

# **AUTOMATIC QUESTION ANSWERING SYSTEM**

- Aashay Singhal, Aman Mehta

**Aim:** Build an Automatic QA system over high school history textbook and improve quality of answers. Our answer is calculated by converting each given sentence of the document in a vector of features and building a sentence ranking algorithm, i.e. an algorithm to select the most relevant sentences to a given question. So we can divide quality improvement of this task into 2 subtasks, namely:

1. Improving KB
  - a. By adding new relevant relations in our KB
2. Improving sentence ranking:
  - a. Finding bugs/scope of improvement in the feature extraction
  - b. Improve training of feature vectors of each sentences

## **Improving KB**

We have worked over an algorithm to extract the parent subject of a pronoun. This parent extraction (Coreference resolution) is done for every pronoun of the document. This has enabled the KB to capture information which was previously missing due to use of pronouns.

For eg:

Aman is 20 years old. He studies at IIIT-H.

CASE QUESTION: Where does Aman study?

**Previous:** KB : {<Aman is 20 years old>, <He studies at IIIT-H>}

The information is not present in our KB.

**Now:** KB : {<Aman is 20 years old>, <Aman studies at IIIT-H>}

We have added the new previously missing information.

Better cases are described under Results and Observations section for better understanding.

## **Improve Sentence Ranking**

1. Previously the feature weights were not trained over question answering dataset.

2. We have used [SQuAD](#) dataset which is a stanford question answering dataset. In this dataset, the answer to each question is a segment of text, or *span*, from the corresponding reading passage. It consists of multiple topics where each topic includes multiple paragraphs. Each of these paragraphs has multiple QA pairs. It has a total of approximately 100,000 QA pairs.
3. We extracted features of 400 questions which make upto around 4000 <question>-<candidate answer> pairs. We used 80:20 split for train and test. Feature extraction of each QA pair happens in order of minutes and hence we could only train over this small portion (taking around 18 hours).

## **Results and Observations (using this trained model to rank sentences):**

Overall: There has been a significant improvement in ranking of sentences. There are around 10% of the total sentences which were rendered useless due to the use of pronouns. We have overcome this problem by mapping each of the pronouns to their respective parent subjects.

And hence our model is now able to rank such sentences better. Let us explain using the following real-case examples.

Note: Don't mind the order of sentences, we just printed these sentences in decreasing order of their relevance with the question (just for the sake of comparison of previous and current sentence ranking algorithm).

### **Case 1: Missing relations are added**

**Q1:** How did forest acts change the lives of pastoralists?

#### **Previous answer:**

Like pastoralists in India, the lives of African pastoralists have changed dramatically over the colonial and post-colonial periods. How did the life of pastoralists change under colonial rule? They change the paths of their annual movement, reduce their cattle numbers, press for rights to enter new areas, exert political pressure on the government for relief, subsidy and other forms of support and demand a right in the management of forests and water resources. The new territorial boundaries and restrictions imposed on them suddenly changed the lives of pastoralists. These Forest Acts changed the lives of pastoralists. But the life history of the poor pastoralists who depended only on their livestock was different. Colonial Rule and Pastoral Life Under colonial rule, the life of pastoralists changed dramatically. When pasturelands in one place was closed to them, they changed the direction of their movement, reduced the size of the herd, combined pastoral activity with other forms of income and adapted to the changes in the modern world. Unlike the mountain pastoralists, it was not the cold and the snow that defined

the seasonal rhythms of their movement: rather it was the alternation of the monsoon and dry season. In the deserts of Rajasthan lived the Raikas. Second, by the mid-nineteenth century, various Forest Acts were also being enacted in the different provinces. When Pastures Dry Drought affects the life of pastoralists everywhere. All of them had to adjust to seasonal changes and make effective use of available pastures in different places. Conclusion So we see that pastoral communities in different parts of the world are affected in a variety of different ways by changes in the modern world. How did pastoral communities elsewhere cope with these changes in the modern world? The timing of their entry and departure was specified, and the number of days they could spend in the forest was limited. Second, a new distinction between the wealthy and poor pastoralists developed. Pastoralists reacted to these changes in a variety of ways. Such changes were not experienced only by pastoral communities in India. Like the Maasai, other pastoral groups were also forced to live within the confines of special reserves. New laws and new borders affect the patterns of their movement. They had to move because the Forest Department permits that had been issued to them now ruled their lives. The frequent bad years led to a steady decline of the animal stock of the pastoralists. Not All were Equally Affected In Maasailand, as elsewhere in Africa, not all pastoralists were equally affected by the changes in the colonial period. The social changes in Maasai society occurred at two levels. These pastoral movements allowed time for the natural restoration of vegetation growth.

### **Answer using new model:**

Like pastoralists in India, the lives of African pastoralists have changed dramatically over the colonial and post-colonial periods. How did the life of pastoralists change under colonial rule? They change the paths of their annual movement, reduce their cattle numbers, press for rights to enter new areas, exert political pressure on the government for relief, subsidy and other forms of support and demand a right in the management of forests and water resources. The new territorial boundaries and restrictions imposed on them suddenly changed the lives of pastoralists. These Forest Acts changed the lives of pastoralists. But the life history of the poor pastoralists who depended only on their livestock was different. Colonial Rule and Pastoral Life Under colonial rule, the life of pastoralists changed dramatically. When pasturelands in one place was closed to them, they changed the direction of their movement, reduced the size of the herd, combined pastoral activity with other forms of income and adapted to the changes in the modern world. They had to move because the Forest Department permits that had been issued to them now ruled their lives. Unlike the mountain pastoralists, it was not the cold and the snow that defined the seasonal rhythms of their movement: rather it was the alternation of the monsoon and dry season. In the deserts of Rajasthan lived the Raikas. Second, by the mid-nineteenth century, various Forest Acts were also being enacted in the different provinces. When Pastures Dry Drought affects the life of pastoralists everywhere. The timing of their entry and departure was specified, and the number of days they could spend in the forest was limited. All of them had to adjust to seasonal changes and make effective use of available pastures in different places. Conclusion So we see that pastoral communities in different parts of the world

are affected in a variety of different ways by changes in the modern world. How did pastoral communities elsewhere cope with these changes in the modern world? Second, a new distinction between the wealthy and poor pastoralists developed. Pastoralists reacted to these changes in a variety of ways. **Most often, they did not have the resources to tide over bad times. They needed a permit for entry. Such changes were not experienced only by pastoral communities in India. They were now prevented from entering many forests that had earlier provided valuable forage for their cattle. In the dry season they moved to the coastal tracts, and left when the rains came.** Like the Maasai, other pastoral groups were also forced to live within the confines of special reserves. New laws and new borders affect the patterns of their movement.

The highlighted portion(bold) is the useful portion w.r.t. the question which was missing previously but is now included in the answer.

Explanation:

1. *They needed a permit for entry* - This sentence previously did not add any value to the question
2. But when we coreferenced this and used it as "*They (Pastoralists) needed a permit for entry*" - This is included in the answer now.

And similarly about other sentences. We have successfully tackled a issue to add missing relations (to some extent).

## Case 2: Existing relations are ranked relatively better

**Example Q2:** Explain the warriors of the Maasai society

**Previous answer:**

First, the traditional difference based on age, between the elders and warriors, was disturbed, though it did not break down entirely. The warriors consisted of younger people, mainly responsible for the protection of the tribe. White settlers and European colonists saw pastoralists as dangerous and savage – people with whom all contact had to be minimised. Through these Acts some forests which produced commercially valuable timber like deodar or sal were declared 'Reserved'. They combined a range of different activities – cultivation, trade, and herding – to make their living. Other forests were classified as 'Protected'. One group of Raikas – known as the Maru (desert) Raikas – herded camels and another group reared sheep and goat. It was seen as 'waste land' that needed to be brought under cultivation. In pre-colonial times Maasai society was divided into two social categories – elders and warriors. Pastoralism in Africa Let us move to Africa where over half the world's pastoral population lives. Further to the east, in Garhwal and Kumaon, the Gujjar cattle herders came down to the dry forests of the

bhabar in the winter, and went up to the high meadows – the bugyals – in summer. Young men came to be recognised as members of the warrior class when they proved their manliness by raiding the cattle of other pastoral groups and participating in wars. Consequently, the traditional authority of both elders and warriors was adversely affected. These contractors tried to extract as high a tax as they could to recover the money they had paid to the state and earn as much profit as they could within the year. First, the colonial state wanted to transform all grazing lands into cultivated farms. Where have the Grazing Lands Gone? So the Raikas combined cultivation with pastoralism. Feeding the cattle became a persistent problem. **They defended the community and organised cattle raids.** How Did these Changes Affect the Lives of Pastoralists? How Did the Pastoralists Cope with these Changes? These Forest Acts changed the lives of pastoralists. In the deserts of Rajasthan lived the Raikas. Pastoralists are not relics of the past. When pasturelands in one place was closed to them, they changed the direction of their movement, reduced the size of the herd, combined pastoral activity with other forms of income and adapted to the changes in the modern world.

#### **Answer using new model:**

First, the traditional difference based on age, between the elders and warriors, was disturbed, though it did not break down entirely. **They defended the community and organised cattle raids.** The warriors consisted of younger people, mainly responsible for the protection of the tribe. They combined a range of different activities – cultivation, trade, and herding – to make their living. It was seen as ‘waste land’ that needed to be brought under cultivation. White settlers and European colonists saw pastoralists as dangerous and savage – people with whom all contact had to be minimised. Through these Acts some forests which produced commercially valuable timber like deodar or sal were declared ‘Reserved’. Pastoralism in Africa Let us move to Africa where over half the world’s pastoral population lives. Other forests were classified as ‘Protected’. One group of Raikas – known as the Maru (desert) Raikas – herded camels and another group reared sheep and goat. In pre-colonial times Maasai society was divided into two social categories – elders and warriors. Further to the east, in Garhwal and Kumaon, the Gujjar cattle herders came down to the dry forests of the bhabar in the winter, and went up to the high meadows – the bugyals – in summer. Young men came to be recognised as members of the warrior class when they proved their manliness by raiding the cattle of other pastoral groups and participating in wars. Consequently, the traditional authority of both elders and warriors was adversely affected. These contractors tried to extract as high a tax as they could to recover the money they had paid to the state and earn as much profit as they could within the year. First, the colonial state wanted to transform all grazing lands into cultivated farms. Where have the Grazing Lands Gone? They needed a permit for entry. So the Raikas combined cultivation with pastoralism. Feeding the cattle became a persistent problem. How Did these Changes Affect the Lives of Pastoralists? How Did the Pastoralists Cope with these Changes? By September they began their return movement. These Forest Acts changed the lives of pastoralists. In the deserts of Rajasthan lived the Raikas. Pastoralists are not relics of the past.

Note the position of highlighted sentence:

1. *They defended the community and organised cattle raids* - If you read the question, it is a very important statement but previously it was ranked very badly, due to the missing relation that “They” in the statement corresponds to “The warriors”
2. Now, due to the transformation - *The warriors(They) defended the community and organised cattle raids* is an extremely important statement for the question and hence is ranked to be the second most important statement.

So, here we can see the improvement in our sentence ranking.

### Case 3: Answers not improved significantly

Let's now take a back seat to understand where things did not show up as per expected

1. In some of the questions, we expected several sure-shot improvements which did not come along. Why?
  - a. Note that the following outputs have evolved after using Hobbs Coreference Algorithm and a XGB Regressor model trained over only 4000 QA pairs (out of the 100,000+ QA pairs)
  - b. The hobbs algorithm can be surely improved over this baseline(which we have used currently) and if training over only 4% of the total QA pairs can give accuracy of 60% over SQuAD dataset and can give significant improvements on Sentence Ranking, then we strongly believe better training would fetch better results.

## Future Work

1. We could work on methods to improve Hobbs coref resolution algorithm:
  - a. By performing preprocessing over english text using Stemming and Stop-words deduction
  - b. By exploring other Coreference Resolution algorithms
2. Better training using SQuAD dataset:
  - a. Once getting server access and shifting entire module to server and adding all the dependencies required, we could keep the model open for training since it requires time in order of days/weeks to train all the QA pairs
  - b. Report results again (after performing step (a.)) to mark improvement
3. We recently read thoroughly about a concept of “[Active Question Answering](#)”. The concept being, consider a query question ‘Q’, now we try to convert Q into 10 other

paraphrased questions  $q_1, q_2, \dots, q_{10}$  and we get 10 different answers for each  $q[i]$ . Later we build an aggregator model to extract out best answer  $A$  out of these  $a_1, a_2, \dots, a_{10}$ .

- a. Idea: Trying to paraphrase question means we try to get all the possible interpretations of a question and get an answer for each of the representation,  $q[i]$ .
- b. Eg:
  - i. *Q. Mention the date, month and year when Louis XVI died.* becomes,
  - ii. *q1. When did Louis XVI die?* Which may fetch better results than original  
Q
- c. The paraphrasing of question is a CNN learnt over a question-paraphrasing dataset which we could also implement to expand our scope of questions.
- d. This paraphrasing technique is essential we feel for LAQs at the least. Because LAQ questions are difficult to be understood using a single representation of query by a QA system.
- e. The Answer aggregator model is a Reinforcement Learning model, which we will have to further read about to understand completely. This model proved to produce 11% better F1 score than the state-of-the-art QA currently.