Language models as a representation of meaning – A Summary

In the context of natural language processing the main goal of language models is to make machine generated text seem human-made. Language models try to achieve this goal by estimating the probability of a sequence of words given an input sequence: For example, if given the input sequence 'kings and __' the language model will compute the most probable output which is 'kings and **queens**'. This summary will focus on the key steps taken by a language model to generate the different probability scores for all words it knows to then estimate the best suited word (or phrase) to follow the input sequence.

A trained language model calculates a prediction in 3 steps: The starting point is the input sequence. The model looks up the embeddings of a given sequence. In the second step the input goes through some computations so that the model can capture the relevant information of the input words and also their relationship between each other. Then it uses its parameters to make predictions for the output as show in figure 1 below:

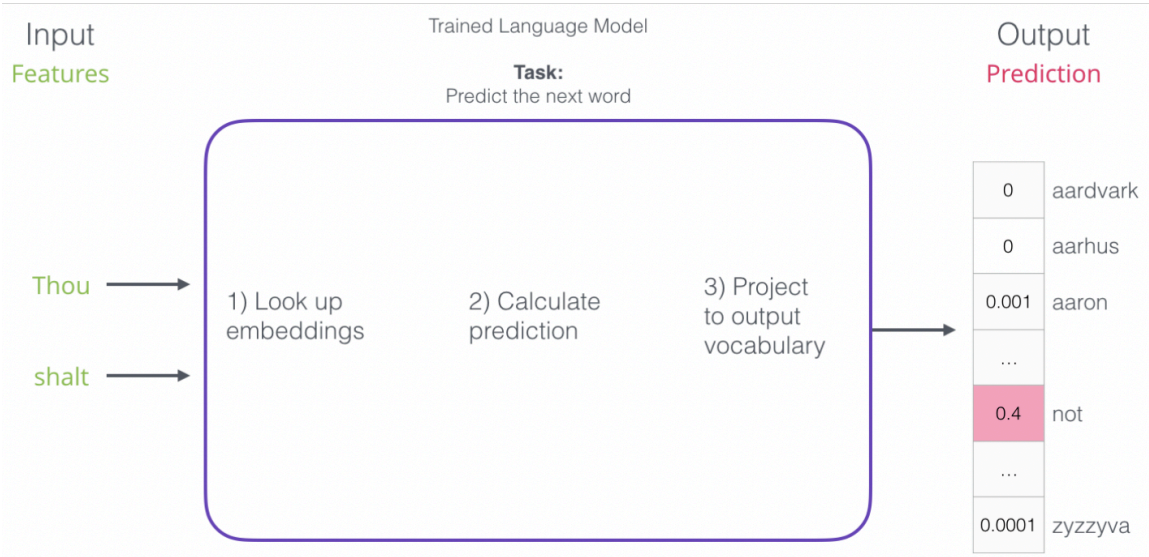

Figure 1: Visualization of a trained language model which calculates a prediction in 3 steps

The loss function of a language model is a method that determines the quality of a model's predictions. The loss function measures the difference between the predicted distribution and the actual distribution of the next word in the sequence by calculating a gradient that represents the direction and magnitude of the error with respect to the model's parameters. During the training of the model the loss function is minimized to be able to predict words more accurately.

Grammar and meaning are the basis for all language models. Grammar focuses on the study of how sentences in a language are constructed; how words change their form and combine with other words to make sentences. In language modeling the goal is to generate syntactically correct sentences. Meaning on the other hand refers to the information conveyed by these sentences. In the context of language models syntactically correct sentences using the meaning of symbols should be generated. A decent language model should be able focus on both these aspects.
Parameters capture the relationships between the input and the output. They are the weights and biases of a neural network which are adjusted during the training to minimize the difference between target and predicted outputs.

Two variants of language models can be differentiated: statistical language models (SLMs) and neural language models (NLMs). Statistical language models use probabilistic methods, such as n-grams, to determine a word based on its context. Neural language models depend on word embeddings and their processing using neural networks. SLMs are computationally more efficient whereas NLMs are more expensive but deliver better results.

The goal of distributional semantics is to represent the meaning of words based on the contexts they appear in. It is assumed that similar words have similar meanings. Language models predict the probability distribution of a word or sequence in a given context. They go beyond distributional patterns as they capture a word's meaning based on its context.