City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.
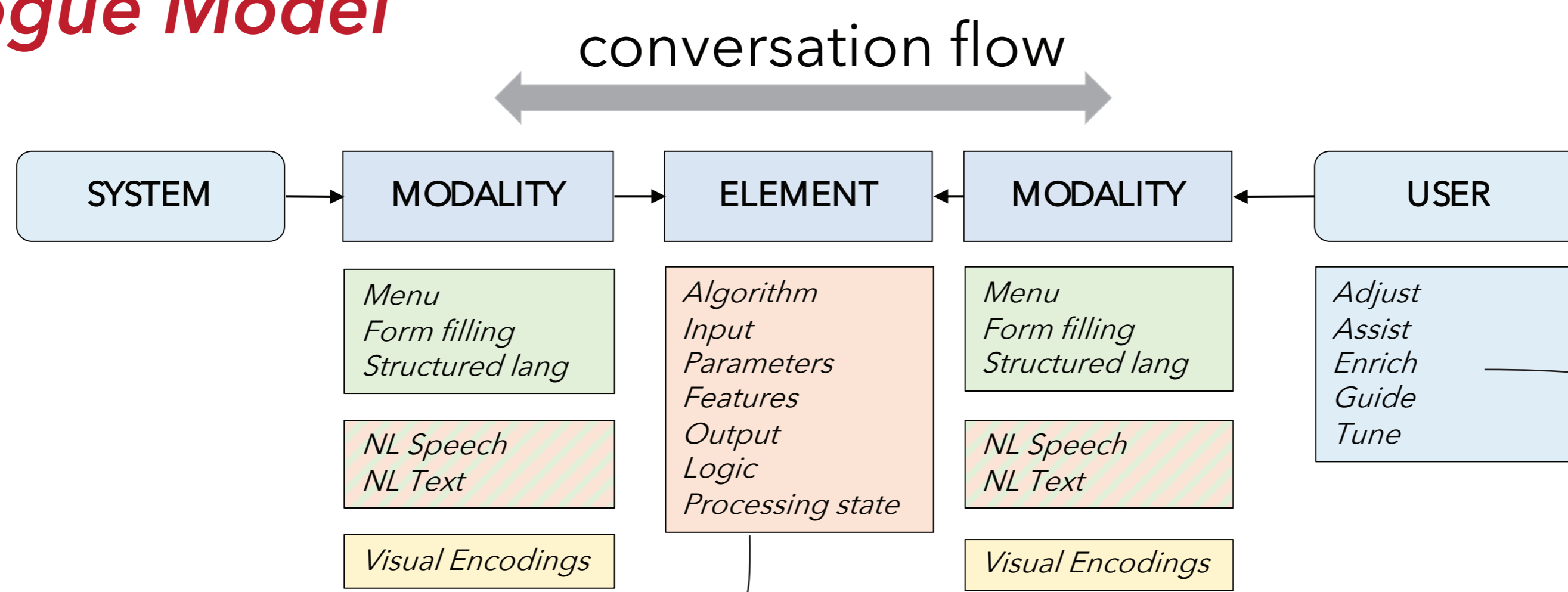
Towards Natural Language Empowered Interactive Data Analysis

Cagatay Turkay, Rafael Henkin
giCentre, City, University of London

www.gicentre.net/nlvis
cagatay.turkay@city.ac.uk
rafael.henkin@city.ac.uk

Multimodal Data Analytics to combine the strengths of visual analytics with **natural language as an expressive channel** of interaction. We discuss a **two-way dialogue model**, potential **enhancements** through natural language and **analytical workflows** as a structured methodology

Dialogue Model

conversation flow

SYSTEM	→	MODALITY	→	ELEMENT	←	MODALITY	←	USER

MODALITY:
Menu
Form filling
Structured lang

NL Speech
NL Text

Visual Encodings

ELEMENT:
Algorithm
Input
Parameters
Features
Output
Logic
Processing state

MODALITY:
Menu
Form filling
Structured lang

NL Speech
NL Text

Visual Encodings

USER:
Adjust
Assist
Enrich
Guide
Tune

Interaction Facets

Contextual aspects/elements where useful interactions can occur between the user and an algorithm

User Roles

The various roles that the user can play in interactive model building as derived and harmonised from existing VA and Human-Algorithm Interaction models

Potential Enhancements

E1: *Enhanced expressiveness for* **guidance**

E2: *Enhanced* **model steering**

E3: *Enhanced* **comprehension and transparency**

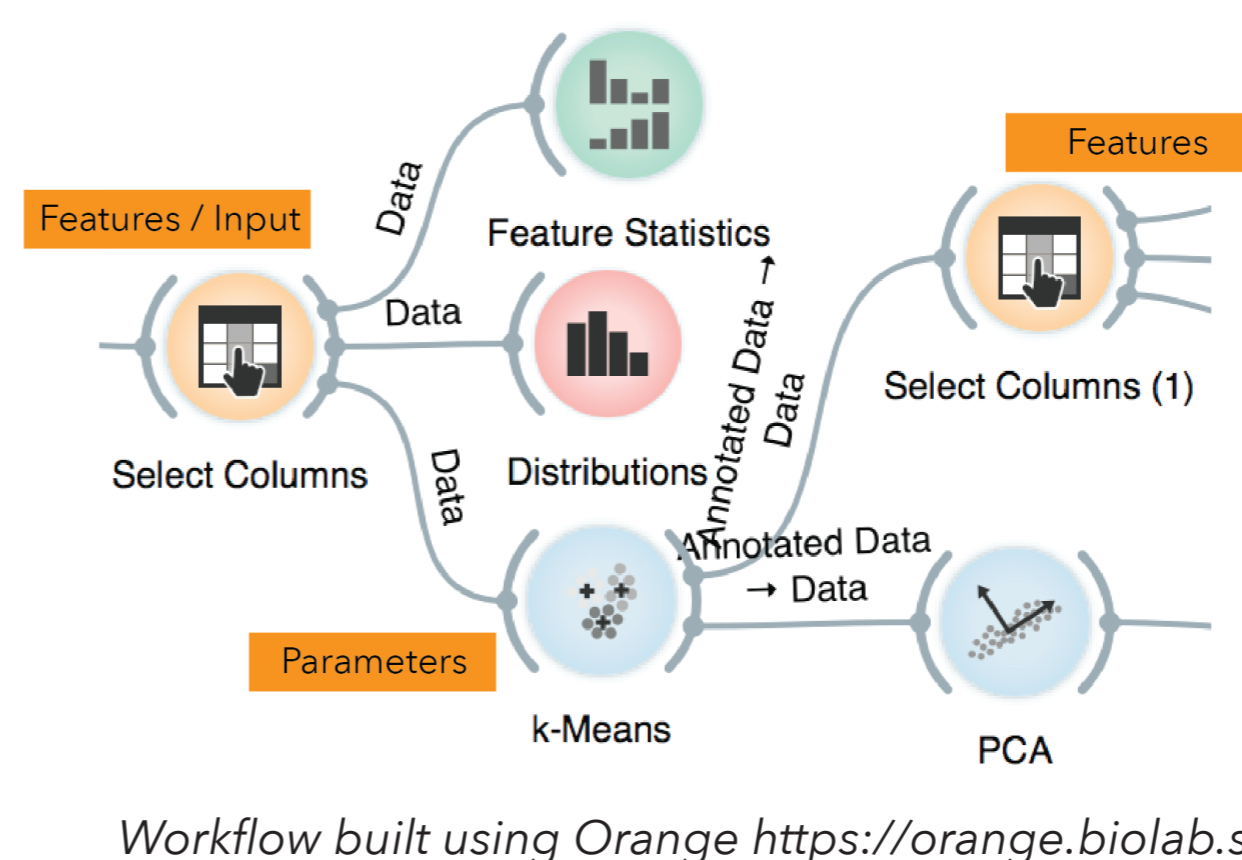E4: *Enhanced* **knowledge elicitation**

Analytical Workflows

Instantiations of the dialogue model for specific algorithms and processes. Provided as interactive analysis templates for systematic evaluation and usage
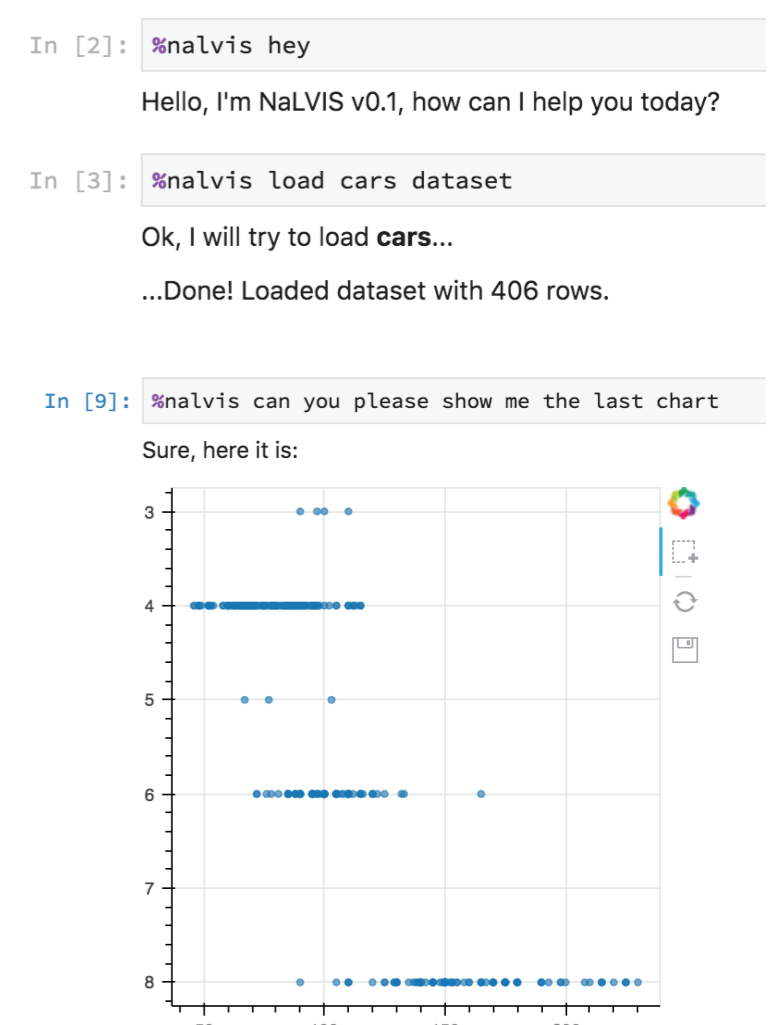
Guide us !!

Mini User-Study - What would you like to say to an algorithm?

Considering an interactive data analysis scenario, what can you imagine yourself telling to an algorithm? Feel free to post-it, pin your utterance?

Features / Input
Feature Statistics
Data
Select Columns
Data
Distributions
Annotated Data → Data
Data
Select Columns (1)
Features
Parameters
k-Means
Annotated Data → Data
PCA

Workflow built using Orange https://orange.biolab.si/

In [2]: **%nalvis** hey
Hello, I'm NaLVIS v0.1, how can I help you today?

In [3]: **%nalvis** load cars dataset
Ok, I will try to load **cars**...
...Done! Loaded dataset with 406 rows.

In [9]: **%nalvis** can you please show me the last chart
Sure, here it is:

Implementation of the workflows in Jupyter Notebook with interactive visuals and an NL agent

EPSRC
Engineering and Physical Sciences Research Council

CITY
UNIVERSITY OF LONDON
EST 1894