Data Science with Orange Toolbox:

Data Science for Everyone

Blaž Zupan
University of Ljubljana, Slovenia &
Baylor College of Medicine, Houston
Data Science

- Linear algebra
- Probability
- Statistics
- Mathematical optimization
- Data & model representation
- Machine learning
- High performance computing
- Data visualization
- Scripting languages
Data Science

2-5 hours to train concepts
Data Science

2-5 years to train theory
2-5 hours to train concepts
data preparation, feature engineering, clustering, modeling, evaluation, data understanding
Data Science

2-5 hours to train concepts

data preparation, feature engineering, clustering, modeling, evaluation, data understanding
def edges_from_osmdb(osmdb, vertex_namespace, slogs, profiledb=None):
    """generates (vertex1_label, vertex2_label, edgepayload) from osmdb""

    street_id_counter = 0
    street_names = {}

    # for each edge in the osmdb
    for i, (id, parent_id, node1, node2, distance, geom, tags) in enumerate(osmdb):

        # Find rise/fall of edge, if profiledb is given
        rise = 0
        fall = 0
        if profiledb:
            profile = profiledb.get(id)
            if profile:
                rise, fall = get_rise_and_fall(profile)

        # insert end vertices of edge to graph
        vertex1_label = "%s-%s"%(vertex_namespace,node1)
        vertex2_label = "%s-%s"%(vertex_namespace,node2)

        # create ID for the way's street
        street_name = tags.get("name")
        if street_name is None:
            street_id_counter += 1
            street_id = street_id_counter
        else:
            if street_name not in street_names:
                street_id_counter += 1
                street_names[street_name] = street_id_counter
                street_id = street_names[street_name]

        # Create edges to be inserted into graph
        s1 = Street(id, distance, rise, fall)
        s2 = Street(id, distance, fall, rise, reverse_of_source=True)
        s1.way = street_id
        s2.way = street_id

        # See if the way's highway tag is penalized with a 'slog' value; if so:
        slog = slogs.get(tags.get("highway"))
        if slog:
            s1.slog = s2.slog = slog

        # Add the forward edge and the return edge if the edge is not one way
        yield vertex1_label, vertex2_label, s1

        oneway = tags.get("oneway")
        if oneway != "true" and oneway != "yes":
            yield vertex2_label, vertex1_label, s2
Purse Analytics
(live demo)
Democratized image analytics by visual programming through integration of deep models and small-scale machine learning


Nature Communications 10, Article number: 4551 (2019) | Cite this article

Abstract

Analysis of biomedical images requires computational expertise that are uncommon among biomedical scientists. Deep learning approaches for image analysis provide an opportunity to develop user-friendly tools for exploratory data analysis. Here, we use the visual programming toolbox Orange (http://orange.biolab.si) to
Key Concepts

Interactive Visualizations

Workflows

“Lego Bricks” for Data Science

Democratization of Data Science

Easy Access to Data

Reproducibility

Customization

Story Telling

Experimentation

Tools for Training
“Lego” Bricks for Data Science
Responsive Workflows
Interactive Visualizations
Interactive Visualizations
Data Integration

- Single Cell Datasets
- t-SNE
- Differential Expression
- GO Browser
- Genes

**Input ID** | **Entrez ID** | **Name** | **Description** | **Synonyms**
--- | --- | --- | --- | ---
IL1B | 3553 | IL1B | interleukin 1 beta | IL-1, IL1-BET...
CD74 | 972 | CD74 | CD74 molecule | DHLAG, ...
FCER1G | 2207 | FCER1G | Fc fragment of ... | FCRG
CD14 | 929 | CD14 | CD14 molecule |
S100A11 | 6282 | S100A11 | S100 calcium ... | HEL-S-43, ...
CFD | 1675 | CFD | complement ... | ADIPSIN, AD...
MND4 | 4332 | MND4 | myeloid cell ... | PYHIN3
BETN | 68720 | BETN | receptor | ADSE, AD772

**GO term** | **Cluster** | **Reference** | **p-value** | **FT**
--- | --- | --- | --- | ---
- biological_process | 85 (94.44%) | 17899 (87.36%) | 0.02283 | 0
- immune system process | 65 (72.22%) | 3068 (14.97%) | 5.4e-34 | 1
- immune response | 55 (61.11%) | 2160 (10.54%) | 4.5e-31 | 7
- leukocyte activation | 41 (45.56%) | 1180 (5.76%) | 6.4e-27 | 5
- immune effector process | 37 (41.11%) | 1196 (5.84%) | 2.6e-22 | 6
- cell activation involved in immune respons... | 30 (33.33%) | 690 (3.37%) | 6.1e-22 | 1
- leukocyte degranulation | 27 (30.00%) | 533 (2.60%) | 2.3e-21 | 3
- leukocyte mediated immunity | 30 (33.33%) | 815 (3.98%) | 6.3e-20 | 7
- myeloid leukocyte mediated immunity | 27 (30.00%) | 550 (2.68%) | 5.0e-21 | 6
- neutrophil mediated immunity | 26 (28.89%) | 499 (2.44%) | 7.2e-21 | 8
- neutrophil degranulation | 26 (28.89%) | 485 (2.37%) | 3.6e-21 | 5
Explanation
Interactive Explanation
Tools for Training
Tools for Training
Training

Data Science Workshop for Public Administration of Slovenia, 2018 & 2019.
Getting Started with Orange 01: Welcome to Orange

271,544 views • Dec 21, 2015

790 likes • 11 dislikes

Orange Data Mining
12K subscribers
Open Problems

Tech
- Response Time & Interactivity
- Servers-Side Compute & Visual Analytics
- Data Fusion

Human
- Simplicity vs. Flexibility
- Gamification
- Interpretation vs. Explanation
biolab

Andrej Čopar, GPUs, data fusion
Niko Colnerič, text mining
Tomaž Curk, bioinformatics
Janez Demšar, core development, visualization
Aleš Erjavec, lead developer
Tomaž Hočevar, algorithms, speed-up
Primož Godec, image analytics, text mining
Pavlin Poličar, clustering, embedding, speed-up
Ajda Pretnar, communication, digital humanities
Martin Stražar, bioinformatics
Vesna Tanko, development, testing
Marko Toplak, spectral analysis, bioinformatics
Robert Cvitković, bug fixes
Blaž Zupan, enjoys
Lan Žagar, code review, testing
Marinka Žitnik, method expert

external

Nejc Ilenič, deep learning, cloud
Nejc Debevc, web
Jaka Kokošar, bioinformatics
Matjaž Pančur, cloud services
Fabio Ricciato, telecommunications
Veljko Pejović, telecommunications
Anže Starič, software development

support

ARRS Slovenia, NIH USA, MVZT Slovenia
EU FP7 & Horizon 2020
Astra Zeneca, Lek, Elletra, Soleil
Fulbright
Google Summer of Code
NumFOCUS