CPSC 599.27 Term Project Proposal: Reverse Jeopardy!

SHAEMUS MELVIN* and VALERIE KIM*, University of Calgary, Canada

ACM Reference Format:
Shaemus Melvin and Valerie Kim. 2023. CPSC 599.27 Term Project Proposal: Reverse Jeopardy!. 1, 1 (December 2023), 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

IBM's Watson defeating Jeopardy! GOAT Ken Jennings is an event that has cemented itself in the annals of AI history, alongside Deep Blue defeating Garry Kasparov and AlexNet using Convolutional Neural Networks to reduce the error rate in the ILSVRC by 10%. While both the domains of Games and ImageNet have since had hundreds of thousands of working hours dedicated to their research, Jeopardy! has enjoyed far less publicity.

Watson's task in Jeopardy! involved receiving the host's prompt in the form of a declarative statement (herein referred to as a clue) in order to reply with the contestant's response in the form of an interrogative statement (herein referred to as an interrogative) that would lead to the clue. Ferrucci et al. states that Watson represented a major integration in many QA technologies including: "parsing, question classification, question decomposition, automatic source acquisition and evaluation, entity and relation detection, logical form generation, knowledge representation and reasoning, and lastly, confidence estimation."[5]

We plan to flip this task on its head creating a system that, when given a prompt in the form of a Jeopardy! interrogative ("What is Germany?"), will output clues ("This European country was split between east and west for much of the 20th century and reunified in 1990"). Our challenge is in traversing knowledge to generate a Jeopardy! domain appropriate clue for any given topic, which must fit a number of constraints: firstly, clue categories will restrict the domain of related concepts to mention; and second, there must be a conceptual distance between what is explicitly mentioned in the clue vs. the interrogative. This model will be comprised of at least two models: firstly a world-knowledge model ("Germany", "Europe", "Berlin" all being closely associated, for example), and a natural language generation model which takes these associations and outputs a Jeopardy! clue.

The language generation model will be trained on an archive of Jeopardy! question and answer pairs going back to 1996, making up 216,930 pairs[1] We have several options for our knowledge representation map. Our current direction is in transfer learning, where we take a model pre-trained on a subset of Wikipedia so that our system can generate Clues in response to Interrogatives of topics unexplored in Jeopardy!.

*Both authors contributed equally to this research.

Authors' address: Shaemus Melvin, shaemus.melvin@ucalgary.ca; Valerie Kim, minjae.kim1@ucalgary.ca, University of Calgary, 2500 University Dr SW, Calgary, Alberta, Canada, T2N 1N4.

2 RELATED WORK

The body of work referring to the question answering format within natural language processing is referred to (understandably) as QA[7]. There has been quite a bit of research and documentation on QA models that answer questions, both before Watson made its public debut and after, and additionally there is work by the creators of Watson/DeepQA on the technology itself, explaining both how the technology functions[6] and how it was built[5]. Additionally, in the years following Watson's appearance some of the creators have further analyzed Watson's performance and speculated on real-life use-cases such as medical and governmental use [2]. There is comparatively little research regarding "reverse QA", where a question is generated from an answer. It is within this gap of research that we will be working, using the other component of the project to bridge the gap.

The other major body of work relevant to the project is Natural Language Generation. This will be the crux of the project, with the QA being the subordinate component. Natural Language Generation has recently become a popular topic of discussion due to the release of ChatGPT, however GPT-3 is itself obviously based on decades of research[4][3].

3 PROPOSED WORK

As mentioned in the introduction, the model will be split into two sub- components: the natural language generation model, and the knowledge representation model. The natural language generation model will be trained exclusively on Jeopardy! questions in order to generate clues in the format we desire, and the knowledge representation model can be trained/generated based on external (read: non-Jeopardy!) data. It is important to note as well that since the language generation model is the most important part of the project, it would be feasible to use a pre-trained knowledge model (E.g. a model trained on SQuAD[8]) without defeating the purpose of the project.

As far as preprocessing goes, it might be desirable to modify the format of the responses (e.g. remove 'what is' from 'what is Germany' to make the response 'Germany'), but it should not affect the overall functionality of the of the system either way, as long as the user knows whether to include 'what is' in the prompt. Additionally, some lemmatization would be useful, at least with knowledge representation, so that the terms 'German city' and 'City in Germany' are both represented the same way in the knowledge map.

Another preprocessing step that could be useful is parts of speech tagging. Specifically, it would be useful to have the input data specify what part of the clue is pointing towards the response (if any are at all - sometimes clues consist simply of a description of the response, e.g. 'A part of the hand, or a numeral', 'What is digit'). Additionally, this would aid in formatting the clues correctly; the last thing that we want to occur is a jumbled mess of topics as a clue that make no grammatical sense ('what is Germany' to 'Europe country World War II Berlin').

REFERENCES

[1] [n. d.]. https://redd.it/1uyd0t

[2] Kevin R. Dixon Zachary Benz Justin Basilico Aaron K. Baughman, Wesley Chuang. 2014. DeepQA Jeopardy! Gamification: A Machine Learning Perspective. *IEEE Transactions on Computational Intelligence and AI in Games* 6, 1 (2014).

[3] Emiel Krahmer Albert Gatt. 2018. Survey of the State of the Art in Natural LanguageGeneration: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research* 61 (2018), 65–170.

[4] Haifan Gong Miaoxin Chen Junxin li Ying Shen Min Yang Chenhe Dong, Yinghui Li. 2022. A Survey of Natural Language Generation. *Association for Computing Machinery Computing Survey* 1, 1 (2022).

[5] David A. Ferucci. 2010. Building Watson: An Overview of the DeepQA Project. *AI Magazine* 31, 3 (2010), 59–79.

[6] David A. Ferucci. 2012. Introduction to "This is Watson". *IBM Research and Development* 56, 3–4 (2012).

[7] T. Mitamura Menggiu Wang, Noah A. Smith. 2007. What is the Jeopardy Model? A Quasi-Synchronous Grammar for QA. (2007).

[8] Konstantin Lopyrev Percy Liang Pranav Rajpurkar, Jian Zhang. 2016. SQuAD: 100,000+ Questions for Machine Comprehension of Text. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (2016), 2383–2392.