Using Analogy to Cluster Hand-Drawn Sketches for Sketch-Based Educational Software

Maria D. Chang, Kenneth D. Forbus

■ *One of the major challenges to building intelligent educational software is determining what kinds of feedback to give learners. Useful feedback makes use of models of domain-specific knowledge, especially models that are commonly held by potential students. To empirically determine what these models are, student data can be clustered to reveal common misconceptions or common problem-solving strategies. This article describes how analogical retrieval and generalization can be used to cluster automatically analyzed hand-drawn sketches incorporating both spatial and conceptual information. We use this approach to cluster a corpus of hand-drawn student sketches to discover common answers. Common answer clusters can be used for the design of targeted feedback and for assessment.*

Sketching and drawing are valuable tools for communicating conceptual and spatial information. When people communicate spatial ideas with each other, drawings and diagrams are highly effective because they lighten working memory load and make spatial inference easier (Larkin and Simon 1987). Visual representations may also be helpful for communicating abstract ideas, even when those ideas are not literally about space (for example, reasoning about probability [Cheng 2011]). Essentially, drawings provide externalized symbol systems that facilitate spatial reasoning, which can be applied to a variety of domains.

Sketching is especially useful for learning and instruction in spatially rich subjects, like science, technology, engineering, and mathematics (that is, STEM fields). For science education, sketching can be used to increase engagement, improve learning, and encourage encoding across different representations (Ainsworth, Prain, and Tytler 2011). Drawing and sketching have the potential to play critical roles in science education, especially considering the importance of spatial skills in STEM disciplines. Data from more than 50 years of psychological research indicate that spatial skills are stable predictors of success in STEM fields (Wai, Lubinski, and Benbow 2009). Individuals with greater spatial skills are more

 Copyright © 2014, Association for the Advancement of Artificial Intelligence. All rights reserved. ISSN 0738-4602

likely to earn advanced STEM degrees and attain STEM careers. As such, it is important that science education make use of spatial tools, like sketching, for teaching the next generation of STEM professionals, educators, and researchers, as well as for more informed citizens.

Advances in intelligent tutoring systems have opened the possibility of creating educational software than can support sketching and take advantage of the many benefits it has to offer (for example, Valentine et al. [2012]). However, building intelligent sketching software is challenging because it requires software that understands sketches in humanlike ways. The noisy nature of sketches makes them difficult to interpret. Consequently, assessing the quality of a student's sketch requires a considerable amount of spatial and conceptual reasoning. With the exception of advanced design sketches, most sketches are rough approximations of spatial information. They are rarely drawn to scale and often require multimodal cues (for example, gestures, speech) to facilitate understanding. For example, a sketch of a map might contain various shapes that represent different landmarks. The shapes may look nothing like the actual landmarks, but may be denoted as landmarks by labels. No one has trouble understanding that a blob can represent something that looks physically different; such visual information is processed with a grain of salt. Somehow, people are able to focus on the spatially and conceptually important information in the sketch and, for the most part, ignore irrelevant information. Building software that can achieve this level of understanding from a sketch is a major challenge for the artificial intelligence community.

Qualitative representations are a good match for sketched data because they carve continuous visual information (for example, two-dimensional location) into discrete categories and relationships (for example, round, right of, and others). These representations enable software systems to reason about sketches using the same structured representations that are hypothesized to be used by people.

Since comparison is prevalent in instruction, a model of visual comparison that incorporates conceptual information is also important for sketch understanding. Analogical comparison using structure mapping (Gentner 1983) allows structured descriptions to be compared to each other to evaluate how similar the two descriptions are. The structure-mapping model of analogy can be used to compare sketches to each other, highlighting qualitative similarities and differences, while adhering to constraints and biases that are supported by psychological research. Computational models of structure mapping have been used to simulate cognitive phenomena (Gentner and Forbus 2011), solve physics problems (Klenk and Forbus 2009; Lockwood and Forbus 2009), and compare text passages to Jeopardy clues for question answering (Murdock 2011). Structure mapping enabled these systems to make more humanlike comparisons. In educational software, such as Sketch Worksheets (Yin et al. 2010), structure mapping generates comparisons that can be used to assess a student's sketch by comparing it to a predefined solution.

A major challenge in any intelligent tutoring system is determining how to coach students. When designing feedback, instructors must hypothesize what will be hard and what will be easy for students. Such hypotheses are not always data driven and can be inaccurate (Nathan, Koedinger, and Alibali 2001). Consequently, most successful intelligent tutoring systems incorporate detailed cognitive models of the task being taught. Building cognitive models requires research on novice misconceptions and strategies (Anderson et al. 1995). Some systems also model the strategies of human tutors, such as intervention techniques and tutoring dialogue (VanLehn et al. 2007). Creating cognitive models for both correct knowledge and common misconceptions for an entire domain is difficult. By contrast, specific exercises can have easily defined misconceptions that can be identified without a full analysis of the domain. By analyzing the work of multiple students on an example, common models (some of which may be misconceptions) can be mined from the data. Although there has been work devoted to assessing student knowledge through sketches (Jee et al. 2009; Kindfield 1992) and mining information about students from learning data (for example, from hand-coded sketches [Worsley and Blikstein 2011]) we are unaware of any efforts to combine automatic sketch understanding and educational data mining. This paper describes an approach for using analogical reasoning over hand-drawn sketches to detect common student answers.

Our hypothesis is that analogical generalization can be used to generate meaningful clusters of hand-drawn sketches. We compare analogical generalization to a *k*-means clustering algorithm and evaluate its performance on a set of labeled (that is, clustered by hand) student sketches. The resulting clusters from the experiments can be inspected to identify the key characteristics of each cluster. These characteristics can be used to identify student misconceptions and to design targeted feedback for students.

Background

Structure mapping is the comparison mechanism of our clustering approach. Here we summarize the computational models for analogical matching, retrieval, and generalization that we use. We then describe Sketch Worksheets, which is our sketch-based educational software system used to collect and encode hand-drawn sketches.

Structure Mapping

The structure-mapping engine (SME) (Faulkenhainer, Forbus, and Gentner 1989) is a computational model of analogy that compares two structured descriptions, a base and a target, and computes one or more analogical mappings between them. Each mapping contains a set of correspondences to indicate which items in the base correspond to which items in the target, a structural evaluation score to measure match quality, and a set of candidate inferences, which are statements that are true in the base and hypothesized to be true in the target.

Three constraints are imposed on the mapping process to prevent all potential mappings from being computed and to account for certain psychological phenomena. The mapping process begins by matching identical relations to each other. This constraint is referred to as *identicality.* Nonidentical relations may end up corresponding to each other in the final match, but only if their correspondence is suggested by higher-order matches. The mapping process also adheres to the 1:1 *constraint,* which ensures that items in one description can have at most one corresponding item in the other description. *Parallel connectivity* requires that arguments to matching relations also match. These constraints make analogical matching tractable.

Importantly, SME has a bias for mappings with greater *systematicity,* which means that it prefers mappings with systems of shared relations, rather than many shared isolated facts. In other words, given two mappings, one with many matching but disconnected (that is, shallow, lower-order) relations, and another with many matching interconnected (that is, deep, higher-order) relations, SME will prefer the latter. The systematicity preference in SME captures the way people perform analogical reasoning. People prefer analogies that have systems of shared relations and are more likely to make analogical inferences that follow from matching causal systems, rather than shallow matches (Clement and Gentner 1991).

To create sketch clusters, we use SAGE (sequential analogical generalization engine), an extension of SEQL (Kuenhe et al. 2000) that computes probabilities for expressions during generalization and retrieves structured descriptions using analogical retrieval. Generalizations are created by incrementally introducing exemplars into a generalization context. Each generalization context consists of a case library that includes both exemplars and generalizations. For each new exemplar, the most similar exemplar or generalization in the generalization context is found via analogical retrieval using MAC/FAC (Forbus, Gentner, and Law 1995). MAC/FAC computes content vectors that measure the relative frequency of occurrence of relations and attributes in structured representations. It finds the maximum dot product of the vector for the exemplar with the vectors of everything in the generalization context. This step may retrieve up to three items and is analogous to a bag of words approach to similarity, albeit with predicates. These items are then compared to the exemplar using SME and the item with the highest structural evaluation score (that is, the most similar case) is returned.

The reminding returned by MAC/FAC is either an exemplar or an existing generalization. If the similarity between the reminding and the exemplar is above a predefined assimilation threshold, then they are merged together. If the best match is a generalization, the new exemplar is added to it. If the best match is another exemplar, then the two are combined into a new generalization. If there is no best match above the assimilation threshold, then the new exemplar is added directly to the case library for that generalization context. It will remain an exemplar in the generalization context until it is joined with a new exemplar or until there are no more exemplars to add.

The resulting generalizations contain generalized facts and entities. Each fact in a generalization is assigned a probability, which is based on its frequency of occurrence in the exemplars included in the generalization. For example, a fact that is true in only half of the exemplars would be assigned a probability of 0.5. Thus, entities become more abstract, in that facts about them "fade" as their probability becomes lower.

Sketch Worksheets

Sketch Worksheets are built within CogSketch (Forbus et al. 2011), our domain-independent sketch understanding system. Each sketch worksheet includes a problem statement, a predefined solution sketch, and a workspace where the student sketches his or her candidate answer. As part of the authoring process, the worksheet author describes the problem, sketches an ideal solution, and labels elements in the solution with concepts that he or she selects from an OpenCyc-derived knowledge base.[1] CogSketch analyzes the solution sketch by computing qualitative spatial and conceptual relations between items in the sketch. Spatial relations that are automatically computed include topological relations (for example, intersection, containment) and positional relations (for example, right of, above), all of which are domain independent. Conceptual relations include relations selected by the worksheet author and can be a combination of domain general relationships and domain-specific relationships. The worksheet author can then peruse the representations computed by CogSketch and identify which facts are important for capturing the correctness of the sketch. For each such fact, the author includes a piece of advice that should be given to the student if that fact does not hold in the student's sketch. When a student asks for feedback, SME is used to compare the student's sketch to the solution sketch. The candidate infer-

ences, which represent differences between the sketches, are examined to see if there are any important facts among them. If there are, the advice associated with that important fact is included as part of the feedback. By identifying important facts and associating advice with them, the worksheet author creates relational criteria for determining the correctness of the sketch.

Worksheet authors can also identify which drawn elements have quantitative location criteria by defining quantitative ink constraints, which define a tolerance region for a particular drawn element. Which elements of a solution sketch correspond to elements of a student's sketch is determined through the correspondences that SME computes. If the student's drawn element falls outside of the tolerance region, it is considered incorrect. If it falls within the tolerance region, it is considered correct. This allows the author to set criteria about the absolute location of drawn elements.

The difference between these two criterion types can be illustrated by two different worksheet exercises. Consider a worksheet that asks a student to draw the solar system. The exact location of the sun does not matter, as long as it is contained by the orbit rings of other planets. In other words, its location is constrained relative to other drawn entities. A worksheet author would capture this by marking a containment fact as important and associating a natural language advice string with that fact. Alternatively, consider a worksheet that asks a student to identify the temporal lobe on a diagram of the human brain. The absolute location of the drawing denoting the temporal lobe would be important. For an element whose location is constrained relative to an absolute frame of reference (for example, a background image), quantitative ink constraints are necessary.

Sketch Worksheets have been used in experiments on spatial reasoning as well as classroom activities in geoscience (at Northwestern University and Carleton College) and elementary school biology. The sketches used in the experiments described in this paper were collected using sketch worksheets.

Clustering Through Analogical Generalization

Clustering is achieved by performing analogical generalization over student sketches. The clustering algorithm adds the sketches in random order, using the SAGE algorithm mentioned above. A single generalization context is used, that is, it operates unsupervised, because the goal is to see what clusters emerge.

Encoding

A major challenge to clustering sketches is choosing how to encode the information depicted in each sketch. Each sketch contains a wealth of spatial information, not all of it relevant for any particular situation.[2] In order to highlight visually and conceptually salient attributes and relationships, we harness information explicitly entered by the student and the worksheet author. More specifically, we filter the underlying representations in each sketch based on the following principles: conceptual information is critical, quantitative ink constraints must constrain analogical mappings, and worksheet authoring should guide spatial and conceptual elaboration.

Conceptual Information

Every sketch worksheet comes equipped with a subset of concepts from an OpenCyc-derived knowledge base. This subset contains the concepts that may be used in the worksheet and are selected by the worksheet author to limit the conceptual scope of the exercise. These concepts are applied by students to elements of their drawing through CogSketch's conceptual labeling interface. This is useful for education because the mapping between shapes and entities is often one to many. While visual relationships are computed automatically by CogSketch, conceptual relationships are entered by sketching arrows or annotations and labeling them appropriately, through the same interface. Thus, the conceptual labels constitute the students' expression of their model of what is depicted. Consequently, conceptual information is always encoded for generalization.

Quantitative Ink Constraints Limit Matches

Another type of information that is entered explicitly by the worksheet author is quantitative ink constraints. Recall that quantitative ink constraints define a tolerance region relative to an absolute frame of reference (for example, a background image). Quantitative ink constraints are defined for entities whose absolute position matters.

When encoding information about entities for which there are quantitative ink constraints, the encoding algorithm computes their position with respect to the tolerance regions, to determine if the entity's location meets the constraint or not. If it does not, we further encode how the constraint was violated (for example, too wide, too narrow) and include that information in the encoding.

Furthermore, each entity that is evaluated with respect to a quantitative ink constraint is associated with that constraint as a location-specific landmark. This association limits the possible analogical mappings by ensuring that entities associated with one landmark cannot map to entities that are associated with a different landmark. This also ensures that entities cannot be generalized across different location-specific landmarks. This approach for using quantitative constraints to limit the analogical mappings has been shown to lead to sketch comparisons that provide more accurate feedback to students (Chang and Forbus 2012).

Spatial and Conceptual Elaboration

Worksheet authors can also specify a subset of the visual relationships computed by CogSketch as important. For example, the core of Earth must be

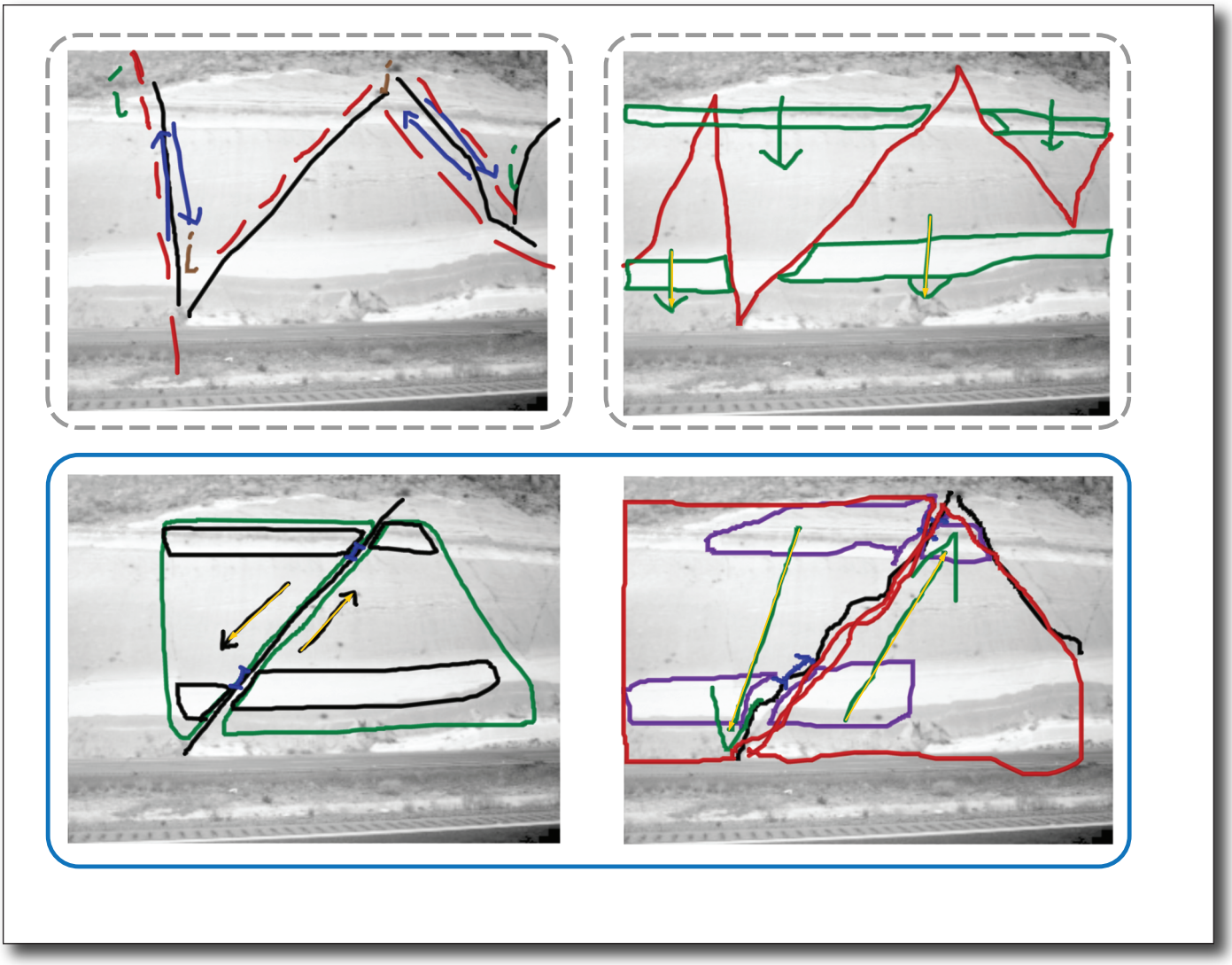


Figure 1. Four Example Student Sketches.

The bottom two are grouped together and considered nearly correct. The top two occupy their own single-member clusters.

inside its mantle. Some conceptual information can also be marked as important, for example, that one layer of rock is younger than another. All facts marked as important by the worksheet author, whether spatial or conceptual, are always included in the encoding for generalization.

Evaluation

To evaluate our clustering algorithm we used a set of fault identification worksheets (for example, figure 1) submitted by students taking an undergraduate geoscience course at Northwestern University. There were 28 sketches in total, spanning three different fault identification exercises. A gold standard was created by hand-clustering the sketches for each exercise separately, taking into account visual features of the sketch (for example, location of drawn entities), what type of entities were included in the sketch, and whether or not the sketch would be considered correct by an instructor. We then ran our generalization algorithm on the unlabeled data for each exercise to evaluate how well the clusters it produced match the gold standard. Because clusters may differ depending on the order in which sketches are selected, we repeated the clustering over 10 iterations. We collected three measures from the resulting clusters: purity, precision, and recall. Purity is a measure of the quality of a set of clusters, defined as the ratio of correctly classified exemplars across clusters to the total number of exemplars. The precision of a cluster is the proportion of exemplars in the cluster that are classified correctly, and the recall of a cluster is the proportion of exemplars that are correctly included in

the cluster. Precision and recall were computed with respect to the gold standard.

k-Means Clustering

To explore the impact of relational structure on generalization behavior, we also compared our approach to a nonstructural way of ascertaining similarity. Specifically, we used the MAC/FAC content vectors (described above) as a cruder, nonrelational form of similarity. While content vectors are still sensitive to the presence of relationships, since those predicates are included in them, content vectors only contain relative frequency information. In other words, "man bites dog" is the same as "dog bites man." We used *k*-means clustering on the same data, where each mean was the content vector of a sketch and the distance measure between means was the inverse dot product of the content vectors being compared. The more overlap between the content vectors (that is, the more overlap between attributes and relationships), the greater the similarity and the smaller the distance. For each *k*-means clustering process we supplied *k* by counting the number of labeled clusters. In this sense, the *k*-means clustering approach had a slight advantage over analogical generalization. The *k*-means clustering algorithm was also repeated 10 times, since the initial *k* means can affect the makeup of clusters.

Results

Table 1 shows the average purity, precision. and recall for each approach across the three worksheet groups, averaged over 10 iterations of each approach. Analogical generalization outperformed *k*-means without analogy for clustering in all measures. Since purity is often high when there are many clusters, it is important to consider the precision and recall measures as well.

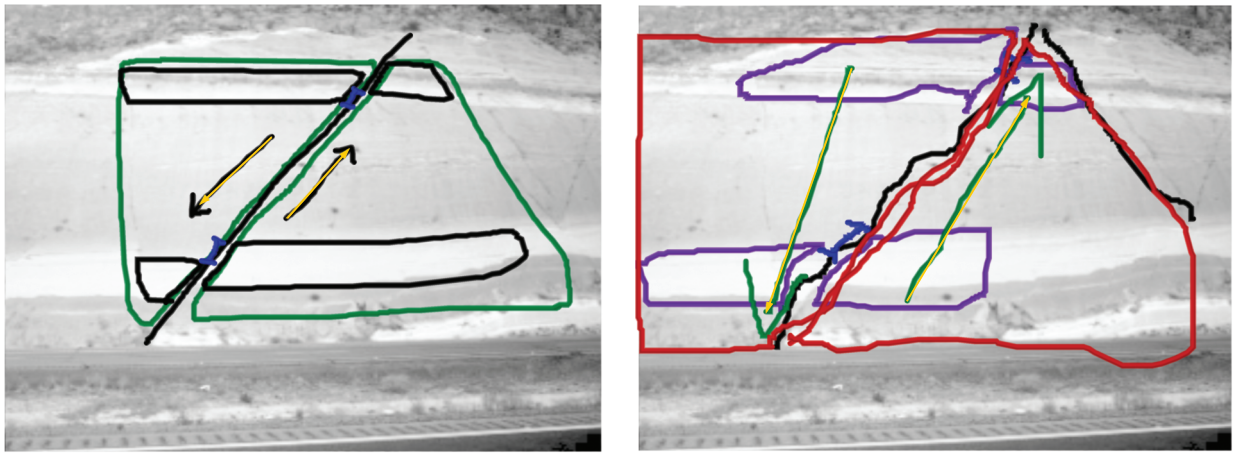
We used independent samples *t*-tests to test for significant differences between purity, precision, and recall for each sketch group separately (for a total of nine comparisons). Each measure was significantly higher for analogical generalization than for *k*-means clustering ($p < 0.005$, Bonferroni corrected for nine comparisons).

Figure 2 shows two sketches that were frequently generalized together. This cluster indicates a common sketching behavior exhibited by students. The high-probability facts in the generalization indicate the defining criteria for the cluster. Most of the high-probability facts in this generalization are concept membership attributes. Other facts refer to the direction of the sketched diagonal arrows in the sketch. These facts were already considered in the feedback design of this worksheet. However, the three high-probability facts shown in figure 2 indicate the potential for more targeted feedback. These facts indicate that three of the four marker beds failed quantitative ink constraints in specific ways. The bold horizontal arrows imposed on the figure point to two marker beds that map to each other in an analogical mapping. Both of these marker beds fall short of the left bounds of their quantitative ink constraints (see first fact in figure 2). Similarly, two other marker beds (unmarked) fall short of the right bounds of the quantitative ink constraints. Without knowing that multiple students would exhibit this common behavior, a worksheet author would have no reason to include targeted feedback about it. Indeed, the authors of these worksheets were surprised to discover that multiple students had engaged in the same interpretation of the image and consequently the same erroneous sketching behavior. Given that multiple students commit this error, targeted feedback about the horizontal extent of marker beds would have been helpful, for example, "Marker bed regions are not just near the fault; they can extend to the edges of the image." This is the type of information that can be leveraged from clustering student data. In turn, this information can be used to design exercises with advice for students that can have a greater impact than the feedback created by the worksheet authors a priori.

	SAGE	*k*-means
Sketch Group 1		
Number of Clusters	6.7	6
Purity**	0.90	0.72
Precision**	0.86	0.56
Recall*	0.85	0.56
Sketch Group 2		
Number of Clusters	5.5	4
Purity**	0.94	0.74
Precision**	0.98	0.61
Recall**	0.82	0.59
Sketch Group 3		
Number of Clusters	6.1	4
Purity**	0.96	0.83
Precision**	0.99	0.67
Recall*	0.80	0.68

Table 1. Clustering Measures for Analogical Generalization (SAGE) and k-Means Clustering (Without Analogy).

All measures are averaged over 10 random restart iterations of the clustering procedure. Asterisks indicate the probability associated with independent samples *t*-tests between SAGE and *k*-means measures: ** $p < 0.001$, * $p < 0.005$



(lessThanQuantInkLeftBound (GenEntFn 5 1 Faults-1) Object-438) (lessThanQuantInkRightBound (GenEntFn 3 1 Faults-1) Object-440)
(lessThanQuantInkRightBound (GenEntFn 4 1 Faults-1) Object-436)

Figure 2. Two Sketches That Are Frequently Clustered Together and Three High-Probability Facts from Their Generalization.

The horizontal block arrows point to the drawn entity that is referenced in the first fact. That entity falls short of the left bound of its quantitative ink constraint.

Related Work

Many researchers have explored misconceptions in domains like algebra, geometry (Anderson et al. 1995), and physics (VanLehn et al. 2007). Each of these research programs answers important questions about the structure of knowledge during learning. These answers have shaped the coaching strategies of various tutoring systems.

Many sketch understanding systems exist but most stick to a single domain because they use sketch recognition (Lee et al. 2007; de Silva et al. 2007; Valentine et al. 2012). No other sketch understanding systems use structure mapping as a model for comparison. Despite this, it may still be possible to apply similar clustering techniques to those systems.

Discussion and Future Work

This article describes a method for clustering sketches to detect common answer patterns. We used models of human analogical processing to cluster hand-drawn sketches completed by undergraduate geoscience students. The analogical clustering approach significantly outperformed a *k*-means clustering algorithm.

This technique can be used to mine common answer patterns from sketches so that they can be used for assessment or for designing targeted feedback. Instructors may use this technique to discover the distribution of answer patterns in their classrooms, some of which may be prevalent misconceptions. This approach enables common answer detection in a data-driven (but tightly scoped) manner, without requiring a cognitive analysis of the entire domain or even the entire task.

One of the limitations to this approach is the understandability of the facts used to describe generalizations. As discussed above, high-probability facts can be used to understand the defining criteria of a cluster. For an instructor to easily interpret these facts would require familiarity with the knowledge representations used there. However, it can be argued that

the instructors may not need those explicit facts. Instead, they can simply view a prototypical member of the cluster and decide on the defining criteria for themselves. With this technique, rather than looking at all the sketches submitted by students, an instructor can inspect only as many sketches as there are clusters.

In the future we plan to continue refining encoding procedures of sketches. The procedures used in this experiment are domain general, but there are likely cases where different filters on conceptual and/or spatial information will be needed. We may also be able to learn more about common problem-solving strategies by including an analysis of a sketch's history in our encoding. This would allow us to create clusters based on sketching behaviors over time, rather than only in the final state of a student's sketch. We also have not yet integrated shape and edge level representations into this encoding procedure (Lovett et al. 2012), as these are only now starting to be integrated into our sketch worksheets. We also plan to add clustering to the grading utilities built into sketch worksheets. Most importantly, we plan to extend our experiments to more sketches in more STEM domains. This will involve continued collaboration with STEM experts and educators at the K–12 and university levels. We anticipate that using a varied corpus of sketches will enable us to converge on encoding procedures that will scale up and become powerful tools for designing environments with impactful instructional feedback.

Acknowledgements

Thanks to Brad Sageman and Andrew Jacobson and their students for providing us with expertise, advice, and data. This work was supported by the Spatial Intelligence and Learning Center (SILC), an NSF Science of Learning Center (Award Number SBE-1041707).

Notes

1. www.cyc.com/platform/opencyc.

2. For example, a student might draw a planet above the sun versus below the sun, a visually salient difference that doesn't matter in most orbital diagrams.

References

Ainsworth, S.; Prain, V.; and Tytler, R. 2011. Drawing to Learn in Science. *Science* 333(6046): 1096–1097. dx.doi.org/10.1126/science.1204153

Anderson, J. R.; Corbett, A. T.; Koedinger, K. R.; and Pelletier, R. 1995. Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences* 4(2): 167–207. dx.doi.org/10.1207/s15327809jls0402_2

Chang, M. D., and Forbus, K. D. 2012. Using Quantitative Information to Improve Analogical Matching Between Sketches. In *Proceedings of the Twenty Fourth Conference on Innovative Applications of Artificial Intelligence,* 2269–2274. Menlo Park, CA: AAAI Press. PMid:23083649

Cheng, P. C.-H. 2011. Probably Good Diagrams for Learning: Representational Epistemic Recodification of Probability Theory. *Topics in Cognitive Science* 3(3): 475–498. dx.doi.org/10.1111/j.1756-8765.2009.01065.x

Clement, C. A., and Gentner, D. 1991. Systematicity as a Selection Constraint in Analogical Mapping. *Cognitive Science* 15(1): 89–132. dx.doi.org/10.1207/s15516709cog1501_3

de Silva, R.; Bischel, T. D.; Lee, W.; Peterson, E. J.; Calfee, R. C.; and Stahovich, T. 2007. Kirchhoff's Pen: A Pen-Based Circuit Analysis Tutor. In *Proceedings of the Fourth Eurographics Workshop on Sketch-Based Interfaces and Modeling*, 75–82. New York: Association for Computing Machinery. dx.doi.org/10.1145/1384429.1384447

Falkenhainer, B.; Forbus, K.; and Gentner, D. 1989. The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence* 41(1): 1–63. dx.doi.org/10.1016/0004-3702(89)90077-5

Forbus, K.; Gentner, D.; and Law, K. 1995. MAC/FAC: A Model of Similarity-Based Retrieval. *Cognitive Science* 19(2): 141–205. dx.doi.org/10.1207/s15516709cog1902_1

Forbus, K. D.; Usher, J.; Lovett, A.; Lockwood, K.; and Wetzel, J. 2011. Cogsketch: Sketch Understanding for Cognitive Science Research and for Education. *Topics in Cognitive Science* 3(4): 648–666. dx.doi.org/10.1111/j.1756-8765.2011.01149.x

Gentner, D. 1983. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science* 7(2): 155–170. dx.doi.org/10.1207/s15516709cog0702_3

Gentner, D., and Forbus, K. 2011. Computational Models of Analogy. *Wiley Interdisciplinary Reviews: Cognitive Science.* 2(3): 266–276. dx.doi.org/10.1002/wcs.105

Jee, B.; Gentner, D.; Forbus, K. D.; Sageman, B.; and Uttal, D. H. 2009. Drawing on Experience: Use of Sketching to Evaluate Knowledge of Spatial Scientific Concepts. In *Proceedings of the Thirty First Annual Conference of the Cognitive Science Society,* 2499–2504. Wheat Ridge, CO: Cognitive Science Society.

Kindfield, A. C. H. 1992. Expert Diagrammatic Reasoning in Biology. In *Reasoning with Diagrammatic Representations: Papers from the AAAI Spring Symposium*, 41–46. AAAI Technical Report SS-92-02. Palo Alto, CA: AAAI Press.

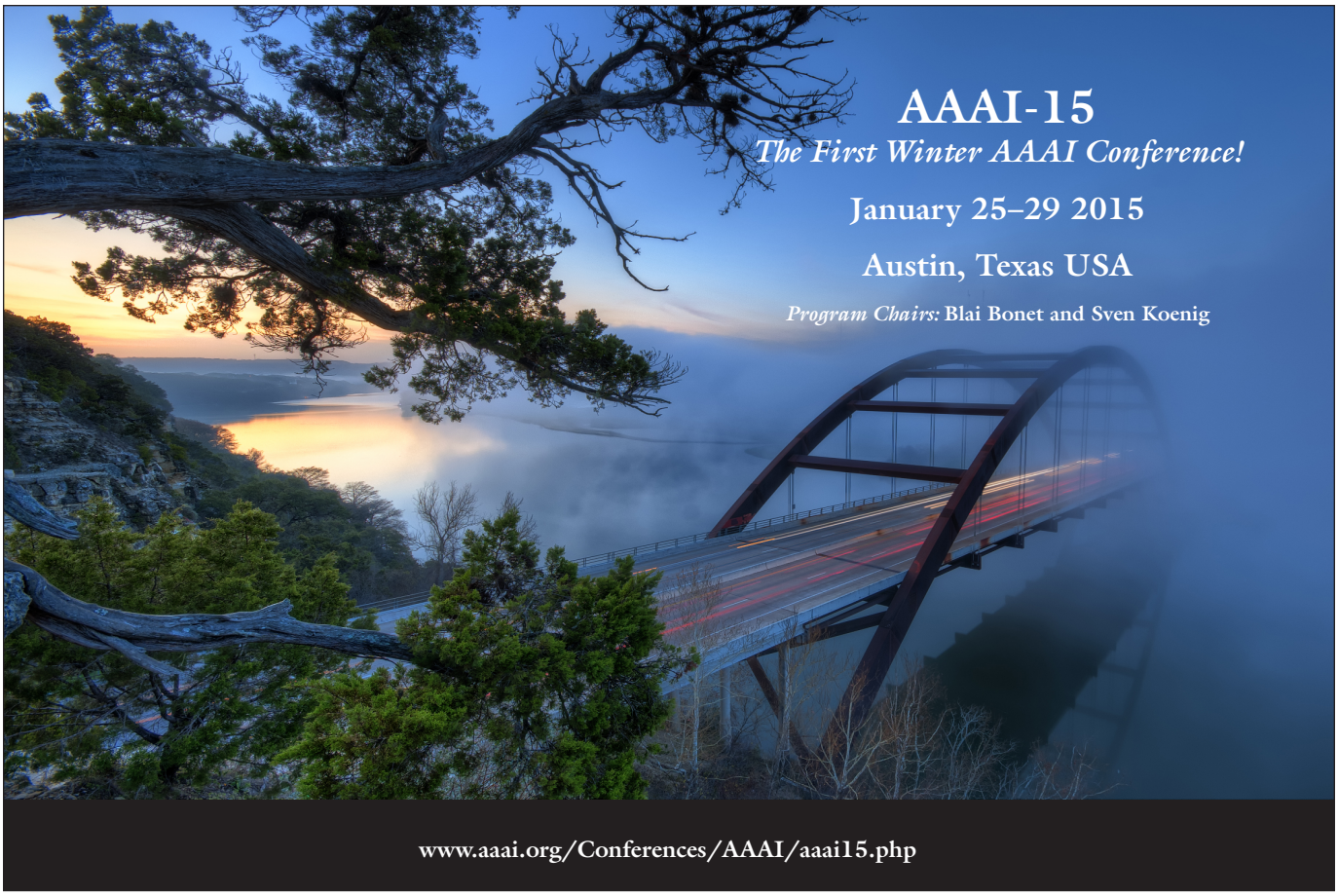
Klenk, M., and Forbus, K. 2009. Analogical Model Formulation for Transfer Learning in AP Physics. *Artificial Intelligence* 173(18): 1615–1638. dx.doi.org/10.1016/j.artint.2009.09.003

Kuehne, S.; Forbus, K.; Gentner, D.; and Quinn, B. 2000. SEQL: Category Learning as Progressive Abstraction Using Structure Mapping. In *Proceedings of the Twenty Second Annual Conference of the Cognitive Science Society,* 770–775. Wheat Ridge, CO: Cognitive Science Society.

Larkin, J. H., and Simon, H. A. 1987. Why a Diagram Is (Sometimes) Worth 10000 Words. *Cognitive Science* 11(1): 65–99. x.doi.org/10.1111/j.1551-6708.1987.tb00863.x

Lee, W.; De Silva, R.; Peterson, E. J.; Calfee, R. C.; and Stahovich, T. F. 2007. Newton's Pen — A Pen-Based Tutoring System for Statics. In *Proceedings of the Eurographics Workshop on Sketch-Based Interfaces and Modeling,* 59–66. New York: Association for Computing Machinery. dx.doi.org/10.1145/1384429.1384445

Lockwood, K., and Forbus, K. 2009. Multimodal Knowledge Capture from Text and Diagrams. In *Proceedings of the Fifth International Conference on Knowledge Capture,* 65–72. New



York: Association for Computing Machinery. dx.doi.org/10.1145/1597735.1597747

Lovett, A.; Kandaswamy, S.; McLure, M.; and Forbus, K. 2012. Evaluating Qualitative Models of Shape Representation. Paper presented at the Twenty-Sixth International Workshop on Qualitative Reasoning. Playa Vista, CA, July 16–18. PMid:22228709

Murdock, J. 2011. Structure Mapping for Jeopardy! Clues. In *Proceedings of the Nineteenth International Conference on Case-Based Reasoning,* 6–10. Berlin: Springer.

Nathan, M.; Koedinger, K. R.; and Alibali, M. 2001. Expert Blind Spot: When Content Knowledge Eclipses Pedagogical Content Knowledge. Paper presented at the 2001 Annual Meeting of the American Educaitonal Research Association, Seattle, WA, April 10–14.

Valentine, S.; Vides, F.; Lucchese, G.; Turner, D.; Rim, H.; Li, W.; Linsey, J.; Hammond, T. 2012. Mechanix: A Sketch-Based Tutoring System for Statics Courses. In *Proceedings of the Twenty Fourth Conference on Innovative Applications of Artificial Intelligence,* 2253–2260. Menlo Park, CA: AAAI Press.

VanLehn, K.; Graesser, A. C.; Jackson, G. T.; Jordan, P.; Olney, A.; and Rosé, C. P. 2007. When Are Tutorial Dialogues More Effective Than Reading? *Cognitive Science* 31(1): 3–62. dx.doi.org/10.1080/03640210709336984

Yin, P.; Forbus, K. D.; Usher, J.; Sageman, B.; and Jee, B. 2010. Sketch Worksheets: A Sketch-Based Educational Software System. In *Proceedings of the Twenty-Second Conference on Innovative Applications of Artificial Intelligence,* 1871–1876. Menlo Park, CA: AAAI Press.

Wai, J.; Lubinski, D.; and Benbow, C. P. 2009. Spatial Ability for STEM Domains: Aligning over 50 Years of Cumulative Psychological Knowledge Solidifies Its Importance. *Journal of Educational Psychology* 101(4): 817–835. dx.doi.org/10.1037/a0016127

Worsley, M., and Blikstein P. 2011. What's an Expert? Using Learning Analytics to Identify Emergent Markers of Expertise Through Automated Speech, Sentiment, and Sketch Analysis. In *Proceedings of the Fourth Annual Conference on Educational Data Mining,* 235–240. Boston: International Educational Data Mining Society.

Maria D. Chang is a Ph.D. candidate in computer science at Northwestern University. Her research interests include knowledge representation and reasoning, human-computer interaction, and the use of artificial intelligence in educational software. She is a member of the Qualitative Reasoning Group and the Spatial Intelligence and Learning Center.

Kenneth D. Forbus is the Walter P. Murphy Professor of Computer Science and a professor of education at Northwestern University. His research interests include qualitative reasoning, analogical reasoning and learning, spatial reasoning, sketch understanding, natural language understanding, cognitive architecture, reasoning system design, intelligent educational software, and the use of AI in interactive entertainment. He is a Fellow of AAAI, the Cognitive Science Society, and the Association for Computing Machinery.