IMPACTS OF THE ARTS: TOPICS & TOOLS FOR ORGANIZATIONAL SENSEMAKING

TOPIC ESTIMATION

Topic prevalence was estimated with the atm package in R (Roberts et al. 2014, 2017). Using the topic search technique from Lee and Mimno (2014), this technique consistently overestimated the number of topics by human interpretation of the most representative and comprehensible number of topics for the data. Held-out likelihood was near uniform for topic numbers above seven. The three most frequent numbers of topics estimated were used as a starting point for manual inspection of the response clustering and adjustment of the topic model to maximize the human interpreted representation, coherence, and exclusivity of underlying topics.

TEAM BASED INTERPRETATION

Manual inspection of the results, topic identification, and “tuning” was assisted with a custom made interactive topic browser which provided decision support for team-based interpretation of each child topic and parent node. Starting with the leaf nodes and working back towards the root of each topic tree, topic labeling and description of the model and clusters was completed by the research team which included domain and community experts for the population sampled and the topic areas under study. The team iteratively performed close reading of representative stems and representative responses (10-50 per topic), examined topic relationships and topic proportions, discussed, interpreted, and revised as needed the topic labels — leading to further development of descriptions, contextualization, and policy implications for key audiences.

EXAMPLES OF OTHER QUESTIONS ANALYZED

How did you feel when you were involved in performing in the arts?
How many hours per week did you spend engaging in the arts?
How did your behavior or thinking change (as a result of engaging in the arts)?
How did you feel after engaging in the arts?
What impact did the arts have on your education?
How did the arts improve your overall well-being?
What impact, if any, did arts have on your professional development?
How did the arts help you to develop and/or identify specific skills?
How could the arts be improved for your needs?

ANALYTIC RESOURCES AND TECHNIQUES

Representative Stems: We use these stems to generally identify the topic. The stems are generated from various metrics based on frequency, exclusivity, and log probabilities which are then clustered through weighted voting. Stems are used instead of words in order to merge nouns, adjectives, and verb tenses; e.g. appreciation, appreciation, appreciation, etc. The number of stems (e.g. 600 such as art, i.e. ones specific to the question phrasing) are removed ahead of time in order to produce more meaningful results.

Representative Documents: Documents contain multiple topics and percentages of the whole response: e.g. the response “it’s too expensive and competitive” might be considered to 49% related to the expressive topic, 49% to the competitive topic, and 2% to the other topics in the model. 10 to 50 documents are displayed in order of document topic proportion, with a minimum of 15% representative for the given topic.

Labeling: Topics and clusters can be labeled, saved to the visualization, and downloaded to reintegrate with the source datasets for subsequent analysis or use elsewhere.

Topic Clusters: The tree visualization intuitively show inter-topic relationships and provides the ability to inspect higher level parent topics based on the identity of child nodes. The 03-based visualization is created from hierarchical clustering on the topic proportion dissimilarity matrix using the rdist R package, and then further customized for our use-case of team-based interpretation.

Representative Cluster Responses: These responses assist with inference of parent node topics, and are generated from cluster constituent responses with a minimum of 10% representation of each topic, and 10% variance between topics.

Node Proportions: Each topic contains a number representing the sum of document topic proportions correspond to that topic. Topic clustering uses the sum of the total document proportions; e.g. if 100 total responses were 50% topic A, then topic A would be 5% of all document proportions.

Packaged as an HTML file. One of the nice things about packaging the topic model, labels, trees, representative documents, and topic descriptions together in an HTML file is that it can be easily shared as an all-in-one package for communicating about the subject. This makes for simpler dissemination to generate enthusiasm among different members of the institutions — all of which have different levels of familiarity with the topics and techniques used to uncover them.

NEXT STEPS

The results of this work are being used to build better models and frameworks for how the arts, design, and interdisciplinary practices create impacts in higher education; to stimulate wider national conversations around the roles and impacts of the arts in student learning and success; and they are expected to support sensemaking around the undergraduate student experience as part of U-M’s Presidential Arts Initiative launched in Fall 2019.

Principal Components Analysis (PCA) that includes dictionary-based cognitive, psychographic, and rhetorical measures of the responses as additional variables—along with topic prevalence—has suggested integrative dimensions in the data. Using the component scores from the PCA, we have further analyzed the variation in those scores and their relationship with demographic factors provided by the survey. This work is leading to more refined research questions, improved design of empirical studies, and better insights for policy makers. We hope to extend these approaches to help other universities and organizations easily ask and interpret their own students’ and stakeholders’ experiences.

Research Tools:

decision support for team-based interpretation of each child topic and parent node.

Research Team:

Gabriel Harp
Deb Mccltoite
Jack Bowman
Mengdan Yuan

department of Art and the Arts Engine at the University of Michigan. The Arts Engine is the home of the Arts Engine platform and chief executive officer of the Arts Engine.

Research partners include the National Endowment for the Arts, the U-M College of Engineering, the A. Alfred Taubman College of Architecture and Urban Planning, the Penny W. Isherwood School of Art & Design, the i-school at Michigan, Theatre & Dance, and the U-M Office of Research.

ARTS ENGINE
UNIVERSITY OF MICHIGAN

IMPACTS OF THE ARTS: TOPICS & TOOLS FOR ORGANIZATIONAL SENSEMAKING

TOPIC ESTIMATION

Topic prevalence was estimated with the atm package in R (Roberts et al. 2014, 2017). Using the topic search technique from Lee and Mimno (2014), this technique consistently overestimated the number of topics by human interpretation of the most representative and comprehensible number of topics for the data. Held-out likelihood was near uniform for topic numbers above seven. The three most frequent numbers of topics estimated were used as a starting point for manual inspection of the response clustering and adjustment of the topic model to maximize the human interpreted representation, coherence, and exclusivity of underlying topics.

TEAM BASED INTERPRETATION

Manual inspection of the results, topic identification, and “tuning” was assisted with a custom made interactive topic browser which provided decision support for team-based interpretation of each child topic and parent node. Starting with the leaf nodes and working back towards the root of each topic tree, topic labeling and description of the model and clusters was completed by the research team which included domain and community experts for the population sampled and the topic areas under study. The team iteratively performed close reading of representative stems and representative responses (10-50 per topic), examined topic relationships and topic proportions, discussed, interpreted, and revised as needed the topic labels — leading to further development of descriptions, contextualization, and policy implications for key audiences.

EXAMPLES OF OTHER QUESTIONS ANALYZED

How did your behavior or thinking change (as a result of engaging in the arts)?
How many hours per week did you spend engaging in the arts?
How did your behavior or thinking change (as a result of engaging in the arts)?
How did you feel after engaging in the arts?
What impact did the arts have on your education?
How did the arts improve your overall well-being?
What impact, if any, did arts have on your professional development?
How did the arts help you to develop and/or identify specific skills?
How could the arts be improved for your needs?

ANALYTIC RESOURCES AND TECHNIQUES

Representative Stems: We use these stems to generally identify the topic. The stems are generated from various metrics based on frequency, exclusivity, and log probabilities which are then clustered through weighted voting. Stems are used instead of words in order to merge nouns, adjectives, and verb tenses; e.g. appreciation, appreciation, appreciation, etc. The number of stems (e.g. 600 such as art, i.e. ones specific to the question phrasing) are removed ahead of time in order to produce more meaningful results.

Representative Documents: Documents contain multiple topics and percentages of the whole response: e.g. the response “it’s too expensive and competitive” might be considered to 49% related to the expressive topic, 49% to the competitive topic, and 2% to the other topics in the model. 10 to 50 documents are displayed in order of document topic proportion, with a minimum of 15% representative for the given topic.

Labeling: Topics and clusters can be labeled, saved to the visualization, and downloaded to reintegrate with the source datasets for subsequent analysis or use elsewhere.

Topic Clusters: The tree visualization intuitively show inter-topic relationships and provides the ability to inspect higher level parent topics based on the identity of child nodes. The 03-based visualization is created from hierarchical clustering on the topic proportion dissimilarity matrix using the rdist R package, and then further customized for our use-case of team-based interpretation.

Representative Cluster Responses: These responses assist with inference of parent node topics, and are generated from cluster constituent responses with a minimum of 10% representation of each topic, and 10% variance between topics.

Node Proportions: Each topic contains a number representing the sum of document topic proportions correspond to that topic. Topic clustering uses the sum of the total document proportions; e.g. if 100 total responses were 50% topic A, then topic A would be 5% of all document proportions.

Packaged as an HTML file. One of the nice things about packaging the topic model, labels, trees, representative documents, and topic descriptions together in an HTML file is that it can be easily shared as an all-in-one package for communicating about the subject. This makes for simpler dissemination to generate enthusiasm among different members of the institutions — all of which have different levels of familiarity with the topics and techniques used to uncover them.

NEXT STEPS

The results of this work are being used to build better models and frameworks for how the arts, design, and interdisciplinary practices create impacts in higher education; to stimulate wider national conversations around the roles and impacts of the arts in student learning and success; and they are expected to support sensemaking around the undergraduate student experience as part of U-M’s Presidential Arts Initiative launched in Fall 2019.

Principal Components Analysis (PCA) that includes dictionary-based cognitive, psychographic, and rhetorical measures of the responses as additional variables—along with topic prevalence—has suggested integrative dimensions in the data. Using the component scores from the PCA, we have further analyzed the variation in those scores and their relationship with demographic factors provided by the survey. This work is leading to more refined research questions, improved design of empirical studies, and better insights for policy makers. We hope to extend these approaches to help other universities and organizations easily ask and interpret their own students’ and stakeholders’ experiences.

Research Tools:

decision support for team-based interpretation of each child topic and parent node.

Research Team:

Gabriel Harp
Deb Mccltoite
Jack Bowman
Mengdan Yuan

department of Art and the Arts Engine at the University of Michigan. The Arts Engine is the home of the Arts Engine platform and chief executive officer of the Arts Engine.

Research partners include the National Endowment for the Arts, the U-M College of Engineering, the A. Alfred Taubman College of Architecture and Urban Planning, the Penny W. Isherwood School of Art & Design, the i-school at Michigan, Theatre & Dance, and the U-M Office of Research.