Brain Knowledge Engine

Hongyin ZHU [1], Yi ZENG [1, 2], Dongsheng WANG [1]

[1] Institute of Automation, Chinese Academy of Sciences, China,

[2] Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, China

zhuhongyin2014@ia.ac.cn

DOI:10.14931/aini2016.ds.12

Brain Knowledge Engine (BKE) is an automatic knowledge extraction, integration, analysis and service engine for understanding the Brain (http://www.linked-brain-data.org/bke). It is currently part of Linked Brain Data. The domain knowledge of the brain is automatically extracted and integrated from various sources containing PubMed, Allen Brain Atlas [1], Neuroscience Information Framework [2], NeuroLex [3], NeuroMorpho [4], Wikipedia, etc. BKE integrates massive brain knowledge and displays them with a user-friendly interface. It mainly contains the relations among brain building blocks, cognitive functions, brain diseases, etc. All the entities related to the target entity can be displayed in different categories respectively. For example, if the query term is "hippocampus", users would get the knowledge including the relations on region and disease, region and cognitive functions, neurons and regions, type of neurons in the specific region, and URIs that are related to this query. BKE provides users a relatively complete set of results from multiple perspectives.

Facts about the brain are always species specific. Knowledge extracted from different species cannot be simply mixed together, and whether it is true for other species needs biological experiment validations. Hence, BKE is designed to provide species meta-data for each knowledge triples extracted from various sources. Species labels can help users to compare the similarities and differences between different species easily.

BKE has a service which can automatically identify the type of neurons in the corresponding brain regions. Namely, it discovers contain relationships among brain regions and different types of neurons. Keyword match to find these relations turns out to be not efficient since it only identify a small number of relations of this kind. Hence, machine learning techniques need to be applied for improvements. However, the labeled training data for type of neurons in the specific brain region is not enough. In order to solve this problem, we manually annotate some training corpus and adopt a supervised learning strategy. During the corpus annotation stage, we summarized some elaborately design rules which can be adopted as features. This rule set consists of two subsets which are the positive set and negative set. The sentences, which are in accordance with the positive rule set, have a higher probability to describe the content of a specific brain region containing the type of neurons. On the contrary, the sentences following the rules of the negative rule set are considered as having little or nothing to do with region containing neurons. We also use the dependency parser [5] to automatically extract the syntax features of every sentence, including the distance of two entities, the length of the dependency path, the distance of every entity to the root node, etc. Then, we adopt the neural network to learn a model to classify the natural language sentences. 85% correct classification rate using 10-fold cross validation is achieved. Compared with the KNN, SVM and logistic regression, the neural network gets the best generalization performance.

BKE can integrate the newly extracted data which are with different formats compared to previous data very easily. In addition, it also integrates the figures that are highly relevant to the target entity. The figures are automatically extracted from Wikipedia and other websites. The integration of brain science related knowledge and the user-friendly interface can help users get relatively complete structured knowledge in seconds.

Citation: Hongyin ZHU, Yi ZENG, Dongsheng WANG (2016). Brain Knowledge Engine. *Advances in Neuroinformatics IV.* AINI 2016 and INCF Nodes Workshop Abstract: Demonstration Session-12. DOI:10.14931/aini2016.ds.12