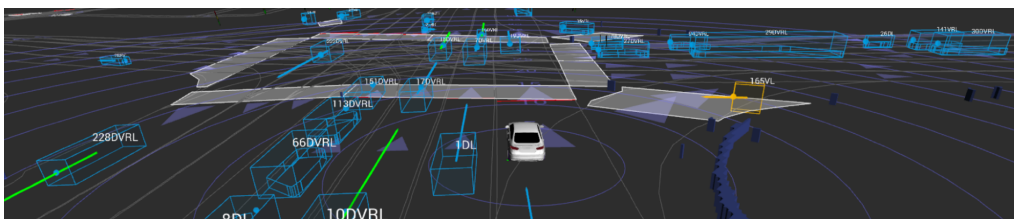
- Историја тог агента $A_h = \{(x_a^t, y_a^t) \mid t \in [-H, -1]\}$, где је H дужина историје трајекторије;
- Историја осталих покретних објеката $\{O_n \mid n \in N\}$, где је O_n истог облика као и A_h , а N је скуп свих суседа;
- Окружење агента E - варира у односу на скуп података. Подаци окружења зависе од скупа алата који се користе за прикупљање података о окружењу. Примери тих алата су камере, радар, *LiDAR* сензори и слично, а од информација се чувају подаци о путевима тј. локацијама сегмената путева у референтном координатном систему, локације пешачих прелаза, локације семафора и њихових стања, ... Подаци о окружењу нису строго дефинисани и разликују се од једног скупа података до другог.

Сам проблем је општији од примене у аутономног вожњи и може да се примењује у другим случајевима. Предвиђање кретања пешака се такође своди на исту поставку проблема, али није од кључног значаја као у системима за аутономну вожњу.

1.3 HD мапе

Мапа са високим нивоом детаља (*eng. HD map*) је прецизна мапа путева са грешком до пар центиметара и високим нивоом познавања окружења: позиције пешачких прелаза, позиције семафора, полигони путева, смерови улица итд... Битно је нагласити да су HD мапе супериорне у односу на GPS у смислу прецизности и разноврсности информација. Саме GPS мапе нису довољно прецизне да може да их користи систем за аутономну вожњу уместо људи који су значајно робуснији на грешке GPS-а (што може да буде и до неколико метара). Аналогија HD мапа са човеком следећа: Случај када је човек вози унапред добро познатим путем одговара рачунару који управља возило коришћењем HD мапа, а случај када човек вози неким путем први пут у животу одговара рачунару који управља возилом без коришћења HD мапа.

Овај тип података даје јако велику количину информација које се могу процесирати у целокупном систему аутономне вожње. Два главна проблема и изазова овог типа података је њихово прикупљање (*LiDAR* системи су скупи) и одржавање, и обрада овако неструктурираних података. Алтернатива је коришћење камера и модела дубоког учења за прикупљање података о окружењу у реалном времену. Ти модели могу да буду истренирани над подацима који су прикупљени коришћењем *LiDAR* система.



Слика 1.1: HD мапа - *Kaggle Lyft* [2]

На слици 1.1 се види пример визуализоване HD мапе. Бело возило је агент, а плави правоугаоници су суседи. Са слике се види да мапе садрже и полигоне пешачких прелаза и још неке додатне податке. Подаци ових мапа су често у 3D формату тј. узима се у обзир и висина. У реалном времену се ови подаци могу комбиновати са улазима из камера.

1.4 Циљ рада

Циљ овог рада је анализа различитих приступа разумевања контекста сцена у саобраћају. Под контекстом се подразумева разумевање статичког дела сцене који обухвата путеве, семафоре, пешачке прелазе и слично, динамичког дела сцене који обухвата историју кретања објеката на сцени, и интеракције свих елеманата у том окружењу.

Сцене се посматрају из птичје перспективе и дате су у формату *HD* мапа. На основу информација добијених из обрађених мапа и историје трајекторије објеката на сцени се врши предвиђање трајекторија агената. Посебна пажња се посвећује анализи већ постојећих техника заснованих на растеризацији сцена и графовских техника које захтевају трансформацију података у структуру графа. Главна инспирација су хијеархијска графовска неуронска мрежа *VectorNet* [3] и *HOME* [4] архитектура која претежно заснована на примени конволутивних мрежа над растеризованим подацима.

Подаци једног сценарија (једна инстанца *HD* мапе, **један случај**) нису нужно дати само у једном облику тј. само једној структури. Историја трајекторија може да се чува у табеларном формату, где се за сваки ред чува временски тренутак, идентификатор агента и координате тог агента у том тренутку. Возно подручје може да буде задато као бинарна матрица где 1 означава да је могућа вожња на тој координати, а 0 да није могућа вожња. Први корак сваке технике подразумева комбиновање и трансформацију података у одговарајући облик, било да је крајњи резултат слика или граф. У раду се презентују детаљи имплементације припреме података у одговарајуће облике. Методе које се анализирају не нуде отворену имплементацију отвореног кода и не залазе у конкретне детаље имплементације.

Квалитет модела зависи од начина на који се примењује доменско знање. Одређене кључне информације са сцене могу да се изведу детерминистичким алгоритмима заснованим на одговарајућим хеуристикама. У раду се презентују једноставне хеуристике које комбинују доменско знање са моделима машинског учења.

Глава 2

Преглед основних градивних елемената

2.1 Неуронске мреже

TODO: Osnovno, MLP

2.2 Конволутивне неуронске мреже

TODO: CNN, Transposed CNN

2.3 Графовске неуронске мреже

TODO: Osnovno, opšte, GCN, GraphSage

2.4 Механизам пажње

TODO: attention

Глава 3

Преглед постојећих приступа

Уколико имамо више објеката на сцени за које треба да се изврши предвиђање трајекторија, онда је лакше да се предвиђање сваке трајекторије изврши из угла конкретног објекта. Тај објекат наставку реферише као агент. Сваки агент може да има свој референтни систем и одговарајуће окружење у том систему тј. елементе сцене у околини тог агента. Скуп релеватних елемената као што су путеви, остала возила, пешаци и слично се разликују од окружења једног агента до другог. Један од главних фактора за релеватност елемената је удаљеност тог елемента у окружењу од агента.

Предвиђање може да се извршава у једном кораку, где излаз модела подразумева више трајекторија (по једна трајекторија за сваког агента), или итеративно за сваког агента посебно. Сам приступ зависи од конкретног модела, али углавном су модели ограничени да итеративно генеришу предикције за сваког агента, што може да буде рачунски захтевније. Посматрано из угла аутономне вожње, скоро увек је интересантан само један агент на сцени који одговара возилу којим се управља.

Технике за предикцију трајекторија више објеката могу да се групишу грубо у четири групе:

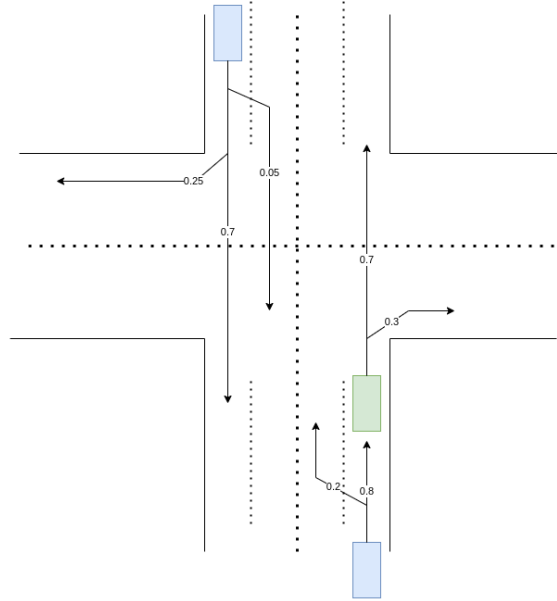
- Технике прилагођене репрезентацијама временских серија;
- Технике засноване на растеризацији;
- Технике засноване на графовским репрезентацијама;
- Хибридне технике.

У наставку ове главе биће укратко описане претходно наведене технике

3.1 Технике прилагођене репрезентацијама временских серија

Из поставке проблема може да се закључи да овај проблем може да се сведе на предвиђање временских серија, јер свака трајекторија може да се представи помоћу временских серија. У том случају је интуитиван иницијални приступ да се користе рекурентне неуронске мреже (*eng. RNN - Recurrent Neural Network*) и конволутивне мреже за једнодимензионе сигнале (*CNN - Convolutional Neural Network*). Често коришћена *LSTM* архитектура рекурентних неуронских мрежа је погодна за кодирање динамике објеката на сцени. Динамика објеката се односи на историју трајекторија агента и осталих објеката на сцени. У случајевима када трајекторија једног објекта зависи од трајекторија осталих објеката, није довољно да се само узима у обзир историја трајекторије агента. У том случају је погодно да се користи механизам пажње као компонента модела која учи релевантне везе између различитих објеката на сцени [5, 6, 7].

Један од главних проблема претходних приступа је неспособност модела да научи мултимодалну природу трајекторија на сцени. На слици је 3.1 је приказан пример сценарија где не постоји само једна валидна предикција трајекторије. Ради успешног краткорочног планирања у аутономној вожњи, морају да се узму у обзир скоро сва могућа будућа стања сценарија. Модели који као излаз дају само највероватнију трајекторију нису увек довољни. Бољи приступ је да се уместо модела који као излаз има највероватнију трајекторију користи модел који као излаз има N ($N > 1$) трајекторија заједно са вероватноћом сваке од тих трајекторија. Модели који се узимају у обзир у овом раду функционишу по том принципу.



Слика 3.1: Милтимодална природа могућих будућих трајекторија - бројеви на трајекторијама представљају вероватноће

Битно је нагласити да трансформација модела који као излаз има једну трајекторију у модел који као излаз има више трајекторија није тако једноставно. Природан избор за функцију грешке током тренирања је одступање предвиђене трајекторије од истините трајекторије. Конкретан пример функције грешке је просечно еуклидско растојање тачака предвиђене трајекторије и истините трајекторије које одговарају истом временском тренутку. Ако се тренира модел који предвиђа једну трајекторију са овом функцијом грешке (под претпоставком да квалитет и квантитет података нису проблем), онда ће он научити да предвиђа просечну трајекторију. Уколико се повећа број трајекторија у излазу модела, а функција грешке остане иста тј. предвиђене трајекторије се упоређују са једном истинитом трајекторијом у сваком сценарију, то не значи да ће модел научити да узима у обзир и остале вероватне трајекторије, већ ће вероватно генерисати само N „дупликата“ трајекторија, где је N број трајекторија у излазу тог модела. Овај проблем се назива колапс мода *eng. mode collapse* и односи се на проблем игнорисања неких мода приликом генерисања предикција [8].

Уместо учења модела који дају предикције трајекторија које су у просеку најбоље у тим случајевима, алтернатива је скуп модела заснованих на условним генеративним моделима који омогућавају генерисање произвољног броја трајекторија кроз узорковање из условне расподеле (расподела је условљена историјом трајекторије, као и окружењем у којем се тај објекат налази). Овакви модели нису имуни на колапс мода, али боље превазилазе тај изазов. Пример таквог модела је *Social GAN* [9] архитектура која узима у обзир претходно наведене услове и генерише „социјално

прихватљиве“ трајекторије пешака. Током процеса предвиђања генеративни модели захтевају напредније алгоритме узорковања над великим бројем генерисаних трајекторија ради оптимизације покривености трајекторијама тј. ради оптимизација односа будућих стања сценарија који је модел узео у обзир и свих могућих будућих стања сценарија. Чак и са одговарајућим алгоритмима, није гарантована оптимална покривеност због стохастичког узорковања из латентног простора [10]. Овакви модели могу да се користе у аутономној вожњи, али су ипак погоднији за проблем предвиђања трајекторија пешака у произвољном окружењу који није нужно улица, већ може на пример да буде и парк који има слабија ограничења на кретање тог пешака.

3.2 Технике засноване на растеризацији

Растеризација подразумева трансформацију *HD* мапа у слике, где слике приказују сцене из птичје перспективе (*eng. BEV - Bird's-eye-view*). То значи да различите структуре података које описују један сценарио морају да се конвертују у облик слике. На пример, историја трајекторија која је углавном задата у табеларном формату мора да се конвертује у слику или да се на неки начин комбинује са растеризованим подацима у оквиру архитектуре модела. Предност слика је могућност разумевања интеракција елемената на мапи коришћењем низа слојева конволутивних неуронских мрежа. Идеја низа конволутивних мрежа је да преводу сирове слике у компактнији облик својстава који углавном има мању резолуцију, али већи број канала. Овако добијена својства се користе за предвиђање трајекторија. Конволутивне мреже нису ограничене да раде са RGB и сличним форматима слика. Сваком пикселу може да се додели вектор својстава. Нека од очигледнијих својстава су: да ли агент заузима тај пиксел, да ли сусед (неагент) заузима тај пиксел, да ли пиксел припада улици, ... Излаз CNN енкодера се углавном накнадно комбинује са осталим компонентама за генерисање резултата.

Излази модела такође могу да се представљају као топлотне мапе (*eng. heatmap*). Структура топлотних мапа одговара слици, где се сваком пикселу додељује вероватноћа да се агент налази на тој локацији. За добијање топлотних мапа се користи енкодер-декодер архитектура, где се сирова слика прво пресликава у компактнији облик својстава низом конволутивних мрежа (енкодер), а онда се применом низа транспонованих конволутивних слојева (декодер), које неформално представљају инверз конволутивних слојева, добија топлотна мапа која је исте резолуције или редуковане резолуције у односу на улазну слику. Како крајњи резултат треба да буде једна или више предикција трајекторија, неопходно је да се примени одговарајући алгоритам за узорковање циљних тачака. Циљна тачка се дефинише као последња

локација предикције трајекторије. Главна мотивација за узорковање циљних тачака је претпоставка да уколико је познато где се трајекторија завршава, онда је лако да се одреди цела та трајекторија тј. естимира кретање агента до крајње тачке¹. Број узоркованих крајњих тачака зависи од алгоритма и може да буде фиксан или да варира од једног до другог сценарија. Пример алгоритма је узорковање N циљних тачака тако да површина које њихове околине покривају има највећу сумарну вредност вероватноћа пиксела топлотне мапе. Околина тачке може да буде круг са фиксном величином полупречника који представља параметар алгоритма узорковања. Када је позната циљна тачка трајекторије, онда може да се користи једноставнији модели неуронских мрежа (нпр. потпуно повезане мреже) [4, 11].

Архитектура *HOME* [4] користи принцип за паралелно кодирање растеризоване сцене (статички део) и трајекторија објеката (динамички део). Резултати статичког и динамичког дела модела се спајају конкатенацијом и прослеђују као улаз у декодер за генерисање топлотне мапе. Узорковањем тачака из топлотних мапа се добија скуп циљних тачака. Различити алгоритми узорковања дају боље перформансе модела по различитим метрикама за евалуацију квалитета трајекторија у аутономној вожњи. Независно од модела топлотне мапе се тренира модел који естимира кретање трајекторије када је позната њена крајња тачка.

Могуће је растеризовати потпуно податке тј. растеризовати и трајекторије као низ слика. Свака слика садржи скуп тачака које представљају локације објеката. Архитектура *CASPNet* [12] примењује CNN енкодер на свако стање историје сценарија и користи ConvLSTM [13] модел за разумевање темпоралних веза тих слика. На овај начин се добијају својства динамичког дела сцена. На сличан начин је могуће и извући својства и за статички део (возне линије, возни површине, ...) користећи класичне конволутивне мреже. Комбинацијом ових података се генеришу трајекторије које су исто дате у растеризованом облику. Овакав приступ је рачунски веома скуп.

3.3 Технике засноване на графовским репрезентацијама

Мапе имају комплексну топологију, а технике засноване на растеризацији користе конволуцију која тешко извлачи потпуно семантику тих мапа. Такође, конволуција је скупа операција. Алтернатива је моделовање мапа графовским структурама. Технике засноване на графовском репрезентацијом као улаз добијају стање мапе кодиране као граф и примењују моделе графовских неуронских мрежа. Две архи-

¹Претпоставка важи у случају када је агент возило које у саобраћају има ограничено кретање, док у случају човека који се креће у неком парку то није толико реална претпоставка

текстуре које представљају основе за већину архитектура ове групе су *LaneGCN* [14] и *VectorNet* [3].

Мапа може да се моделује као скуп повезаних сложених линија (*eng. polylines*), где сваком објекту одговара једна усмерена сложена линија. *VectorNet* је хијерархијска графовска неуронска мрежа која као улаз добија мапу која је моделована као скуп сложених линија, а као резултат даје предикцију једне трајекторије за сваког агента.² Идеја је да се прво извуку својста из сложених линија појединачних објеката, а онда пронађу одговарајуће међусобне везе између објеката и везе између објеката и возних линија. За проналазак међусобних веза се користи механизам пажње [7, 3]. Мотивација је следећа: Нису сви објекти у околини значајни за конкретног агента, јер они објекти који су далеко од њега и иду супротном линијом не утичу толико на његово кретање. Нису ни сви путни сегменти значајни за агента. Свакако су значајни путни сегменти који се поклапају са историјом трајекторије агента и путни сегменти по којим агент може потенцијално да се креће у скорој будућности. Задатак модела је да закључи који елементи сцене су значајни.

Архитектура *LaneGCN* нуди варијанту конволутивних графовских мрежа (*GCN - Graph Convolution Network*) [15] која је специјализована за графове путних сегмената које имају различите типове веза. Користе се различите матрице повезаности за суседе, претходнике, следбенике (леви и десни) у контексту путних сегмената. За сваку матрицу повезаности може да се примени класичан GCN, а комбиновањем тих елемената је добија један *LaneGCN* слој [14]. Сам модел се своди на *GCN* са више матрица повезаности.

3.4 Хибридне технике

Хибридне технике користе комбинацију структура графова и BEV слика. Модел *GOHOME* је модификована верзија *HOME* модела који уместо CNN енкодера и растеризоване слике као улазне податке, користи *LaneGCN* архитектуру као енкодер и податке у облику графа. Заправо ансамбл ова два модела *HOME* и *GOHOME* даје најбоље резултате по *MR* метрици.³

3.5 Технике засноване на облацима тачака

Последња група нешто одступа од осталих по броју објављених радова, али и даље даје добре резултате. Подаци се посматрају као облаци тачака (*eng. point cloud*) и примењују се технике намењене за такву структуру података. Основна архитектура

²У овом контексту се издвајају агенти као објекти на сцени за које се врши предвиђање трајекторије.

³Објашњење за ову метрику се налази у секцији о евалуацији

је *TPCN* [16] која је заснована на *PointNet* [17]. Већина осталих техника су „изведене“ из ње.

Глава 4

Метрике за евалуацију модела

Неке од стандардних метрика за евалуацију квалитета предикције трајекторија су „просечна грешка одступања“ (*eng.* ADE - *Average Displacement Error*) и „грешка последњег одступања“ (*eng.* FDE - *Final Displacement Error*). У наставку се користе енглеске скраћенице ADE и FDE . Метрика ADE се рачуна као просечно еуклидско растојање између тачака предикције и истините трајекторије у свим временским тренуцима који одговарају тој трајекторији. Метрика FDE узима у обзир само последњу тачку тј. еуклидско растојање између последње тачке реализације и предикције. Што су метрике ADE и FDE мање након евалуације модела, то је модел квалитетнији. [9, 5]. У наставку се налазе формуле у случају да се посматра тачно један објекат (нпр. само агент):

$$ADE = \frac{1}{T} \sum_{k=1}^T \sqrt{(x_k - \hat{x}_k)^2 + (y_k - \hat{y}_k)^2}$$

$$FDE = \sqrt{(x_{last} - \hat{x}_{last})^2 + (y_{last} - \hat{y}_{last})^2}$$

Метрике се једноставно уопштавају у случајевима где постоји више објеката на сцени:

$$ADE = \frac{1}{T \times N} \sum_{n=1}^N \sum_{k=1}^T \sqrt{(x_k^n - \hat{x}_k^n)^2 + (y_k^n - \hat{y}_k^n)^2}$$

$$FDE = \frac{1}{N} \sum_{n=1}^N \sqrt{(x_{last}^n - \hat{x}_{last}^n)^2 + (y_{last}^n - \hat{y}_{last}^n)^2}$$

Ако узмемо у обзир претпоставку да се релативно детерминистички може одредити трајекторија ако је дата њена последња тачка (што је у случају возила у саобраћају разумна претпоставка), онда су ове две метрике корелисане у смислу да модел који има добар FDE , има и добар ADE (и супротно).

Ове једноставне метрике су погодне под претпоставком да је расподела трајекторија унимодална у односу на дату историју тј. да за сваку историју трајекторије

постији тачно једна смислена предикција. Скупови података трајекторија могу да имају јачу стохастичку природу због природе самог проблема или због непотпуних информација о окружењу. Пример таквог скупа података је скуп трајекторија пешака. Пешак који је прешао пешачки прелаз, може у том тренутку да скрене лево или десно. У том случају имају два вероватна сценарија за исту историју трајекторије, али углавном немамо информације о циљевима самог пешака [9, 18]. У случају возила је ситуација мало блажа због већег броја ограничења, али проблем и даље постоји. На слици 3.1 је већ приказан пример једног таквог сценарија у саобраћају.

Скуп „најбољи из групе“ (*eng.* „*Best of Many*“) метрика узимају у обзир мултимодалну природу расподела трајекторија. Модел може да генерише неколико различитих предикција трајекторија и да за сваку трајекторију да одговарајућу вероватноћу (поузданост). За рачунање грешке се узима предикција која је најбоља по дефинисаном критеријуму. Критеријум не мора да се поклапа са самом мером која се користи тј. не мора нужно да се изабере трајекторија која је најбоља по тој мери, већ за тај избор може да се користи друга мера. [18, 5]. Претходно наведене метрике ADE и FDE се уопштавају у $minADE$ и $minFDE$. Због једноставности узимају се у обзир облици са једним објектом: [19, 18]

$$minADE = ADE(\arg \min_{\hat{A}_k} FDE(\hat{A}_k, A), A), k \in \{1, \dots, K\}$$

$$minFDE = \min FDE(\hat{A}_k, A), k \in \{1, \dots, K\}$$

Овде је A_k истинита трајекторија, а \hat{A}_k трајекторија из скупа предикција. Обе метрике $minFDE$ и $minADE$ се свде на избор трајекторије која најмање одступа од истините трајекторије у последњем тренутку тј. бира се трајекторија која има најмањи FDE .

Уколико модел генерише више од K трајекторија, узима се и обзир првих K по поузданости. У специјалном случају када је $K = 1$, онда $minFDE$ постаје FDE , а $minADE$ постаје ADE . Проблем са $minADE$ и $minFDE$ је у томе што не узимају у обзир остале трајекторије поред најбоље и самим тим се не прави разлика између предикције са свим добрим трајекторијама и предикције са једном добром трајекторијом [19]. Друга замерка овим метрикама је што не узимају у обзир поузданост предикција након филтрирања K трајекторија. Уколико је најбоља трајекторија прецизна, желимо и даље да знамо да ли је модел сигуран или је добар резултат последица „среће“. Модификоване метрике [4]:

$$p-minADE = ADE(\hat{A}_{min}, A) - \ln P(\hat{A}_{min}|E)$$

$$p\text{-minFDE}_{prob} = FDE(\hat{A}_{min}, A) - \ln P(\hat{A}_{min}|E)$$

Овде је \hat{A}_{min} трајекторија која има најбољу FDE оцену, а $p(\hat{A}_{min}|E)$ је условна вероватноћа те трајекторије \hat{A}_{min} у односу на стање окружења E . Уколико је мала грешка по метрици ADE (FDE) за одговарајућу трајекторију, али њена вероватноћа има малу вредност, онда негативан логаритам те вероватноће има велику вредност [5]. У импементацији се ова вероватноћа ограничава са доње стране, како не би дошло до прекорачења због велике апсолутне вредности након примене логаритма на веома мале вредности.

Одступање крајње тачке које је 1 или 2 метра даље од реализације није толико релевантно у односу на одступање од неколико метара [4]. Метрика MR_K^α (*eng. Miss Rate - „проценаи промашаја“*) даје информацију колико су процентуално чести случајеви великих одступања тј. ситуација у којима је промашена мода. Ова метрика се дефинише на следећи начин: Нека је дат мултимодални скуп података димензије N и нека се за сваки елемент из скупа података генерише k предикција (трајекторија). Ако у M случајева крајња тачка најбоље од тих K трајекторија одступа за више од α од истините крајње тачке, онда је $MR_K^\alpha = \frac{M}{N}$. Најбоља трајекторија од K предвиђених је она која је најбоља по FDE метрици тј. крајња тачка те трајекторије је најближа крајњој истинитој тачки. За параметар α се узима вредност $2m$ за *Argoverse* скуп података [5]. Дефиниција метрике се компактно записује на следећи начин:

$$MR_K^\alpha = \frac{\sum_{n=1}^N \min(I(\|\hat{A}_n^k - A_n\|_2 > \alpha) \mid k \in \{1..K\})}{N}$$

Такође постоји верзија метрике која узима у обзир вероватноћу и кажњава предикцију модела ако је добра, а модел је ипак несигуран. [5] Обична MR_K^α метрика рачуна само број тачака трајекторије које одступају много од одговарајуће тачке истините трајекторије, док $MR_K^\alpha_{prob}$ додатно кажњава ситуације у којима не постоји велико одступање, али модел није довољно сигуран.

Све до сада наведене метрике су опште примене на било које објекте за које се предвиђају трајекторије. Пошто је агент углавном возило, може да се анализира да ли предвиђена трајекторија скреће са пута. Због тога се уводи метрика „сагласност са возним подручјем“ (*eng. DAC - Drivable Area Compliance*), која одређује учесталост трајекторија које нису скренуле са пута од изабраних K трајекторија [5]:

$$DAC = \frac{DAC_{occurrences}}{T}$$

У евалуацији модела се узимају у обзир све метрике. За параметар K за $minADE$ и $minFDE$ се узима вредност 6.

Глава 5

Припрема података

Основни скуп података за тренирање и тестирање техника предвиђања трајекторија је *Argoverse Motion Forecasting* скуп података који се састоји од 324 хиљаде сценарија у саобраћају сачиуваних као *HD* мапе. Постоје две групе *HD* мапа за градове Питсбург и Мајами. Коришћењем аутономних возила су генерисани сценарији који представљају неколико узастопних слика сцена (у табеларном формату) на деловима мапа. Сви детаљи о овом скупу података се могу пронаћи на адреси www.argoverse.org [5].

Кључне информације које се издвајају из сваког сценарију су:

- Мапа сценарија (Питсбург или Мајами);
- Трајекторије агената;
- Трајекторије осталих објеката на сцени (суседи);
- Путни сегменти.

У наставку се описује процес прве фазе припреме података. Формат података који се добија након ове фазе је погодан за даљу трансформацију у графовску структуру или слику (растеризован облик).

5.1 *Argoverse* интерфејс

Сирови подаци сваког сценарија се векторизују и чувају у полу-структурираном формату. За парсирање и обраду улазних података се користи *argoverse API* интерфејс. Сваки град се састоји из скупа повезаних путних сегмената који чине један усмерен граф. Функције које се користе из *Argoverse* интерфејса се односе на операције над тим графом. Скуп функција које се користе:

- Одређивање свих путних сегмената који се налазе у околини агента на основу последње тачке у историји трајекторије.
- Одређивање следбеника, претходника и суседа у графу у односу на посматрани путни сегмент.
- Читање метаподатака о путним сегментима: Да ли се путни сегмент налази у неком пресеку, да ли постоји контрола саобраћаја и да ли је путни сегмент усмерен лево, десно или је прав.
- Примена претраге у дубину над графом у односу на посматрани путни сегмент који представља почетни чвор у алгоритму претраге.
- Одређивање путног подручја града (матрица).

Циљ овог корака процесирања података је векторизација података у формат који је независан од модела који се користи тј. обједињавање заједничког дела припреме података.

5.2 Припрема трајекторије агента

Трајекторија агента¹ се дели на два дела: историја (својства) и реализација (будуће вредности). Реализација се састоји од N_r вредности x и y координата тј. облик реализације је $(N_r, 2)$. Историја се аналогно формира да садржи историју N_h опажања x и y координата. Овај део трајекторије иде непосредно пре реализације. Посматрамо следеће случајеве:

- Постоји више од $N_h + N_r$ опажања: Одбацује се реп трајекторије (првих неколико вредности хронолошки гледано);
- Постоји мање од $N_{hmin} + N_r$ опажања: Сценарио се одбацује (сматра се да је невалидан);
- Постоји између $N_{hmin} + N_r$ и $N_h + N_r$ опажања: реп трајекторије који се односи на историју трајекторије се допуњава до димензије $N_h + N_r$ посматрано као да објекат мирује у тим тренуцима.

Овде је N_{hmin} минимални број опажања у историји трајекторије неопходан за вршење предвиђања трајекторија. Уколико је историја трајекторије скоро потпуно непозната, онда није реално да се очекују добре перформансе модела на том сценарију. Филтрирање се односи на скуп података за учење, а током предвиђања је

¹Низ (x, y) тачака, где је приближна временска разлика између две тачке око 0.1 секунди у *Argoverse* скупу података.

свакако неопходно да се да неки резултат. Параметар N_r је постављен на 30, а N_h је постављен на 20. Ове вредности су изабране тако ради усаглашавања са осталим радовима. Параметар N_{hmin} је постављен на 5, али избор вредности је флексибилан. Коначно, облик историје је $(N_h, 3)$, где трећа вредност означава да ли је опажање право (1) или допуњено (0).

Све координате се нормализују тако да су релативне у односу на последње опажање историје агента. Последње опажање историје агента има координату $(0, 0)$ у новом координатном систему, а све остале тачке се транслирају за вектор $-C$, где је C оригинална позиција последњег опажања. Координатни систем се такође ротира тако да се вектор који је одређен првом и последњом тачком историје трајекторије, што чини наивно апроксимиран правац трајекторије, поклапа са y -осом слично као и у оригиналном *Vectornet* раду [3]. Последња трансформација је скалирање свих координата са $\frac{1}{25}$, где је ова вредност добијена анализом стандардних девијација свих тачака на сценаријима (узета је просечна вредност свих сценарија). Ове трансформације су кључне за перформансе модела, јер је сам проблем који се решава једноставнији:

- Применом транслације модел не мора да учи где се агент налази, јер је то увек $(0, 0)$ координата;
- Применом ротације моделу је олакшано учење усмерења трајекторије тј. моделу је лакше да одреди оријентацију агента;
- Применом скалирања се добијају стабилнији градијенти.

Оригинална позиција последњег опажања и угао ротирања се чувају за сваки сценарио. На овај начин је могуће извршити инверзно пресликавање у оригинални координатни систем. За све метрике наведене у претхођој секцији сем DAC је довољно да се само изврши инверзно скалирање, јер су транслација и ротација изометрије које чувају растојања. За метрику DAC је неопходно да се изврши усклађивање са мапом путног подручја.

5.3 Припрема трајекторија суседа

Трајекторије суседних објеката се деле на два дела аналогно трајекторији агента. Неопходно је да се синхронизују трајекторије суседних објеката по временским ознакама (eng. *timestamp*) са трајекторијом агента, јер не постоји у сваком тренутку исти број објеката на сцени. Након синхронизације се трајекторије деле на историју и реализацију и проверава се да ли дужине тих делова задовољавају критеријуме:

- Уколико је дужина трајекторије историје краћа од N_{homin} , онда се објекат одбацује;
- Уколико је дужина трајекторије реализације краћа од N_{romin} , онда се објекат исто одбацује.

Вредности параметара N_{homin} и N_{romin} се постављају на 5 и 3.

Као додатна провера, за сваки сусед се провера растојање од агента. Уколико је сусед превише далеко, онда се он одбацује. Критеријум за одбацивање суседа узима у обзир брзину агента (по x и y оси одвојено) и растојање њихових последњих опажања у трајекторији историје. Уколико неки од следећих услова није испуњен, онда се сусед игнорише у сценарију:

- $\frac{O_n^x}{V_s^x} \leq T_{steps}$
- $\frac{O_n^y}{V_s^y} \leq T_{steps}$

где је O_n^x (O_n^y) нормализована x (y) координата суседа, V_s^x (V_s^y) је наивно апроксимирана брзина агента по x (y) оси² и T_{steps} је параметар толеранције. Трајекторије се секу или допуњавају до фиксног облика (као што је то у случају припреме трајекторије агента). Векторизован облик: $(N_n, N_h, 3)$ за историје трајекторија и $(N_n, N_r, 3)$ за реализацију трајекторија, где је N_n број судедних објеката. Трећа вредност је информација о томе да ли је та тачка допуњена или не.

5.4 Припрема путних сегмената

На основу локације агента се издвајају путни сегменти који су у околини агента са радијусом D_{lsinit} по $L1$ растојању. Уколико нема пронађених сегмената централних линија, онда се вредност за D_{lsinit} множи са K_{ls} ³ до највише D_{lsmax} . Ако и даље нема сегмената, онда се сценарио одбацује, јер није пронађен ниједан путни сегмент у околини. За сваки сегмент се чува низ од 20 (x, y) координата приширених са метаподацима:

- *is_intersection* - да ли се сегмент сече са неким сегментом,
- *turn_right* - да ли је у питању скретање у десно,
- *turn_left* - да ли је у питању скретање у лево,
- *turn_none* - да ли нема стретања,

²Брзина се апроксимира као просек промена координата у трајекторији историје

³ D_{lsinit} и K_{ls} су фиксне вредности у *argoverse* интерфејсу

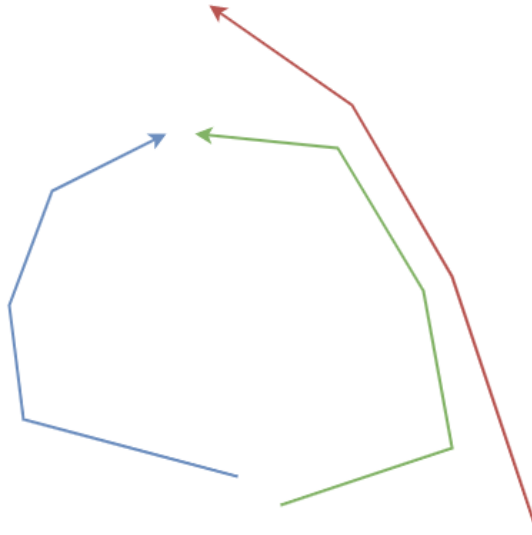
- *is_traffic_control* - да ли постоји контрола саобраћаја.

Коначан облик је $(N_{ls}, 20, 7)$ тј. низ N_{ls} путних сегмената дужине 20, где свака тачка путног сегмената има своје 2 координате и 5 вредности за метаподатаке.

5.5 Припрема кандидата путних сегмената

Постоји коначан број путних сегмената по којој објекат може да се креће у скорој будућности на задатој мапи. Због тога је корисно да се информације о кандидатима путних сегмената користе као улаз у модел или као хеуристике. *Argoverse* интерфејс нуди своју имплементацију за проналазак кандидата путних сегмената. Дат је поједностављен алгоритам 5.1 у *Python* псеудокоду који описује укратко идеју алгоритма.

Овај алгоритам у великом броју случајева погађа кандидате путних сегмената. У теорији је главна мана што узима у обзир само прву и последњу тачку трајекторије агента. На слици 5.2 је направљен вештачки пример где алгоритам ради лоше. Нека је зелена стрелица трајекторија агента и нека су плава и црвена стрелица путни сегменти који се узимају у обзир. Црвени путни сегмент је сличнији трајекторији агента од плавог путног сегмента, али крајеви плавог путног сегмента и крајеви трајекторије агента се скоро савршено поклапају. То значи да се приоритет даје плавом сегменту иако је визуално више има смисла црвени путни сегмент.



Слика 5.2: Пример где *Argoverse* алгоритам за проналазак кандидата не даје тачне одговоре

```

def argoverse_algoritam(trajektorija, avm, dfs_radijus, min_duz, max_duz):
    # trajektorija: istorija trajektorije agenta
    # avm: Argoverse mapa (interfejs ka grafu mape)
    # dfs_radijus: uslov zaustavljanja dfs algoritma
    # min_duz: Minimalna duzina poklapanja trajektorije i putnog segmenta
    # max_duz: Maksimalna duzina poklapanja trajektorije i putnog segmenta

    prva_tacka = trajektorija[0]
    poslednja_tacka = trajektorija[-1]
    putni_segmenti = [] # trenutni putni segmenti
    radijus = 2 # trenutni radijus

    # radijus povecava dok se ne pronadju neki putni segmenti
    # u okolini po L1 rastojanju (okolina je kvadrat)
    while len(putni_segmenti) == 0:
        putni_segmenti = avm.pronadji_putne_segmente_u_okolini(
            poslednja_tacka, radijus)
        radijus *= 2

    # Za svaki putni segment se pronalaze njegovi prethodnici
    # i sledbenici u grafu koji se onda objedinjuju
    # (svaki prethodnik sa svakim sledbenikom)
    potpuni_putni_segmenti = []
    for ps in putni_segmenti:
        sledbenici = avm.dfs(ps, dfs_radijus)
        prethodnici = avm.dfs(ps, dfs_radijus, unazad=True)
        for p in prethodnici:
            for s in sledbenici:
                potpuni_putni_segmenti.append(objedini_segmente(p, s))

    # svi putni segmenti se aproksimiraju parametrizovanom krivom P(t)
    aproksimirani_putni_segmenti = aproksimacija_ps(potpuni_putni_segmenti)

    filtrirani_putni_segmenti = []
    # pronalaze se putni segmenti koji se najbolje poklapaju po duzini
    # sa istorijom trajektorije agenta
    for aps in aproksimirani_putni_segmenti:
        t_pocetak = aps.projekcija(prva_tacka)
        t_kraj = aps.projekcija(poslednja_tacka)
        duzina = t_kraj - t_pocetak
        if duzina >= min_duz and duzina <= max_duz:
            filtrirani_putni_segmenti.append(aps)
    return filtrirani_putni_segmenti

```

Слика 5.1: Поједностављен *Argoverse* алгоритам за проналазак кандидата путних сегмената

Током тренирања *VectorNet* модела који ће бити објашњен у наредној секцији, анализом података је примећено да је један од главних разлога лошијих перформанси заправо слаб квалитет узорковања кандидата путних сегмената. Није ретко да се изабере путни сегмент који припада супротној линији вожње и да због тога предикције буду јако лоше. Из тих разлога је одлучено да се имплементира једноставнији алгоритам који се своди на коришћење *DTW* (*Dynamic Time Warping*) методе и апроксимације угла између трајекторије и путног сегмента. приказан је алгоритам 5.3 као *Python* псеудокод.

Идеја је следећа: Уместо да се узима у обзир само последња тачка, упоређују се све тачке путног сегмента и трајекторије (путни сегмент је такође трајекторија). Коришћење *ADE* метрике није идеално за поређење временских серија, јер то подразумева јака ограничења за асоцијације тачака. Асоцијација тачака се односи на то која тачка прве трајекторије може да се упари са којом тачком друге трајекторије. У случају *ADE* метрике, свака тачка прве трајекторије може да се упари само са једном тачком друге трајекторије. Метода *DTW* одбацује то ограничење и омогућава да свака тачка једне трајекторије може да се упари са било којом тачком друге трајекторије. Остају ограничења да једна тачка једне трајекторије може да се упари са највише једном тачком друге трајекторије (а може и мање) и немогуће је испреплетано упаривање (ако се тачка 5 упари са тачком 8, онда тачка 4 не може да се упари са тачком 9). Метода *DTW* има разне примене над временским серијама, а алгоритам је уопштење Левештајновог растојања за ниске. На слици 5.4 се илуструје начин функционисања методе. У алгоритму се узима најбољих K_c кандидата пре примене *DFS*-а.

```

def dtw_algoritam(trajektorija, avm, dfs_radijus, putni_segmenti,
    k = 3, max_ugao = 90):
    # trajektorija: istorija trajektorije agenta
    # avm: Argoverse mapa (interfejs ka grafu mape)
    # dfs_radijus: uslov zaustavljanja dfs algoritma
    # putni_segmenti: koriste se putni segmenti dobijeni iz prethodnog koraka
    # k: broj najboljih putnih segmenata koji se uzimaju u obzir kao kandidati
    # max_ugao: Maksimalni ugao uzmedju trajektorije
    #           i kandidata putnog segmenta

    # racuna se rastojanje putnog segmenta od trajektorije
    # koristi se dtw metoda sa L2 metrikom
    rastojanje_putni_segment = []
    for ps in putni_segmenti:
        rastojanje = dtw(trajektorija, ps)
        rastojanje_putni_segment.append((rastojanje, ps))

    # sortiranje po rastojanju
    rastojanje_putni_segment = sorted(rastojanje_putni_segment)
    # rastojanje vise nije bitno
    putni_segmenti = [ps for _, ps in rastojanje_putni_segment]
    putni_segmenti = putni_segmenti[:k] # k najboljih

    # za svakog od putnih segmenata se pronalaze sledbenici dfs algoritmom
    kandidati = []
    for ps in putni_segmenti:
        sledbenici = avm.dfs(ps, dfs_radijus)
        kandidati.extend(sledbenici)

    kandidati = list(set(kandidati)) # uklanjanje duplikata

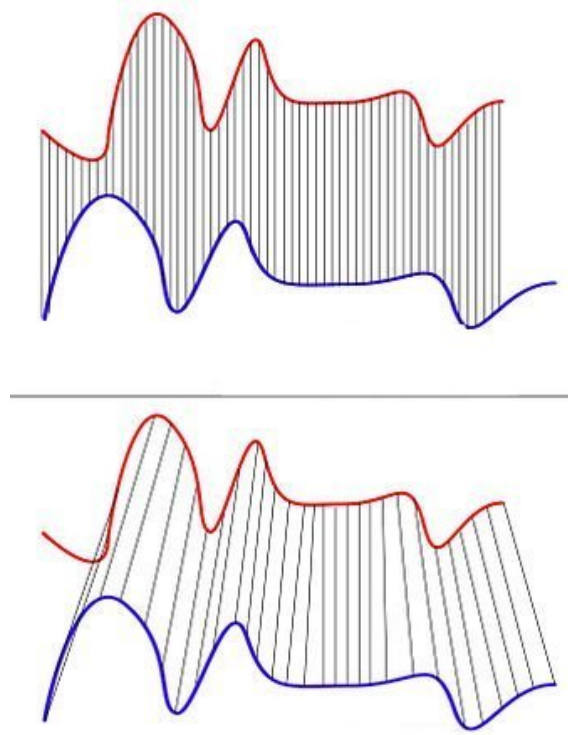
    # kandidati se filtriraju na osnovu ugla izmedju putnog segmenta
    # i trajektorije
    filtrirani_kandidati = []
    traj_pravac = izracunaj_pravac_trajektorije(trajektorija)
    traj_ugao = izracunaj_ugao_pravca(traj_pravac)
    for ps in kandidati:
        ps_pravac = izracunaj_pravac_trajektorije(ps)
        ps_ugao = izracunaj_ugao_pravca(ps_pravac)
        ugao = abs(traj_ugao - ps_ugao)
        if ugao > min_ugao:
            continue

        filtrirani_kandidati.append(ps)

    return filtrirani_kandidati

```

Слика 5.3: Алгоритам заснован на *dtw* метрици и хеуристици угла између правца трајекторија



Слика 5.4: Упоредбање серија помоћу класичног еуклидског растојања (горе) и *DTW* методе (доле)

Упоредбање трајекторија на овај начин није довољно. Проблем прави трајекторија агента на сценама где се он не креће или се веома споро креће. Пошто се *DTW* своди на еуклидско растојање, онда у том случају може да се деси упаривање са најближим путним сегментом. Проблем код тога је што је путни сегмент одређен централном трајекторијом и ако агент није потпуно усклађен са његовим путним сегментом (нпр. искаче мало из своје траке), онда је најближи путни сегмент заправо онај суседни који има супротну траку. Узимање путног сегмента са супротном траком је катастрофално, јер применом *DFS*-а се врши претрага у потпуно погрешном смеру. Због тога се додаје хеуристика која упоређује углове између трајекторија кандидата путних сегмената добијених *DTW* методом и угла трајекторије агента. Угао се одређује као угао вектора правца трајекторије, а вектор правца је одређен првом и последњом тачком трајекторије. Уколико је угао између трајекторије агента и трајекторије путног сегмента већи од θ_{min} степени, онда се тај путни сегмент одбацује. За K_c се узима вредност 3, а за θ_{min} се узима вредност 90.

Коначан векторизован облик је: $(N_c, N_r, 3)$, где је N_c број пронађених кандидата, N_r дужина трајекторије реализације. Пошто се путеви допуњавају по потреби до димензије N_r , користи се трећа координата за маску.

5.6 Преглед дефинисаних параметара

На табели 5.2 се види преглед свих параметара процеса. Параметри N_r и N_h су изабрани тако да се поклапају са осталим радовима. Ако би ти параметри били другачији, онда поређење метрика нема толико смисла. Остали параметри су изабрани интуицијом. Престпоставка је да претрагом параметара може да се повећа квалитет података за учење, а самим тим и квалитет модела. Како је та опција рачунски јако скупа, у овом случају се прескаче.

| Ознака параметра | Објашњење |
|------------------|---|
| N_r | Дужина трајекторије реализације (део који се предвиђа) |
| N_h | Дужина трајекторија историје |
| N_{hmin} | Минимална дужина трајекторије историје пре допуњавања |
| N_{hmin} | Минимална дужина трајекторије историје суседа пре допуњавања |
| N_{rmin} | Минимална дужина трајекторије реализације суседа пре допуњавања |
| T_{steps} | Умножак максималног растојања до сегмента пута |
| D_{lsmax} | Максимално растојање до пута |
| K_c | Број најбољих путних сегмената по DTW методи |
| θ_{min} | Најмањи угао између правца путног сегмента и правца трајекторије агента |

Табела 5.1: Преглед параметара припреме података (TODO: proveriti da neki parametar nije preskocen)

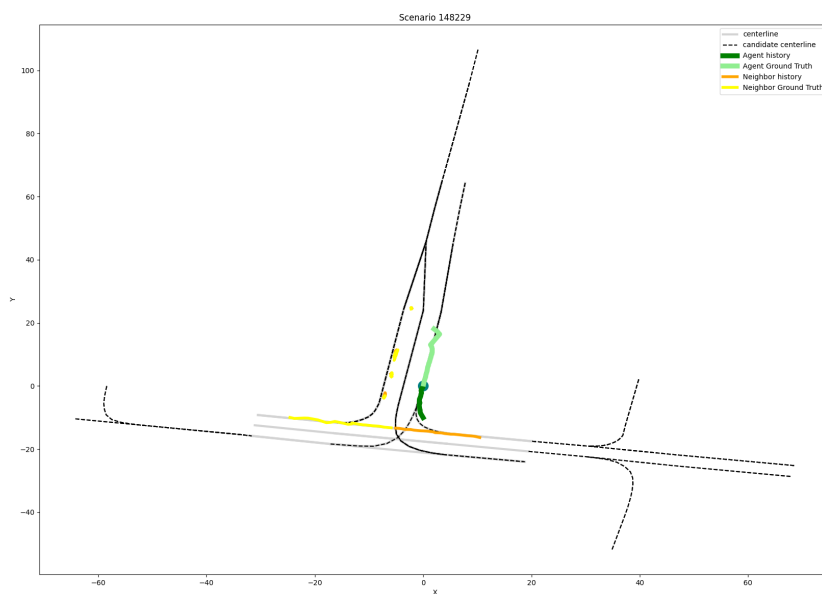
| Ознака параметра | Објашњење |
|------------------|---|
| N_r | Дужина трајекторије реализације (део који се предвиђа) |
| N_h | Дужина трајекторија историје |
| N_{hmin} | Минимална дужина трајекторије историје пре допуњавања |
| N_{hmin} | Минимална дужина трајекторије историје суседа пре допуњавања |
| N_{rmin} | Минимална дужина трајекторије реализације суседа пре допуњавања |
| T_{steps} | Умножак максималног растојања до сегмента пута |
| D_{lsmax} | Максимално растојање до пута |
| K_c | Број најбољих путних сегмената по DTW методи |
| θ_{min} | Најмањи угао између правца путног сегмента и правца трајекторије агента |

Табела 5.2: Преглед параметара припреме података (TODO: proveriti da neki parametar nije preskocen)

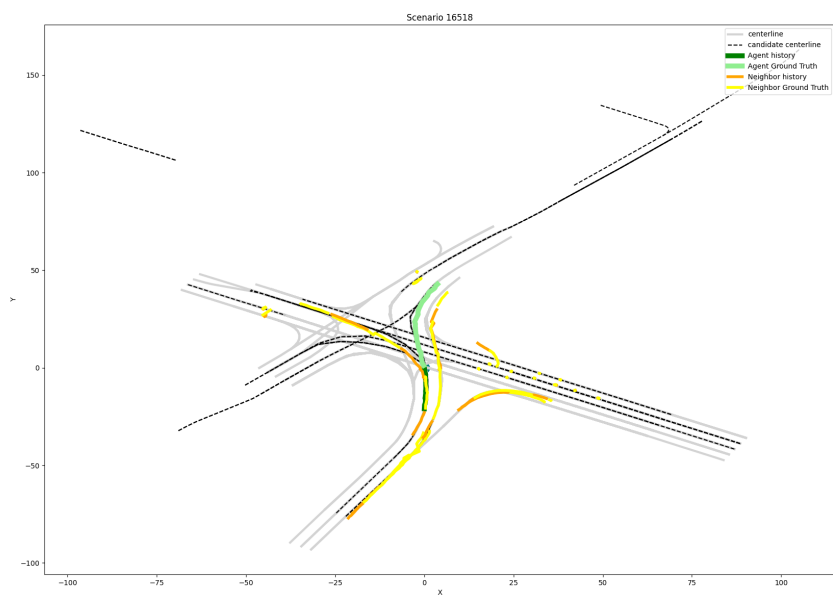
5.7 Визуализација

На сликама 5.5 и 5.6 се налазе примери два визуализована сценарија након претходне припреме. У овом формату нису прикази делови сцене на којој је могућа вожња, али постоје више путања по којим возила могу да се крећу са истом истори-

јијом трајекторије (зависно од циља). Изузеци су у случају неких скретања, промени линија, ...



Слика 5.5: Визуализација припремљених података - Пример 1 (TODO: Ажурирати слике)



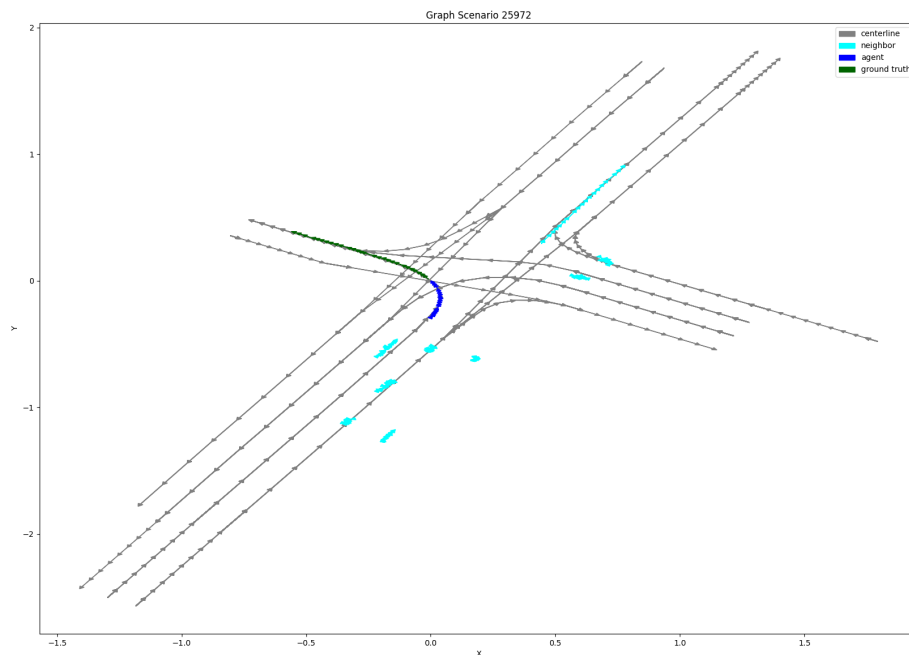
Слика 5.6: Визуализација припремљених података - Пример 2 (TODO: Ажурирати слике)

Глава 6

Техника заснована на разумевању контекста обрадом сцене представљене графом

Независно од конкретног скупа података *HD* мапа, сваки сценарио може да се представи графовском структуром која повезује тачке на сцени, где свака тачка има своја својства. Свака трајекторија може да се посматра као усмерена сложена незатворена линија (*eng. polyline*) тј. низ тачака тако да су сваке две суседне тачке у низу спојене једном усмереном дужи. Тој сложеној линији одговара граф у којем свака тачка чини чвор, а усмерене дужи чине гране. Како већина елемената *HD* мапа може да се представи сложеном линијом, граф чини идеалну структуру за представљање једног сценарија. Граф може да се посматра на два нивоа, где се први, нижи ниво односи на топологији сложених линија, а други, виши ниво се односи на на везе између сложених линија тј. сложене линије чине чворове у том графу. На овај начин је дефинисана хијерархија графа. У наставку се граф на првом нивоу назива подграф, а граф веза између сложених линија се назива глобални граф интеракција (односи се на интеракције између различитих сложених линија на сцени).

Сценарио из *Argoverse* скупа података се састоји из трајекторија агента, трајекторија суседа, путних сегмената и кандидата путних сегмената. Како путни сегмент чини специјалну врсту трајекторија, јасно је да сви елементи сценарија могу да се представе сложеним линијама, а самим тим цео сценарио може да се представи као претходно поменути граф. На слици 6.1 се налази пример једног сценарија представљен преко сложених линија.



Слика 6.1: Визуализација репрезентације сложеним линијама

6.1 Модел *VectorNet*

VectorNet је хијерархијска графова неуронска мрежа која као улаз добија претходно дефинисани граф на два нивоа, а као резултат даје предикције трајекторија за све изабране агенте. Архитектура може да генерише предикције за више агената у исто време и то је погодно искористити у случају да координатни систем сценарија није већ прилагођен конкретном агенту. У супротном је погодније да се користи итеративни приступ генерисања предикција.

У наставку ове и осталих секција ће бити описана комплетна архитектура и начин тренирања ове неуронске мреже. Модел се састоји из три компоненте, чији називи одговарају структури комплетног графа који се користи [3]:

- Подграф: Агрегација сложених линија у један вектор који се даље посматра као чвор у глобалном графу;
- Глобални граф интеракција: Моделовање интеракција на високом нивоу у глобалном графу помоћу механизма пажње;

- Предикција трајекторија за чворове глобалног графа који одговарају агентима.

Подграф

Свака сложена линија на сцени се посматра као посебан граф. Применом варијанте графовске неуронске мреже се издвајају битна својства и везе између чворова. Резултат се на крају агрегира како би се добио један вектор који представља репрезентацију те сложене линије.

Модел подграфа је варијанта *GCN* архитектуре [15] са више слојева. Визуализација једног слоја се види на слици 6.2. Сваки слој може да се представи следећом формулом:

$$v_i^{(l+1)} = \rho_{rel}(g_{enc}(v_i^{(l)}), \rho_{agg}(A, \{g_{enc}(v_j^{(l)})\})))$$

Овде је $v_i^{(l)}$ вектор својства чвора из претходног слоја ($v_i^{(0)}$ су улазна својства), g_{enc} слој за кодирање својства чвора, ρ_{agg} операција агрегације порука добијених од суседних чворова, A је матрица повезаности сложене линије, а ρ_{rel} операција обједињавања агрегираних порука суседа и својства самог чвора. Избор операција је произвољан.

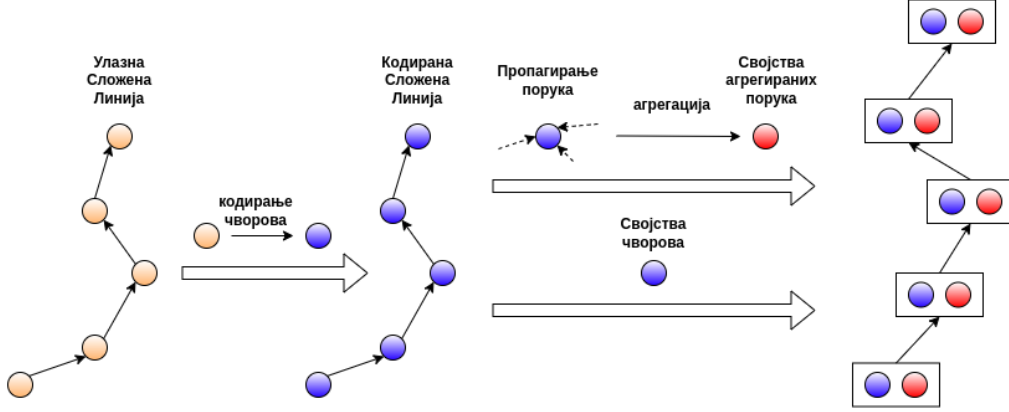
Конкретно [3]:

- g_{enc} је потпуно повезана неуронска мрежа;
- ρ_{agg} је максимум свих порука од суседа (*max pool*) (TODO: тренутно је *avg pool*);
- ρ_{rel} је конкатенација својста чвора и агрегираних порука;
- A је у оригиналном раду матрица повезаности потпуно повезаног графа. Алтернативе су повезаност једног чвора са следећим или свим следећим у сложеној линији.

Вектор сложене линије се добија агрегацијом добијеног графа последњег слоја. За агрегацију се опет користи *max pool* функција:

$$P_{feat} = \rho_{agg}(\{v_i^{(L)}\})$$

На слици 6.2 је визуализован цео процен трансформације који се примећује над подграфом. Процес на слици се односи на само једну итерацију трансформације, али може бити више идентичних итерација.



Слика 6.2: Визуализација једног слоја

Из угла имплементације, улаз у ову компоненту модела је вектор димензије (B, P, T, F) , где је B димензија подскупа података једног корака, P је број сложених линија у једном графу, T је дужина сложене линије, а F је број својстава. Након L слојева се добија вектор димензије $(B, P, T, F \cdot 2^L)$ који се агрегира на нивоу чворова на вектор димензије $(B, P, F \cdot 2^L)^1$. Агрегирану сложену линију посматрамо као један чвор у глобалном графу.

Глобални граф интеракција

Глобални граф интеракција се посматра као класичан граф са матрицом повезаности A преко које може да се дефинише хеуристика, као што је на пример удаљеност чворова на сцени. У оригиналном раду се ради једноставности глобални граф посматра као потпуно повезан граф без тежина [3].

Циљ глобалног графа интеракција је разумевање интеракција између агената и осталих објеката на сцени. Механизмом пажње се примењује над чворовима глобалног графа и проналазе везе између чворова графа [7]. Интуиција је да се на овај начин проналазе путни сегменти који најбоље одговарају будућој трајекторији агента и одређивање суседних објеката који утичу на само кретање агента. Не постоји ограничење на избор неуронске мреже која се примењује на глобални граф.

$$\{P_i^{(l+1)}\} = GNN(\{P_i^{(l)}\}, A)$$

Конкретно у имплементацији се користи следећи облик. Скалирање је опционо.

$$GNN = softmax(\frac{P_Q P_K^T}{\sqrt{d_K}}) P_V$$

¹Претпоставка је да свака сложена линија има исти број својстава и да сваки граф има исти број сложених линија

Овај и претходни слој модела чине енкодер, чији је циљ разумевања контекста сцене на високим нивоу. Последњи слој је декодер који се односи на предикцију трајекторија.

Предикција трајекторија

За предикцију трајекторија се филтрирају чворови агената из глобалног графа интеракције. За сваког агента се примењује произвољан декодер модел чији је излаз предикција трајекторије. Наједноставнији приступ је коришћење потпуно повезане неуронске мреже под претпоставком да су тачке трајекторије међусобно независне. Овај корак се замењује напреднијим приступом у модификованој верзији заснованој на узоркованим циљним тачкама.

Функција грешке

За функцију грешке предикције трајекторија L_{traj} се користи *huber* функција грешке тј. комбинација $L1$ и $L2$ функције грешке (TODO: opisati huber funkciju greske negde.). У процесу тренирања може да се дода задатак **естимације недостајућих чворова глобалног графа интеракција**. Насумично се бирају чворови у графу и замаскирају се његова својства (множењем са нулом). Сваком чвору се додају две „нове“ вредности које су једнаке минималној вредности свакој од координата почетних тачака сложене линије². Нове вредности представљају идентификаторе тих сложених линија који се користе приликом тренирања модела за естимацију недостајућих својстава чворова. Циљ овог задатка је форсирање бољег разумевања веза између трајекторија и генерално веза између сложених линија у графу. Функција грешке овог задатка (L_{node}) се имплементира коришћењем $L1$ функције грешке [3].

$$L = L_{traj} + \alpha \cdot L_{node}$$

6.2 Модификована верзија *VectorNet* са узоркованим крајњим тачкама трајекторија

Претходно поменути верзија *VectorNet* модела може да се унапреди додавањем доменског знања. Један приступ је да се користе унапред дефинисани предлози трајекторија који се користе као основе за генерисање предикција трајекторија. Предлози су фиксни током процеса предвиђања и бирају се на основу анализе података и кластеровањем свих трајекторија [20]. Ова метода је аналогна примени предлога („сидра“ - *eng. anchors*) која се користи код детекције објеката.

²Опис својстава сложених линија је описан у секцији за припрему података

ГЛАВА 6. ТЕХНИКА ЗАСНОВАНА НА РАЗУМЕВАЊУ КОНТЕКСТА ОБРАДОМ СЦЕНЕ ПРЕДСТАВЉЕНЕ ГРАФОМ

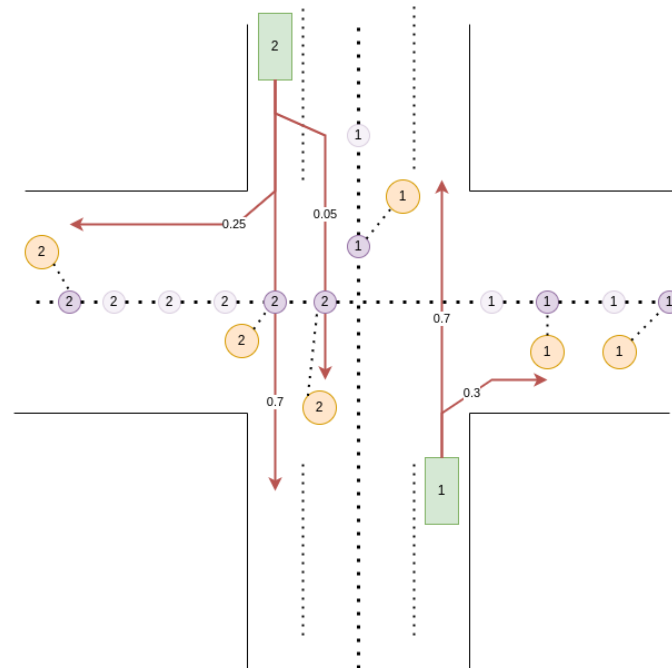
Алтернатива која се исто заснива на предлозима („сидрима“) је да се уместо предлога трајекторија користе предлози крајњих тачака трајекторија тј. циљних тачаја. Модел који генерише трајекторију на основу циљне тачке и својства из *VectorNet* се тренира слично као и основни модел, али квалитет целог система доста зависи од квалитета узорковања тих предлога. Алгоритам за узорковање предлога је већ објашњен у секцији за иницијалну припрему података.

Архитектура *TNT: Target-driveN Trajectory prediction* [10] се заснива баш на тој идеји и подразумева следеће кораке:

1. Разумевање контекста помоћу *VectorNet* модела (основа архитектуре);
2. Генерисање корекција (eng. offsets) и поузданости за сваки предлог циљне тачке трајекторије;
3. Узорковање модификованих предлога на основу поузданости;
4. Естимација трајекторије до сваког изабраног предлога крајње тачке.
5. Естимација вероватноће за сваку од добијену трајектору (TODO: Trenutno se koriste verovatnoce krajnjih tacaka);
6. Филтрирање трајекторија на основу вероватноћа (TODO).

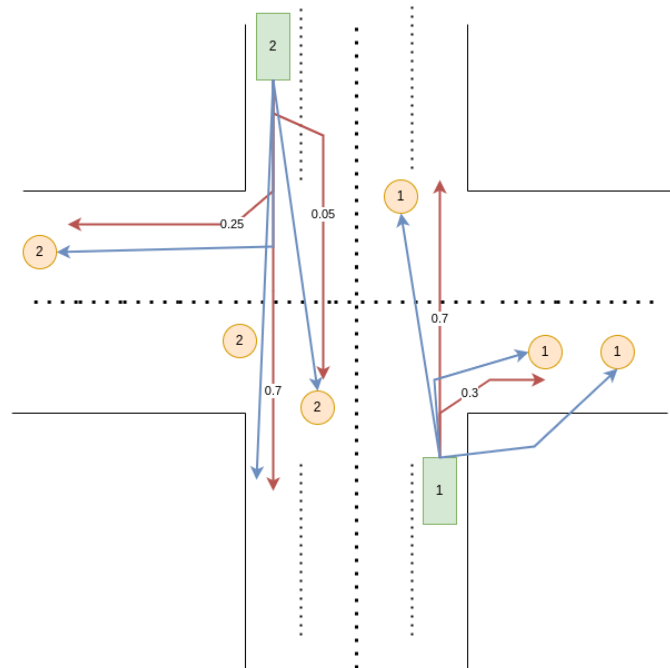
На слици 6.3 су представљени кораци 2 и 3 под претпоставком да је корак 1 претходно извршен. На сцени се налазе два агента 1 и 2. Мали љубичасти кругови се односе на предлоге циљних тачака који су узорковани са кандидата путних сегмената у процесу припреме података (TODO: Opis pripreme је на крају, али вероватно треба гатије...). Сви се налазе на централним линијама, јер се узоркују са центра путних сегмената. Сви кругови су означени тако да је јасно ком агенту припадају. На основу поузданости је извршено филтрирање иницијалних предлога, због чега су неки кругови избељени. За остале кругове постоје корекције³ (наранџасти кругови спојени са иницијалним предлогом).

³У имплементацији се у једном кораку генеришу корекција и поузданост за све узорковане циљне тачке, али овде су због прегледности избачене корекције за одбачене предлоге.



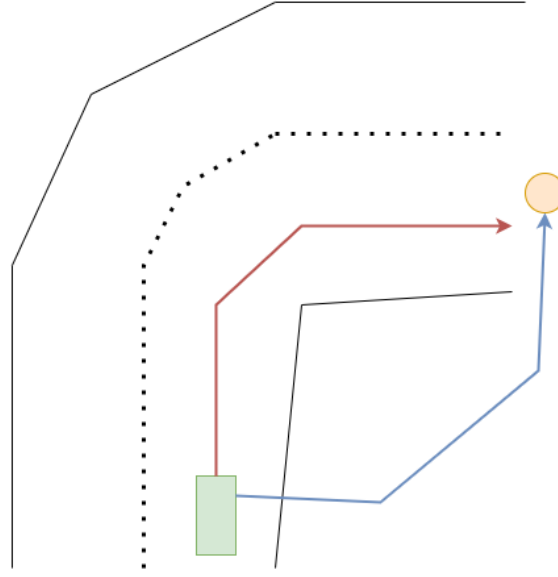
Слика 6.3: Визуализација *TNT* - фаза 1

На основу ових коначних циљних тачака се врши предвиђање трајекторија и то је приказано на слици 6.4. Крај предвиђене трајекторије не мора нужно да се поклапа са циљном тачком.



Слика 6.4: Визуализација *TNT* - фаза 2

Уместо последња два корака могу да се користе вероватноће (поузданости) предлога крајњих тачака, али из добре крајње тачке не следи нужно квалитетна трајекторија (нпр. уколико предложена трајекторија прелази преко тротоара). На слици 6.5 је дат такав пример. Овде је зелени правоугаоник агент, црвена трајекторија је истинита вредност трајекторије која се предвиђа, наранџасти круг је циљна тачка, а плава трајекторија је предикција.



Слика 6.5: Пример случаја где је добра циљна тачка, а лоша трајекторија

Овде *VectorNet* чини језгро архитектуре, а сви наредни кораци могу да се имплементирају потпуно повезаним неуронским мрежама уз комбинацију једноставних детерминистичких алгоритама. Главни изазов ове архитектуре је у алгоритму за генерисање предлога⁴ и балансирању параметара функција грешака приликом учења модела.

Ако се не учи *есџимације недостиђајућих чворова глобалног графа интеракција* уз *VectorNet*, онда функција грешке има следећи облик:

$$L = \lambda_1 \cdot L_{offsets} + \lambda_2 \cdot L_{tarconf} + \lambda_3 \cdot L_{trajde} + \lambda_4 \cdot L_{trajconf}$$

Нека је $P_{anchors}$ скуп свих предлога крајњих тачака трајекторија и нека је P_{gt} истинита крајња тачка трајекторије. Тада се изваја из скупа $P_{anchors}$ елемент $P_{closest}$ који је најближи истинитој крајњој тачки⁵. Тада је:

$$L_{offsets} = H_{delta}(P_{closest_offset}, \hat{P}_{closest_offset})$$

где је H_{delta} *huber* функција грешке са параметром $delta$, $P_{closest_offset} := P_{gt} - P_{closest}$, а $\hat{P}_{closest_offset}$ је предикција тог одступања. Циљ је да баш тај најближи предлог има највећу поузданост, па се функција грешке за поузданост дефинише на следећи начин:

⁴Алгоритам је укратко описан у секцији за иницијалну припрему података

⁵Овде не узимамо у обзир поправке, већ нетрансформисане, узорковане вредности. Уколико се у обзир узимају вредности са поправкама, онда процес учења постаје тежак због шума (мења се индекс најближе тачке). Ово можда не би био толики проблем да су остале трајекторије другачије дефинисане тј. независне од промене изабране, најближе тачке.

$$L_{tarconf} = BCE(P_{closest_onehot}, \hat{P}_{confs})$$

где BCE је бинарна унакрсна ентропија, $P_{closest_onehot}$ индекс елемента $P_{closest}$ у $onehot$ формату⁶ и \hat{P}_{confs} поузданост модела за сваки предлог (за сваки предлог се даје поузданост из интервала $[0, 1]$). Функција грешке за трајекторије је аналогна као и за предлога крајњих тачака. Састоји се из функције грешке за одступање трајекторија од реализације и оцене поузданости модела за сваку од тих трајекторија.

$$L_{trajde} = H_{delta}(T_{traj}, \hat{T}_{traj})$$

где је T_{traj} истинита вредност трајекторије, а \hat{T}_{traj} је њена предикција. Приликом учења се за естимацију трајекторије \hat{T}_{traj} узима истинита крајња тачка трајекторије како би учење било стабилније. Ова техника се зове „Учитељско форсирање“ [21]. Последња компонента се односи на поузданост модела за сваку естимацију трајекторија. За сваку естимирану трајекторију се рачуна максимално растојање између свих упарених тачака естимираних трајекторије и истините вредности трајекторије тј. $D(T_{traj}, \hat{T}_{traj}) := \max(\|T_{traj}^k - \hat{T}_{traj}^k\|_2^2)$. Тада се „истинита расподела“ P_{traj_confs} одређује као $softmax$ ових негативних вредности.

$$L_{trajconf} = BCE(P_{traj_confs}, \hat{P}_{traj_confs})$$

6.3 Припрема података

Припрема података у овој секцији се наставља на иницијални процес претпроцесирања података и своди се на трансформацију свих елемената HD мапа у сложене линије. Да би било могуће спојити више сценарија у један подскуп података за једну итерацију тренирања, неопходно је да се испуне одређена ограничења:

- Сваки сценарио мора да има исти број сложених линија (N_p). Ако је број сложених линија већи од N_p , онда се вишак сложених линија одбацује, при чему се води рачуна о приоритету сложених линија: агент, остали објекти, путеви, путеви кандидати. Ако је број сложених линија мањи од N_p , онда се скуп допуњава сложеним линијама са нула чворовима.
- Свака сложена линија мора да има исти број чворова (N_n). Ово се решава аналогно броју сложених линија.
- Сваки чвор сложених линија мора да има исти број својстава (N_f). Сваки чвор има иста својства, са тим да се својства допуњавају нулама ако немају смисла за тај тим сложених линија (агент и путеви немају иста својства).

⁶Индекс се представља као низ димензије броја класа (предложених крајњих тачака у овом случају) где су свуда нуле сем на локацији која одговара вредности тог индекса.

Сваки чвор сложене линије има следећа својства:

- Координате x и y : Просек претходне и тренутне тачке трајекторије - 2 скалара;
- Смер по x и y : Разлика тренутне и претходне тачке трајекторије - 2 скалара;
- Тип објекта као *onehot* вектор (агент, сусед - објекат, пут, пут кандидат) - 4 скалара;
- Метаподаци путних чворова (нуле у случају трајекторија агента и објеката суседа)
 - Да ли се сегмент пресеца са неким другим сегментом - 1 скалар;
 - Да ли постоји контрола саобраћаја - 1 скалар;
 - Смер као *onehot* вектор (нема, десно, лево) - 3 скалара
- Да ли је чвор прави (1) или вештачки (0) - 1 скалар.

Коначан формат улазних података у *VectorNet* је $(B, N_p, N_n, 14)$, где је B број сценарија у једном подскупу података, а N_p и N_n су параметри.

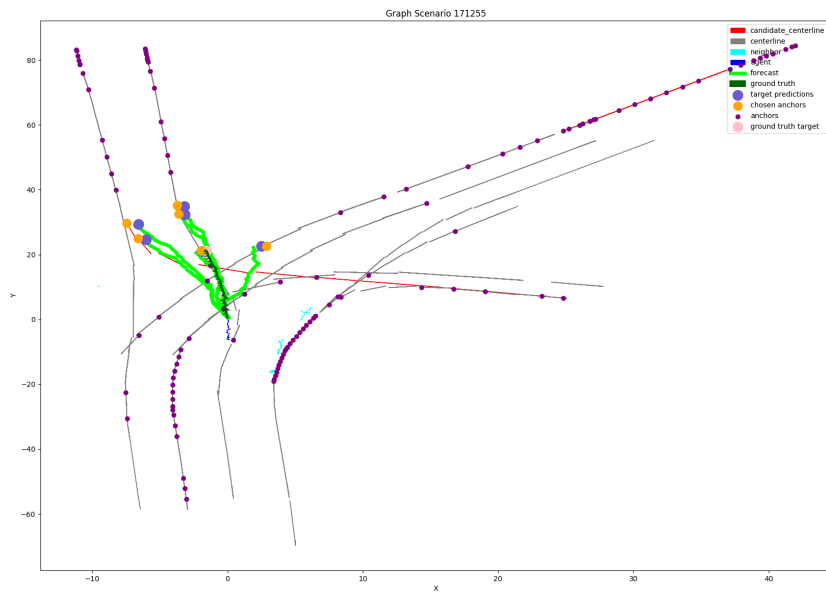
6.4 Експерименти и резултати

На слици 6.6 је приказан пример примене *TNT-VectorNet* архитектуре.

ГЛАВА 6. ТЕХНИКА ЗАСНОВАНА НА РАЗУМЕВАЊУ КОНТЕКСТА
ОБРАДОМ СЦЕНЕ ПРЕДСТАВЉЕНЕ ГРАФОМ

| Назив модела | $minADE$ (6) | $minFDE$ (6) |
|--|--------------|--------------|
| <i>TNT-original</i> без оцене трајекторија | 0.877 | 1.632 |
| <i>TNT-original</i> са оценом трајекторија | 0.728 | 1.292 |
| <i>TNT-custom</i> без оцене трајекторија | 1.19 | 2.25 |

Табела 6.1: Резултати



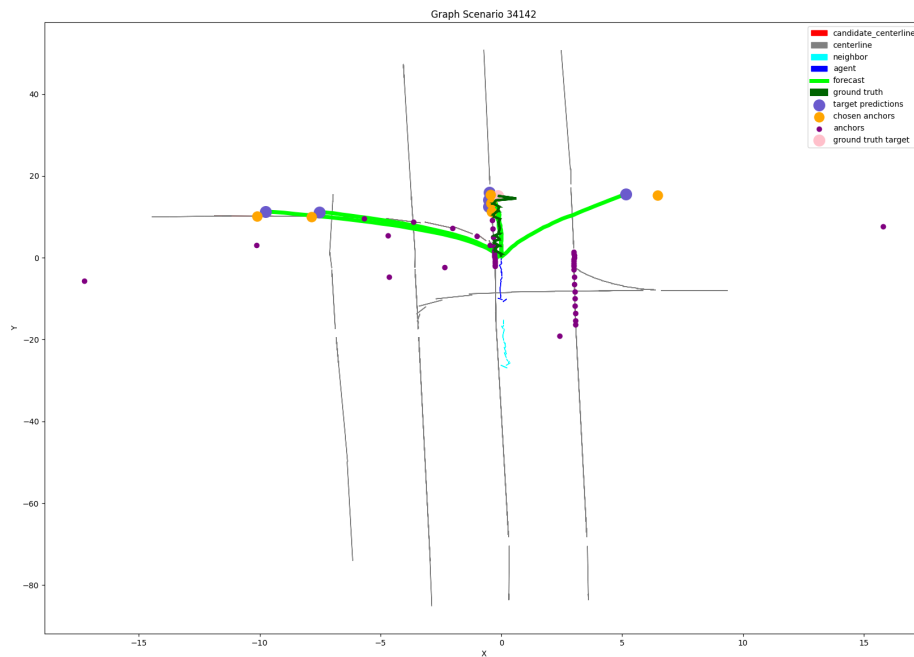
Слика 6.6: Пример резултата за сценарио 171255 - $minADE_6 = 0.6$, $minFDE_6 = 1.0$. Сивом бојом су обојени путеви, тамно плавом је обојена историја трајекторија агента, светло плавом су обојене историје суседних објеката, црвеном линијом су обојени путеви кандидати. Мали љубичасти кругови су узорковани предлози, велики плави кругови су изабрани узорковани предлози са највећом поузданошћу, а велики наранџасти кругови су модификовани изабрани предлози. Светло зеленом бојом су представљене предикције за сваку од изабраних крајњих тачака, а тамно зеленом бојом су обојени истинити резултати.

У табели Table 6.1 се налазе резултати репродукованих и модификованих верзија (TODO) *TNT-VectorNet* модела. TODO: Тренутни модел је трениран укупно 1h над свим подацима...

У наставку су изабране слике које укратко описују неке од врлине и мана основног модела. На слици 6.7 модел даје 6 трајекторије за 6 изабраних по моделу најпоузданих тачака од 50 генерисаних (исти параметри су и за остале слике). Три трајекторије се завршавају у околини истините крајње тачке, што даје добре резултате на овом конкретном сценарију. Поред тога постоје 2 трајекторије које одговарају

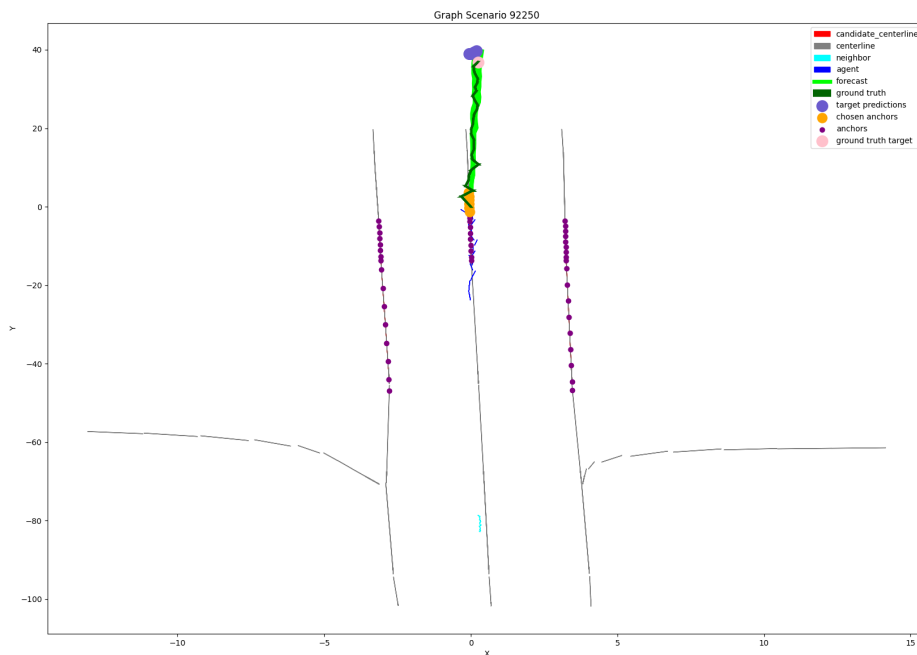
ГЛАВА 6. ТЕХНИКА ЗАСНОВАНА НА РАЗУМЕВАЊУ КОНТЕКСТА ОБРАДОМ СЦЕНЕ ПРЕДСТАВЉЕНЕ ГРАФОМ

случају скретања и то представља добар одговор на изазов мултимодалности. Последња трајекторије нема смислену крајњу тачку и скреће са пута. Трајекторије су углавном исправне под условом да је крајња тачка добро изабрана.



Слика 6.7: Предикције на сценарију *MIA-34142*

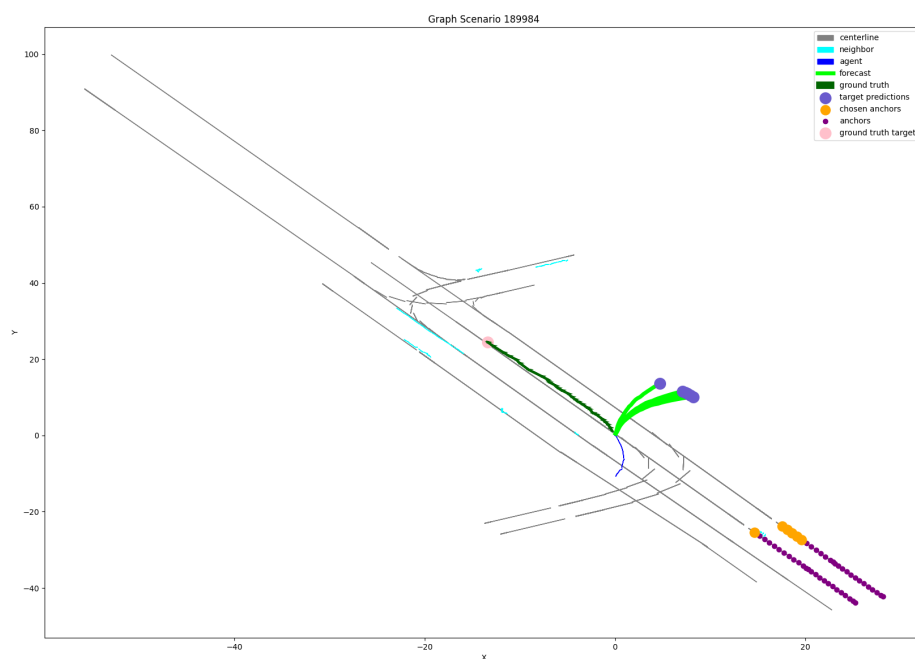
Сценарио на слици 6.8 приказује случај где су предлози крајњих тачака јако лоши (део сценарија је лоше исечен), али модел и даље успева да модификује те крајње тачке (претпоставка је да је то на основу трајекторије историје) и да добар резултат.



Слика 6.8: Предикције на сценарију *MIA-92250*

Сценарио на слици 6.9 приказује јако лоше одговоре који су последица лошег алгоритма узорковања предлога (кандидата). Мана наивног приступа, где се за путни сегмент агента одређује онај који је најближи последњој тачки трајекторије. У случаје грешке се добијају кандидати на путевима који су усмерени на супротну страну. Квалитет модела значајно зависи од алгоритма узорковања иницијалних предлога крајњих тачака.

ГЛАВА 6. ТЕХНИКА ЗАСНОВАНА НА РАЗУМЕВАЊУ КОНТЕКСТА
ОБРАДОМ СЦЕНЕ ПРЕДСТАВЉЕНЕ ГРАФОМ



Слика 6.9: Предикције на сценарију *MIA-189984*

Глава 7

Техника заснована на разумевању контекста обрадом растеризоване сцене

Алтернативни приступ структурирања података *HD* мапа је њихова растеризација. Како су коришћени подаци сцене из птичје перспективе тј. дводимензионалном облику, онда је природно да се мапе посматрају као слике. Стандардне слике се представљају као матрица пиксела, где сваки пиксел има једну вредност (један канал) у случају црно белих слика или три вредности (три канала) у случају да су слике у *RGB* формату којим се представља боја пиксела. У случају *HD* мапа, не користе се боје, већ се сваком пикселу додељују својства сцене као што су: да ли се на датој позицији налази агент, да ли постоји контрола саобраћаја, да ли је могућа возња на тој локацији и слично. Процес трансформације *HD* мапе у слику се у наставку реферише као растеризација *HD* мапе.

Када су подаци у облику слика, тада је могућа примена конволутивних неуронских мрежа са циљем извлачења контекста из растера. У овој секцији се примећује *HOME* [4] архитектура која имплементира предикцију трајекторија кроз у три фазе: генерисање топлотне мапе вероватноћа крајњих тачака трајекторије, узорковање крајњих тачака на основу топлотних мапа и естимација трајекторије до задате крајње тачке.

У наставку се описује припрема података. Након припреме података је детаљно описана архитектура.

7.1 Растеризација *HD* мапа

Овај корак припреме података се наставља на иницијалну припрему и секцији 5 исто као и припрема података за *Vectornet* модел. Овај корак припреме захтева онлајн приступ припреме због димензије података након конверзије.¹ Растеризован облик чини низ ретких матрица (већина вредности је нула), због чега он може да буде и до 200 пута веће димензије од основног облика. За чување целог свих слика је потребно око $4TB$. Овакав начин припреме не утиче значајно на дужину трајања једне епохе током тренирања. Иницијална припрема се и даље врши офлајн. Исти приступ је могуће искористити и за *VectorNet*, али у том случају није од толиког значаја.

У *Argoverse* скупу података су познате комплетне сцене градова, а за генерисање предикција конкретног агента у конкретном тренутку је потребан само један исечак града у облику квадрата центриран у односу на агента. Димензија тог квадрата је значајан параметар за припрему података и сам модел. Квадрат треба да буде довољно велики да обухвати скоро сваку трајекторију у скупу података. Дужина већина трајекторија је краћа од 56 метара што чини 112 пиксела у *Argoverse* скупу података. Због тога се за исечак узима квадрат димензије (224, 224) центриран у односу на последњу познату локацију агента. Овај исечак са издвојеним одговарајућим својствима у сваком пикселу чини улаз у модел за генерисање топлотне мапе вероватноћа крајњих тачака. У имплементацији страница квадрата представља параметар који је подесив.

Коначан облик једне слике је $(9, 224, 224)^2$, где је 9 низ својстава пиксела. Сваки пиксел садржи следеће информације:

- Да ли локација пиксела припада путу тј. да ли је могућа вожња на тој локацији. Информација се бележи бинарно, где 1 означава да је на тој локацији могућа вожња, а 0 да није могућа вожња.
- Растеризована трајекторија агента: Трајекторија агента се растеризује тако што се посматра локација агента у сваком тренутку који је забележен у историји трајекторије агента. Тој локацији тј. тим координатама се додељује правоугаоник димензије (S_x, S_y) . Сваки пиксел који припада том квадрату добија вредност 1 за својство растеризоване трајекторије агента. У оригиналној имплементацији се за сваки временски тренутак чува засебна трајекторија. Ради уштеде меморије се користи само једно заједничко својство за све тренутке.
- Растеризоване трајекторије суседа: Аналогно агенту се растеризују трајекторија осталих објеката на сцени.

¹онлајн припрема подразумева да се подаци припремају упоредо са тренирањем модела.

²Слика се представља у (C, H, W) што одговара *PyTorch* формату.

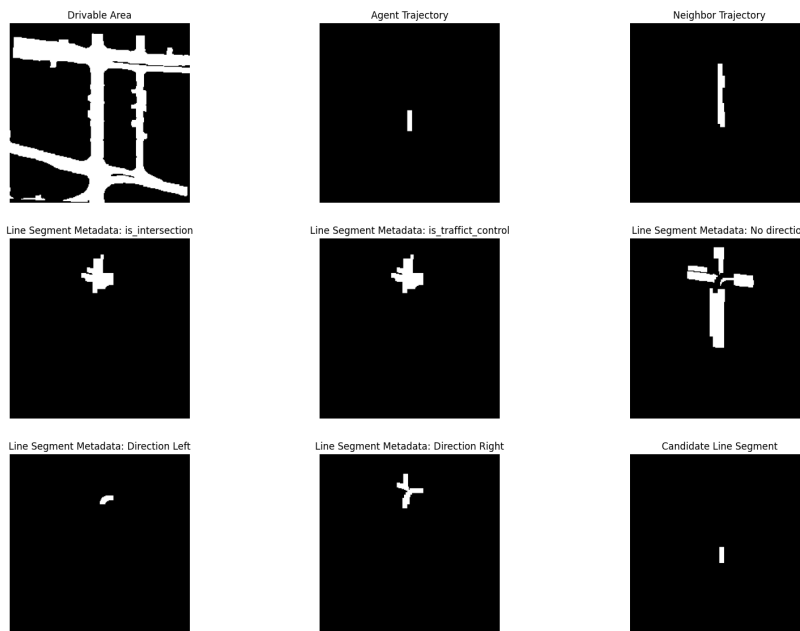
ГЛАВА 7. ТЕХНИКА ЗАСНОВАНА НА РАЗУМЕВАЊУ КОНТЕКСТА ОБРАДОМ РАСТЕРИЗОВАНЕ СЦЕНЕ

- Метоподаци путних сегмената: За сваки путни сегмент је дато 5 бинарних својстава:
 - да ли путни сегмент припада пресеку путних сегмената;
 - да ли путни сегмент обухвата контролу саобраћаја;
 - да ли путни сегмент подразумева скретање у десно;
 - да ли путни сегмент подразумева скретање у лево;
 - да ли путни сегмент не подразумева скретање;

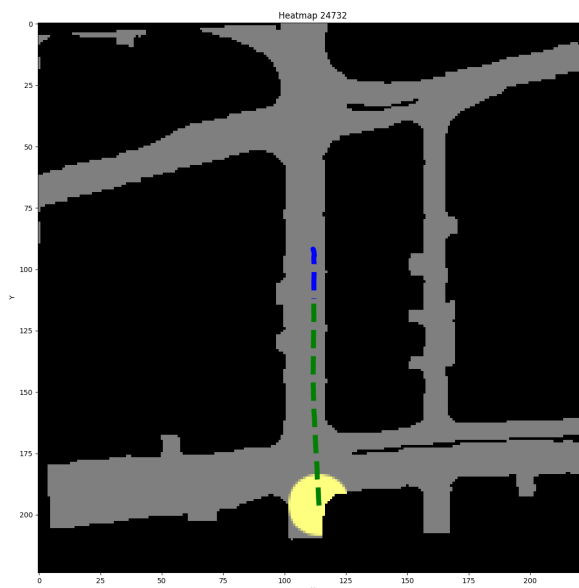
Путни сегменти се посматрају као специјална врста трајекторија, па се растеризују аналогно трајекторији агента.

- Растеризовани кандидата путних сегмената: Кандидати путних сегмената на основу којих се узоркују предлози крајњих тачака за *VectorNet* се у овом случају користе као својства.

На слици 7.1 је визуализована свако од претходно наведених својстава. Белом бојом су обојени пиксели који имају вредност 1, а необојени тј. црни пиксели имају вредност 0. Возни пут је такође ротиран тако да одговара координатном систему агента. На слици 7.2 је приказан пример топлотне мапе истог сценарија.



Слика 7.1: Растеризована својства на сценарију *MIA-24732*



Слика 7.2: Топлотна мапа крајње тачке трајекторије агента на сценарију *MIA-24732*

Поред припреме улазних података, се припрема и истините вредности топлотних мапа вероватноћа крајњих тачака. Истинита крајња тачка трајекторије агента припада тачно једном пикселу, али непосредно суседни пиксели такође чине добре предикције. У том случају да се користи само један пиксел за истиниту вредност, модел се подједнако кажњава ако изабаре најгори могући пиксел као и када изабере непосредно суседни пиксел. Из тог разлога се примећује гауовом кернел након постављене истините крајње тачке топлотне мапе. Облик топлотне мапе је $(1, 224, 224)$ тј. аналогно улазним подацима. Опционо, топлотна мапа која се добија као излаз модела може да буде мање димензије, а да се накнадно скалира на одговарајућу димензију.

Архитектура *HOME* поред слике користи и историју агента и осталих објеката у облику временских серија. Ове две структуре података се комбинују.

7.2 Модел *HOME*

Модел *HOME* (*Heatmap Output for future Motion Estimation*) [4] као улаз узима растеризоване слике сцене и историју трајекторије агента, а као излаз даје произвољан број предикција за тај сценарио са њиховим вероватноћама. Процес предвиђања се састоји из три корака:

ГЛАВА 7. ТЕХНИКА ЗАСНОВАНА НА РАЗУМЕВАЊУ КОНТЕКСТА ОБРАДОМ РАСТЕРИЗОВАНЕ СЦЕНЕ

- Генерисање топлотне мапе вероватноћа крајње тачке трајекторије агента;
- Узорковање крајњих тачака са топлотне мапе;
- Естимација трајекторије на основу узоркованих крајњих тачака.

Основна мотивација се поклапа са *TNT-VectorNet* моделом тј. сматра се да ако се одреди крајња тачка трајекторије, онда применом једноставног модела који узима у обзир ту крајњу тачку и историју трајекторије може тачно да се естимира трајекторија до те крајње тачке. Зато се и последњи корак тј. естимација трајекторија на основу узоркованих крајњих тачака поклапа.

Генерисање топлотне мапе вероватноћа крајње тачке трајекторије агента

Прва фаза подразумева генерисање топлотне мапе вероватноћа крајње тачке трајекторије агента. Овај приступ је инспирисан *CenterNet* моделу за детекцију објеката. Укратко, идеја *CenterNet* модела је да да као излаз топлотну мапу вероватноћа центара објеката тј. сваком пикселу је додељена вероватноћа да се ту налази центар неког објеката. Узорковањем мода топлотне мапе се добијају центри објеката, а одговарајући правоугаоници који обухватају објекат се добијају естимирањем ширине и висине. Аналогно, у овом случају пиксели дају вероватноћу крајње тачке трајекторије уместо центра неког објекта на слици.

Архитектура је енкодер-декодер, где се паралелно обрађују растеризована својства *HD* мапе и својства историје трајекторија у основном облику. Обрада растеризованих својства одговара растер енкодеру тј. енкодеру статичког контекста сцене, док обрада историје трајекторија одговара темпоралном енкодеру тј. енкодеру динамичког контекста сцене.

(TODO. izopstiti brojeve)

Растер енкодер се састоји из низа парова слојева конволутивне неуронске мреже и *maxpool* операције. Конволутивним мрежама се издвајају тополошка својства са сцене који се агрегирају *maxpool* операцијом након које је димензија слике дупло мања до ширине и висине. Сваки пар ових. Улаз у енкодер је је припремљен растер димензије (9, 224, 224) на који се четири пута примењује пар конволутивног слоја и *maxpool* слоја што као излаз даје слику димензије (256, 14, 14). Последњи слој растер енкодера је конволутивни слој који даје слику димензије (512, 14, 14). Излаз из растер енкодера треба да обухвати статичке информације сцене. [4]

(TODO: dodati mehanizam paznje)

Енкодер трајекторија је *LSTM* рекурентна неуронска мрежа након које следи један линеарни слој. Улаз у енкодер трајекторија је трајекторија димензије (20, 3), где је 20 дужина историје трајекторије агента, а 3 је број својстава који садржи x ,

ГЛАВА 7. ТЕХНИКА ЗАСНОВАНА НА РАЗУМЕВАЊУ КОНТЕКСТА ОБРАДОМ РАСТЕРИЗОВАНЕ СЦЕНЕ

у координату заједно са информацијом да ли је трај временски тренутак допуњен (бинарна вредност) у случају да је оригинална дужина трајекторије мања од 20. Излаз из мреже је облика (128) који треба да обухвати динамичке информације сцене. [4]

Излаз растер енкодера и енкодера трајекторија се комбинују тако што се излаз из енкодера трајекторија понавља 14×14 пута како би се добила димензија (128, 14, 14). Овако добијен облик може да се налепи на излаз растер енкодера чиме се добија слика димензије (640, 14, 14). Укратко, сваком пикселу се додају динамичке информације сцене добијене из енкодера трајекторија.

Спојени излаз енкодера представља излаз у декодер који као излаз даје топлотну мапу димензије (1, 224, 224), где сваки пиксел одговара вероватноћи да се на тој локацији налази крајња тачка трајекторије агента. Декодер се имплементира као низ транспонованих конволутивних слојева, где сваки слој увећава димензију претходног улаза два пута по ширини и дужини. Слагањем 4 оваква слоја се добија излаз који је 16 пута већи по дужини и ширини тј. добија је жељени излаз (1, 224, 224). На крајњи излаз се такође примењује сигмоид функција како би вредност сваког пиксела припадала интервалу $[0, 1]$ тј. како би вредности излаз представљале валидне вероватноће које се у овом случају моделују.

Функција грешке за генерисање топлотних мапа

За учење се користи модификована *Focal Loss* функција грешке као и у *CenterNet* раду [11, 22]. Ова функција грешке представља модификацију унакрсне ентропије која је погоднија у случају екстремног небалансираног скупа података између позадинских и непозадинских класа. У основном облику, функција грешке унакрсне ентропије не разликује негативне и позитивне класе. У случају небалансираног скупа података, често постоји велики број лакших негативних примера који доминирају утицајем над резултајућим графицијентом функције грешке током учења. Стандардни облик функције грешке унакрсне ентропије је:

$$CE(\hat{p}, p) = \begin{cases} -\log(\hat{p}), & p = 1 \\ -\log(1 - \hat{p}), & p \neq 1 \end{cases}$$

Функција грешке *Focal Loss* додаје фактор $(1 - \hat{p})^\alpha$ за позитивну класу и фактор \hat{p}^β за негативну класу, где су α и β хиперпараметри. [22]

$$FL(\hat{p}, p) = \begin{cases} -(1 - \hat{p})^\alpha \cdot \log(\hat{p}), & p = 1 \\ -\hat{p}^\beta \cdot \log(1 - \hat{p}), & p \neq 1 \end{cases}$$

ГЛАВА 7. ТЕХНИКА ЗАСНОВАНА НА РАЗУМЕВАЊУ КОНТЕКСТА ОБРАДОМ РАСТЕРИЗОВАНЕ СЦЕНЕ

На конкретном примеру је јасан ефекат ових параметара. Нека је $\alpha = 4, p = 1$. У табели 7.1 су приказане вредности обе функције грешке за различите излазе модела. У случају великог броја лаких позитивних примера, очекује се да модел може врло брзо да научи да препознаје те позитивне примере и да буде веома сигуран приликом предвиђања тј. да вредности вероватноћа буду велике. Ти примери у случају *Focal Loss* функције грешке у даљем тренирању имају значајно мањи утицај на вредности градијента (погледати последње редове у табели). Модел и даље наставља да учи над позитивним примерима које је класификовао као негативне (погледати прве редове у табели).

| \hat{p} | $CE(\hat{p}, p)$ | $FL(\hat{p}, p)$ |
|-----------|------------------|------------------|
| 0.1 | 2.3026 | 1.5107 |
| 0.2 | 1.6094 | 0.6592 |
| 0.3 | 1.2040 | 0.2891 |
| 0.4 | 0.9163 | 0.1188 |
| 0.5 | 0.6931 | 0.0433 |
| 0.6 | 0.5108 | 0.0131 |
| 0.7 | 0.3567 | 0.0029 |
| 0.8 | 0.2231 | 0.0004 |
| 0.9 | 0.1054 | 0.0000 |

Табела 7.1: Анализа утицаја *Focal Loss* фактора на функцију грешке

У случају проблема генерисања топлотних мапа крајњих тачака трајекторија, постоји тачно један пиксел који одговара истинитој трајекторији агента и који припада позитивној класи, док сви остали пиксели припадају негативној класи (нула пиксели). Ако је слика топлотне мапа 224×224 , онда је однос 1 према 50176 што значи да однос позитивних и негативних класа екстремно небалансиран. Такође, модел може врло лако у пар итерација да научи да класификује већину нула пиксела тачно тј. постоји велики број лаких негативних класа. На основу овога је закључено да је погодно да се користи *Focal loss* функција грешка.

Функција грешке унакрсне ентропије и основни *Focal Loss* дају исту вредност грешке за пиксел који је непосредно поред правог пиксела као и за пиксел на слици који је најдаљи од правог пиксела. То није толико погодно, јер пиксел који је непосредно поред правог пиксела такође представља јако добру предикцију за крајњу тачку. Због тога се примећује Гаусов кернел ради ублажавања грешака у околини правог пиксела. Вредност правог пиксела остаје иста.

Применом Гаусовог кернела над истинитом топлотном мапом се добијају „меке” класе тј. вредности су сада реални бројеви из интервала $[0, 1]$. Основна дефиниција *Focal Loss* није погодна у том случају, јер је намењена за класе 0 и 1. Архитектура *HOME* користи поједностављену верзију *CenterNet Focal loss* функције грешке и има

следећи облик:

$$FL_{home}(\hat{p}, p) = \begin{cases} -\log(\hat{p}), & p = 1 \\ -(1-p)^4 \cdot \log(1-\hat{p}), & p \neq 1 \end{cases}$$

Смисао $(1-p)^4$ је умањивање грешке унакрсне ентропије у случају да је модел пикселима у околини правог дао вероватноће које су веће од 0. Коначна функција грешке узима у обзир и средњеквадратно растојање:

$$L_{home}(\hat{p}, p) = (p - \hat{p})^2 \cdot \begin{cases} -\log(\hat{p}), & p = 1 \\ -(1-p)^4 \cdot \log(1-\hat{p}), & p \neq 1 \end{cases}$$

Узорковање крајњих тачака трајекторија

Над добијеним топлотним мапама се примењује алгоритам за узорковање крајњих тачака трајекторија. Избор алгоритма у овом случају је произвољан и може да се мења у зависности од метрике која треба да се минимизује. Алгоритам који је оптималан по FDE не значи да је оптималан по MR и обрнуто.

Произвољној трајекторији агента може да се додели вероватноћа користећи топлотну мапу крајњих тачака. Свакој трајекторији се додељује вероватноћа као интеграл над површином која одговара околини крајње тачке на топлотној мапи. За околину може да се користи круг или квадрат са задатим радијусом. Круг у овом случају можда има више смисла, али квадрат је погоднији јер одговара примени оптимизоване конволуције са филтером који има све јединице. Ако је циљ оптимизација MR_k^α метрике, онда је то еквивалентно избору K трајекторија тако да је збир њихових вероватноћа највећи. Проналазак оптималних K тачака није лак проблем. Уместо тога може да се користи похлепни алгоритам који налази K приближно оптималних тачака. Похлепни алгоритам 7.3 је приказан у наставку.

ГЛАВА 7. ТЕХНИКА ЗАСНОВАНА НА РАЗУМЕВАЊУ КОНТЕКСТА ОБРАДОМ РАСТЕРИЗОВАНЕ СЦЕНЕ

```
def mr_uzorkovanje(t_mapa, k, radijus):  
    # t_mapa: toplotna mapa  
    # k: broj tacaka koje se uzorkuju  
    # radijus: radijus okoline za aproksimaciju verovatnoce tacke  
  
    suma_kernel = napravi_kernel_jedinica(radijus)  
    traj_vrv = primeni_kernel(t_mapa, suma_kernel)  
  
    tacke = []  
    for _ in range(k):  
        # pronalazak tacke sa najvecom vrv  
        red, kolona = traj_vrv.argmax()  
        # brisanje okoline uzorkovane tacke  
        # tj. postavljanje vrednosti piksela u okolini na 0  
        traj_vrv = azuriraj_traj_vrv(traj_vrv, t_mapa, red, kolona, radijus)  
        # koordinatni sistem se ne poklapa koordinatnim sistemom toplotne  
        # mape zbog primene konvolucije  
        red, kolona = korekcija_koordinata(red, kolona, radijus)  
        tacke.append((red, kolona))  
  
    return tacke
```

Слика 7.3: Алгоритам за узорковање

Узорковане тачке се користе као улаз у модел за естимацију трајекторија.

Естимација трајекторије до предложене крајње тачке

У овом тренутку се претпоставља да су дате крајње тачке трајекторија и да на основу њих треба да се естимирају трајекторије. За естимацију трајекторија се користи једноставан модел који се састоји само од потпуно повезаних слојева. Над историјом трајекторије агента се примењује потпуно повезан слој. Излаз тог слоја се спаја са координатама предложене крајње тачке, а након спајања се примењује још један потпуно повезан слој којим се добија предвиђена трајекторија.

Глава 8

Закључак

У изради...

Библиографија

- [1] Y. Huang and Y. Chen, “Autonomous driving with deep learning: A survey of state-of-art technologies,” in *2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 2020. Arxiv preprint: 2006.06091.
- [2] “Lyft motion prediction for autonomous vehicles - kaggle competition.” <https://www.kaggle.com/c/lyft-motion-prediction-autonomous-vehicles>.
- [3] J. Gao, C. Sun, H. Zhao, Y. Shen, D. Anguelov, C. Li, and C. Schmid, “Vectornet: Encoding hd maps and agent dynamics from vectorized representation,” in *CVPR*, 2020. Arxiv preprint: 2005.04259.
- [4] T. Gilles, S. Sabatini, D. Tsishkou, B. Stanciulescu, and F. Moutarde, “Home: Heatmap output for future motion estimation,” 2021. Arxiv preprint: 2105.10968.
- [5] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays, “Argoverse: 3d tracking and forecasting with rich maps,” in *CVPR*, 2019. Arxiv preprint: 2103.11624.
- [6] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social lstm: Human trajectory prediction in crowded spaces.”
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *NeurIPS*, 2017. Arxiv preprint: 1706.03762.
- [8] K. Mangalam and R. Garg, “Overcoming mode collapse with adaptive multi adversarial training,” 2021. <https://www.bmvc2021-virtualconference.com/assets/papers/0690.pdf>.
- [9] Gupta, Agrim, Johnson, Justin, Fei-Fei, Li, Savarese, Silvio, Alahi, and Alexandre, “Social gan: Socially acceptable trajectories with generative adversarial networks,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, no. CONF, 2018.

- [10] H. Zhao, J. Gao, T. Lan, C. Sun, B. Sapp, B. Varadarajan, Y. Shen, Y. Shen, Y. Chai, C. Schmid, C. Li, and D. Anguelov, “Tnt: Target-driven trajectory prediction,” in *CVPR*, 2020. Arxiv preprint: 2008.08294.
- [11] X. Zhou, D. Wang, and P. Krahenbuhl, “Objects as points (centernet),” 2019. Arxiv preprint: 1904.07850.
- [12] M. Schafer, K. Zhao, M. Buhren, and A. Kummert1, “Context-aware scene prediction network (casnet),” 222. Arxiv preprint: 2101.06933.
- [13] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. kin Wong, and W. chun Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” 2015. Arxiv preprint: 1506.04214.
- [14] M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, and R. Urtasun, “Learning lane graph representations for motion forecasting,” in *ECCV*, 2020. Arxiv preprint: 2007.13732.
- [15] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *Proceedings of the 5th International Conference on Learning Representations*, 2016. Arxiv preprint: 1609.02907.
- [16] M. Ye, T. Cao, and Q. Chen, “Tpcn: Temporal point cloud networks for motion forecasting,” in *CVPR*, 2021. Arxiv preprint: 2103.03067.
- [17] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in *CVPR*, 2017. Arxiv preprint: 1612.00593.
- [18] M. F. Apratim Bhattacharyya, Bernt Schiele, “Accurate and diverse sampling of sequences based on a “best of many” sample objective,” in *CVPR*, 2018. Arxiv preprint: 1806.07772.
- [19] Chen, Guangyi, Li, Junlong, Zhou, Nuoxing, Ren, Liangliang, Lu, and Jiwen, “Personalized trajectory prediction via distribution discrimination,” in *ICCV*, 2021.
- [20] Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, “Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction,” in *CoRL*, 2019. Arxiv preprint: 1910.05449.
- [21] R. J. Williams and D. Zipser, “A learning algorithm for continually running fully recurrent neural networks. neural computation,” in *Neural Computation (Volume: 1, Issue: 2, June 1989)*, 1989. ISSN 0899-7667.
- [22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *CVPR*, 2017. Arxiv preprint: 1708.02002.

Биографија аутора

Момир Аџемовић У изради...