EuroVis Workshop on Visual Analytics (2021)
J. Bernard and K. Vrotsou (Editors)

Talk2Hand: Knowledge board interaction in augmented reality easing analysis with machine learning assistants

Y. Hong[†1] and B. Watson[‡1] K. Thompson[2§] D. Paul [2¶]

Abstract

Analysts now often use machine learning (ML) assistants, but find them difficult to use, since most have little ML expertise. Talk2Hand improves the usability of ML assistants by supporting interaction with them using knowledge boards, which intuitively show association, visually aid human recall, and offer natural interaction that eases improvement of displayed associations and addition of new data into emerging models. Knowledge boards are familiar to most and studied by analytics researchers, but not in wide use, because of their large size and the challenges of using them for several projects simultaneously. Talk2Hand uses augmented reality to address these shortcomings, overlaying large but virtual knowledge boards onto typical analyst offices, and enabling analysts to switch easily between different knowledge boards. This paper describes our Talk2Hand prototype. (see http://www.acm.org/about/class/class/2012)

CCS Concepts

• ***Human-centered computing*** → ***Visual analytics;*** *Mixed / augmented reality;* • ***Theory of computation*** → *Semi-supervised learning;*

1. Introduction

Analysts have always struggled with the volume of their data. In the past, they used physical displays called knowledge boards to aid them, but these boards were too large and unwieldy to rely on regularly, particularly when analysts worked on several projects simultaneously. As technology increased data flows further, analysts began to use technology to deal with them, using machine learning (ML) during analysis. Unfortunately, training ML models can often add difficulty and require analysts to have ML expertise.

We describe *Talk2Hand*, a new interface designed to broaden access to ML analytic assistants. Rather than explicitly labeling data and controlling ML models, analysts communicate with them implicitly by manipulating familiar knowledge boards. To fit these knowledge boards into typical workplaces and workflows without compromising privacy, we implement them in augmented reality (AR). The resulting system increases the utility of ML and knowledge boards in sensemaking.

2. Related work

Due to the complexity of analytic workflows in intelligence [GM18] and the growing data flows analysts are facing, machine learning is taking on a growing role during analysis. Yet as surveys of this work note [ERT*17,SKKC18], analyst-ML interaction must become much richer and more transparent, moving beyond simple labeling schemes and avoiding ML concepts and jargon so that collaboration with ML assistants does not disrupt workflows.

Figure 1: *Analog knowledge board [Vor09]*

Knowledge boards (e.g. Figure 1) are well-known tool in analysis of all kinds, which predate not only ML but even computers [SIMM95]. They make a natural match to the early stages of intelligence sensemaking [AR96, PC05], in which data are organized into groups or “shoeboxes.” They also bring many of the significant advantages of large displays [CSR*03, AEN10], and could help address the spatial and layout problems observed in ML analytic systems such as CHISSL [ASW*19]. Indeed, although it does not incorporate ML, the ForceSPIRE project and its notion of “se-

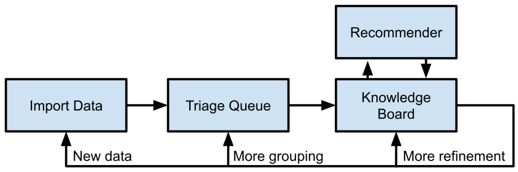
† North Carolina State University
‡ North Carolina State University
§ Laboratory for Analytic Sciences
¶ Laboratory for Analytic Sciences

DOI: 10.2312/eurova.20211090

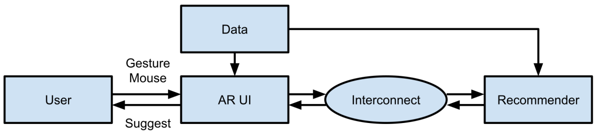
mantic interaction" [EFN12a, EFN12b] have been a special inspiration in our work, by making knowledge boards a central part of the analytic process.

Yet use of knowledge boards in analysis remains limited. Analog boards are difficult to take down and later restore, making them a poor fit to busy analyst workflows. Digital knowledge boards can address this problem, but like analog boards require a great deal of space, making them challenging to use in typical office environments. For example, BumpTop [AB06] requires a large touchscreen inter, while Endert et al. [EFN12a] used eight high resolution monitors. Both would be difficult to fit into every analyst's cubicle!

Immersive technologies such as virtual reality (VR) and AR can extend the workplace by adding virtual display and interaction space, but the discomfort of wearing headsets has been a pervasive challenge [SLR*20]. Interestingly, Kemeny et al. have shown that including afixed outside world anchor in the view can significantly decrease cybersickness [KGMC17], hinting that AR users may suffer less discomfort than VR users. VR can also introduce social challenges in office environments, isolating users from their colleagues, and their own bodies [FZMA20]. We are not aware of any systems integrating AR with analytic knowledge boards.



(a) *Experience Flow*



(b) *System UI*

Figure 2: *(a) The flow of Talk2Hand's sensemaking experience. Analysts import raw data into Talk2Hand, triage it into groups, refine those groups, and iterate. (b) Talk2Hand's system architecture. The AR front end accepts data and user input, and communicates with the recommender back end, which suggests categories.*

3. Contribution

Our primary contributions:

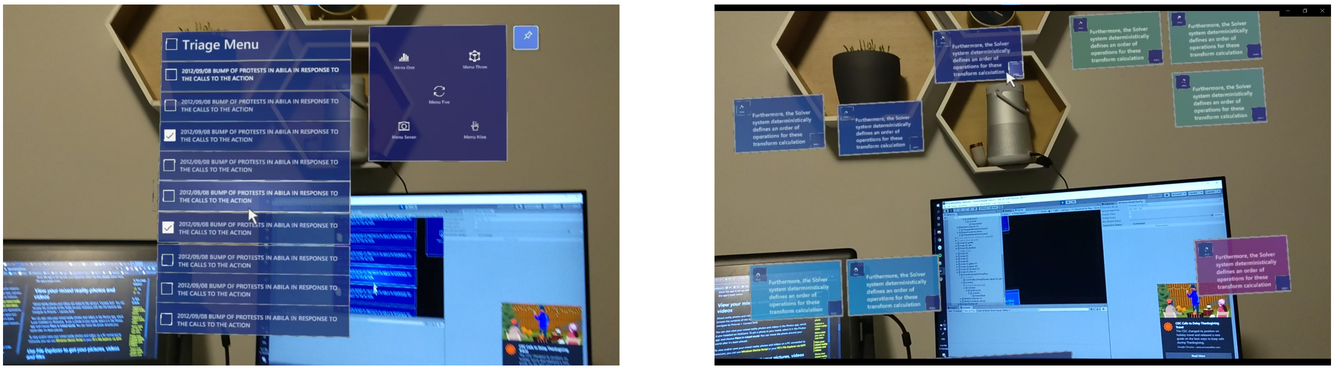
- make it easier for analysts to use knowledge boards, by using augmented reality to overlay virtual knowledge boards onto existing workplaces. Analysts need not fit large displays next to their desk, isolate themselves from their colleagues, nor use a display they share with their colleagues.
- use knowledge boards to simplify interaction with ML assistants without disrupting analysis. While knowledge boards have been proposed for analysis [EFN12a, EFN12b], they have not to our knowledge been coupled with ML assistants.

4. Sensemaking Experience

In *Talk2Hand*, analysts seek knowledge in their data by moving iteratively through three stages (Fig. 2a): data import, triage, and refinement. *Talk2Hand* accepts input data in a tabular (e.g. CSV) format consisting of attributes that include search relevance, date, title, and textual summary. We anticipate that analysts will gather this data using their existing search tools. *Talk2Hand* also makes use of user feedback, provided as mouse and gestural input. Gestures are fairly simple, and include tapping, pointing with a ray, and selecting by pinching with thumb and index finger. Both forms of input allow analysts to categorize data items into groups and navigate through the *Talk2Hand* workflow, with mouse input offering seated desktop interaction supporting intense longer-term work; and gestural interaction offering less constrained and often standing interaction supporting shorter-term work. The availability of both physical mouse and virtual gesture input modalities in *Talk2Hand* offers an intuitive transition for analysts between the familiar WIMP interface and *Talk2Hand*'s novel gestural interface. Allowing multiple input modalities can also allow the user achieve different tasks with the best suited action [BSES17].

Talk2Hand first presents a representative subset of the gathered data in the *triage queue*, which is designed to help analysts quickly filter their information, turning "one pile" into "a few piles." The triage queue presents the data as a scrolling list floating above their desktop, with one row per item, showing minimal information such as relevance, title, and date, enabling quick decisions (Fig. 3a, see also supplementary video). Analysts can sort the data (e.g. by relevance or date), then begin separating their data into categories they define, such as "leaders," "resources," "goals," and "irrelevant." The queue automates interaction to ease this triage. The queue begins by highlighting the first item. After viewing the item, analysts can examine it in detail with a click or tap (Figure 4a) or categorize it with a simple 2D swipe, with the direction of the swipe indicating the category. After each swipe, the triage queue plays a tone matching the category, and automatically moves the highlight to the next row. This process continues until analysts choose to exit or triage the entire queue.

Next, analysts see the *knowledge board*, which is inspired by real-world knowledge boards and affinity diagrams, and allows analysts to refine their "piles," and perhaps create new ones. The board replaces the queue above the desktop and is much larger than it, spanning a large portion of the analysts' available space. All triaged data items appear on the board as cards, now revealing a portion of the data item summary ("irrelevant" items appear only in collapsed form, with a single group label representing the entire set of irrelevant cards). To represent the analysts' categories, the board groups cards visually with distinct hues and spatial clustering (e.g. grouped at corners). Figure 3b shows the knowledge board. To refine data categories, analysts can view data item details (Figure 4a), annotate data items (Figure 4c), change category membership by dragging cards from one group and dropping them into another, and evaluate categories by visually comparing groups. Analysts can also indicate relationship strength by moving cards within groups,



(a) *Triage queue* **(b)** *Knowledge board*

Figure 3: *(a) Triage queue, used to form initial data categories with minimal interaction. (b) Knowledge board, used to refine data categories.*

and postpone decisions about certain data items by placing their cards on the physical desk (Figure 4b).

As they refine categories in the knowledge board, *Talk2Hand*'s *recommender* will observe the groups that analysts define, and attempt to generalize them to the remainder of the imported data. Analysts experience this as cards appearing automatically at the periphery of existing groups. To further distinguish these suggested categories from those analysts have chosen themselves, peripheral cards have the same hue as the group, but are less saturated, with lower saturation indicating that the recommender is less confident. Despite the knowledge board's large size, the recommender will often have many more suggestions than can be displayed. It can accommodate these additional suggestions by organizing them into card stacks, through which analysts can cycle. Analysts can confirm a suggestion by moving the card away from the periphery to the center of the group, which changes the card from a suggestion to an analyst-approved categorization, with full saturation. To reject a suggestion, analysts can move the card into a different group (including the "irrelevant" group). To reflect this and other analyst activity, analysts can manually request suggestion adjustments. Alternatively, the recommender can periodically adjust its suggestions. In this case, to avoid suggestions disappearing before analysts can act on them, the recommender replaces only the oldest suggestions, representing a minority of those not yet confirmed.

As they continue their effort to glean knowledge from their data, analysts can repetitively refine groups within the knowledge board, categorize additional imported data by moving back to the triage queue, or integrate additional data by importing more. A small menu floating over analysts' desks (Figure 4b) allows them to navigate between these three functions.

5. System

Talk2Hand uses the Microsoft Hololens 2 as its primary input and output device, along with a traditional mouse. Hololens 2 can run applications on its own, but we use it with a host PC that generates interactive imagery: a Windows 10 machine with Unity 2019.4, 16GB DDR4 RAM, Intel Core i7 9700K, and a GeForce RTX 2080 Super 8GB graphics card. We developed our application using the Mixed Reality Tool Kit (MRTK), an open-source SDK for AR maintained by Microsoft. It supports both 3D virtual interaction, along with more familiar 2D window-style interaction (embedded in 3D). A previous version of *Talk2Hand* used a Magic Leap 1 headset and its Lumin SDK. We ported from the Magic Leap to the HoloLens to enable better integration with existing Windows applications. To implement the workflow described in Section 4, *Talk2Hand* offers several virtual 2D interactive surfaces, embedded at arm's length in the real 3D space surrounding analysts' desktop machines. The triage queue and knowledge board are the largest, the navigation menu and uncategorized (postponed) cards are others.

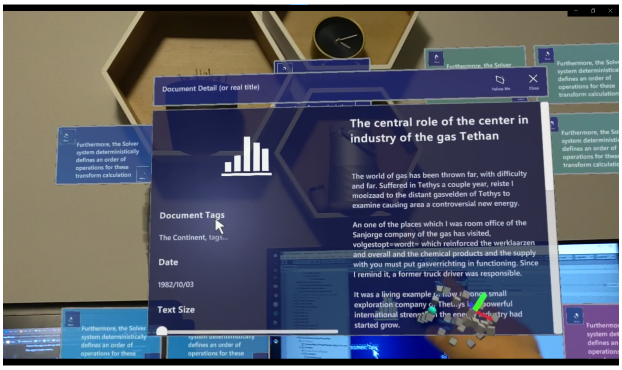
After the user has labeled a small subset of the data (typically, one triage menu), the recommender begins training itself to categorize unlabeled data items. When analysts confirm or reject the recommender's suggested categorizations, or label more data with additional triage, such input will be added to training data for the next training iteration. For the initial implementation of our recommender, we chose semi-supervised learning with self-training [CBZ10]. New self-training iterations begin whenever analysts change the set of categorized data items.

The front and back ends are loosely coupled, so that they are platform-independent. We achieved this using Flask to implement a RESTful API. In JSON format, the AR front end sends group labels and spatial location for each data item to the recommender back end, which returns suggested labels and confidence to the front end.

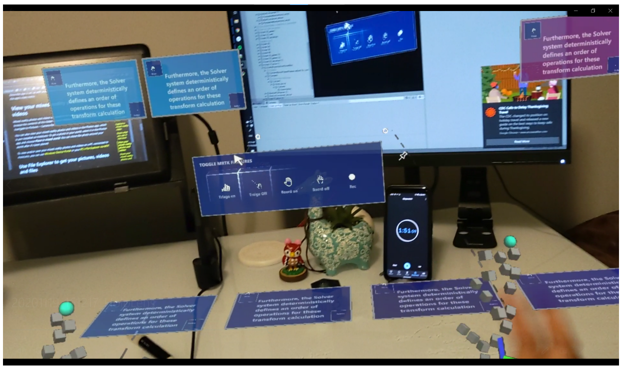
6. Data, Preprocessing and Training

Talk2Hand was designed for use by intelligence analysts, and its development was driven by a matching textual dataset from 2014's IEEE VAST Challenge [Com]. The dataset contains several mini challenges spanning textual, geospatial, and visual data. We focused on textual data as a first step and detail our textual work here, but *Talk2Hand* could be trivially adapted to other textual data, and its knowledge board interface could be adapted to other data sets and types; indeed one of the strengths of knowledge boards is their ability unify heterogeneous data into a cognitively cohesive whole.

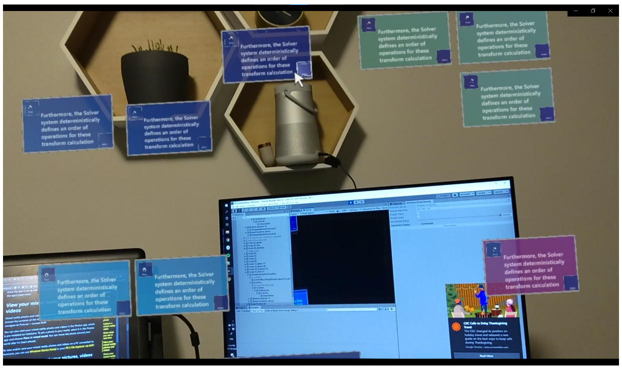
In order to maintain flexibility, we preprocess each raw textual



(a) *Detail view*



(b) *Navigation menu*



(c) *Annotation view*

Figure 4: *(a) Detail view, used for closer study of data items (b) Navigation menu, analysts use it to iterate by categorizing more data, or adding more data. (c) Annotation view, used for memos about data items*

data entry to save crucial information (filename, dates, publisher, author, and content) in our data structure. To preprocess the text, we used the Natural Language Toolkit [DS17], which tokenizes, converts to lowercase, etc. To allow sorting in triage mode, we also transformed dates into a uniform $(yyyy/mm/dd)$ format. We vectorized documents with Doc2vec [ŘS10], which we trained until stable with shuffled documents using PV-DM $dm = 1$. We use the resulting vectors as data features, and calculate them only once. As analysts iteratively describe relatedness by arranging cards into groups, the recommender back end regularly receives updated data item positions and labels. Each time, the classifier is retrained by targeting the updated labels, and stacking (x,y) positions with the document vectors to add two more dimensions. We use `SGDClassifier` from `scikit-learn` [PVG*11]. The classifier returns prediction confidence along with labels, since it decides which suggestions the front end displays to analysts.

7. Conclusion and Future Work

Talk2Hand is a spatial user interface for ML assistants inspired by knowledge boards. By arranging data tokens spatially, analysts can implicitly describe the emerging structures they find in their data, and collaborate with ML assistants more intuitively. Because physical knowledge boards are unwieldy, we implement them in AR.

In future work, we will perform comparative usability studies with intelligence analysts to confirm that *Talk2Hand* eases interaction with ML assistants, without reducing analytic insight [Nor06]. While we will gather quantitative measures of efficency and analytic success, our goals will be primarily formative. We will use a talk-aloud protocol, in which analysts work several small tasks while verbalizing their thoughts, enabling us to find and address both interface strengths and pain points, and well as gain a good understanding of analyst insights.

We also plan to improve the utility of *Talk2Hand* by supporting new data types (e.g. geo, image, audio, video), diagrams (e.g. timelines, social networks), inputs (e.g. eye tracking, voice), functions (e.g. histories, multiple/hierarchical knowledge boards), and ML algorithms (e.g. unsupervised learning for initial triage, and active learning for subsequent iterations).

Talk2Hand already integrates the physical and the virtual by augmenting the analyst's workspace with a larger display, without precluding use of the familiar mouse, nor cutting analysts off from their colleagues. We plan to investigate tighter integration of AR, desktop and physical interfaces that might allow analysts to drag objects from their traditional display to *Talk2Hand*'s virtual one, and to position virtual tokens on physical surfaces in their office. We will also explore applications of *Talk2Hand* outside analytics.

References

[AB06] AGARAWALA A., BALAKRISHNAN R.: Keepin' it real: Pushing the desktop metaphor with physics, piles and the pen. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2006), CHI '06, Association for Computing Machinery, p. 1283–1292. 2

[AEN10] ANDREWS C., ENDERT A., NORTH C.: Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2010), pp. 55–64. 1

[AR96] AGRE P., ROSENSCHEIN S.: *Computational Theories of Interaction and Agency*. Artificial intelligence. MIT Press, 1996. 1

[ASW*19] ARENDT D., SALDANHA E., WESSLEN R., VOLKOVA S., DOU W.: Towards rapid interactive machine learning: evaluating tradeoffs of classification without representation. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (2019), pp. 591–602. 1

[BSES17] BADAM S. K., SRINIVASAN A., ELMQVIST N., STASKO J.: Affordances of input modalities for visual data exploration in immersive environments. In *2nd Workshop on Immersive Analytics* (2017). 2

[CBZ10] CHAPELLE O., BERNHARD S., ZIEN A.: *Semi-supervised learning*. MIT Press, 2010. 3

[Com] COMMUNITY V. A.: Vast challenge 2014. Mini Challenge 1 (MC1). 3

[CSR*03] CZERWINSKI M., SMITH G., REGAN T., MEYERS B., ROBERTSON G. G., STARKWEATHER G. K.: Toward characterizing the productivity benefits of very large displays. In *Interact* (2003), vol. 3, pp. 9–16. 1

[DS17] DENNY M., SPIRLING A.: Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *When It Misleads, and What to Do about It (September 27, 2017)* (2017). 4

[EFN12a] ENDERT A., FIAUX P., NORTH C.: Semantic interaction for sensemaking: inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2879–2888. 2

[EFN12b] ENDERT A., FIAUX P., NORTH C.: Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2012), pp. 473–482. 2

[ERT*17] ENDERT A., RIBARSKY W., TURKAY C., WONG B. W., NABNEY I., BLANCO I. D., ROSSI F.: The state of the art in integrating machine learning into visual analytics. In *Computer Graphics Forum* (2017), vol. 36, Wiley Online Library, pp. 458–486. 1

[FZMA20] FREEMAN G., ZAMANIFARD S., MALONEY D., ADKINS A.: My body, my avatar: How people perceive their avatars in social virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2020), CHI EA '20, Association for Computing Machinery, p. 1–8. 2

[GM18] GARNER G., MCGLYNN P.: *Intelligence analysis fundamentals*. CRC Press, 2018. 1

[KGMC17] KEMENY A., GEORGE P., MÉRIENNE F., COLOMBET F.: New vr navigation techniques to reduce cybersickness. *Electronic Imaging 2017*, 3 (2017), 48–53. 2

[Nor06] NORTH C.: Toward measuring visualization insight. *IEEE computer graphics and applications 26*, 3 (2006), 6–9. 4

[PC05] PIROLLI P., CARD S.: The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (2005), vol. 5, McLean, VA, USA, pp. 2–4. 1

[PVG*11] PEDREGOSA F., VAROQUAUX G., GRAMFORT A., MICHEL V., THIRION B., GRISEL O., BLONDEL M., PRETTENHOFER P., WEISS R., DUBOURG V., VANDERPLAS J., PASSOS A., COURNAPEAU D., BRUCHER M., PERROT M., DUCHESNAY E.: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830. 4

[ŘS10] ŘEHŮŘEK R., SOJKA P.: Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (Valletta, Malta, May 2010), ELRA, pp. 45–50. `http://is.muni.cz/publication/884893/en`. 4

[SIMM95] SHIPMAN III F. M., MARSHALL C. C., MORAN T. P.: Finding and using implicit structure in human-organized spatial layouts of information. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1995), pp. 346–353. 1

[SKKC18] SACHA D., KRAUS M., KEIM D. A., CHEN M.: Vis4ml: An ontology for visual analytics assisted machine learning. *IEEE transactions on visualization and computer graphics 25*, 1 (2018), 385–395. 1

[SLR*20] STANNEY K., LAWSON B. D., ROKERS B., DENNISON M., FIDOPIASTIS C., STOFFREGEN T., WEECH S., FULVIO J. M.: Identifying causes of and solutions for cybersickness in immersive technology: Reformulation of a research and development agenda. *International Journal of Human–Computer Interaction 36*, 19 (2020), 1783–1803. 2

[Vor09] VORA P.: File:wmfui-affinity horiz.jpg - wikimedia commons, Apr 2009. 1