

Video-game data: test bed for data-mining and pattern mining problems

Mehdi Kaytoue

GT IA des jeux - GDR IA

December 6th, 2016



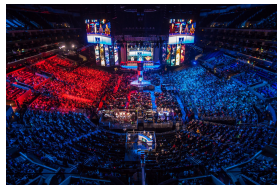
The video game industry

- Millions (billions!) of players worldwide,
- at any-time on any device

The rise of eSports and Streaming

- Teams and sponsors
- Twitch.tv and TVs

Challenge: games shall be hard for pros, enjoyable for casual players



G. Cheung and J. Huang.

Starcraft from the stands: understanding the game spectator.

In *SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 763–772.



M. Kaytoue, A. Silva, L. Cerf, W. Meira Jr. et C. Raïssi

Watch me playing, i am a professional: a first study on video game live streaming.

In *WWW 2012 (Companion Volume)*, pages 1181–1188. ACM, 2012.



T. L. Taylor

Raising the Stakes:E-Sports and the Professionalization of Computer Gaming.

In *MIT Press*, 2012.

Video game data, interesting benchmark for pattern mining

- 1 Discovering the habits and weaknesses of a MOBA player
- 2 Studying balance issues in RTS Games
- 3 Identifying players from game traces

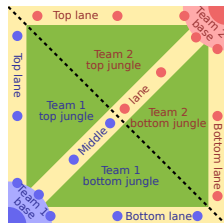
Discovering the habits and weaknesses of a MOBA player

Multi-player Online Battle Arena games

- For this talk: DOTA2
- 2 teams playing some kind of rugby
- Equilibrium gets easier to break with time
- Large heroes pool with different roles and style

Requires practice, knowledge ... and advice

- Positioning
- Build order, items
- Experience and gold rates
- Trigger/coordinate team fights, estimating enemy positions
- Micro management



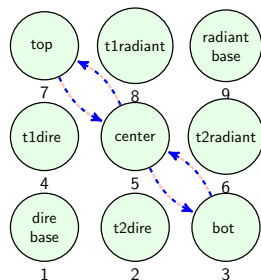
How can I learn from my mistakes?

Can I discover weaknesses from my enemy?

Computing a reference behavior graph for DOTA2

Principle

- Select a set of references player game traces
- Select a set of POIs (towers, shops,...)
- Compute the movement frequencies
- Filter out unfrequent edges
- Store the resulting graph



Leagues Of Legends



Mirana (DOTA2)



Pudge (DOTA2)

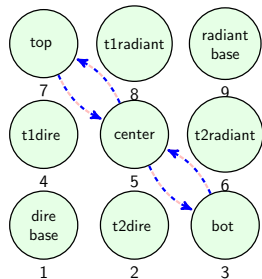
Computing the deviation from a reference model

pid	Trajectory a	Description	Description	Outlier Score	Victory?
1	$\langle 1, 4, 7, 5, 7, 5, 7 \rangle$	$\{buy_X, buy_Y\}$	$\{ab_{A_1}, ab_{B_2}\}$	0.33	yes
2	$\langle 1, 2, 3, 5, 3, 5, 3 \rangle$	$\{buy_X, buy_Y\}$	$\{ab_{A_1}, ab_{B_2}\}$	0.33	yes
3	$\langle 1, 5, 7, 5, 7, 5 \rangle$	$\{buy_X\}$	$\{ab_{A_1}, ab_{B_2}\}$	0.40	yes
4	$\langle 1, 2, 3, 5, 3, 6, 3 \rangle$	$\{buy_X, buy_Z\}$	$\{ab_{A_1}, ab_{C_2}\}$	0.66	no
5	$\langle 1, 2, 3, 5, 6, 3 \rangle$	$\{buy_Z\}$	$\{ab_{A_1}, ab_{C_2}\}$	0.80	no

Given a trace t and a Reference Model matrix representation M , the outlier score is defined as:

$$\mu(t, M) = \frac{\sum_{i=0}^{|trajectory(t)|-1} M(t_i, t_{i+1})}{|trajectory(t)| - 1}$$

where $|\cdot|$ counts the number of POIs



Mining emerging patterns

$$\mathcal{D}^+ = \{description(t) \mid t \in \mathcal{T}, \mu(t, M) \leq \theta\}$$

$$\mathcal{D}^- = \{description(t) \mid t \in \mathcal{T}, \mu(t, M) > \theta\}$$

$$\phi(X) = \frac{|supp_{\mathcal{D}^+}(X)| - |supp_{\mathcal{D}^-}(X)|}{|supp_{\mathcal{D}^+}(X)| + |supp_{\mathcal{D}^-}(X)|}$$

pid	Description	Description	class
1	{buy _X , buy _Y }	{ab _{A₁} , ab _{B₂} }	+
2	{buy _X , buy _Y }	{ab _{A₁} , ab _{B₂} }	+
3	{buy _X }	{ab _{A₁} , ab _{B₂} }	+
4	{buy _X , buy _Z }	{ab _{A₁} , ab _{C₂} }	-
5	red{buy _Z }	{ab _{A₁} , ab _{C₂} }	-

Example

With $\theta = 0.5$: $\mathcal{D}^+ = \{d(t_1), d(t_2), d(t_3)\}$ and $\mathcal{D}^- = \{d(t_4), d(t_5)\}$. With $min_sup = 2$, $X_1 = \{buy_X\}$, $X_2 = \{buy_Z\}$, $X_3 = \{buy_X, buy_Y\}$ are frequent

$$\phi(\{buy_X\}) = (3 - 1)/(3 + 1) = 0.5$$

$$\phi(\{buy_Z\}) = (0 - 2)/(0 + 2) = -1$$

$$\phi(\{buy_X, buy_Y\}) = (2 - 0)/(2 + 0) = 1$$

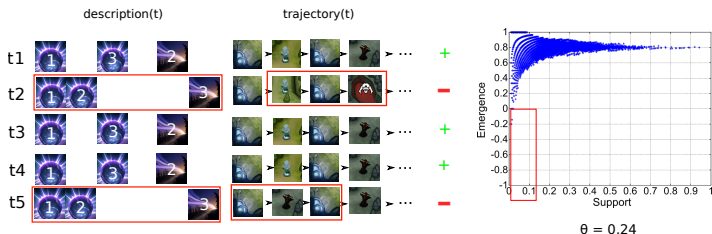
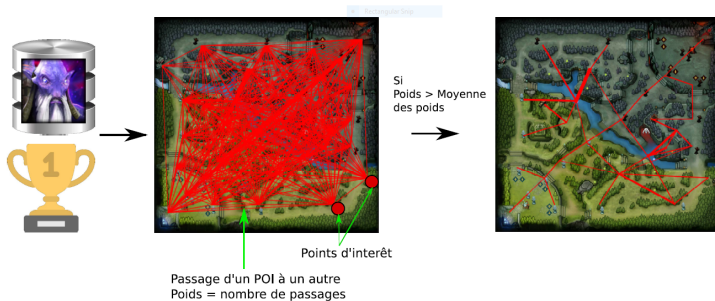


G. Dong, J. Li

Efficient mining of emerging patterns: discovering trends and differences.

KDD 1999

Scenario



- 1 Discovering the habits and weaknesses of a MOBA player
- 2 Studying balance issues in RTS Games
- 3 Identifying players from game traces

RTS Games: A lot of challenges.

StarCraft 2

- Two players are battling against each other on a map
- Each chooses a faction (Zerg, Terran, Protoss)
- Goal: use units to gather resources, to create buildings that can produce units ... establish a strategy (choose the right buildings and army composition) to destroy your opponent.

Security issues, Bugs, cheaters, **Balance issues**, Fun vs challenging agents, Profiling & prediction, Match preparation, Playground for AI research Deepmind vs FAIR, **Discover strategies automatically from a large set of games ; Evaluate their capacity to win**



S. Ontanon, G. Synnaeve, A. Uriarte, F. Richoux, D. Churchill, and M. Preuss,

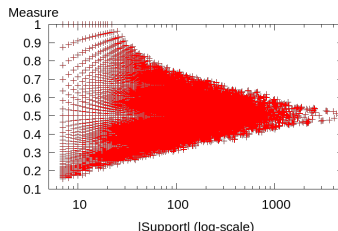
A survey of real-time strategy game ai research and competition in starcraft.

Computational Intelligence and AI in Games, IEEE Transactions on, vol. 5, no. 4, pp. 293–311, 2013.

Encoding data & Mining strategic patterns

Sequence
$\langle (a, +) \{ (b, +) (c, +) (c, -) \} \{ (a, -) (d, +) \} (b, -) \rangle$
$\langle (a, +) \{ (b, +) (c, +) (d, +) \} \{ (b, -) (c, -) \} (d, -) \rangle$

$$balance(s) = \frac{|support_{\mathcal{D}}(s)|}{|support_{\mathcal{D}}(s)| + |support_{\mathcal{D}}(\tilde{s})|}$$



- Symmetric axis: $y = 0.5$
- Non perfect symmetry: if a sequence s is frequent, it does not imply that \tilde{s} is frequent too
- Pattern with highest support: well-known strategies (balanced)

Example of discovered patterns

PvZ - A well-known opening (Forge-Expand)

- $s = \langle \{(Nexus, 5, +)\} \{ (Gateway, 6, +) (PhotonCannon, 6, +) \} \rangle - balance(s) = 0.52$
- $s = \langle \{(Nexus, 5, +)\} \{ (PhotonCannon, 6, +) (Assimilator, 6, +) \} \rangle - balance(s) = 0.52$

A balance issue (TvZ - Bunker rush)

$s = \langle \{(Barracks, 1, S, 1)\}, \{(SpPool, 4, F, 1)\}, \{(Bunker, 6, S, 1), (SpCrawler, 6, F, 1)\} \rangle$ $balance(s) = 0.61$

Corrected in May 2012 by the game editor (a Zerg counter unit as been slightly improved and bunker timing is longer).

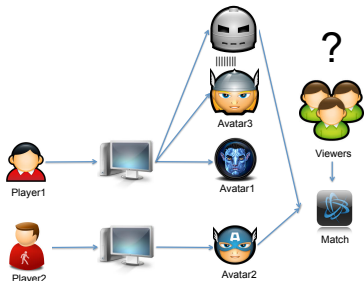
<https://github.com/guillaume-bosc/BalanceSpan>

- 1 Discovering the habits and weaknesses of a MOBA player
- 2 Studying balance issues in RTS Games
- 3 Identifying players from game traces

The problem

Players and teams observe game records of others

- Complete game logs are available, rankings (ATP)



순위	플레이어
상위 16	
1번째	플레이어
2번째	플레이어
3번째	플레이어
4번째	플레이어
5번째	[토스는종말이] lifeisrisk
6번째	플레이어
7번째	플레이어
8번째	플레이어
9번째	플레이어
10번째	플레이어
11번째	플레이어
12번째	플레이어
13번째	플레이어
14번째	플레이어
15번째	플레이어
16번째	[Imp] yoeFWLeenock

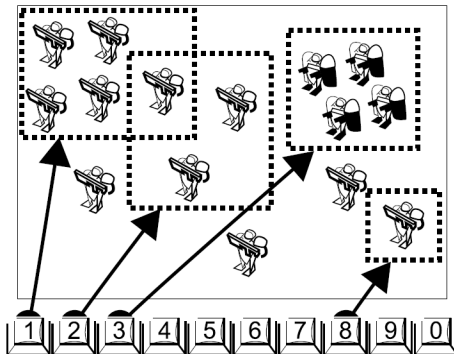
<http://leagueoflegends.wikia.com/wiki/Smurf>

https://www.reddit.com/r/starcraft/comments/3gkfs0/sc2_who_is_that_smurf/

Behavioural data as replay files

The RTS game StarCraft 2: to improve strategy execution, players

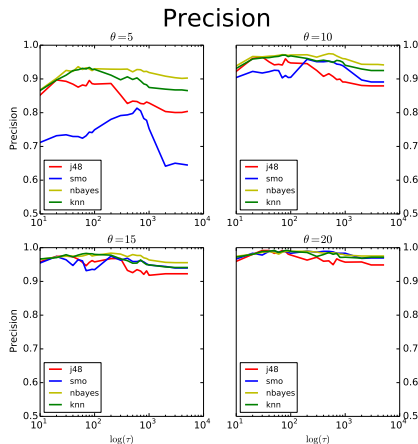
- assign control groups to units and buildings,
- bind them to keyboard hotkeys (1, 2, ..., 9, 0),
- use them intensively along with the mouse.



Source: Yan et al., SIGCHI2015

Avatar	Game trace	Outcome
Ror0	s,s,hotkey4a,s,hotkey3a,s,hotkey3s, ...	Lose
TAiLS	Base,hotkey1a,s,hotkey1s,s,hotkey1s, ...	Win

Predictive models with high accuracy



Hotkeys hide unique patterns

- 20 first seconds of the game are enough
- 20 games are enough

We found a similar result, but considering on purpose dataset without avatar aliases, since precision drastically drops



Eddie Q. Yan, Jeff Huang, Gifford K. Cheung.

Masters of Control: Behavioral Patterns of Simultaneous Unit Group Manipulation in StarCraft2.

In *CHI 2015, Crossings, Seoul, Korea 37–11*, 2015.

Unscrambling models to identify aliases

Idea: two avatars of the same player should draw a high confusion

	l_1	l_2	l_3	l_4	l_5
l_1	0.6	0.4	0	0	0
l_2	0.4	0.55	0.05	0	0
l_3	0	0	0.8	0.15	0.05
l_4	0	0.05	0	0.7	0.25
l_5	0	0	0	0.5	0.5

We are searching for pairs of labels that concentrate the fusion

- $\tilde{C}_{ij}^p \simeq \tilde{C}_{ji}^p \simeq \tilde{C}_{ii}^p \simeq \tilde{C}_{jj}^p$
- $\tilde{C}_{ij}^p + \tilde{C}_{ji}^p + \tilde{C}_{ii}^p + \tilde{C}_{jj}^p \simeq 2$

	...	l_i	l_j	...
...	...			
l_i	...	$C_{i,i}$	$C_{i,j}$...
l_j	...	$C_{j,i}$	$C_{j,j}$...
...	...			

We proposed a method in (fuzzy) formal concept analysis that highlight good results when comparing to a ground truth

Conclusion

Supervised pattern discovery

- Language (itemset, sequence, graphs, ...)
- Quality measure: deviation between the models induced by (i) the objects covered by the pattern, (ii) the full dataset
- Algorithm: exhaustive search (constraint pattern mining, upper bounds) vs heuristic search (beam search, pattern sampling, MCTS)
- Anytime pattern mining and expert knowledge incorporation
- Expert preference learning

JF Boulicaut, Marc Plantevit, Céline Robardet (LIRIS Lyon)

but also, Amedeo Napoli, Chedy Raïssi (INRIA Nancy)

Bruno Crémilleux, François Rioult, Albrecht Zimmerman (GREYC Caen),
and many others

Logistic data generator (EPCIS Protocol)

Game data


- Easily available (to some extent)
- Many industrial problems transferable into video games





[https://github.com/AnesBendimerad/
EPCIS-Events-Generator-Based-On-OpenTTD](https://github.com/AnesBendimerad/EPCIS-Events-Generator-Based-On-OpenTTD)

A few references


Avatar prediction and “smurf” detection in StaCraft II

-  O. Cavadenti, V. Codocedo, J.-F. Boulicaut, M. Kaytoue
When Cyberathletes Conceal Their Game:
Clustering Confusion Matrices to Identify Avatar Aliases.
[IEEE DSAA 2015](#)

Discovering and describing balance issues in StaCraft II

-  G. Bosc, C. Raïssi, J.-F. Boulicaut, P. Tan, M. Kaytoue
A Pattern Mining Approach to Study Strategy Balance in RTS Games
[IEEE Transactions on Computational Games and Artificial Intelligence \(in press\)](#).
-  C. Low-Kam, C. Raïssi, M. Kaytoue, J. Pei
Mining Statistically Significant Sequential Patterns.
[International Conference on Data Mining \(ICDM\) 2013](#).

Discovering and understanding deviant mobility behaviors

-  O. Cavadenti, V. Codocedo, J.-F. Boulicaut, M. Kaytoue
What did I do Wrong in my MOBA Game?:
Mining Patterns Discriminating Deviant Behaviours.
[IEEE DSAA 2016](#)