Analyzing NHL Goalie Stats (03-04—07-08) Using the Self-Organizing Map

By: Chuck Crittenden

"In hockey, goaltending is 75 percent of the game. Unless it's bad goaltending. Then it's 100 percent of the game, because you're going to lose."

~ Gene Ubriaco (NHL forward)

Overview

- Previous Problem
- Data
- Algorithm
- Self-Organizing Map

Overview

- Specific Maps
- Alternate Paths
- Conclusion
- Extensions

Previous Problem

- NHL Goaltending Statistics by Team
  - (03-04 through 07-08)
- Average Standings for each Team
- Use Self-Organizing Map
  - Find natural clusters

Previous Problem

- Stats – GAA, SV %, GA, GF, DIFF
- Standings and Levels

	overall standings	
Detroit	1.25	high
San Jose	5.75	high
New Jersey	7	high
Ottawa	7.25	high
Dallas	7.75	high
Buffalo	10.5	medhigh
Anaheim	10.75	medhigh
Nashville	10.75	medhigh
Calgary	11.75	medhigh
Colorado	12	medhigh
Montreal	12.5	medhigh
Philadelphia	14	med
Vancouver	14	med
New York R	14.5	med
Minnesota	14.75	med
Carolina	15.25	med
Tampa Bay	15.75	med
Toronto	16	med
Boston	16.25	med
Pittsburgh	18	medlow
Edmonton	18.75	medlow
Atlanta	20	medlow
New York I	20.25	medlow
Florida	22	medlow
St. Louis	23.5	low
Washington	23.5	low
Los Angeles	24.5	low
Columbus	25.5	low
Phoenix	25.5	low
Chicago	25.75	low

The Result

- 15x15 Map

Pittsburgh	x	x	x	Carolina	x	x	x	x	LosAngeles	x	x	StLouis	x	Chicago
x	x	Toronto	x	x	x	Florida	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	NewYorkI	x	Washington	x	x	x	x
Vancouver	x	x	TampaBay	x	x	x	x	x	x	x	x	x	Phoenix	x
x	x	x	x	x	x	x	x	x	Atlanta	x	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NewJersey	x	x	Calgary	x	x	x	x	x	x	x	x	x	Boston	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
x	Colorado	x	x	x	x	x	Montreal	x	x	Edmonton	x	x	x	x
x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Buffalo	x	x	x	x	x	x	x	x	x	x	x	x	x	Philadelphia
x	x	x	x	x	x	Anaheim	x	x	x	x	x	x	x	x
x	Ottawa	x	x	x	x	x	x	x	Minnesota	x	x	x	x	x
x	x	x	SanJose	x	Dallas	x	x	x	x	x	x	x	x	x
Detroit	x	x	x	x	x	x	Nashville	x	x	x	x	x	NewYorkR	x

The Data

TEAM	GAA	SV%	GA	GF	DIFF
Anaheim	2.4	0.915	199	222	22.8
Atlanta	2.98	0.903	247	234	-13
Boston	2.77	0.91	230	213	-17
Buffalo	2.74	0.906	227	261	34.8
Calgary	2.45	0.912	202	230	28
Carolina	2.76	0.902	228	231	3.25
Chicago	2.98	0.898	246	207	-40
Colorado	2.66	0.906	220	252	32.3
Columbus	2.83	0.906	234	195	-40
Dallas	2.3	0.907	191	225	34.5
Detroit	2.26	0.91	187	265	78.5
Edmonton	2.75	0.901	228	221	-7.3
Florida	2.77	0.915	230	220	-9.5
Los Angeles	3	0.898	247	224	-23
Minnesota	2.33	0.92	194	215	20.8
Montreal	2.66	0.913	220	236	16
Nashville	2.56	0.915	211	241	29.5
New Jersey	2.27	0.917	189	213	24
New York I	2.82	0.906	233	222	-12
New York R	2.56	0.908	212	224	11.8
Ottawa	2.53	0.91	209	280	70.5
Philadelphia	2.84	0.904	235	238	2.5
Phoenix	2.88	0.901	248	213	-35
Pittsburgh	2.88	0.908	238	248	9.25
San Jose	2.36	0.909	196	239	43.5
St. Louis	2.9	0.898	239	201	-38
Tampa Bay	2.83	0.895	233	239	5.5
Toronto	2.84	0.898	240	245	4.75
Vancouver	2.52	0.912	209	228	19.3
Washington	3.08	0.902	255	222	-33

	overall standings	
Detroit	1.25	high
San Jose	5.75	high
New Jersey	7	high
Ottawa	7.25	high
Dallas	7.75	high
Buffalo	10.5	medhigh
Anaheim	10.75	medhigh
Nashville	10.75	medhigh
Calgary	11.75	medhigh
Colorado	12	medhigh
Montreal	12.5	medhigh
Philadelphia	14	med
Vancouver	14	med
New York R	14.5	med
Minnesota	14.75	med
Carolina	15.25	med
Tampa Bay	15.75	med
Toronto	16	med
Boston	16.25	med
Pittsburgh	18	medlow
Edmonton	18.75	medlow
Atlanta	20	medlow
New York I	20.25	medlow
Florida	22	medlow
St. Louis	23.5	low
Washington	23.5	low
Los Angeles	24.5	low
Columbus	25.5	low
Phoenix	25.5	low
Chicago	25.75	low

Data

- GAA – Goals Against Average

$$\frac{\text{Goals Allowed}}{\text{Number of Minutes Played}(1/60)}$$

- SV% – Save Percentage

$$\frac{\text{Goals Allowed}}{\text{Shots Allowed}}$$

- GA – Goals Allowed
- GF – Goals Scored
- DIFF – Goal Differential

DIFF = Goals Scored – Goals Allowed

The Algorithm

- Self-Organizing Map (SOM)
  - Artifical Neural Network
- Clusters in 2-dimensional map

What is Needed?

- A .bat file containing the reference to the executables and the specifics of the map.
- The executables randomly initialize, run the algorithm, and calibrate the label onto the points.

```
randinit -din nhl.dat -cout nhl.cod -xdim 15 -ydim 15 -topol rect -neigh bubble -rand 0
vsom -din nhl.dat -cin nhl.cod -cout nhl.cod -rlen 10000 -alpha 0.05 -radius 15
vsom -din nhl.dat -cin nhl.cod -cout nhl.cod -rlen 1000000 -alpha 0.02 -radius 5
vcal -din nhl_label.dat -cin nhl.cod -cout nhl_label.cod
```

- som_mapper.exe

Initial Map

- Randomly intialized.
- Each team (p) compared to each point on the map (q) with Euclidean distance.

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

- Whichever point the specific team is closest to.
  - That point is trained accordingly.
  - Other points around it are also trained, just not as much.

SOM

- Process repeats for a set number of times.
- The labels are pasted on to each instance.
- The Map is made.

Team-Specific Maps

- Using only randinit and vsom
- Use a specific team's data only
  - Use vcal to attach the labels of each season
- Allows monitoring of team's progress

Boston Bruins

03-04	Boston	2.2	0.919	184	209	25
05-06	Boston	3.04	0.904	252	228	-24
06-07	Boston	3.32	0.9	275	210	-65
07-08	Boston	2.52	0.916	209	206	-3
08-09	Boston	1.97	0.935	54	92	38

Point Totals

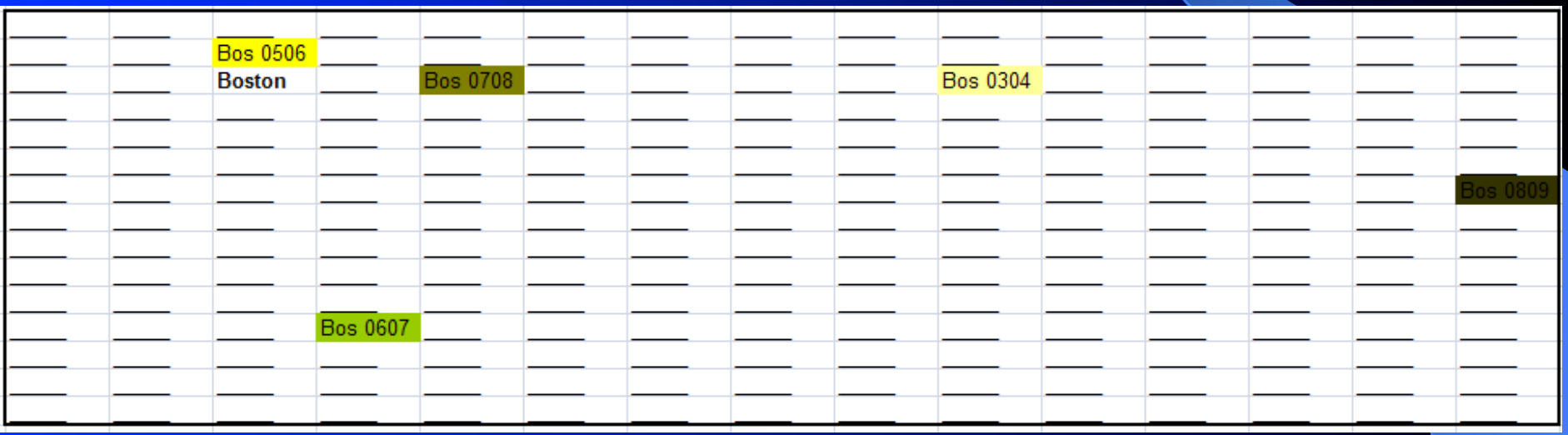
03-04 104

05-06 74

06-07 76

07-08 94

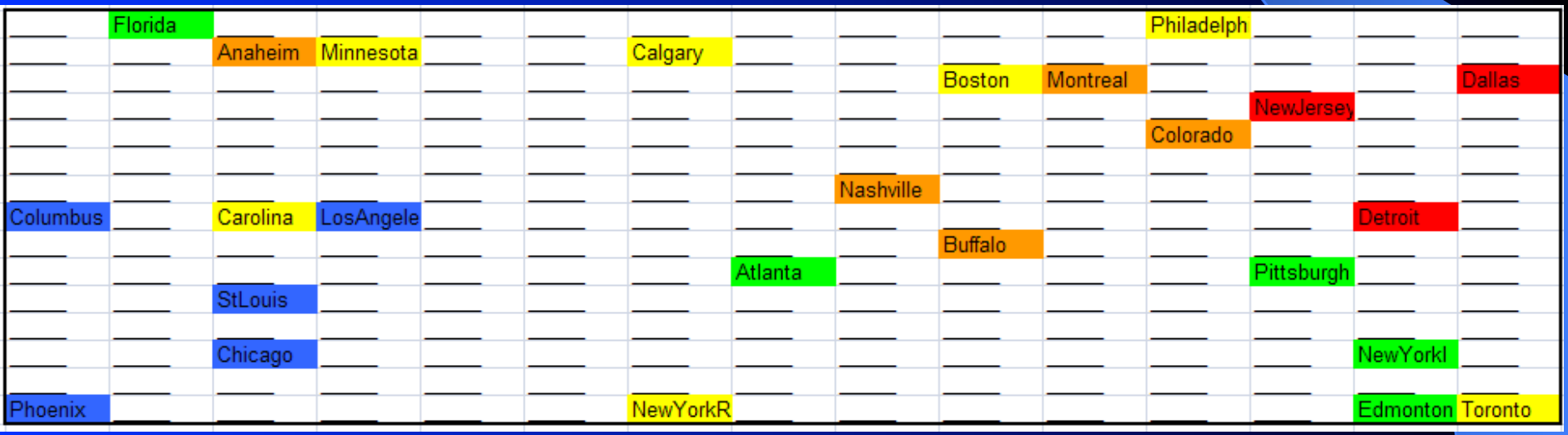
Boston’s Map



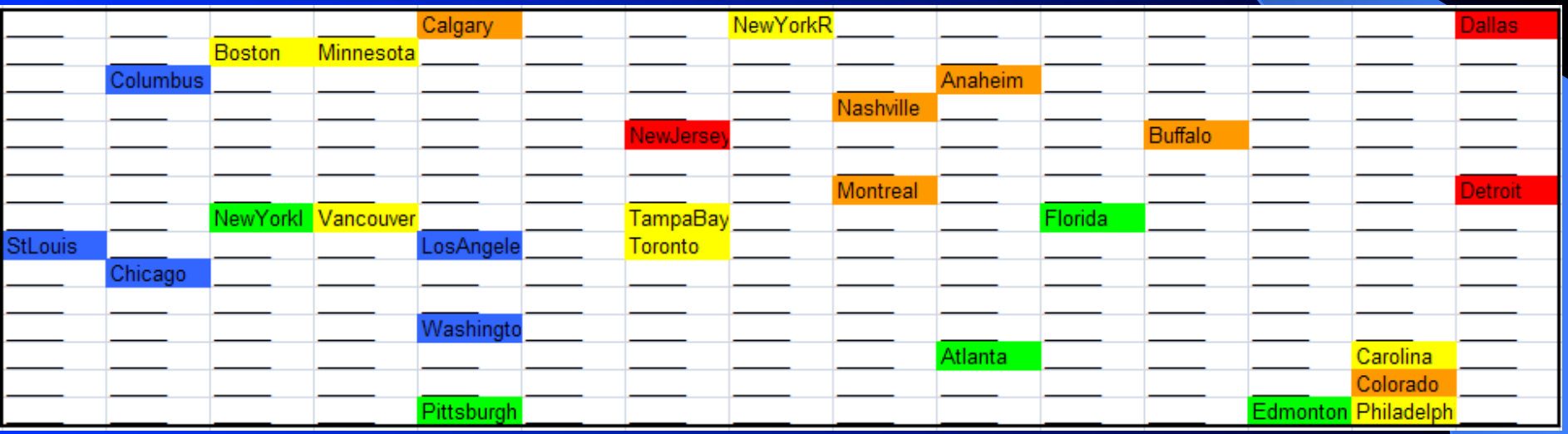
Year-Specific Maps

- Using only randinit and vsom
- Use a specific season's data only
  - Use vcal to attach the labels of each team
- Allows monitoring of every team's performance when maps put consecutively

2003-2004 Map



2005-2006 Map



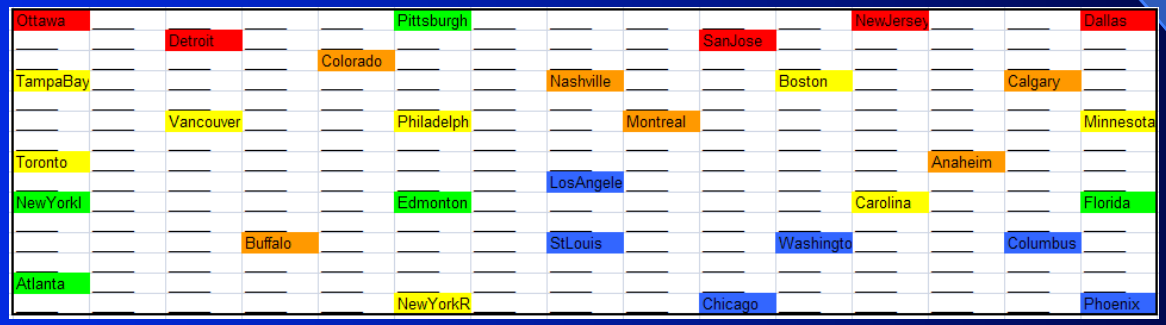
Alternate Means

- Rather than use same map as base
- Use a seed for the randomization process
  - In theory will force better teams into the same section for all maps

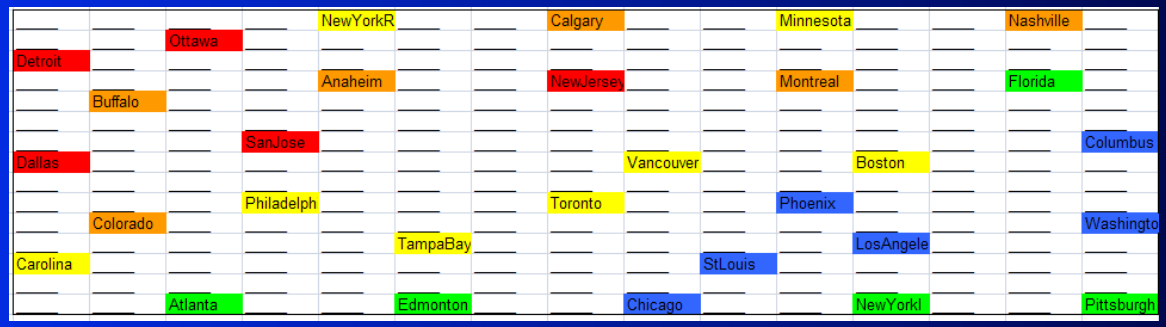
Randomization

- Didn't work out as planned.

03-04



05-06



Conclusion

- In SOM using a map with all of the data is superior to a seed
  - Assuming data is representative
- Is possible to monitor team's progression

Extensions

- This same idea can be used to track a single goalie
  - Removing GA, GF, and DIFF
  - Using only their data matched against all of the data in the league
- Compare two or more teams in separate years
- Use more attributes to compare individual players

Summary

- Previous Problem
- Data
- Algorithm
- Self-Organizing Map

Summary

- Specific Maps
- Alternate Paths
- Conclusion
- Extension

Sources

Aleshunas, John. Retrieved Apr. 17, 2008. "Self-Organizing Map (SOM)" from:
http://mercury.webster.edu/aleshunas/MATH%203210/MATH%203210%20Source%20Code%20and%20Executables.html

Aleshunas, John. Retrieved Dec. 9, 2008. "Crittenden – NHL Goalie SOM" from:
http://mercury.webster.edu/aleshunas/Support%20Materials/SOM/Crittenden%20-%20NHL%20Goalie%20SOM.doc

Goaltender's Annex. Retrieved May 5, 2008. Ubriaco Quote from:
http://www.angelfire.com/sk/goalieannex/quotes02.html

NHL.com. Retrieved Apr. 16, 2008. "Goalie Statistics and Team Standings" from:
http://www.nhl.com/nhlstats/app

Yahoo Sports. Retrieved Apr. 16, 2008. "Goalie Statistics and Team Standings" from:
http://sports.yahoo.com/nhl/teams/___/stats (Replace ___ with each team's abbreviation).

Wikipedia. Retrieved Apr. 17 2008. "Stepping through the Algorithm" from:
http://en.wikipedia.org/wiki/Self-organizing_map - Stepping_through_the_algorithm

Wikipedia. Retrieved May 6, 2008. "Euclidean Distance" from:
http://en.wikipedia.org/wiki/Euclidean_distance