Constructing Models For Diagnosing Rare Diseases

Brian S. Mathiasen — soborg@diku.dk
Henrik G. Jensen — henne@diku.dk

31st January 2011

1 Disposition

The use of the Internet for aiding in diagnoses is an important part in the daily practice of physicians. It is therefore important to develop systems that will aid physicians, with particular focus on rare diseases.

The goal of the project is to develop a system for building language models for rare diseases. The language models will be based on prior knowledge provided by Orphanet[1]. It will be expanded through automated iterations on popular search engines. The information gathered from the results of these searches will be validated based on a correlation with the prior knowledge and new searches will be based on the posterior language model. We will use machine learning and natural language processing to extract the relevant information with a focus on symptoms and other important features.

During the project, we will evaluate the system by it's ability to rank a relevant disease (the precision of the system) compared with previous similar projects. A ranking is produced by entering a case report or a list of symptoms of a rare disease into the system.

2 Learning Goals

- Utilize natural language processing techniques.

- Utilize intermediate and advanced algorithms for machine learning.

- Utilize efficient data structures for data mining.

- Build language models for rare diseases based on results from popular search engines such as Google and PubMed.

---

[1] http://orpha.net/

- Build a model for disease/symptom harvesting.

- Utilize background knowledge of selected rare diseases to determine the validity and accuracy of the model.

- Compare validity and accuracy to previous work on support decision systems on diagnosing rare diseases.

3 Project Time Frame

Week	Activity	Subtasks
1	Programming	Select data format, arrange prior data
2	Programming	preliminary test searches, begin model coding
3	Programming	model and preliminary results
4	fine-tuning	fine tune programming
5	define test	.
6	tests, results	possible fine-tuning.
7	further test or buffer, report	.
8	report or buffer	.
9	report	.