

УДК 004

ОБЗОР МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ ДВИЖЕНИЯ ДЛЯ АВТОНОМНЫХ ТРАНСПОРТНЫХ СРЕДСТВ

Порохин Ю.М., студент гр. ИТб-221, IV курс

Научный руководитель: Сыркин И.С., к.т.н., доцент

Кузбасский государственный технический университет имени Т.Ф. Горбачёва,
г. Кемерово

В контуре управления автономного транспортного средства блок прогнозирования движения отвечает за оценку множества правдоподобных будущих состояний окружающих участников движения. В отличие от задачи детекции или трекинга, где требуется восстановить текущее положение объекта, в задаче прогнозирования необходимо оценить условное распределение будущей траектории по наблюдаемой истории, геометрии дорожной сцены и поведению соседних агентов [1]. Практическая ценность такого прогноза определяется не только низкой средней ошибкой, но и способностью модели заранее выделять конфликтные сценарии: перестроение в плотном потоке, левый поворот через встречную полосу движения, слияние полос. Поэтому современные предикторы, как правило, формируют не единственную траекторию, а набор гипотез с вероятностями, пригодный для последующего планирования [1, 3].

Постановка задачи и требования к модели. Формально прогнозирование движения можно рассматривать как задачу оценки распределения $p(Y|X, M)$, где X обозначает историю состояний агента и его окружения, M – данные высокоточной карты, а Y – будущую траекторию на конечном горизонте прогноза. С инженерной точки зрения к модели предъявляются сразу несколько требований. Прежде всего, ей необходимо быть мультимодальной, потому что для одной и той же дорожной ситуации возможные разные, но одинаково правдоподобные варианты маневра. Кроме того, модели прогнозирования важно опираться на ограничения, задаваемые самой дорожной сценой: конфигурацию полос, разметку, разрешенные направления движения и текущий сигнал светофора. Не менее значим и учет взаимодействия между участниками движения, поскольку неверная интерпретация намерений или приоритета движения другого агента может привести к ошибкам на этапе планирования. Наконец, существенную роль играет корректная калибровка вероятностей: даже точные по геометрии траектории мало полезны на практике, если установленные им вероятности не отражают реальный уровень риска [1, 8].

Первые подходы к краткосрочному прогнозированию движения в основном опирались на кинематические и динамические модели. К ним относились схемы с постоянной скоростью или ускорением, движение с постоянной угловой скоростью, а также фильтры Калмана или интерактивный многомодельный алгоритм (ИММ). Их преимущества заключаются в том, что они просты, вычисли-

тельно недороги и хорошо интерпретируются. На небольшом горизонте прогноза и на относительно простых участках дороги такие методы до сих пор остаются полезными как базовый ориентир или резервный механизм. Однако их возможности быстро заканчиваются там, где нужно понять намерение водителя или учесть реакцию на других участников движения. Иными словами, классические модели неплохо продолжают уже наблюдаемое движение, но заметно хуже работают в ситуациях выбора, а именно такие сценарии особенно важны для автономного транспорта.

Появление обучаемых моделей позволило перейти от простой экстраполяции к использованию статистических закономерностей реального дорожного движения и более полно учитывать влияние окружения. Одним из ранних заметных примеров стала модель Social LSTM, в которой траектория каждого агента описывается рекуррентной сетью, а сведения о соседях объединяются через «социальное объединение» (social pooling) [2]. Эта работа показала, что даже относительно простое совместное рассмотрение нескольких участников дает более качественный прогноз, чем независимое моделирование каждого из них. Позднее появились вероятностные архитектуры, способные формировать несколько вариантов поведения. Так, в MultiPath используется набор якорных (anchor) гипотез, для каждой из которых оценивается вероятность, что удобно для последующей интеграции с планировщиком [3]. В Trajectron++ взаимодействия уже задаются через динамический граф сцены и вариационный механизм CVAE, благодаря чему модель становится гибче и лучше учитывает неоднородность участников движения [4]. Вместе с тем у таких подходов есть характерное ограничение: многое зависит от того, как задано пространство модальностей. Слишком грубое описание приводит к потере реалистичных сценариев, а слишком широкое – к росту числа лишних маловероятных траекторий.

Серьезный шаг вперед в задачах прогнозирования был связан с переходом от чисто растрового описания окружения к векторному представлению карты и траекторий. В работе VectorNet дорожная сеть, геометрия полос и история движения агентов задаются как совокупность связанных геометрических примитивов, что позволяет сохранить структуру сцены без лишней дискретизации и потери топологии [5]. Для автономного вождения это особенно важно, поскольку поведение автомобиля часто лучше объясняется именно геометрией дороги – формой полосы, кривизной участка, положением стоп-линии или конфигурацией съезда, – чем локальным изображением сцены. Векторные модели, как правило, эффективнее используют данные HD-карт и лучше масштабируются на протяженные дорожные участки. Дополнительно их преимущество состоит в том, что они позволяют естественно связать пространственные отношения между агентами с семантикой карты, что особенно полезно при прогнозе на более длинном горизонте.

Goal-conditioned методы. Отдельный класс решений сначала оценивает высокоуровневую цель маневра, а уже затем восстанавливает подробную траекторию. Логика здесь достаточно прозрачна: на сложном перекрестке основная неопределенность часто связана не с локальной гладкостью пути, а с выбором конечной цели – например, с решением ехать прямо, повернуть направо или перестроиться в соседнюю полосу. DenseTNT реализует именно такую декомпозицию: модель оценивает плотное множество целей-кандидатов и затем достраивает траекторию с учетом наиболее вероятных вариантов [6]. Подход оказался удачным по двум причинам. Во-первых, он улучшает мультимодальность, так как разные маневры естественно связываются с разными целями. Во-вторых, уменьшается длина причинно-следственной цепочки: модель сперва определяет направление намерения, а потом уточняет кинематику движения. Ограничение у goal-conditioned предикторов тоже очевидно: ошибка на этапе выбора цели переносится на весь дальнейший прогноз, поэтому качество системы во многом зависит от полноты и точности целевого пространства.

Графовые и трансформерные архитектуры. По мере роста сложности сценариев стало ясно, что одной лишь истории целевого агента недостаточно. На перекрестках, в плотном потоке и при множественных перестроениях необходимо учитывать дальние зависимости между участниками движения и дорожными объектами. Именно поэтому в последних поколениях моделей широко используются графовые сети и attention-механизмы. Scene Transformer рассматривает сцену как единую систему агентов, дорожных элементов и временных срезов; благодаря маскированию запросов одна и та же архитектура применяется и для marginal prediction, и для joint prediction, и для conditioned prediction [7]. Иерархические варианты трансформеров, например HiVT, объединяют локальные взаимодействия соседних агентов с более глобальным контекстом сцены, что позволяет лучше улавливать как непосредственные конфликты траекторий, так и дальние топологические зависимости [9]. Сильная сторона этого направления – способность описывать scene-consistent гипотезы на длинном горизонте. Однако цена за такое качество выражается в большом объеме данных, высокой вычислительной нагрузке и необходимости особенно аккуратно работать с вероятностной калибровкой результата.

Интерактивное прогнозирование как отдельная постановка. Для автономного вождения принципиально важно не просто предсказать движение каждого объекта по отдельности, а сформировать согласованную картину будущего всей локальной сцены. Если два автомобиля подходят к конфликтной зоне, независимые marginal predictions могут оказаться геометрически правдоподобными по отдельности, но совместно образовать невозможную ситуацию: взаимное наложение траекторий, одновременный проезд в точке конфликта или нереалистичное игнорирование приоритета. По этой причине в литературе усилился интерес к joint-подходам, которые моделируют распределение сразу для нескольких

агентов [1, 7]. Однако здесь возникает комбинаторная проблема: число совместных гипотез растет значительно быстрее, чем при индивидуальном прогнозе, и потому прямое моделирование полного joint-space становится вычислительно тяжелым. На этом фоне особенно ценными оказываются методы, которые сохраняют совместную согласованность, но не требуют полного перебора всех комбинаций.

Факторизованный интерактивный прогноз на примере M2I. Один из наиболее показательных подходов такого типа представлен в работе M2I: From Factored Marginal Trajectory Prediction to Interactive Prediction [8]. Авторы исходят из идеи, что в паре взаимодействующих агентов часто можно выделить влияющего участника (influencer) и реагирующего участника (reactor). Вместо прямого моделирования совместного распределения $p(Y_A, Y_B|X, M)$ предлагается его факторизация в виде $p(Y_I|X, M) p(Y_R|X, M, Y_I)$, где сначала строится набор маргинальных траекторий для влияющего агента, а затем для каждой такой гипотезы генерируется условный ответ реагирующего агента [8]. До этапа прогнозирования relation predictor определяет тип отношения между агентами – pass, yield или none, после чего система выбирает соответствующую схему формирования совместных гипотез (рис. 1,2). В результате M2I сочетает интерпретируемость и вычислительную приемлемость: модель явно указывает предполагаемую структуру взаимодействия и при этом избегает полного перебора совместного пространства траекторий.

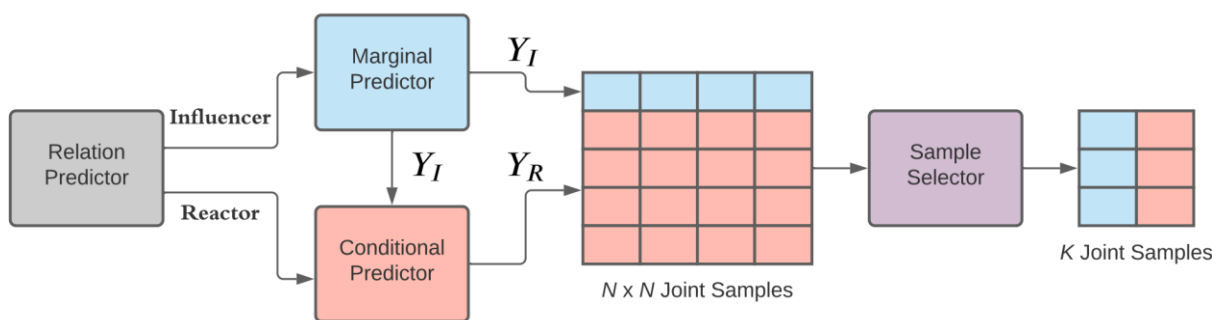


Рисунок 1 – Общая схема метода M2I.

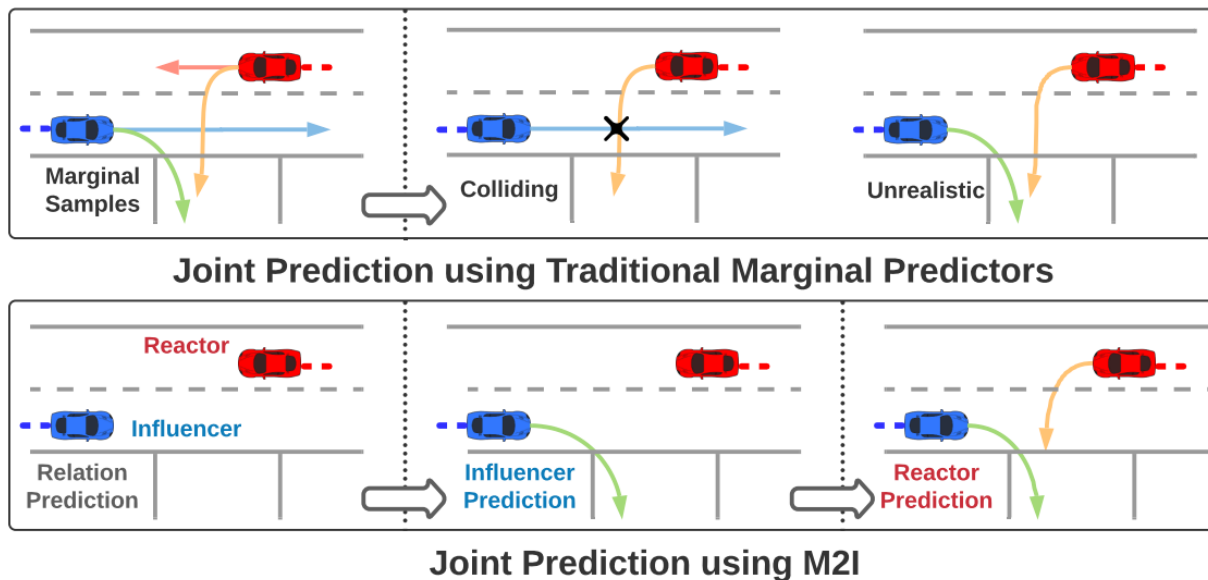


Рисунок 2 – Сравнение прогнозирования классического Marginal Predictors и M2I.

Достоинства и ограничения M2I. Практическая ценность M2I хорошо видна в конфликтных сценариях – при слиянии, опережении, выполнении левого поворота и других взаимодействиях между участниками движения. В таких случаях модель не ограничивается независимым прогнозом двух траекторий, а связывает ответ одного агента с предполагаемым будущим другого. Это делает итоговые гипотезы более согласованными и уменьшает число нереалистичных комбинаций [8] (рис. 3).

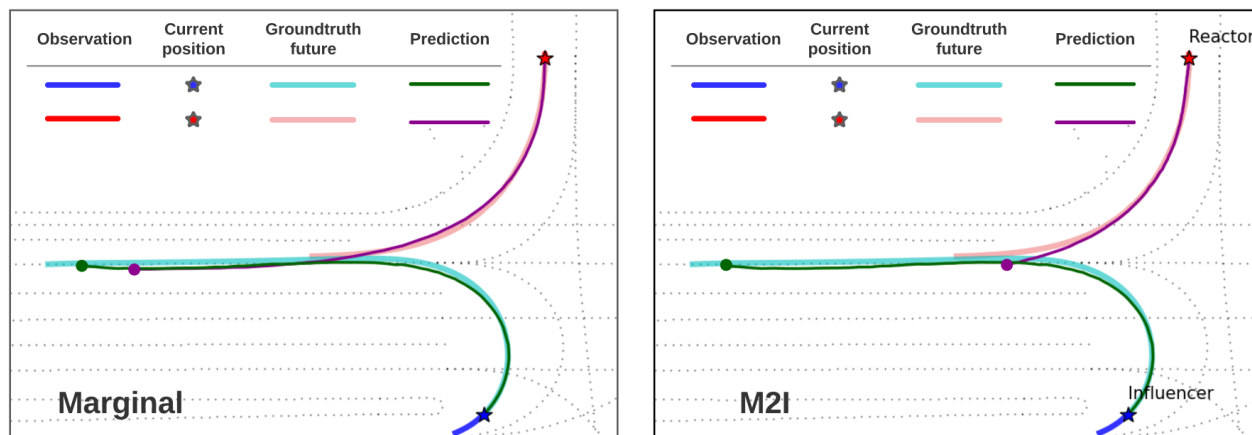


Рисунок 3 – Качественное сравнение Marginal baseline и M2I в интерактивном сценарии.

Важным достоинством подхода является и его модульность: relation predictor, marginal predictor и conditional predictor могут опираться на разные backbones (опорные сети), а значит, M2I не привязан к единственной архитектуре кодировщика сцены. Вместе с тем метод опирается на предположение о направленном влиянии, которое выполняется не всегда. В плотном городском трафике

нередки ситуации взаимной адаптации, когда оба участника одновременно подстраиваются друг под друга. Кроме того, ошибка *relation predictor* непосредственно ухудшает итоговый прогноз, а при переходе от пары агентов к мультиагентной сцене возникает риск циклических зависимостей. Следовательно, M2I следует рассматривать не как универсальное решение, а как сильный компромисс между полной совместной моделью и независимым *marginal forecasting*.

Сравнение классов методов с прикладной точки зрения. Если рассматривать прогнозирование движения как часть промышленного стека автономного вождения, выбор модели нельзя сводить только к метрикам *minADE* или *minFDE*. Кинематические методы дают быстрый и устойчивый базовый прогноз, но почти не помогают в задачах переговорного поведения. Вероятностные рекуррентные модели лучше отражают мультимодальность, однако чувствительны к способу задания латентных модальностей [2–4]. Векторизованные *goal-conditioned* модели сегодня выглядят наиболее рациональным компромиссом для массового *marginal forecasting*, поскольку хорошо учитывают структуру карты и сохраняют приемлемую вычислительную стоимость [5, 6]. Трансформерные и *joint*-подходы особенно ценны там, где ошибка совместной согласованности опаснее, чем небольшой рост геометрической ошибки [7, 9]. Факторизованные методы наподобие M2I занимают промежуточное положение: они уступают идее полного *joint*-моделирования по общности, но часто выигрывают по интерпретируемости и вычислительной практичности [8].

Метрики и практические критерии оценки. Для сравнения современных предикторов недостаточно проверять только среднее отклонение ближайшей гипотезы от истинной траектории. В автономном транспорте не менее важно, насколько полно набор предсказаний покрывает реалистичные исходы и насколько корректно распределены вероятности между ними. Именно поэтому в бенчмарках *Waymo Open Motion Dataset* используются не только *minADE* и *minFDE*, но также *miss rate*, *overlap rate* и *mAP* [1]. Первая группа метрик оценивает геометрическую точность, вторая – способность избежать пропуска правдоподобного сценария, а вероятностные показатели характеризуют качество ранжирования гипотез. Для реального планировщика именно такое комплексное измерение оказывается наиболее полезным: слишком «смелая» модель, уверенная в одной красивой траектории, может быть опаснее более осторожного предиктора, который предлагает несколько вариантов с адекватно распределенными весами.

Перспективные направления развития. Анализ публикаций показывает, что развитие *Motion Prediction* идет в сторону гибридных решений. Наиболее успешные модели сочетают три идеи: структурное описание сцены через карту и полилинии, явный учет взаимодействий между агентами и вероятностное представление будущего [5–9]. Для практического внедрения этого уже недостаточно. Следующий шаг связан с калибровкой неопределенности, обработкой редких

и аномальных сценариев, устойчивостью к ошибкам восприятия и с тесной интеграцией прогноза с модулем планирования. Иначе говоря, задача постепенно смещается от построения «средней лучшей траектории» к формированию надежной картины риска, полезной для принятия решений в реальном времени. На этом фоне особенно интересны подходы, которые способны учитывать топологию дороги, взаимодействия между участниками и внутреннюю логику самого маневра.

В целом развитие методов прогнозирования движения для автономных транспортных средств идет от простых кинематических схем к более сложным мультимодальным моделям, которые используют карты, графовые структуры и механизмы внимания. Рекуррентные и генеративные подходы впервые позволили явно учитывать влияние одних участников движения на других. Векторные и goal-conditioned модели заметно улучшили качество прогноза за счет более точного описания дорожной среды, а трансформенные архитектуры приблизили исследователей к согласованному сценарию прогноза на уровне всей сцены. Особый интерес представляет M2I, поскольку этот метод пытается совместить согласованность современного прогноза с приемлемой вычислительной сложностью, разделяя взаимодействие на influencer и reactor [8]. Для практических систем автономного вождения наиболее разумным выглядит многоуровневый подход: быстрые модельные предикторы используются как страховочный слой и средство проверки здравого смысла, а основное множество гипотез формируется мультимодальной моделью, которая явно учитывает взаимодействия между агентами. Такое сочетание важно не только для повышения средней точности, но и для более надежного поведения автомобиля в сложных дорожных ситуациях.

Список литературы

1. Ettinger S., Cheng S., Caine B. et al. Large Scale Interactive Motion Forecasting for Autonomous Driving: The Waymo Open Motion Dataset // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. URL: https://openaccess.thecvf.com/content/ICCV2021/html/Ettinger_Large_Scale_Interactive_Motion_Forecasting_for_Autonomous_Driving_The_Waymo_ICCV_2021_paper.html (дата обращения: 23.03.2026).
2. Alahi A., Goel K., Ramanathan V. et al. Social LSTM: Human Trajectory Prediction in Crowded Spaces // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. URL: https://openaccess.thecvf.com/content_cvpr_2016/html/Alahi_Social_LSTM_Human_CVPR_2016_paper.html (дата обращения: 23.03.2026).
3. Chai Y., Sapp B., Bansal M., Anguelov D. MultiPath: Multiple Probabilistic Anchor Trajectory Hypotheses for Behavior Prediction // Proceedings of the Confer-

- ence on Robot Learning. 2020. URL:
<https://proceedings.mlr.press/v100/chai20a.html> (дата обращения: 23.03.2026).
4. Salzmann T., Ivanovic B., Chakravarty P., Pavone M. Trajectron++: Multi-Agent Generative Trajectory Forecasting with Heterogeneous Data for Control // European Conference on Computer Vision. 2020. URL: <https://arxiv.org/abs/2001.03093> (дата обращения: 23.03.2026).
 5. Gao J., Sun C., Zhao H. et al. VectorNet: Encoding HD Maps and Agent Dynamics from Vectorized Representation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. URL:
https://openaccess.thecvf.com/content_CVPR_2020/html/Gao_VectorNet_Encoding_HD_Maps_and_Agent_Dynamics_From_Vectorized_Representation_CVPR_2020_paper.html (дата обращения: 23.03.2026).
 6. Gu J., Sun C., Zhao H. DenseTNT: End-to-End Trajectory Prediction from Dense Goal Sets // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. URL:
https://openaccess.thecvf.com/content/ICCV2021/html/Gu_DenseTNT_End-to-End_Trajectory_Prediction_From_Dense_Goal_Sets_ICCV_2021_paper.html (дата обращения: 23.03.2026).
 7. Ngiam J., Caine B., Vasudevan V. et al. Scene Transformer: A Unified Architecture for Predicting Future Trajectories of Multiple Agents // International Conference on Learning Representations. 2022. URL:
<https://openreview.net/forum?id=Wm3EA5OIHsG> (дата обращения: 23.03.2026).
 8. Sun Q., Huang X., Gu J. et al. M2I: From Factored Marginal Trajectory Prediction to Interactive Prediction // arXiv. 2022. URL: <https://arxiv.org/abs/2202.11884> (дата обращения: 23.03.2026).
 9. Zhou Z., Ye L., Wang J., Wu K., Lu K. HiVT: Hierarchical Vector Transformer for Multi-Agent Motion Prediction // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. URL:
https://openaccess.thecvf.com/content/CVPR2022/html/Zhou_HiVT_Hierarchical_Vector_Transformer_for_Multi-Agent_Motion_Prediction_CVPR_2022_paper.html (дата обращения: 23.03.2026).