Prediction of Significant Price Changes in Magic: The Gathering Cards

Matthew Pawlicki, Joseph Polin, Jesse Zhang

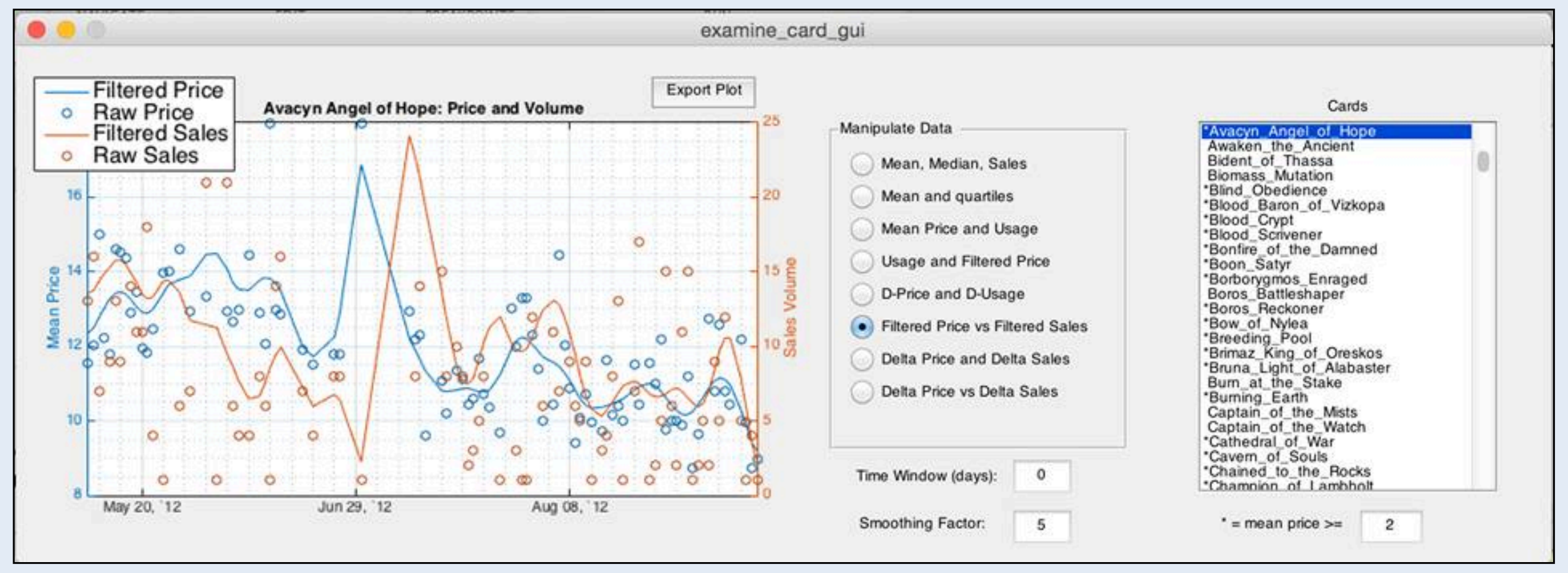
Motivation

Magic: The Gathering is a popular trading card game with roots going back to 1993[4], and the growing MTG community is currently composed of over 12 million players. The prices of individual cards fluctuate daily with the printing of new sets of cards and the evolving competitive tournament scene. Commonly played cards can sell for upwards for $50 per copy. The goal of this study was to find a correlation between carefully chosen features of a card and the change in its mean price over the course of the upcoming week.

Data

Due to limitations in the dataset, only dates after May 2012 were processed.

- **Price and daily sales history**: obtained from curator of *mtgprice.com*, a website that processes internet price/ inventory data for every MTG card[2]
- **Card popularity**: obtained by data-mining online sources that list all cards used in tournament winning decks
- **Card attributes**: obtained from a card database API



examine_card_gui

Export Plot

Manipulate Data

○ Mean, Median, Sales
○ Mean and quartiles
○ Mean Price and Usage
○ Usage and Filtered Price
○ D-Price and D-Usage
● Filtered Price vs Filtered Sales
○ Delta Price and Delta Sales
○ Delta Price vs Delta Sales

Time Window (days): 0

Smoothing Factor: 5

Cards

*Avacyn_Angel_of_Hope
Awaken_the_Ancient
Bident_of_Thassa
Biomass_Mutation
*Blind_Obedience
*Blood_Baron_of_Vizkopa
*Blood_Crypt
*Blood_Scrivener
*Bonfire_of_the_Damned
*Boon_Satyr
*Borborygmos_Enraged
Boros_Battleshaper
*Boros_Reckoner
*Bow_of_Nylea
*Breeding_Pool
*Brimaz_King_of_Oreskos
*Bruna_Light_of_Alabaster
Burn_at_the_Stake
*Burning_Earth
Captain_of_the_Mists
Captain_of_the_Watch
*Cathedral_of_War
*Cavern_of_Souls
*Chained_to_the_Rocks
*Champion_of_Lambholt

* = mean price >= 2

Figure 1. MATLAB GUI for visualizing temporal card data

An example labeling scheme was:

$$\text{Label 1 if } \mu_d^{(i)} = \frac{1}{7}\sum_{i=1}^{7} P_{d+i}^{(i)} \geq P_d + 0.5,\ \text{0 otherwise}$$

Features

The temporal data was processed to produce differences in sales, price, and popularity over time along with price history and specific card attributes. Each data point represented a specific card on a specific date and contained the following 28 features:

Index	Feature	Description
1-7	$P_d \;\; P_{d-1} \cdots \;\; P_{d-6}$	Prices on day *d* through *d-6*
8-13	$(P_d - P_{d-1}) \;\; (P_d - P_{d-2}) \;\; \ldots \;\; (P_d - P_{d-6})$	Changes in prices
14-19	$(U_d - U_{d-1}) \;\; (U_d - U_{d-2}) \;\; \ldots \;\; (U_d - U_{d-6})$	Changes in popularities
20-25	$(S_d - S_{d-1}) \;\; (S_d - S_{d-2}) \;\; \ldots \;\; (S_d - S_{d-6})$	Changes in sales
26	M	Card mana cost
27	R_d	Days until card becomes obsolete
28	σ_d^2	Variance of past week's prices

Support Vector Machine with L_2 Regularization[3]

Let $x^{(i)} \in \Re^{n\times 1}$ represent the i^{th} sample, which has *n* features, and $\theta = [\, b \;\; w^T \,]^T \in \Re^{(n+1)\times 1}$ be the vector of parameters.

$$h(\theta^T x^{(i)}) = 1\left\{ w^T x^{(i)} + b > 0 \right\}$$

$$w = \min_w \; \frac{1}{2}\| w \|^2 + C\sum_{i=1}^{m}\left(\max(0, 1 - y^{(i)}\left\langle w, x^{(i)} \right\rangle) \right)^2$$

L_2 Regularized Logistic Regression (LR)[1]

Here, $x^{(i)} \in \Re^{(n+1)\times 1}$ because an intercept term is appended to each training sample.

$$\theta = \operatorname*{argmax}_\theta \sum_{i=1}^{m} \log p(y^{(i)} \mid x^{(i)};\theta) + \lambda \| \theta \|_2^2$$

$$p(y^{(i)} \mid x^{(i)};\theta) = h(\theta^T x^{(i)})^{y^{(i)}} (1 - h(\theta^T x^{(i)}))^{1-y^{(i)}} \qquad h(\theta^T x^{(i)}) = 1\left\{ \frac{1}{1+e^{-\theta^T x^{(i)}}} > 0.5 \right\}$$

Solve for θ using gradient descent by iterating the following:

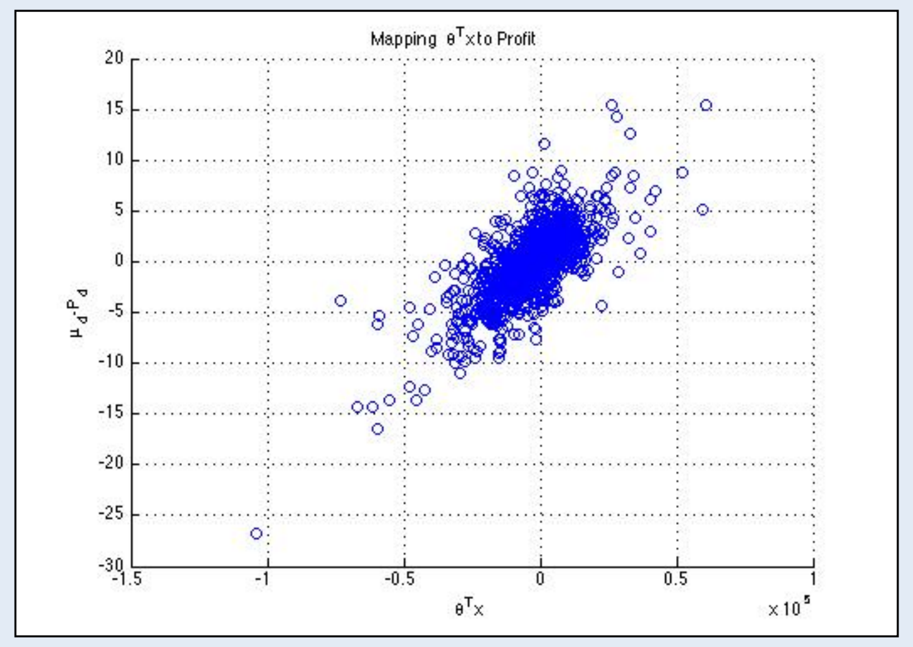
$$\theta_0 : \text{increment by } -\alpha \frac{1}{m}\sum_{i=1}^{m}(h(\theta^T x^{(i)}) - y^{(i)})x_0^{(i)}$$

$$\theta_j : \text{increment by } -\alpha \left[\frac{1}{m}\sum_{i=1}^{m}(h(\theta^T x^{(i)}) - y^{(i)})x_j^{(i)} + \frac{\lambda}{m}\theta_j \right]$$

References

1. Andrew Ng. CS 229 Class Lecture, Topic: "Logistic Regression." Stanford University, 2014.
2. "MTGPrice.com - Magic: The Gathering Value and Price Guide for Ebay, Amazon and Hobby Stores!" *Magic: The Gathering Value and Price Guide*. N.p., n.d. Web. 07 Dec. 2014.
3. R.E. FAN et al. 2008. "LIBLINEAR: A library for large linear classification." *The Journal of Machine Learning Research.*
4. Y. LeJacq. 2013. "At 20, 'Magic: The Gathering' still going strong." *NBC.com.*

Results

	Training (m = 10213)	Testing (m = 3405)
SVM	Profit ratio: 0.85; E = 15%	Profit ratio: 0.85; E = 15% Confusion Matrix: $\begin{bmatrix} 199 & 337 \\ 163 & 2706 \end{bmatrix}$
LR	Profit ratio: 0.88; E = 12%  Figure 2. Price increase v. $\theta^T x$	Profit ratio: 0.93; E = 11% Confusion Matrix: $\begin{bmatrix} 252 & 255 \\ 106 & 2792 \end{bmatrix}$

Discussion

The quality of a classification model is evaluated by the ratio of captured profit to maximum possible profit on a training set:

$$profit(\theta) = \frac{1}{p_{\max}} \sum_{i=1}^{m} h(\theta^T x^{(i)})\left(\mu_d^{(i)} - P_d^{(i)}\right)$$

$$\text{where } profit_{\max} = \sum_{i=1}^{m} 1\left\{ \mu_d^{(i)} - P_d^{(i)} > 0 \right\}\left(\mu_d^{(i)} - P_d^{(i)}\right)$$

When looking at individual features, the features that generated the highest profit per buy and the highest true-positive to false-positive ratio were (U_d-U_{d-3}) and (S_d-S_{d-2}), respectively. As expected, the change in price of a card were most heavily correlated with price, popularity, and sales changes one to three days beforehand. LR performed better than SVM because LR factors in a degree of "confidence," which correlates to price difference (as seen in Figure 2).

Future Work

There are many types of metrics with which to measure the quality of the classification (e.g. profit per buy, true-positive to false-positive ratio). Recasting these metrics as convex optimization problems and maximizing them would result in a model that prioritizes generating profit over simply being accurate.