

Introduction to RSSE

Michael Gohde

June 4, 2015

1 Overview

RSSE (Really Simple Syndication for Experiments) is a new protocol that should facilitate several types of machine learning, especially that of Open Set Learning. As indicated in its title, it draws some inspiration from RSS, which is presently used to distribute content across the web using small XML messages. RSSE similarly distributes small XML messages across the web containing experiments, labels, and classes rather than just updates for various sites.

While this may seem pointless, what has just been described is only one facet of RSSE's design. The true power of the specification is revealed when considering what could be implemented when datasets and experiments are automatically distributed to interested parties. Firstly, it would allow for datasets to continuously expand with ease. Secondly, the URLs distributed could be cached so that experimenters can try several algorithms on the same data without using an excessive amount of internet bandwidth. Finally, XML transfers could be bidirectional, so that experiment servers can collect results from all participants in order to label new classes and provide for true open set recognition.

Each of these features is implemented in RSSE through three distinct modules.

1.1 Experiment Server

The first module is that of the Experiment Server (usually referred to as the EC in all documentation.) The ES distributes URLs and can collect results from clients using a very simple registration scheme (ie. a client can send an experiment name and get a randomly generated ID number) implemented using the

XML transfers mentioned earlier.

1.2 Experiment Client

The Experiment Client is, as its name suggests, a client used to automatically subscribe to, fetch URLs from, and post results back to an Experiment Server. Once URLs are received, the Experiment Client depends on a Cache Module to fetch its data from the internet and provide it locally.

1.3 Cache Module

The Cache Module provides the caching infrastructure in RSSE. Whenever a new URL is sent out by an ES, one or more Experiment Clients (usually referred to as EC in the documentation) request that the Cache Module provide them with the data specified. This allows for both a reduction in internet bandwidth consumption and protection against the constant churn in URLs that the internet has had since its inception. One additional possibility for a CM implementation (not to be present in the first revisions of RSSE) is automatic reporting of which URLs failed to be resolved. Such reporting would allow those managing Experiment Servers to quickly respond to dead links.

2 Documentation in this Directory

This is just the first of several documents. Each of the following subsections is a document in this directory with a description of what it contains.