

Learning to Navigate in Cities Without a Map

Соколов Роман

Национальный Исследовательский Университет Высшая Школа Экономики

30 ноября 2018 г.

План

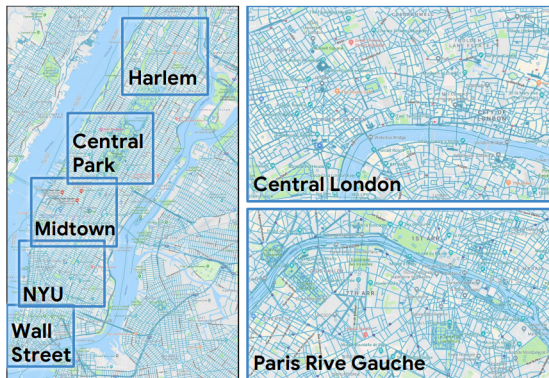
1. Задача
2. Среда и агент
3. Обучение
4. Результаты
5. Вывод
6. Заключение

Задача



Diverse views and corresponding local maps in Street View.

Задача



Street View regions used in this study.

Среда и агент

Среда и агент

- ▶ Какую информацию агент получает от среды?

Среда и агент

- ▶ Какую информацию агент получает от среды?
- ▶ Как агент взаимодействует со средой?

Среда и агент

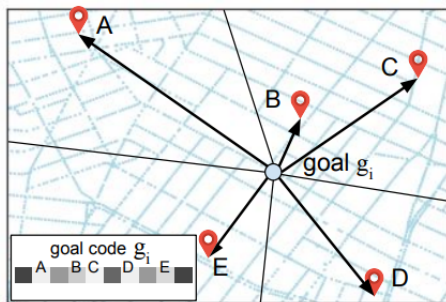
- ▶ Какую информацию агент получает от среды?
- ▶ Как агент взаимодействует со средой?
- ▶ Как подавать агенту на вход цель?

Среда и агент

- ▶ Какую информацию агент получает от среды?
- ▶ Как агент взаимодействует со средой?
- ▶ Как подавать агенту на вход цель?
- ▶ $g_{t,i} = \exp(-\alpha d_{t,i}) / \sum_k \exp(-\alpha d_{t,k})$, $\alpha = 0.002$

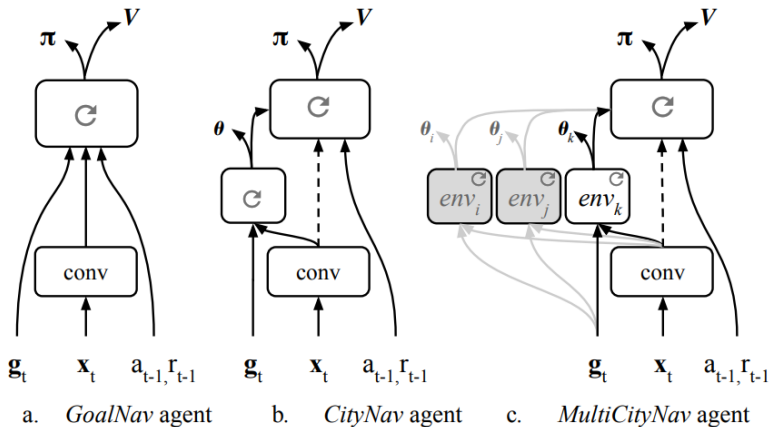
Среда и агент

- ▶ Какую информацию агент получает от среды?
- ▶ Как агент взаимодействует со средой?
- ▶ Как подавать агенту на вход цель?
- ▶ $g_{t,i} = \exp(-\alpha d_{t,i}) / \sum_k \exp(-\alpha d_{t,k})$, $\alpha = 0.002$



Goal description using landmarks.

Среда и агент



Обучение

- ▶ Обучаем как A3C агента

Обучение

- ▶ Обучаем как A3C агента
- ▶ На самом деле используем IMPALA

Обучение

- ▶ Обучаем как A3C агента
- ▶ На самом деле используем IMPALA
- ▶ Используем дополнительную задачу для предсказания угла

Обучение

- ▶ Обучаем как A3C агента
- ▶ На самом деле используем IMPALA
- ▶ Используем дополнительную задачу для предсказания угла
- ▶ Награда агента пропорциональна исходному оптимальному расстоянию до цели

Обучение

- ▶ Обучаем как A3C агента
- ▶ На самом деле используем IMPALA
- ▶ Используем дополнительную задачу для предсказания угла
- ▶ Награда агента пропорциональна исходному оптимальному расстоянию до цели
- ▶ Curriculum learning

Обучение

- ▶ Обучаем как A3C агента
- ▶ На самом деле используем IMPALA
- ▶ Используем дополнительную задачу для предсказания угла
- ▶ Награда агента пропорциональна исходному оптимальному расстоянию до цели
- ▶ Curriculum learning
- ▶ Дополнительные стратегии формирования наград

Обучение

- ▶ Обучаем как A3C агента
- ▶ На самом деле используем IMPALA
- ▶ Используем дополнительную задачу для предсказания угла
- ▶ Награда агента пропорциональна исходному оптимальному расстоянию до цели
- ▶ Curriculum learning
- ▶ Дополнительные стратегии формирования наград
 - ▶ Ранние награды

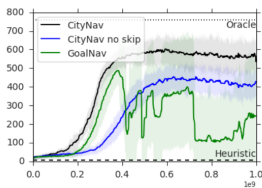
Обучение

- ▶ Обучаем как A3C агента
- ▶ На самом деле используем IMPALA
- ▶ Используем дополнительную задачу для предсказания угла
- ▶ Награда агента пропорциональна исходному оптимальному расстоянию до цели
- ▶ Curriculum learning
- ▶ Дополнительные стратегии формирования наград
 - ▶ Ранние награды
 - ▶ $r_t = \max(0, \min(1, (d_{ER} - d_t^g)/100)) \times r^g$

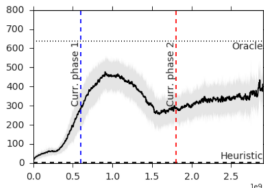
Обучение

- ▶ Обучаем как A3C агента
- ▶ На самом деле используем IMPALA
- ▶ Используем дополнительную задачу для предсказания угла
- ▶ Награда агента пропорциональна исходному оптимальному расстоянию до цели
- ▶ Curriculum learning
- ▶ Дополнительные стратегии формирования наград
 - ▶ Ранние награды
 - ▶ $r_t = \max(0, \min(1, (d_{ER} - d_t^g)/100)) \times r^g$
 - ▶ Монетки

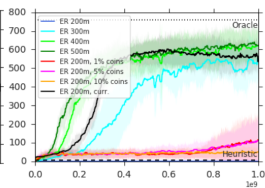
Результаты



(a) NYU (New York City)



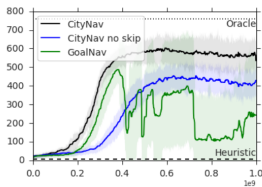
(b) Central London



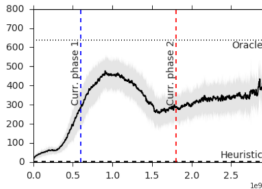
(c) Effect of reward shaping

- CityNav агент обучается существенно лучше GoalNav

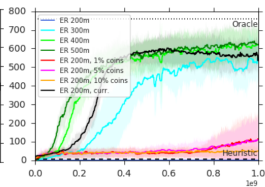
Результаты



(a) NYU (New York City)



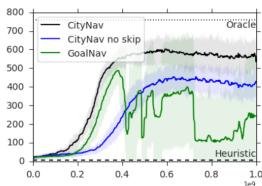
(b) Central London



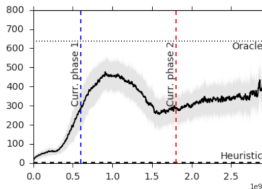
(c) Effect of reward shaping

- ▶ CityNav агент обучается существенно лучше GoalNav
- ▶ CityNav без добавления визуальных признаков к блоку стратегии показывает себя хуже

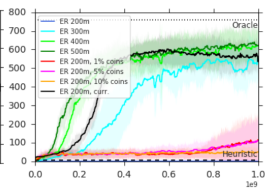
Результаты



(a) NYU (New York City)



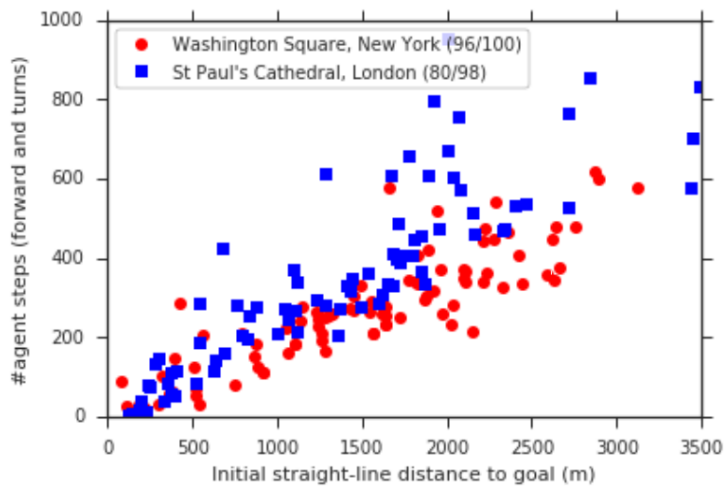
(b) Central London



(c) Effect of reward shaping

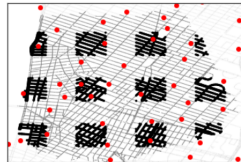
- ▶ CityNav агент обучается существенно лучше GoalNav
- ▶ CityNav без добавления визуальных признаков к блоку стратегии показывает себя хуже
- ▶ Формирование награды сильно влияет на обучение агента, лучше всего себя показывают ранние награды и curriculum learning

Результаты



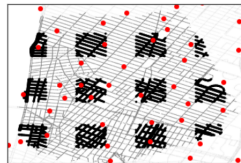
Результаты

GRID SIZE	TRAIN REW	REW	TEST FAIL	$T_{\frac{1}{2}}$
FINE	655	567	11%	229
MEDIUM	637	293	20%	184
COARSE	623	164	38%	243



Результаты

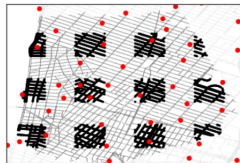
GRID SIZE	TRAIN REW	REW	TEST FAIL	$T_{\frac{1}{2}}$
FINE	655	567	11%	229
MEDIUM	637	293	20%	184
COARSE	623	164	38%	243



- Модели сложно понимать где находятся новые цели

Результаты

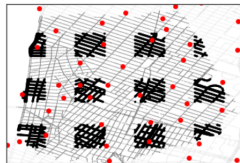
GRID SIZE	TRAIN REW	REW	TEST FAIL	$T_{\frac{1}{2}}$
FINE	655	567	11%	229
MEDIUM	637	293	20%	184
COARSE	623	164	38%	243



- ▶ Модели сложно понимать где находятся новые цели
- ▶ И становится тем сложнее чем меньше целей из этой области она видела ранее

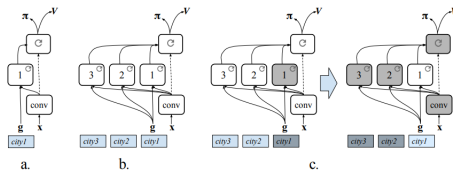
Результаты

GRID SIZE	TRAIN REW	REW	TEST FAIL	$T_{\frac{1}{2}}$
FINE	655	567	11%	229
MEDIUM	637	293	20%	184
COARSE	623	164	38%	243

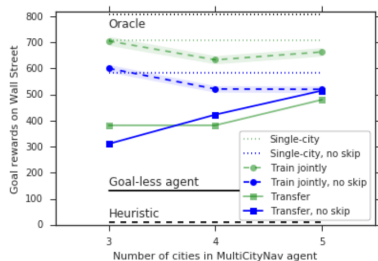


- ▶ Модели сложно понимать где находятся новые цели
- ▶ И становится тем сложнее чем меньше целей из этой области она видела ранее
- ▶ Но агент всё равно способен дойти до середины пути за сравнимое время

Результаты

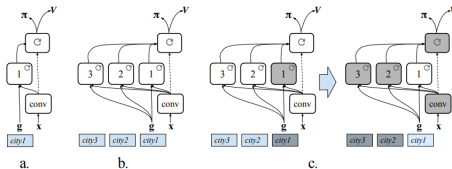


(a) Diagram of transfer learning experiments.

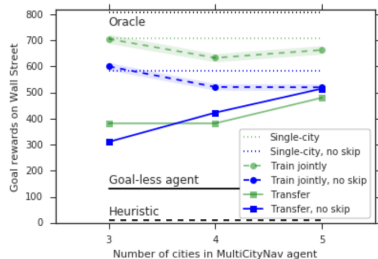


(b) Transfer learning performance.

Результаты



(a) Diagram of transfer learning experiments.



(b) Transfer learning performance.

- ▶ Агент без добавления визуальных признаков к блоку стратегии лучше дообучается для новых городов

Заключение

- ▶ В статье была представлена новая среда для обучения агентов для навигации в реальном мире

Заключение

- ▶ В статье была представлена новая среда для обучения агентов для навигации в реальном мире
- ▶ Был представлен агент с блочной архитектурой, способный дообучаться для навигации в новых городах

Заключение

- ▶ В статье была представлена новая среда для обучения агентов для навигации в реальном мире
- ▶ Был представлен агент с блочной архитектурой, способный дообучаться для навигации в новых городах
- ▶ Были исследованы различные подходы для обучения данного агента

Заключение

- ▶ В статье была представлена новая среда для обучения агентов для навигации в реальном мире
- ▶ Был представлен агент с блочной архитектурой, способный дообучаться для навигации в новых городах
- ▶ Были исследованы различные подходы для обучения данного агента
- ▶ И предоставлена основа для дальнейших исследований

Вопросы

1. Что такое curriculum learning и как этот алгоритм помогает обучению нашего агента?
2. Как мы задаём цель на карте нашему агенту и какие преимущества у этого подхода?
3. Опишите архитектуру агента CityNav, как его можно дообучить для навигации в новом городе?