

DeepHack.Turing



Команда Turing Quest
Артём Коренев и Олег Судаков

DeepHack.Turing

- Отборочный тур

Очный этап:

- 7 дней в МФТИ (Долгопрудный)
- Основное состязание
- 3-5 часов лекций ежедневно
- Главный приз – поездка на NIPS
- Дополнительные номинации



Отборочный тур

- Определить бот или человек
- Метрика: ROC AUC
- Train: ~7.2 млн. реплик (классы сбалансированы)
- Validation: ~500 тыс. реплик
- Test: ~500 тыс. реплик
- Язык: английский



[Competitions](#)

[Create a competition](#)

[Blog](#)

[Kaggle](#)

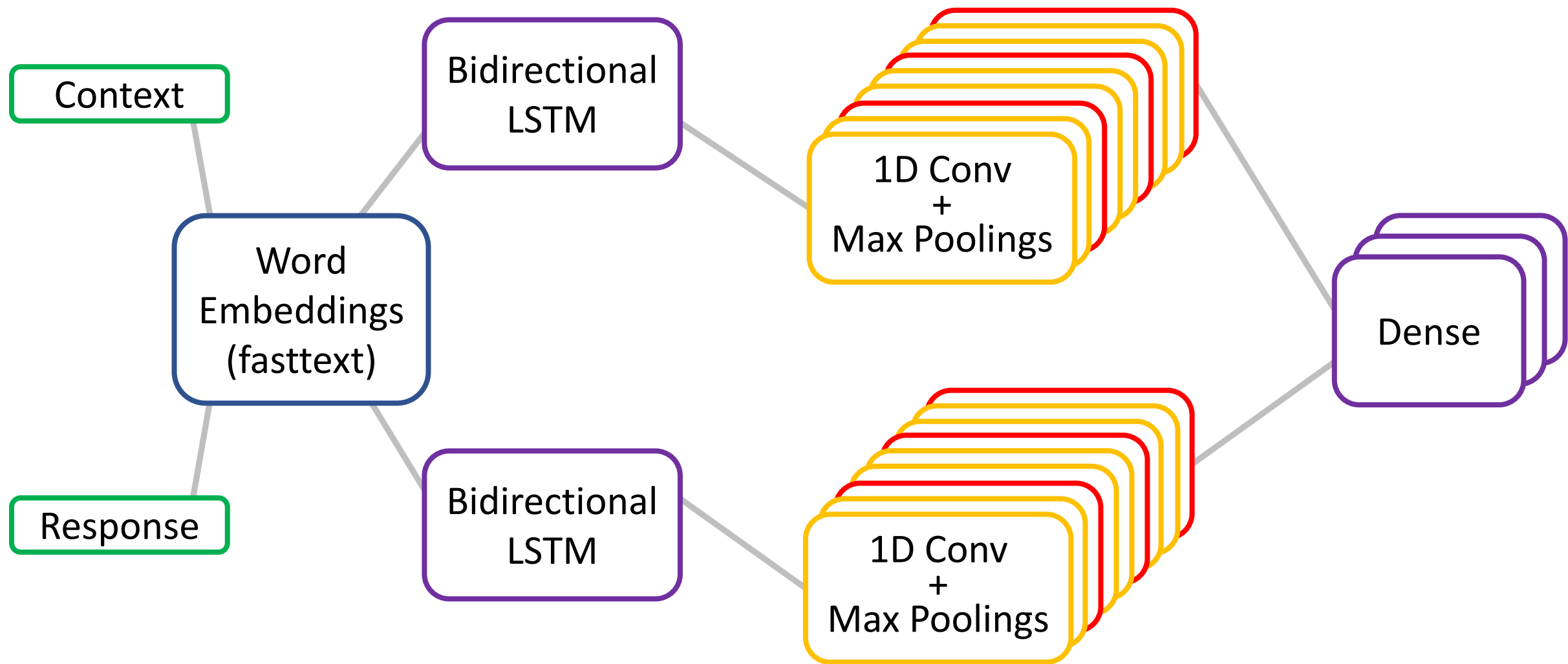


Completed • Knowledge • 78 teams

Human or Machine Generated Text?

Tue 13 Jun 2017 – Sun 9 Jul 2017

Решение отборочного тура



** еще много батчнормов и дропаутов*

Результаты

16	—	Valyaev Ilya	0.83122	6	Sat, 08 Jul 2017 15:59:21 (-30.9h)
17	↑1	monty ponty	0.82952	2	Sat, 08 Jul 2017 02:57:03
18	↓1	Turing Quest 👤	0.82904	18	Fri, 07 Jul 2017 20:24:02
19	—	Insilico 👤	0.82829	3	Fri, 07 Jul 2017 06:19:02
20	↑1	Algis	0.82613	14	Sun, 25 Jun 2017 05:27:39


25	—	Conundrum 👤	0.81355	10	Sun, 09 Jul 2017 21:36:44 (-2h)
1	—	aphex34	0.90148	13	Fri, 07 Jul 2017 20:16:49
2	—	Scitator	0.88906	10	Thu, 06 Jul 2017 21:44:59
3	—	Newbies 👤	0.88700	43	Sat, 08 Jul 2017 05:29:45
4	—	XL-shell 👤	0.85494	34	Sun, 09 Jul 2017 23:23:47 (-0.2h)

Очный этап. День 1

- Лекции
- Генерация данных участниками
- 80% для обучения (и лидерборда)
- 20% для финального тестирования
- Задача: бот или человек?
- Метрика: ROC AUC







Генерация данных


 **convai-bot** 4:31:07 PM
(system msg): Please wait for your partner.




(system msg): The ethnogenesis of the Greek nation is linked to the development of Pan-Hellenism in the 8th century BC. According to some scholars, the foundational event was the Olympic Games in 776 BC, when the idea of a common Hellenism among the Greek tribes was first translated into a shared cultural experience and Hellenism was primarily a matter of common culture. The works of Homer (i.e. Iliad and Odyssey) and Hesiod (i.e. Theogony) were written in the 8th century BC, becoming the basis of the national religion, ethos, history and mythology. The Oracle of Apollo at Delphi was established in this period.

(peer msg):
Hi


  


 **Artem** 4:31:24 PM
Hello


 **convai-bot** 4:31:50 PM
(peer msg):
Do you know anything about Olympic games?


  


(system msg): Chat is finished, please evaluate the overall quality 5:18:55 PM


 **Artem** 5:19:01 PM
4

 **convai-bot** 5:19:01 PM
(system msg): Please evaluate the breadth

 **Artem** 5:19:04 PM
5

 **convai-bot** 5:19:04 PM
(system msg): Please evaluate the engagement

 **Artem** 5:19:04 PM
5

 **convai-bot** 5:19:04 PM
(system msg): Thank you! It was great! Please choose [/begin](#) to continue evaluation.

* бот ставит человеку оценку 0

Очный этап. Дни 2-5

- Лекции
- Генерация данных участниками
- 80% для обучения (и лидерборда)
- 20% для финального тестирования
- Задача: бот или человек?
- Метрика: ROC AUC

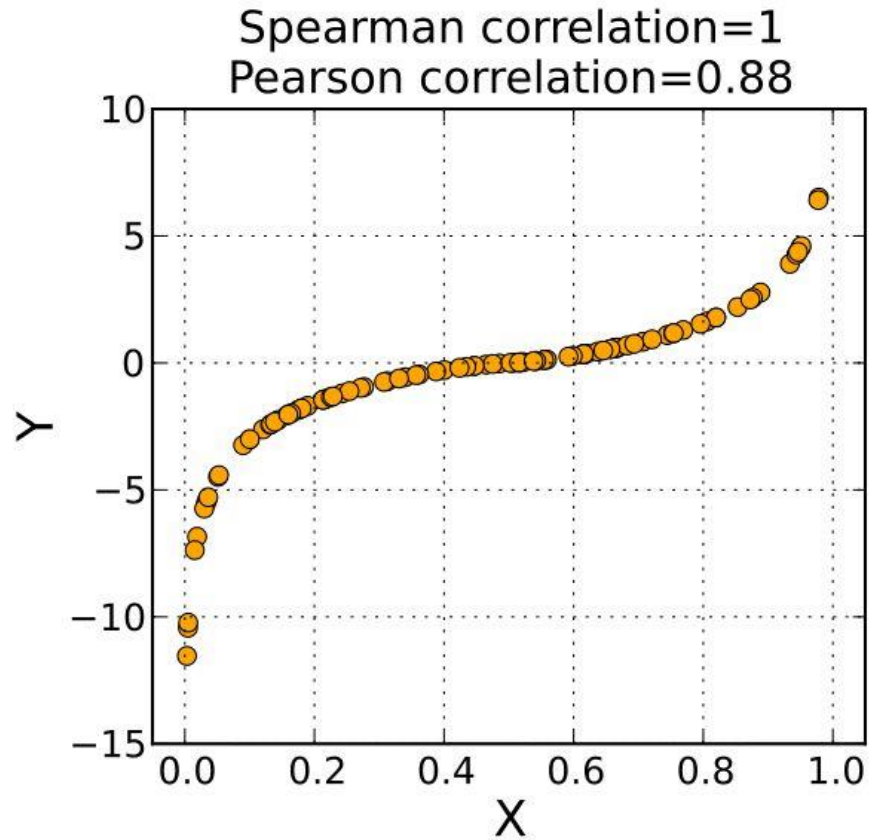


Очный этап. Дни 2-5

- Лекции
- Генерация данных участниками
- 80% для обучения (и лидерборда)
- 20% для финального тестирования
- ~~Задача: бот или человек?~~
- ~~Метрика: ROC AUC~~
- Задача: определить «engagement» собеседников
- Метрика: Spearman Rank-Order Correlation



Spearman rank-order correlation



$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

XGBoost

- 40 различных признаков
 - Стандартные (длина, уник. слова, ...)
 - Особенности датасета (смайлики, капс, ...)
- Учим регрессию на «engagement»
- Учитываем фичи обоих участников диалога
- Корреляция на validation: ~0.67

dmlc
XGBoost

Нейронные сети

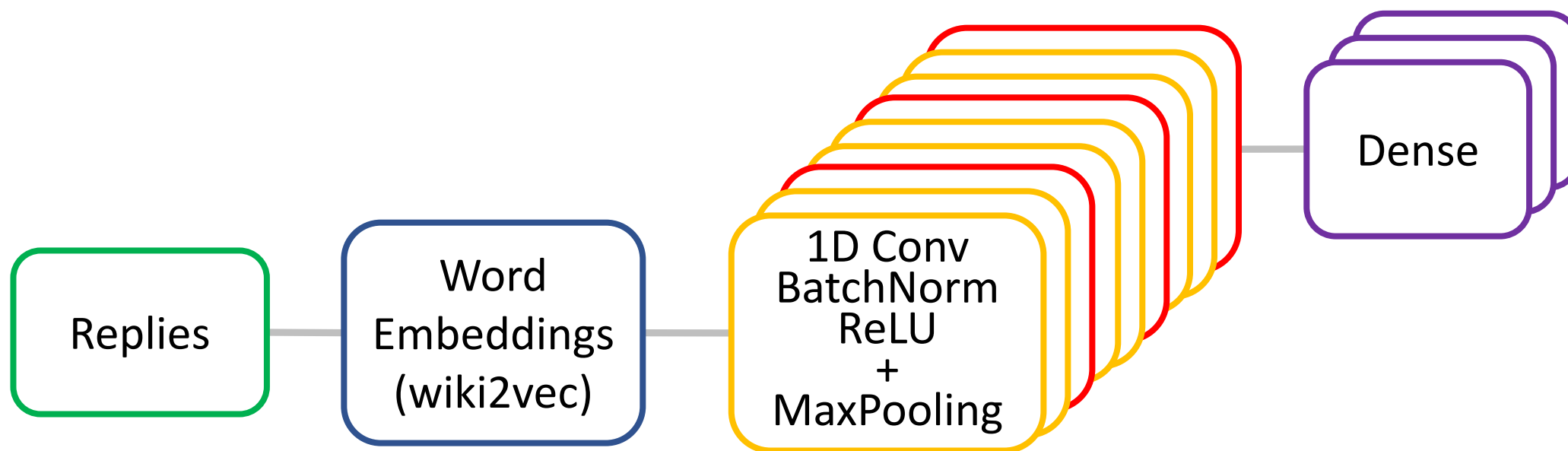
- Контекст слишком большой в сравнении с диалогом
- Мало данных для рекуррентных сетей (<1300 диалогов)
- Мало данных, чтобы уловить взаимодействие участников
- По ответам участника предсказываем как его оценят

Нейронные сети

- Контекст слишком большой в сравнении с диалогом
- Мало данных для рекуррентных сетей (<1300 диалогов)
- Мало данных, чтобы уловить взаимодействие участников
- ~~• По ответам участника предсказываем как его оценят~~
- По ответам участника предсказываем как он оценит другого

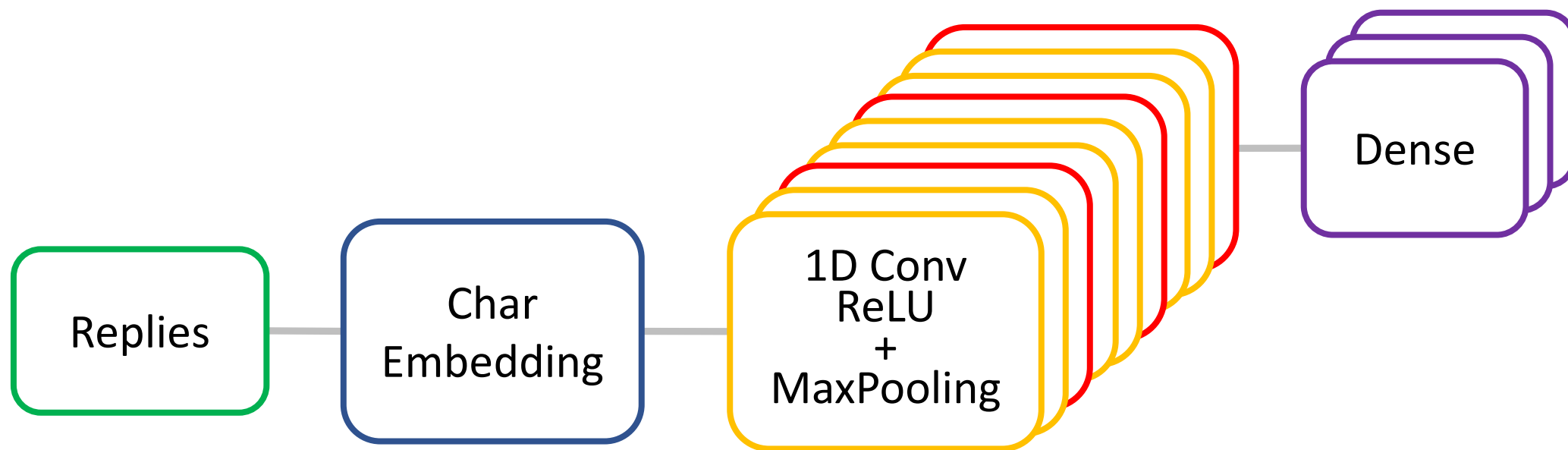
Word-based

- Предобученный wiki2vec
- Опять много дропаутов и батчнормов
- Корреляция на validation: ~ 0.70



Char-based

- Почти копипаста с Word-based
- Корреляция на validation: ~ 0.67



Ансамбли

- Ансамбли очень хорошо работали
- Можно аккуратно и правильно сделать стэкинг

Ансамбли

- Ансамбли очень хорошо работали
- Можно аккуратно и правильно сделать стэкинг
- Ну или обучить по 5 моделей и...

$$\frac{2}{5}CharNN + \frac{2}{5}WordNN + \frac{1}{5}XGBoost$$

Результаты

1. Conundrum	0.771795053835
2. Turing Quest	0.738044289758
3. newbies+	0.723465738142
4. Пластмассовый мир	0.721105580335
5. I have no mouth and I must chat	0.702077144408
6. XL-shell	0.649120693585
7. DATA Siegt	0.644054335294
8. Warp Drive	0.570006706766
9. Agent Smith	0.564292050511
10. TEAM	0.482635480313
11. StackingOverflow	0.357716557906
12. Выперли	0.00436548348506
13. chatme	-0.0584275671832

Другие решения

1 место – Conundrum

- XGBRegressor на оценку
- XGBClassifier (бот или нет)

Признаки:

- Статистические
- Грамматические
- Близость по w2v и doc2vec
- Особенности данных

3 место – newbies+

- LightGBM регрессор на оценку
- Нейронная сеть – и на регрессию и на классификацию
- LightGBM классификатор (бот или нет)

Номинация: лучший бот

1 место – DATA Siegt

- WoW и LHS по текущей реплике
- Поиск самой похожей реплики, после которой ответил человек
- Средняя оценка: 2.36 (из 5)

ИТОГИ

Минусы:

- Данные
- Ежедневный скоринг
- Изменение задачи

Плюсы:

- Пицца
- Много хороших лекторов
- Интересные участники



Спасибо за внимание