Jericho McLeod

CSI-777 Response Paper: Visual Data Mining by Wegman

The first thing that was noteworthy was the amount of improvement in technology from 2003 to 2019. O(n^2) algorithms begin to fail around some threshold on off-the-shelf hardware, and the point Wegman chose as being just beyond this happens to be almost the exact size of the dataset Jay and I used in this class, with off-the-shelf hardware, in an O(n^2) algorithm.

I also notice that when he discusses X features in parallel coordinates, he also mentions that it could be an artefact of taking a single static view. His suggestion is to examine the same data in a Grand Tour visualization to ensure the same characteristics are present to determine whether there is a 5-dimensional hole in the dataset or not.

I would also suggest orienting axes; for data that is ordinal, interval, or ratio, the directionality of the sort should be consistent across dimensions. For nominal data this becomes more problematic to resolve, but also less important- an X feature across two nominal dimensions is meaningless as the order itself carries no information.

Having read through most of the paper, the most pressing question I have for Dr. Wegman is why use grand tour outside of parallel coordinates? Is there any particular reason to rotate through orthogonal dimensions between the true dimensions in the sample space? It seems that one would keep the same number of orthogonal dimensions, and the fact that we’re rotating through combinations of the original dimensions isn’t something that needs to be directly visualized. Or, alternatively, if they do need to be tracked, display a visualization based on the eigenvectors of the original dimensions to show the components of the current state.

![A close up of a logo

Description automatically generated]()

As an example, take the diagram above. D\_o1 is a label for the first of the orthogonal dimensions, and D1 is the first dimension of the original parallel coordinates plot. As the new visualization moves through all possible orthogonal dimensions from the dataset like a grand tour, the relationship to the original dimensions would be displayed at right by some color scheme. In my example, the first orthogonal dimension, D\_o1, is primarily related to D2, with some relationship with D1, and little with D3; this is shown by darker values in the grid on the right.

To make this a bit easier to understand it may also be useful to display the original dimensions in a parallel coordinates plot at the same time. I am undecided as to whether this should be orthogonal to the grand tour dimensions, but I also cannot immediately think of a way to show them in parallel while preserving the grid that shows the relationships between them.



An Introduction to Magic: the Gathering

Players each have a deck, or collection of 60 or more cards. These cards can be of several types:

Land Creature Instant

A screen shot of a person

Description automatically generated  

**Types of cards:**

**Lands:** Provide resources that players use to take certain other actions. This resource typically renews each turn.

**Creatures**: These are the primary means of interaction with other players. Creatures have a cost at the top-right corner of the card; this is paid by using Lands. Creatures also have attributes called “Power” and “Toughness”, shown as a ratio at the bottom right of the card (though it is not truly a ratio). Power indicates how much damage a creature can inflict, and toughness reflects how much damage it can sustain before it is removed from active play.

**Instants**: These are another way a player can interact with opponents and their creatures. Like creatures, Instants have a cost that is indicated at the top right. The action that is taken when this is played is indicated in text below the image of the card; in the example above it would allow a player to temporarily alter the power and toughness of a creature. These may be cast any time a player has ‘priority’.

**Sorceries**: (not shown) Similar to instants; the only difference is in when they can be played

**Enchantments**: (not shown) Similar to Instants and Sorceries, but these remain in play and are not simply temporary actions.

**Artifacts**: (not shown) Similar to enchantments

**Gameplay overview:**

1. Players select who will play first at random
2. Each player draws 7 cards from the top of their deck, which is shuffled and kept face down. Players then decide if they will keep them or ‘mulligan’; an action by which they can replace the initial cards, though at a cost for each additional mulligan
3. Each player begins with 20 ‘life points’
4. The first player takes their turn, then players alternate turns until one player reaches 0 life. The player who’s turn is current in progress is the ‘active’ player

**Turn Structure:**

1. Untap Step: costs that are paid using lands cause the land to be ‘tapped’, or rotated 90 degrees to indicate it has been used. Additionally, creatures that attacked an opponent on the prior turn are also tapped. During this step, all tapped items the active player controls are untapped.
2. Draw Step: the active player draws 1 card from the top of their deck
3. Upkeep Step: Relatively obsolete
4. 1st Main Phase: The active player can cast creatures, sorceries, enchantments, and artifacts during this phase.
5. Combat Phase: The active player may declare attackers. If they do, the defending player may block with any untapped creatures he or she controls. After attackers and blockers are declared, damage resolves; any unblocked creatures deal damage to the defending player. Any creatures that have taken fatal damage are moved to the graveyard.
6. 2nd Main Phase: Same as 1st main phase, but this allows players some freedom decisions around how combat resolves, bluffing, etc.
7. End of Turn: cleanup step- not terribly important for a basic understanding of the game.

**Priority:**

1. The active player has priority at the start of the turn.
2. At the end of each step, priority is passed to the opponent for instants, and then returns to the active player at the start of the next step or phase
3. Whenever a player with priority casts a spell, he or she may cast any number of additional instants.
4. When they are finished, priority passes to the opponent for casting their own spells.

**The Stack**: a list of things to be resolved

1. When a player casts a spell (such as an instant or a creature, but not a land), the spell goes on the stack
2. All subsequent spells cast at instant speed prior to the resolution of the stack are placed on the stack ‘on top’ of the existing items
3. When all players pass priority (meaning no further spells are cast), the stack resolves *from the top down-* the first thing cast resolves last.

As an example of the stack: if you in play a Grizzly Bear, the 2/2 creature shown above, and then you cast Giant Growth, also shown above, the Giant Growth goes on the stack. Your opponent can then cast Lightning Bolt, shown below, targeting the Grizzly Bear. Since this is above the Giant Growth on the stack, it resolves first, and the Grizzly Bear dies. Then, the Giant Growth resolves, but it no longer has a target, so it has no real effect.



Conversely, if I had cast Lightning Bolt (on the right) targetting Grizzly Bear first, and then you cast Giant Growth before my spell resolves, Grizzly Bear would become a 5/5 creature until the end of the turn. It would then take 3 damage, effectively becoming a 5/2, until the end of the turn.

Additional Notes:

* Many cards change the rules of the game. When in conflict, the card win
* Some cards have abilities that are either items with a cost (to be paid with Lands) or are triggered (take place whenever the specified condition occurs).
* Complete rules are available here: <https://magic.wizards.com/en/game-info/gameplay/rules-and-formats/rules>
* Note that the first set of example cards were all green, and the card above is red. There are 5 colors in Magic: the Gathering:
  + White
  + Blue
  + Black
  + Red
  + Green
* These are relevant to understanding differences between decks, but not necessarily the basics of the rules

**Relating Research to Rules:**

Our current methodology was unrelated to the rules of the game; we merely required sets of 60 objects to have specific names that acted as a clustering mechanism. We did, however, relate future research and some of our findings to things that required knowledge of the game. One of these is the use of a cost function. Now that you have a basic understanding of what the card types do it will make more sense to discuss change the measurements of distance between decks to one that treats each type of card differently. For instance, exchanging a land for a land is a change, while exchanging a land for a creature is a larger change.

Understanding how to create these costs requires an intimate understanding of how these changes impact gameplay. If we can describe how to encode knowledge into a cost function for the adjacency matrix prior to clustering, it allows this method to be useful in other applications. For instance, if we know that 400 level engineering classes are relatively interchangeable compared to 400 level classes in Music or Accounting, we can more effectively cluster education programs.

An additional point of interest going forward is the use of discriminators between decks. If we can identify some subset of cards that is sufficient to cluster them, we can save time and computational power on running this type of analysis. That may carry its own information, as well, especially toward naming clusters independently of their existing names altogether. This may also have broader applications: if there are discriminant courses between academic programs, they are the true differences between degrees.

Finally, further research needs to account for card selections that are not truly differences. For instance, the green creature I showed above was Grizzly Bears. Consider, though, the cards below. Barbary Apes has the same functional parameters as Grizzley Bears, while Terrain Elemental has slightly better parameters. Barbary Apes should be considered the same card as Grizzly Bears, and Terrain Elemental should have a reduced distance. Often, cards are chosen that are slightly worse, like choosing Grizzley Bears over Terrain Elemental, due to player budgets and the dollar value of cards. This needs to be encoded in the algorithm somehow.

 

The goal of this research is to abstract ways of encoding these types of a priori knowledge into the algorithm so that it is more effective and has broader applicability than merely a card game. However, given that this game is one of the only things in which I have decades of experience, my knowledge base is not easy to dismiss or to replicate in another field of study. This makes it an appealing starting point from which to research algorithms.