Electronic Supplementary Material (ESI) for Digital Discovery.

Extracting Structured Seed-Mediated Gold Nanorod Growth Procedures from Literature with GPT-3

June 30, 2023

1 Supplemental Materials

1.1 Detailed Results Analysis

Table 1: Model F1-scores and accuracies on gold nanorod characterizations. The support numbers in parentheses account for only the true positives used for the accuracy calculation.

	Placement			Transcription	Combined	
	Precision	Recall	F1-Score	Accuracy	Adj. F1	Support
Aspect Ratio	0.81	0.85	0.83	1.00	**0.83**	24 (17)
Length	0.76	0.89	0.82	0.90	**0.75**	23 (16)
Width	0.74	0.70	0.72	0.93	**0.67**	25 (14)
LSPR	0.80	0.57	0.67	0.94	**0.62**	16 (8)
TSPR	1.00	0.50	0.67	1.00	**0.67**	8 (4)

A closer look at the extraction performance on measurements of gold nanorod dimensions is shown in Table 1. It is immediately clear that the model performs the best for extracting aspect ratios and lengths. This is not particularly surprising since these are perhaps the most common and most important measurements presented for gold nanorods in literature. By comparison, the width tends to suffer because while it may be presented in the text, it is less often identified explicitly and exclusively as the width of the nanorods, as the dimension is often shared with the size of the seeds in the seed solution. The SPR measurement extraction performance is even worse, though they have higher precisions than recalls, which indicates a tendency for errors to be false negatives rather than false positives. This disparity in precision and recall is not observed to nearly the same extent in the other measurements. One explanation for this possibly lies in a tendency for the SPR measurements to be mentioned in the context of measurements for other gold nanoparticle morphologies, which is generally not the case for the other measurements. For example, one can report SPR measurements for gold nanospheres, but this would not be the case for aspect ratios or combinations of length and width. This pattern may be comparatively difficult for the model to learn in complex cases, especially since there are fewer examples in the literature.

Table 2: Model F1-scores and accuracies on seed and growth solution information. The support numbers in parentheses account for only the true positives used for the accuracy calculation.

	Seed Solution				Growth Solution			
	F1	Accuracy	Adj. F1	Support	F1	Accuracy	Adj. F1	Support
Age	1.00	0.96	**0.96**	14 (14)	0.94	1.00	**0.94**	18 (16)
Stir Rate	1.00	0.88	**0.88**	8 (8)	0.71	1.00	**0.71**	9 (5)
Temperature	0.87	0.95	**0.83**	13 (10)	0.87	1.00	**0.87**	13 (10)
Precursor Volume	0.92	0.95	**0.87**	55 (47)	0.95	0.92	**0.87**	107 (96)
Precursor Concentration	0.96	0.97	**0.93**	59 (54)	0.90	0.99	**0.88**	95 (77)
Precursor Mass					1.00	1.00	**1.00**	2 (2)
Seed Shape	1.00	1.00	**1.00**	2 (2)				
Seed Size	0.93	0.90	**0.84**	8 (7)				

A more detailed analysis of the seed and growth solution components is shown in Table 2. For both the seed and the growth solutions, precursor volumes and concentrations dominate the statistics with relatively high support. For the seed solution, both the temperature and the seed size perform relatively poorly, however. The issues with the temperature likely come down to the multitude of different temperatures described in a typical seed solution preparation, such as heating or cooling of the cationic surfactant used or the temperature of the sodium borohydride used. The model is vulnerable to confusing these temperatures for the true target aging temperature, especially if the aging temperature is not provided while other temperatures are provided. This is less common for the growth solution, which is reflected in superior model performance. The sizes of the seeds in the seed solution are often not presented in the text and seem to be easily confused with either the width of resultant nanorods or the sizes of other gold nanoparticle morphologies produced in the growth process, particularly spheres. For the growth solution, the model tends to perform very poorly on the stir rate. This is due to the more complex procedures often presented for the growth solution, which may specify stirring multiple times, such as when combining the gold source and the cationic surfactant or when adding the silver source or acids.

Overall, however, performance is exceptional for the precursors and their associated quantities, which is very encouraging. Some common errors with precursors include mistakes such as when multiple precursors share the same volume. Furthermore, the static nature of the synthesis template tends to dissuade the model from adding precursors that are not present in the training set. Additionally, the model tends to make mistakes when multiple recipes are separately presented in a paragraph (i.e. multiple volumes for the same precursor are distributed throughout the text rather than being adjacent to each other). In these cases, the model will often only extract one recipe and, in at least one case, mistook a second seed solution recipe for a growth solution recipe, perhaps due to confusing it for a common pattern in which the growth solution recipe immediately proceeds after the seed solution recipe.