What’s Cooking

Bernd Wiswedel

KNIME
What’s Cooking

• Enhancements to the software planned for the next feature release
• Actively worked on
• Available in Nightly build

https://www.knime.com/form/nightly-build
What’s Cooking

• New Integrations & Utility Nodes
• Deep Learning
• Enhancements Big Data Extensions
• KNIME Server
New Integrations & Utility Nodes
Git Client Nodes

- Remember this?
- Extract Statistics from Git repositories
- Push, Pull, Tag, Commit History & Details

KNIME Analytics Platform – Source Code on GitHub

- Source Code available on GitHub & BitBucket

https://github.com/knime/
Evaluate Expressions

- (Name of the node is under discussion)
- Complete type support
- Multiple expressions
- More nodes to come in future version
  - Row Filter
  - GroupBy
  - ...
Outlier Handling

• Handle extremely large or small values in numeric columns
• Replace Values or Filter Rows containing outliers
• Usual pattern of Compute-and-Apply (two nodes)
KNIME Deep Learning
– Christian Dietz –
### Learning Behavior

- **Terminate on NaN loss**
- **Terminate on training stagnation (early stopping)**
- **Reduce learning rate on plateau**

<table>
<thead>
<tr>
<th>Monitored Quantity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation loss (total)</td>
<td></td>
</tr>
<tr>
<td>Min. delta</td>
<td>0.0</td>
</tr>
<tr>
<td>Patience</td>
<td></td>
</tr>
<tr>
<td>Factor</td>
<td>0.1</td>
</tr>
<tr>
<td>Patience</td>
<td>10</td>
</tr>
<tr>
<td>Epsilon</td>
<td>1.0E-4</td>
</tr>
<tr>
<td>Cooldown</td>
<td></td>
</tr>
<tr>
<td>Min. learning rate</td>
<td>0.0</td>
</tr>
</tbody>
</table>
KNIME Deep Learning

- Nodes for easy network editing and exporting
- Performance improvements
- Integration of custom layers and loss functions
- More Deep Learning and utility nodes for common tasks (Seq-to-Seq, GANs, Text,...)
Enhancements Big Data Extensions
– Björn Lohrmann –
KNIME Big Data Connectors

Better support for big data file formats:

- **Why?** Efficiency & Speed
- **Which formats?** Parquet, ORC and Avro

![New Connectors](image)

**Updated Connectors**

- Hive Loader
- Impala Loader

- Spark to Hive
- Spark to Impala

_MOCKUP_
KNIME Extension for Apache Spark

• New nodes:
  – Data preparation
    • Spark Pivoting
    • Spark Row Filter
    • Spark Missing Value (Apply)
  – Data mining: Association rule mining nodes (Spark MLlib)

• Infrastructure: Support for Apache Livy
  – Gives out-of-the-box Spark compatibility with:
    • Hortonworks HDP
    • Amazon EMR
    • Microsoft HDInsight

• Running standard KNIME nodes on Spark
  – Public preview coming up
Local Big Data Environment

Create Local Big Data Environment

Hive
HDFS
Spark

Provides everything to
• get started with big data easily
• rapidly prototype workflows
• leverage Spark on multi-core machines

Batteries included:
• no further software installation required
• no cluster required
• no jobserver required
Server-managed customizations

How do I install database drivers?

How do I configure Python?
Server-managed customizations

How do I install database drivers?
How do I configure Python?
How do I install database drivers?
How do I configure Python?
How do I configure Python?
Server-managed customizations

Profiles can include:
Preferences, drivers, and more

KNIME Server

Profiles sent to Analytics Platform installations
Server-managed customizations

KNIME Server

Preferences

Customization Profiles

Here you can choose which KNIME Server to use when getting server managed profiles. The list shows which servers are available. It might take a few seconds until all the available servers are listed. Then you can select one server from the list or enter an address manually.

Enable server-managed customizations

Server

knime-server

https://localhost:8443/knime/rest

Here you can select one or more profiles that are provided by the selected server. You can also change the profile order if you select more than one profile. Preferences in selected profiles further down in the list will override preferences from profiles further up in the list.

Note that changes are applied upon the next start.

Use

Profile

default
windows
linux

Reset ALL preferences to defaults

OK Cancel
Server-managed customizations

- Profiles set on KNIME Server (can include e.g. Database JDBC drivers)
- Set in knime.ini to automatically pick up preferences
- Or manually set in preferences
Remote Job View (Preview)

- Preview functionality.
- Help debug a job that is failing
- See progress of a current running KNIME Server job
- More functionality to follow
Remote Job View (Preview)

Why was there a problem?
Remote Job View (Preview)

What’s my workflow doing now?

Simple Model Training for Classification

Task. From adult data set (census data) predict income group from demographic attributes.

Learning Hub http://www.knime.org/learning-hub contains useful links, e.g. tutorials, videos, books, slides, etc ... for each kind of data analytics topic.

Data Reading
Read adult data set file. One row is one person, with demographic info and income group. The file is located in Supplementary file.

Train a Model
Here a decision tree is trained. Other Learner nodes train other models. Most Learner nodes output a PMML model (blue square output port).

Save the Model
Save the model so that it can be deployed via the KNIME Server REST API.

File Reader
Reading adult.csv

Random Forest Learner
Node 2

Model Writer
Node 3
KNIME Server Distributed Executors

• Need more workflow horsepower?

• KNIME Server currently supports ‘Scale Up’

• KNIME Server Distributed Executors allows ‘Scale Out’

https://upload.wikimedia.org/wikipedia/commons/5/53/Server-multiple.svg
KNIME Server Distributed Executors

Currently available as preview

Requests to execute workflows

KNIME Server

RabbitMQ

Executor 1

Executor 2

Executor 3

KNIME Server + Distributed Executors
Auto-scaling KNIME Server Distributed Executors

Preview: Planned for release in July 2018

Request to execute workflow

KNIME Server

RabbitMQ

Scripted launch on AWS CloudFormation

Executor 1

KNIME Server + Distributed Executors
Auto-scaling KNIME Server Distributed Executors

More requests to execute workflows

KNIME Server

RabbitMQ

Executor 1

CPU load increases

Launching Executor 2

AWS Auto-scaling Group

KNIME Server + Distributed Executors
Auto-scaling KNIME Server Distributed Executors

Requests to execute workflows

KNIME Server + Distributed Executors

AWS Auto-scaling Group

Executor 1

Executor 2

RabbitMQ

KNIME Server
Auto-scaling Distributed Executors

Launch KNIME Server and an Auto-scaling group for the distributed executors
Thanks

Need more info? Get to talk to us at the booth
Acknowledgements

- **Big Data:**
  Björn Lohrmann, Mareike Höger

- **Python Scripting Integration:**
  Clemens von Schwerin

- **Integrations (Google, Athena, Redshift):**
  Ole Ostergaard

- **Azure/Microsoft SQL Server R Services:**
  Jonathan Hale

- **KNIME Server:**
  Thorsten Meinl, Rene Damyon, Martin Horn

- **Date & Time Rewrite:**
  Simon Schmid

- **Logistic Regression:**
  Adrian Nembach

- **Utility Nodes (many...):**
  Moritz Heine

- **DeepLearning:**
  Marcel Wiedenmann

- **Outlier Handling:**
  Mark Ortmann

- **Guided Analytics:**
  Christian Albrecht, Oleg Yasnev

- **Math Formula:**
  David Kolb