The performance of a robust Bayesian approach to the 20 questions game under different patterns of noise

Author: Anton Donle
Supervisor 1: B. B. Kruit
Supervisor 2: S. Schlobach
Institute: Vrije Universiteit Amsterdam

Objective

Playing the 20 questions game and finding the correct entity when exposed to Noise.

20 Questions Game

The goal is to find the entity that an answerer is thinking of by asking up to 20 yes or no questions.

Background

In today's data-driven society, where large amounts of information are accessible to a wide audience, the ability to efficiently and accurately query for information has become increasingly important. Developing game-like approaches to database querying can be used to improve the way users can retrieve specific information from databases. This paper focuses on designing a search algorithm that can handle input errors arising from different sources.

Introduction

Interacting with Knowledge Graphs introduces challenges, especially if the user is not accustomed to its structure. Errors can occur because of modelling choices, open/closed-world assumptions and different beliefs. In this research the 20 Questions game with a Knowledge Graph is used to develop a robust querying algorithm, which can find the correct entity in a Knowledge Graph even if noise in the form of wrong answers is introduced. A non-robust algorithm would aim to split the remaining entities in half with every question, ruling out the ones that do not match the given answer. Robustness is achieved by a combination of Naïve Bayes classifier combined with an entropy formula, ensuring that the probability of an entity never drops to 0. Figure 1 shows how the robust algorithm would play the game, compared to a 'conventional' search algorithm with a simple example.

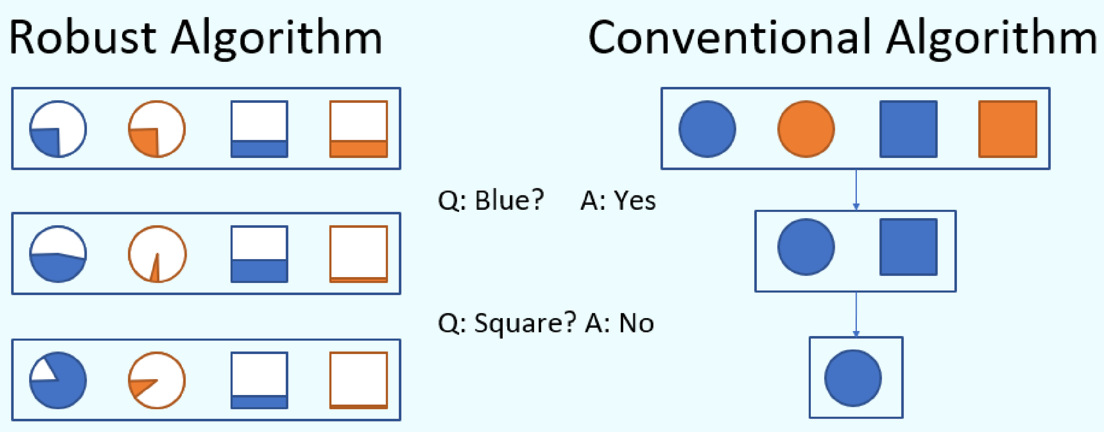


Figure 1. 20 Questions Game comparison of conventional and robust algorithm

Methods

To answer the research question an engine and a bot were developed. The engine uses the Yago data set [1] in combination with the FB15K database [2], reduced to the 134 most famous human entities according to Wikipedia in 2021. The answering bot exposes the algorithm to different patterns and numbers of wrong answers. The amount of correctly identified entities out of those 134 is measured as performance.

For playing the 20 questions game, the Naïve Bayes Classifier assigns an equal probability to all entities. An adjusted version of the Shannon Entropy Formula [3] is used to calculate the entropy loss of all possible splits (questions). The question with the largest entropy loss provides the largest estimated information gain. It is asking this question to the answering bot.

Implementation Details

Attributes: For querying the knowledge graph an attribute is defined as a pair of predicate and object (e.g. Type Human / Occupation Actor / Nationality United States)

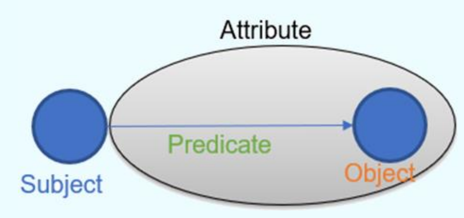


Figure 3. An Attribute in a knowledge graph

Naïve Bayes Classifier: Updates probabilities after each question and normalizes them. Predicts the most likely entity after 20 questions. An Error Rate of 0.1 is added to avoid probabilities dropping to 0.

Weighted Entropy Formula: Find the split that produces the largest estimated entropy loss H(X), based on entity probabilities.

Random Noise: The answering bot will answer N random questions wrong, since this is not deterministic, the average performance over 10 runs is used

Chunk Noise: The answering bot will answer N subsequent answers wrong starting in position X for all possible positions. This is deterministic.

Random + Specific Noise: The answering bot will answer the question in position X wrong and the other N answers are given in random positions. This is averaged over 10 runs.

Formulas

$$P(E|\mathcal{X}) = P(E)\prod_{i=1}^{|\mathcal{X}|} P(\mathcal{X}_i|E)$$

Equation 1. Naïve Bayes

$\mathcal{E}$ the set of all entities
E a specific entity E
X a specific attribute X
$\mathcal{X}$ the set off all defined attributes
$\mathcal{X}_n$ the n-th defined attribute
$\mathcal{E}_{\neg X}$ the set of all entities without attribute X
$\mathcal{E}_X$ the set of all entities with attribute X
$|\mathcal{X}|$ the amount of defined attributes

$$H(X) = -[W(X)\log_2(W(X)) + [W(\neg X)\log_2(W(\neg X))]$$

Equation 2. Entropy loss

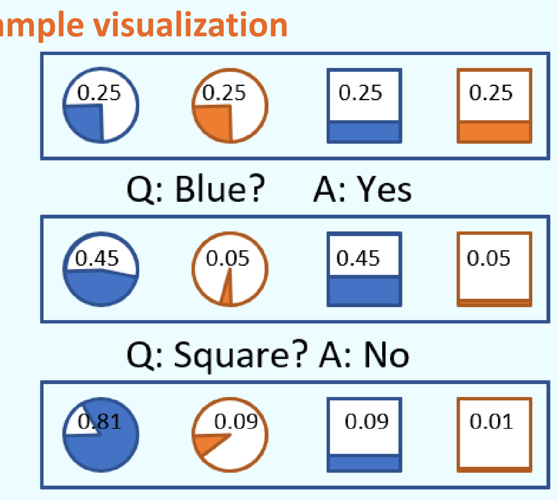
$$W(X) = \frac{\sum P(\mathcal{E}_X)}{\sum P(\mathcal{E})}$$

Equation 3. Proportion of Entities with attribute X

$$W(\neg X) = \frac{\sum P(\mathcal{E}_{\neg X})}{\sum P(\mathcal{E})}$$

Equation 4. Proportion of Entities without attribute X

Example visualization



Attribute	Sum Prob Yes	Sum Prob No	Entropy Loss
Blue	0.5	0.5	1
Square	0.5	0.5	1
Orange + Circle	0.25	0.75	0.811

Attribute	Sum Prob Yes	Sum Prob No	Entropy Loss
Blue	0.9	0.1	0.49
Square	0.5	0.5	1
Orange + Circle	0.05	0.95	0.29

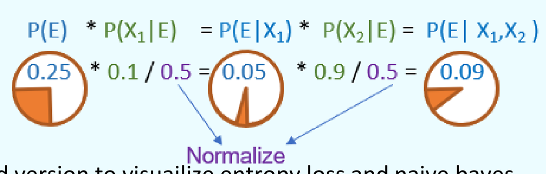


Figure 2. 20 Questions Game simplified version to visuailize entropy loss and naive bayes

Results

Type of Noise	Wrong Answers	N=0	N=1	N=2	N=3	N=4	N=5	N=6
Random Noise	Performance:	100	97.9	92.5	75.4	46.3	19.2	3.2
	STD:	0	1.6	2.1	4.1	3.7	3.4	1.6
Chunk Noise	Performance:	100	98.5	92.3	80.2	64	45.9	23.8
	STD:	0	1.8	6.5	17.3	24.2	23.9	18.0
Specific + Random	Performance:	100	98.5	92.8	76.9	46.2	18.7	3.6
	STD:	0	1.8	3.5	6.4	7	4.9	1.7

Figure 4. Performance under different types of noise

Under both Random and Chunk Noise the performance decreases with more wrong answers. However, this trend is stronger under Random Noise. For 0<N<3 this difference is not statistically significant. For N>3 it is significant.

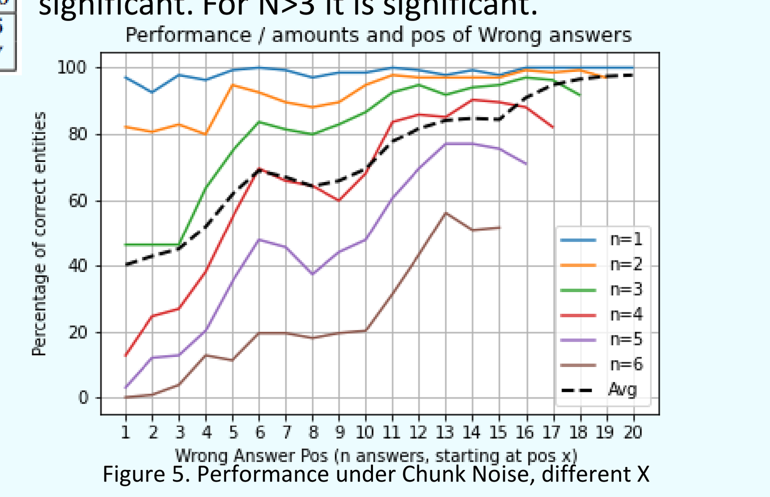


Figure 5. Performance under Chunk Noise, different X

Chunk Noise

The plot shows that the performance increases strongly the later the wrong answers occur, with a local maximum for different N at position 5/6.

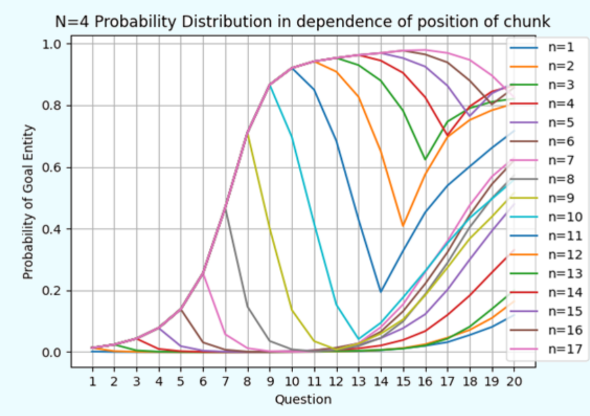


Figure 6. Average Goal Entity Probability Chunk Noise

Plotting the average goal entity probability under Chunk Noise shows that the probability recovers faster the later the wrong answer occurs.

When wrong answers occur early, the goal entity becomes relatively low and the algorithm focuses on a different search space and for later wrong answers the goal entity has a higher probability score, hence it has a better chance to remain the most likely one.

Random + Specific Noise

In Figure 6 it can be seen that the Performance increases slightly with later wrong answers. However, this trends is much weaker than under Chunk Noise and it does not reproduce the local maxima at X=5/6.

The performance and standard deviation are very similar to Random Noise, which suggests that position of a wrong chunk is more relevant for performance than position of one wrong answer.

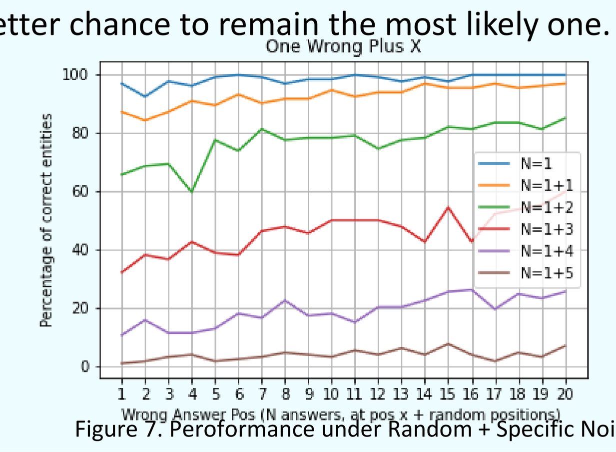


Figure 7. Peroformance under Random + Specific Noise

Conclusion & Future Work

The performance of the robust bayes approach depends on the pattern of noise. Statistical differences between Chunk Noise and Random Noise could be found. The general trends observed in the experiments were that the performance decreases significantly with more wrong answers and whenever the wrong answers occur early. The local maxima in performance at X=5/6 under Chunk Noise could not be fully explained and might be caused by either the dataset or the algorithm itself. No benchmarking with other algorithms was performed, because there was no comparable research in a similar enough domain.

A method that could be further explored in this context is Unequal Error Protection (UEP) to prevent the algorithm from decreasing the goal entity probability too much for early wrong answers. Testing the algorithm on a different dataset could help understand the observed phenomena. Different strategies to reduce computational complexity could be explored. Instead of calculating the entropy of all attributes, this could be done for a random subset of all possible attributes.

References

[1] Pellissier Tanon, T., Weikum, G., Suchanek, F.: YAGO 4: A Reason-able Knowl-edge Base. In: Harth, A., Kirrane, S., Ngonga Ngomo, A.C., Paulheim, H., Rula, A., Gentile, A.L., Haase, P., Cochez, M. (eds.) The Semantic Web. pp. 583–596. Lecture Notes in Computer Science, Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-49461-2_34

[2] Dou, J., Tian, B., Zhang, Y., Xing, C.: A Novel Embedding Model for Knowledg Graph Completion Based on Multi-Task Learning. In: Jensen, C.S., Lim, E.P., Yang, D.N., Lee, W.C., Tseng, V.S., Kalogeraki, V., Huang, J.W., Shen, C.Y. (eds.) Database Systems for Advanced Applications. pp. 240–255. Lecture Notes in Computer Science, Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-73194-6_17

[3] Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal 27(3), 379–423 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[4] Chung, H.W., Sadler, B.M., Zheng, L., Hero, A.O.: Unequal Error Protection Querying Policies for the Noisy 20 Questions Problem. IEEE Transactions on Information Theory 64(2), 1105–1131 (Feb 2018). https://doi.org/10.1109/TIT. 2017.2760634, http://arxiv.org/abs/1606.09233, arXiv:1606.09233