Sentiment Analysis Studio Features

Dr. Goutam Chakraborty

Objectives

- SAS has multiple products that can do sentiment analysis
  - SAS Text Miner (document level using Rule Builder), SAS Visual Text Analytics in Viya (document level), Sentiment Analysis Studio (document or sentence level)
- Describe functionalities and features of SAS Sentiment Analysis Studio.
- Demonstrate the use of SAS Sentiment Analysis Studio.

Architecture of SAS Sentiment Analysis Studio
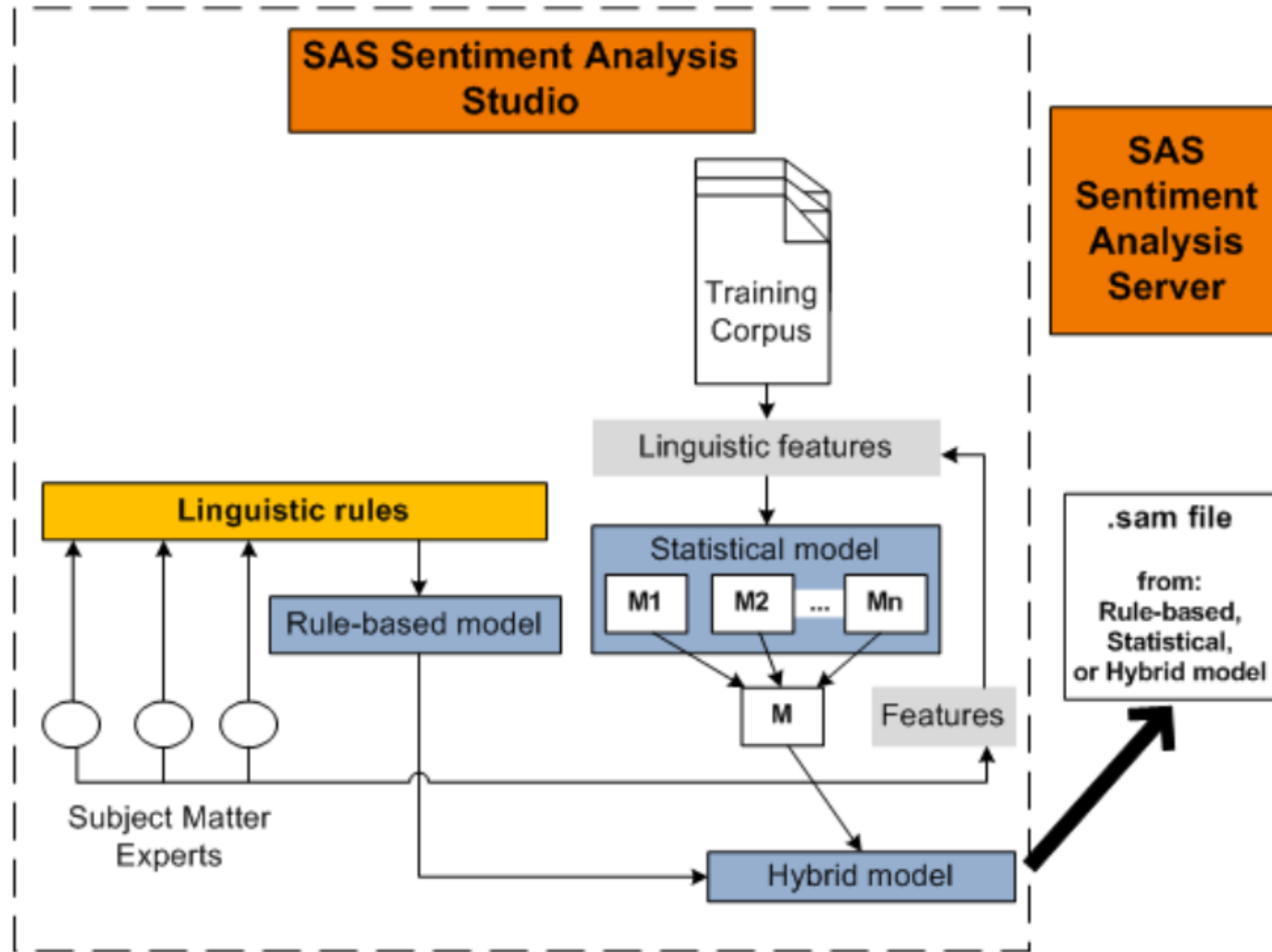
Types of Statistical Models in Sentiment Analysis Studio

- By default, the algorithms use 80% of the data for training and the remaining 20% for validation. (These percentages can be changed.)

- Four different text normalization methods are available (relative frequency, smoothed relative frequency, Okapi BM25, and pivoted length normalization)

- Four versions of feature-ranking algorithms are available (none, risk ratio, chi-square, information gain).

- Simple models combine the text normalization and feature ranking algorithms above.

- Advanced models enable you to define and customize your own model.

Rule-Based versus Statistical Models

- Statistical models identify sentiment at the document level.

- Rule-based models can determine overall sentiment, as well as for individual features or attributes.

- A project can contain multiple statistical models, but only a single rule-based model.

Process for Writing Rules

- Review a sample of documents to do the following:
  - Understand how the products and features are referenced
  - Identify the positive and negative keywords used in the documents
  - Determine how expressions of sentiment appear with respect to products and features
- Build a statistical model (with default options)
  - Import keywords based on classifier type rule from this model
- Write more rules
  - Test and refine the rules

What Does a Rule Do?

- A rule is written to achieve the following:
  - Match specific words or strings
  - Match any word using _w and _cap markers
  - Reference parts of speech
  - Reference defined entities with the _def marker
  - Use Boolean operators such as **AND, OR, NOT, SENT**
  - Use regular expressions (REGEX) to match patterns of characters, digits, or both

Rule Types

- CLASSIFIER
- CONCEPT
- C_CONCEPT
- CONCEPT_RULE
- PREDICATE_RULE
- REGEX

Hybrid Models

- A hybrid model enables you to combine features of both statistical and rule-based models.

- General steps:

1. Build a statistical model.

2. Build a rule-based model.

3. Build the hybrid model.

4. Test the hybrid model.

SAS Text Miner and SAS Sentiment Analysis Studio

- SAS Sentiment Analysis Studio is a *stand-alone* product. It is designed to handle all of the tasks needed to run sentiment analysis on a corpus of documents.

- But, in my opinion, the application of text mining on the same corpus helps analysts build better sentiment analysis models. In particular, I find the following nodes of Text Miner helpful before running Sentiment Analysis Studio.

  - Text Parsing node
  - Text Cluster node
  - Text Topic node
  - Text Rule Builder node
  - Text Profile node