Allen AI Science Challenge Winners

February 16th, 2016

First Place ($50K)

Score: 59.308%
Name: Chaim Linhart (Kaggle User "Cardal")
Location: Hod-Hasharon, Israel
Occupation: Sr. Researcher at TaKaDu, an environmental services company

Second Place ($20K)

Score: 58.344%
Kaggle Team: "poweredByTalkwalker"
Location of all team members: Luxembourg
Occupation: Team consists of employees from Trendiction/TalkWalker (a social media monitoring and analytics company)
Team leader: Benedikt Wilbertz (Kaggle User "Benedikt")
Other team members:

- Sébastien Wagener
- Christopher Schank
- Pol Gleis
- Mechel Conrad
- Rui Wang
- Maxime Marchès
- Rimbaud Van Eetvelde

Third Place ($10K)

Score: 58.257%
Name: Alejandro Mosquera (Kaggle User "amsqr")
Location: Reading, United Kingdom
Occupation: Sr. Principal Research Engineer at Symantec, a cyber security company

Method summaries

First Place - Chaim Linhart

Chaim’s model is a combination of several Gradient Boosting ensembles. He generated two types of features for the GB models: "basic" features like the length of each answer and the propensity of a word to part of a correct answer, as well as "search" features, based on searching for words from the question and answers in external corpora. The most important aspect of the model according to Chaim is the data resources he used, and the various ways he utilized them.

Second Place - poweredByTalkwalker team

poweredByTalkwalker built a large corpus that produced a total of 180GB in lucene indices. They used three main strategies: IR-based features, PMI-style features, and FeatureHashing. In total they arrived at around 900 features, from which they learned a final model using the XGBoost library.

Third Place - Alejandro Mosquera

Alejandro used careful knowledge base selection, data normalization, and a strong IR approach. His model consists of a logistic regression ranking over interactions between question/answer pairs. Model features are based on word2vec, IR and a few heuristics in order to categorize questions in a set of categories.