1. Introduction

Dashboards, often described as a “visual display of the most important information needed to achieve one or more objectives” [Few06] are ubiquitous across domains [SCB*18]. Despite their prevalence, the process of authoring dashboards remains challenging, requiring expertise in both data analysis and visualization design. Dashboard authors often need to explore and understand the data, author individual charts or views, and then combine the individual views into a coherent dashboard. Recent work has shown that going beyond individual view recommendations (e.g., [MHS07, WMA*15, SST*21]) and recommending collections of views and filtering widgets based on dashboard intents or objectives can effectively assist dashboard authoring [PSS23]. However, expressing dashboard objectives (e.g., creating a dashboard that summarizes data changes over time) and identifying content relevant to those objectives remains an open challenge.

Natural language interfaces (NLIs) have shown promise in helping users express their data-related goals as natural language (NL) utterances to generate visualizations [SSL*22]. NLIs can be particularly helpful in a scenario like dashboard authoring, where a set of complex, multi-step analytical inquiries can be simply expressed in plain language. An NLI for dashboard authoring could enable authors to express their high-level dashboard objectives as NL utterances (e.g., “I want to show an overview of the company performance,” “Display an analysis of highest grossing movies”) and recommend a relevant set of visualizations in response.

Building upon this premise, we explore the idea of using NL input for dashboard authoring with a prototype interface, BOLT. Specifically, we first formulate a list of design goals for supporting NL-based dashboard authoring by reviewing prior work on visualization recommendations and NLIs for visualization. We then describe the design and implementation of BOLT, an NLI for dashboard authoring that maps input utterances to popular dashboard objectives, generating a set of dashboard recommendations. Using BOLT as a design probe, we conducted a preliminary user study to validate the proposed idea of NL-based dashboard authoring. Participant feedback confirmed that NLIs like BOLT can support fast and flexible dashboard authoring and assist authors with varying levels of expertise with visualization tools.

2. Related Work

NLIs for visualization. Numerous research efforts have explored how NLIs can support analytical inquiry by producing visualizations [SLJL10, GDA*15, HSTD17, STD19, TS19, YS19, NSS20, SNL*21]; a detailed review of which can be found in survey manuscripts such as [SS17, SSL*22]. While these NLIs present different capabilities, they typically share the common approach of generating individual visualizations using a combination of data attributes and low-level analytic tasks [AES05] inferred from an input utterance. We extend this line of research on NLIs for visualization by investigating a novel use case of dashboard authoring. Specifically, we contribute a system that maps utterances to higher-level dashboard authoring objectives (e.g., change analysis, comparison) compared to low-level analytic tasks (e.g., sorting, finding correlation) in prior work [SLJL10, GDA*15, SBT*16, NSS20]. In this paper, we describe linguistic challenges and considerations for interpreting NL queries for these higher-level dashboard objectives. Additionally, by focusing on dashboards (as opposed to singleton charts in prior systems), we also illustrate how NLIs can present multiple view and widget recommendations in response to a single
query and allow users to mix-and-match system recommendations to customize the output.

**Dashboard and Multi-View Visualization Recommendation Systems.** Despite being an established topic of interest among visualization practitioners [WSC17, Few06], studies have highlighted a scarcity of research on dashboard design and systems [SCB*18, TBFGC21, BFA*22]. In response, a recent surge of systems has explored the idea of content recommendations to aid dashboard authoring. For instance, MultiVision [WWZ*21] and Dashbot [DWQW22] recommend a group of views based on underlying data patterns and any user-selected data attributes. Medley [PSS23] recommends collections of views and widgets for a given combination of data attributes and/or dashboard intents (e.g., measure analysis, change analysis). Supporting rapid dashboard design, LADV [MMG*20] helps author dashboards by generating templates based on example images and roughly-drawn sketches. We contribute to this growing body of work on intelligent dashboard authoring tools by designing an interface that recommends dashboards based on NL input. Specifically, the proposed NL-based approach helps authors explore data during the dashboard authoring process by interpreting their high-level dashboard objectives expressed in plain language and returning relevant sets of views.

3. **BOLT**

3.1. **Design Goals**

We reviewed prior work on NLIs for visualization (e.g., [SLJL10, GDA*15, STD19, SNL*21]) and multi-view visualization systems and guidelines (e.g., [QH17, WBWK00, DWQW22, PSS23]) to compile a set of four goals that motivate BOLT’s initial design.

**DG1. Recommend multiple candidate dashboards and views.** Dashboard authoring can be a subjective task where a predefined single collection of views and widgets may not always exactly map to the author’s goals [PSS23]. Rather than recommending one single ‘best’ dashboard for an analytical inquiry, the interface should recommend a set of dashboard candidates and views for the author to choose from or further refine.

**DG2. Support linguistic variations and ambiguity.** Prior work has shown that people tend to use a variety of phrasings when expressing their intent in NL utterances and that these utterances are often ambiguous [GDA*15, STD19, SNL*21]. Hence, the interface should accommodate phrasing variations (e.g., “Sales Overview” vs. “Generate a dashboard showing an overview of the company’s sales” should generate the same dashboard) and ambiguous utterances (e.g., an utterance like “Company performance overview” is ambiguous because the term ‘performance’ can map to multiple attributes such as Sales, Profit, etc.), showing the provenance of how these utterances are interpreted.

**DG3. Provide reasonable defaults in the generated dashboards.** Users of dashboard authoring tools, particularly novices, often lack knowledge about good dashboard design practices [BFA*22, PSS23]. The interface should reduce friction in the dashboard authoring process by implicitly incorporating interaction and design guidelines for dashboards and multi-view displays [QH17, WBWK00, Rob98] (e.g., using consistent colors and scales across charts, optimizing the layout, enabling multi-view coordination).

**DG4. Enable multi-modal repair and refinement.** Prior work on dashboard authoring tools has shown that authors appreciate the ability to fine-tune the dashboard design [WWZ*21, PSS23, DWQW22]. Studies with NLIs have also shown that making authors aware of the interpretation logic or errors and giving them the ability to rectify these errors through interactive widgets can improve system accuracy and user confidence [GDA*15, SBT*16]. In addition to recommending multiple candidate dashboards incorporating popular design guidelines, the interface should also provide direct manipulation affordances to update its default responses.

3.2. **System Overview and Examples**

Figure 1 shows BOLT’s interface comprising of a data panel on the left, the main canvas, and the recommendation panel on the right. Typing an NL utterance in the text box at the top of the canvas (Figure 1B-top) and pressing the “enter” key invokes the system’s interpretation of the utterance. The recommendation panel (Figure 1C) shows previews of the system recommendations, and a dashboard author can add content from the recommendation panel to the canvas (DG1). Once items are on the canvas, authors can refine the dashboard by dragging, resizing, or removing objects (DG4) and also export (🔗) the dashboard as an interactive web page.

**BOLT** is implemented as a web application that accepts any tabular CSV dataset as input. Visualizations in the tool are created using Vega-Lite [SMWH16]. Below we briefly describe how BOLT generates dashboards based on NL utterances.

**Dashboard Objectives and Collection Templates.** BOLT currently supports five dashboard objectives derived from Pandey et al.’s dashboard intent framework [PSS23]. These include:

- **Overview of Measure(s).** This dashboard objective typically involves 1–3 quantitative attributes (measures) and displays a set of views summarizing those measures across other categorical, temporal, and geographic attributes in the dataset. Figure 1A illustrates a dashboard focusing on this objective.
- **Change Summary.** Dashboards corresponding to this objective contain a set of views and widgets that highlight changes in a
measure’s value over time across other attributes in a dataset. Figure 2A shows an example of such a dashboard that displays the year-over-year change in the measure Sales.

- **Top N.** This objective is facilitated through dashboards that showcase the top data categories for a specific measure. An example is shown in Figure 2B, where the dashboard displays top categories for the Gross measure.

- **Dimension Comparison.** Dashboards focusing on this objective typically contain a set of views and widgets that facilitate comparison between values for a categorical attribute (dimension). Figure 2C presents an example dashboard for this objective comparing the values for the dimension Gender.

- **Measure Comparison.** Dashboards that map to this objective often present a set of views that enable comparison between two measures to understand their relationship and underlying data trends. Figure 2D-top illustrates an example of this type of dashboard comparing the measures Discount and Profit.

Each objective is mapped to collection templates comprising a set of views and filtering widgets. We defined the templates based on an existing survey of popular dashboards created with mainstream visualization tools like Tableau and Microsoft Power BI [PSS23]. For instance, as shown in Figure 1B, the template for the Overview of a single measure objective contains a data summary view, a map, one line chart, three bar charts, and two filter widgets.

**Interpreting NL Utterances.** BOLT’s parser removes stopwords (e.g., ‘a,’ ‘the’) and conjunctions/disjunctions (e.g., ‘and,’ ‘or’) from the utterance and extracts a list of N-grams (e.g., “Company Performance Overview” → [Company], [Performance], [Company Performance], etc.). To detect attribute tokens in the utterance, the system compares the N-grams to available data attributes looking for both syntactic (e.g., misspellings) and semantic similarities (e.g., synonyms) using the Levenshtein distance [YB07] and the Wu-Palmer similarity score [WP94], respectively.

The parser uses a combination of a grammar- and lexicon-based approach to infer dashboard objectives. Specifically, BOLT performs parts-of-speech tagging and generates a dependency tree to understand relationships between words in the input utterance. The extracted tags and dependencies are then matched against a predefined set of grammar patterns (e.g., “summarize changes in <measure>,” “<measure> versus <measure> analysis”) and keywords (e.g., ‘change,’ ‘compare,’ ‘top’) corresponding to the five objectives listed above.

BOLT can thus interpret a wide range of dashboard specification utterances spanning different phrasings (DG2) and accommodate misspellings and synonyms for data attributes, as highlighted through utterances in Figure 2A-C. For instance, the phrase ‘Summarize changes’ in the utterance in Figure 2A matches the lexicon for the Change Summary objective, resulting in the system recommending a yearly change analysis dashboard for the measure Sales (also referenced in the utterance). Similarly, the word ‘Top’ in Figure 2B helps BOLT infer a Top N objective, resulting in the displayed dashboard containing a series of sorted bar charts. Lastly, the grammar of the utterance in Figure 2C matches a phrasing template for the Dimension Comparison objective, and the dependencies between terms enable the system to infer Gender as the focus dimension and Gold as the focus measure.

There may, however, be cases where the input utterance does not contain terms matching data attributes or is phrased such that they do not map to predefined objectives. BOLT also includes safeguards for such scenarios and “guesses” attributes or objectives based on ambiguous utterances (DG2). For instance, when an objective is detected from an utterance but there is no reference to attributes, the system guesses the primary attribute type that a dashboard for the inferred objective focuses on. Consider the utterance “Company Performance Overview.” Here, the term ‘overview’ suggests the Overview of measure(s) objective but there is no attribute mentioned in the utterance as the words ‘company’ and ‘performance’ do not match any attributes listed in Figure 1A. In such cases, BOLT infers the primary attribute type and uses the default objective (e.g., Gender or Sales) to generate a dashboard.

Figure 2: BOLT’s responses for utterances in the context of different datasets, including sales (A, D), movies (B), and Olympics medal winners (C). Utterances in (A)-(C) convey both attributes and objectives, whereas the utterance in (D) only references attributes and does not clearly state the dashboard objective.
cases, similar to prior data- and insight-based recommendation sys-
tems [DWQW22, PSS23], BOLT computes a series of statistical
metrics (e.g., the standard deviation to characterize value distri-
butions, z-scores to detect trends) for all possible measures (e.g.,
Profit, Sales, Discount) that can fit the objective’s collection
template and selects the attribute yielding the highest score (in
this case, Sales). However, since the system’s interpretation may
not match the user’s mental model, BOLT also provides a direct ma-
nipulation widget for authors to update the default attributes and get
an updated set of recommendations (DG4, Figure 1D).

Now consider an alternative utterance “Do discounts yield higher
profits?” that explicitly references data attributes (Discount, Profits) but does not map to a predefined objective. Detecting
the two measures, BOLT infers Measure Comparison and Overview
of Two Measures as potential objectives the author is interested in,
recommending relevant collections (Figure 2D). In cases where
BOLT is unable to detect both attributes and objectives (e.g., “Show
a dashboard for this dataset”), the interface responds with a mes-
sage noting that it could not interpret the utterance and asks authors
to try a different utterance.

Generating Dashboards. When BOLT infers attributes and objec-
tives from an utterance, the interface automatically adds the first
recommended collection to the canvas. In cases like Figure 2C
where the utterance includes value references such as ‘USA;’ as
with the Country dropdown menu at the bottom-right, BOLT also
filters the dashboard by the specified value if a corresponding wid-
get is available in the collection. For ambiguous utterances, the in-
terface only populates the recommendation panel, allowing authors
to first assess the recommendations and then choose the content
they want to add to the canvas (DG1, DG4).

When content is added to the canvas, BOLT computes a grid lay-
out that shows views most relevant to the input utterance first (e.g.,
difference bar chart in Figure 2A), maximizes space for maps and
charts with a large number of categories (e.g., map and grouped bar
chart in Figures 2A, 2C, respectively), and attempts to group filter-
ing widgets in one location to the right of the views. The interface
also implicitly configures interactive links between the views to fa-
cilitate actions like brushing and linking. While BOLT applies these
design rules as smart defaults, authors can manually override both
the layout and the links between views (DG4).

4. User Feedback and Future Work
Using BOLT as a design probe, we conducted a preliminary user
study with six participants (P1-P6) having varying levels of expe-
rience with dashboard authoring (three novices, one intermediate,
and two experts). We had two goals for the study: 1) elicit feedback
on the idea of NL-based dashboard authoring and 2) assess BOLT’s
current design and identify areas for improvement. Sessions were
conducted remotely and lasted approximately 30 minutes. partici-
ipation in the study was voluntary and participants were recruited
via Slack channels at a data analytics software company.

To avoid biasing participants and understand how they might nat-
urally interact with an NLI, we only gave a brief introduction to the
tool and asked participants to use NL to author dashboards in the
context of three datasets (attached in supplementary material), dis-
cussing their reactions to system responses. After the trial phase,
we demonstrated any unexplored system capabilities using addi-
tional utterances and concluded with a semi-structured interview
focusing on the two study goals.

Participants issued a total of 14 utterances (2-3 per session),
with 10 utterances being phrased as dashboard titles or keyword-
style queries, three questions, and one command. All but two ut-
erances that did not yield a response (“Generate an interesting
dashboard” and “GBR Metrics”) were successfully interpreted by
BOLT. Among the 12 interpreted utterances, Measure compar-
ison was the most frequently referenced objective (5), followed by
change summary (4), measure overview (2), and top N (1). For
feedback on the overall idea, participants were unanimously posi-
tive about the idea of using NL for dashboard authoring for: 1) bootstrapping users new to a dataset or visualization tools and 2)
rapid ideation and prototyping as the key use cases. For instance,
P5 said, “When you’re given a new dataset and you’re not sure ex-
actly about the direction you want to go, being able to ask simple
questions or give it a dashboard title and do some of that quick
ad-hoc analysis is great and is even faster than drag-and-drop.”

Alluding to the potential for fostering ideation and rapid proto-
typing, P2 compared BOLT to language generation models like
GPT [BMRR20] and said, “This is a great ideation tool for the an-
alyst. They already know how to make individual views, but which
are the interesting ones? What are interesting points that I can jump
off of for a dashboard? This is a great way to play with options.”
Similarly, P3, a dashboard expert, commented, “I could see myself
using this to quickly generate a canvas to see what the data looks
like and which views might go well together and I might also use
this as an inspiration tool when I don’t know much about the data.”

Discussing areas for improvement, all participants noted that
they would benefit from example utterances to help understand the
system’s capabilities. One direction for future work is to explore
the design and generation of utterance recommendations [SS21]
to enhance the discoverability and learnability of NL input during
dashboard authoring. Five participants noted that they see BOLT’s
recommendations as a starting point and wanted the ability to spec-
ify or fine-tune individual views (e.g., adjust encodings, swap axis
orientation). We intentionally did not support manual view specifi-
cation in BOLT to specifically solicit feedback on NL input; inte-
grating the NL experience with a manual view specification envi-
ronment would enable more multimodal utility of the interface.
We employed a heuristics- and template-based logic for NL inter-
pretation and dashboard recommendation in the context of five dash-
board objectives (Section 3.2), as this afforded a deterministic set
of output for testing and feedback. However, future work is needed
to generalize the system to a wider array of dashboard objectives
and designs. Such generalization would entail updating the system
heuristics (e.g., adding utterance templates, updating objective-to-
collection mappings) and/or exploring machine- and deep learning-
based techniques for NL interpretation [LTL+21, MS23] and dash-
board design [LLW+22].

In conclusion, through the implementation and evaluation of BOLT,
we illustrate the feasibility and potential of an NLI for dashboard
authoring. We hope the work can guide future research on design-
ning more expressive and intelligent dashboard authoring tools.
References


