bAbI

## The bAbI project

Trang này thu thập các tài nguyên liên quan đến dự án bAbI của Facebook AI Research được tổ chức hướng tới mục tiêu của sự hiểu biết và lý luận văn bản tự động. Các bộ dữ liệu mà chúng tôi đã phát hành bao gồm:

* The (20) QA bAbI tasks
* The (6) dialog bAbI tasks
* The Children’s Book Test
* The Movie Dialog dataset
* The WikiMovies dataset
* The Dialog-based Language Learning dataset
* The SimpleQuestions dataset
* HITL Dialogue Simulator

## The bAbI tasks

Phần này trình bay 20 tasks cho kiểm thử text hiểu và lý luận trong bAbI project. Các nhiệm vụ được mô tả chi tiết trong bài báo:

…

Vui lòng xem các trang trình bày sau:

…

Mục đích là mỗi bài kiểm tra một khía cạnh độc đáo của văn bản và lý luận, và do đó kiểm tra khả năng khác nhau của mô hình học tập. Các nhiệm vụ khác được lên kế hoạch trong tương lai để nắm bắt nhiều khía cạnh hơn.

**Training Set Size** Kích thước Tập huấn: Đối với mỗi công việc, có 1000 câu hỏi cho đào tạo, và 1000 để thử nghiệm. Tuy nhiên, chúng tôi nhấn mạnh rằng mục tiêu là sử dụng càng ít dữ liệu càng tốt để làm tốt nhiệm vụ (nghĩa là nếu bạn có thể sử dụng ít hơn 1000 thậm chí còn tốt hơn)- và không cần phải sử dụng các thủ thuật cụ thể cho nhiệm vụ kỹ thuật mà sẽ không khái quát hóa các nhiệm vụ khác, vì chúng có thể không được sử dụng nhiều sau đó. Lưu ý rằng mục tiêu trong quá trình đánh giá là sử dụng \_same\_ learner qua tất cả các nhiệm vụ để đánh giá kỹ năng và khả năng của mình.

**Supervision Signal Tín hiệu giám sát: Hơn nữa trong khi MemNN kết quả trong giấy sử dụng** full supervision **(bao gồm cả các sự kiện hỗ trợ) kết quả với sự giám sát yếu kém cũng sẽ được ưa thích hơn vì loại dữ liệu này được dễ dàng hơn để thu thập.**

Vì những lý do trên đây, hiện có một số thư mục:

1. en/ — the tasks in English, readable by humans.
2. hn/ — the tasks in Hindi, readable by humans.
3. shuffled/ — các tác vụ tương tự với các chữ cái lộn xộn để chúng không thể đọc được bởi con người, và cho các trình phân tích cú pháp hiện tại và tagger không thể được sử dụng trong straight-forward fashion để tận dụng thêm nguồn lực – trong trường hợp này learner bị ép buộc hơn để dựa trên dữ liệu huấn luyện. sự bắc chước của learner đầu tiên được trình bày bằng ngôn ngữ và phải học từ con số 0.
4. en-10k/ shuffled-10k/ and hn-10k/ 3 task có định dạng tương tự, nhưng với 10,000 tập huấn luyện, và hơn 1000 tập huấn luyện. chú ý kết quả sự dụng 1000 tập huấn luyện

format cho mỗi task bên dưới:

…

Id bắt đầu từ và tăng dần, khi mà ID trong file trở về 1 có nghĩa là bắt đầu 1 câu chuyện mới. Hỗ trợ ID thực tế chỉ bao giờ tham khảo các câu trong một "câu chuyện".

Versions: Một số cập nhật nhỏ kể từ khi bản phát hành gốc được thực hiện (xem README trong phần tải xuống dữ liệu để biết thêm chi tiết). Bạn cũng có thể nhận được v1.0 và v1.1 ở đây.

**Data Stats: một vì dữ liệu thông kê bao gồm trùng lập giữ train và test (khá nhỏ) có thể bị tìm thấy ở …, Code được tạo có sãn trên github, chúng tôi hy vọng nó sẽ cỗ vũ cộng đồng machine learning làm việc và phát triển hơn nữa.**

## The (6) dialog bAbI tasks

Phần này trình bày bộ 6 nhiệm vụ để thử nghiệm hệ thống hộp thoại đầu cuối trong miền nhà hàng được mô tả trong bài báo:

…

Mỗi nhiệm vụ kiểm tra một khía cạnh độc đáo của hộp thoại. Nhiệm vụ được thiết kế để bổ sung cho 20 nhiệm vụ bAbI để hiểu được câu chuyện của phần trước.

Đối với mỗi task, có 1000 hộp thoại để đào tạo, 1000 cho phát triển và 1000 để thử nghiệm. Đối với nhiệm vụ 1-5, chúng tôi cũng bao gồm một bộ thử nghiệm thứ hai (có hậu tố -OOV.txt) chứa các hộp thoại bao gồm các thực thể không có trong tập huấn luyện và phát triển.

Định dạng tệp cho mỗi tác vụ như sau:

…

Các ID cho một hộp thoại nhất định bắt đầu từ 1 và tăng lên. Khi các ID trong một tập tin reset lại 1 bạn có thể xem xét các câu sau như là một hộp thoại mới. Khi bot nói hai lần trong một hàng chúng tôi đã sử dụng mã thông báo đặc biệt “<SILENCE>” để điền vào cho người sử dụng thiếu lời. Xem thêm chi tiết trong README kèm theo bộ dữ liệu. Mục đích của nhiệm vụ là dự đoán các phát biểu của bot, có thể là các câu hoặc các cuộc gọi API (các câu bắt đầu với mã thông báo đặc biệt "api\_call"). Đây là một ví dụ về hộp thoại (từ Tác vụ 1):

…

## The Children’s Book Test

Phần này trình bay Children’s Book Test (CBT), được thiết kế để đo trực tiếp các mô hình ngôn ngữ có thể khai thác được phạm vi ngôn ngữ rộng lớn hơn. CBT được xây dựng từ những cuốn sách được cung cấp tự do nhờ Project Gutenberg. Thông tin chi tiết và các kết quả cơ sở của tập dữ liệu này có thể được tìm thấy trong bài báo:

…

Sau khi phân phát sách để đào tạo, xác nhận hoặc kiểm tra bộ, chúng tôi đã hình thành các câu hỏi 'câu hỏi' từ các chương trong cuốn sách bằng cách liệt kê 21 câu liên tiếp. Trong mỗi câu hỏi, 20 câu đầu tiên tạo thành ngữ cảnh và một từ được xóa khỏi câu thứ 21, câu này trở thành query. Các mô hình phải xác định từ trả lời trong số 10 câu trả lời ứng viên xuất hiện trong các câu ngữ cảnh và câu truy vấn. Đối với phân tích hạt mịn hơn, chúng tôi đã đánh giá bốn loại câu hỏi bằng cách loại bỏ các loại từ khác nhau: Các thực thể được đặt tên, (danh từ) chung chung, động từ và giới từ

Dưới đây là ví dụ cho câu hỏi (context + query) từ Alice in Wonderland by Lewis Carroll:

…

## The Movie Dialog dataset

Phần này trình bày Movie Dialog dataset (MDD), được thiết kế để độ hiệu quả của models có thể đặt được với mục đích và không mục đích về chủ đề phim ảnh( trả lời câu hỏi, khuyến nghị và thảo luận).

Chi tiết và baseline result dựa trên dataset này có thể được tìm thấy tại:

…

Định dạng tệp lại giống như trong các tác vụ bAbI. Các ID cho một hộp thoại nhất định bắt đầu từ 1 và tăng lên. Mỗi ID bao gồm một lượt cho mỗi người nói (một "trao đổi"), được chia tab. Khi các ID trong một tập tin thiết lập lại cho 1 bạn có thể xem xét các câu sau đây như là một cuộc trò chuyện mới.

…

## The WikiMovies dataset

Điều này chỉ bao gồm phần QA của bộ dữ liệu Dialog phim, nhưng sử dụng ba kiến ​​thức khác nhau: sử dụng cơ sở tri thức truyền thống (KB), sử dụng Wikipedia như là nguồn kiến ​​thức, hoặc sử dụng IE (information extraction khai thác thông tin) qua Wikipedia. Điều này cho phép kiểm tra khả năng của các mô hình trực tiếp đọc các tài liệu để trả lời các câu hỏi, và để so sánh nó với KBs truyền thống trong cùng một môi trường. Xem giấy để biết thêm chi tiết:

…

## The Dialog-based Language Learning dataset

Phần này trình bày bộ dữ liệu học ngôn ngữ dựa trên Dialog, được thiết kế để đo lường mô hình có thể thực hiện như thế nào khi học tập với tư cách là một sinh viên được trả lời bằng văn bản của giáo viên đối với câu trả lời của sinh viên (cũng như có thể nhận được một tín hiệu thưởng bên ngoài). Thông tin chi tiết và các kết quả cơ sở của tập dữ liệu này có thể được tìm thấy trong bài báo:

…

Đây là một hộp thoại ví dụ, số cuối cùng (0 hoặc 1) là phần thưởng bên ngoài:

## The SimpleQuestions dataset

Phần này đề xuất SimpleQuestions, một bộ dữ liệu thu thập được để nghiên cứu trả lời câu hỏi tự động với các câu hỏi do con người tạo ra. Thông tin chi tiết và các kết quả cơ sở của tập dữ liệu này có thể được tìm thấy trong bài báo:

…

Bộ dữ liệu SimpleQuestions bao gồm tổng cộng 108,442 câu hỏi được viết bằng ngôn ngữ tự nhiên bằng các chú thích nói tiếng Anh của con người mỗi lần kết hợp với thực tế tương ứng, định dạng như (đối tượng, mối quan hệ, đối tượng), cung cấp câu trả lời mà còn là một lời giải thích đầy đủ. Sự thật đã được trích ra từ Cơ sở Kiến thức Freebase. Chúng tôi ngẫu nhiên xáo trộn những câu hỏi này và sử dụng 70% trong số đó (75910) làm tập huấn luyện, 10% là tập hợp xác nhận (10845) và 20% còn lại là tập kiểm tra.

Dưới đây là một số ví dụ về các câu hỏi và sự kiện:

…

## HITL Dialogue Simulator

Mô phỏng đối thoại Human-in-the-loop (HTIL) cung cấp khuôn khổ để đánh giá khả năng của một bot để học cách cải thiện hiệu suất của nó trong một cài đặt trực tuyến sử dụng phản hồi từ đối tác hộp thoại của nó. Bộ dữ liệu chứa các câu hỏi dựa trên bộ dữ liệu bAbI và WikiMovies, nhưng bây giờ với phản hồi từ đối tác đối thoại Chúng tôi bao gồm cả hộp thoại mô phỏng và con người. Các hộp thoại được thực hiện theo cùng một mẫu như trong bộ Tập dữ liệu về Ngôn ngữ học Dựa trên Dialog, nhưng bây giờ phụ thuộc vào mô hình của dự đoán thông qua trình mô phỏng.

Chi tiết đầy đủ về mô phỏng này có trong bài báo sau: