Using Author Types to Predict Review Ratings

Julian Chan, Laurel Hart, and Ruth Morrison

Goal

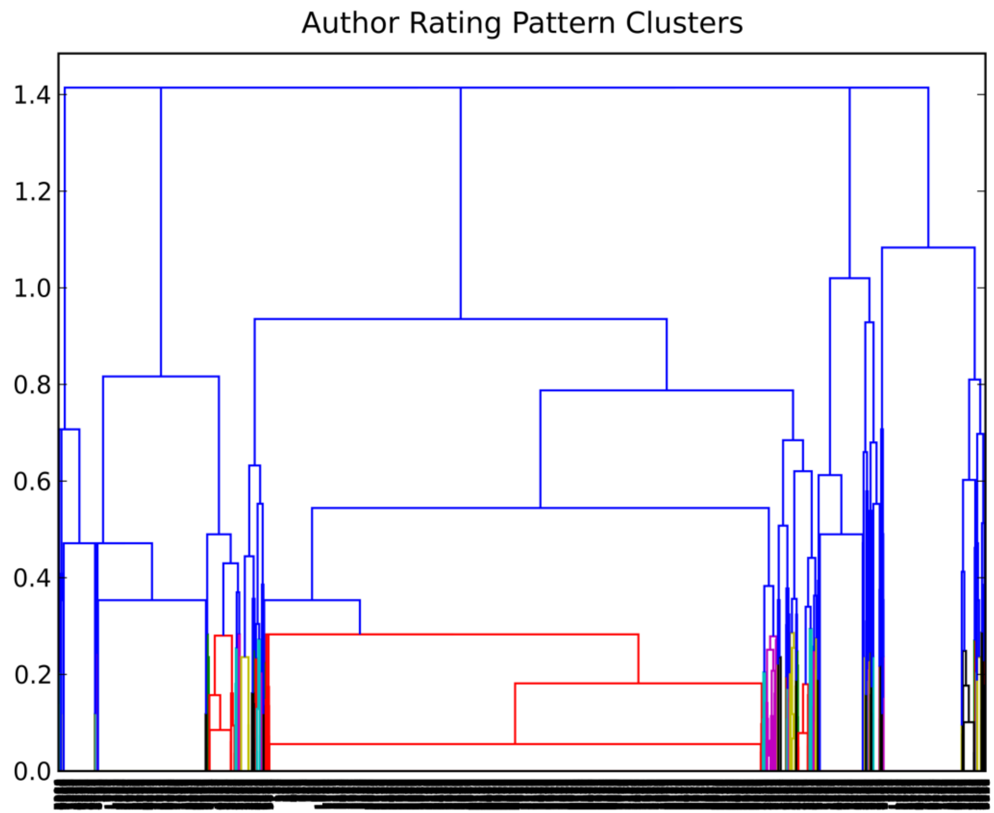
- Predict rating of review based on review text
- Intuition: “dogs of the same street bark alike” -- authors with similar styles will rate similarly
- Amazon review corpus (Bing Liu et. al)
- Mallet for classification (MaxEnt classifier)

Features

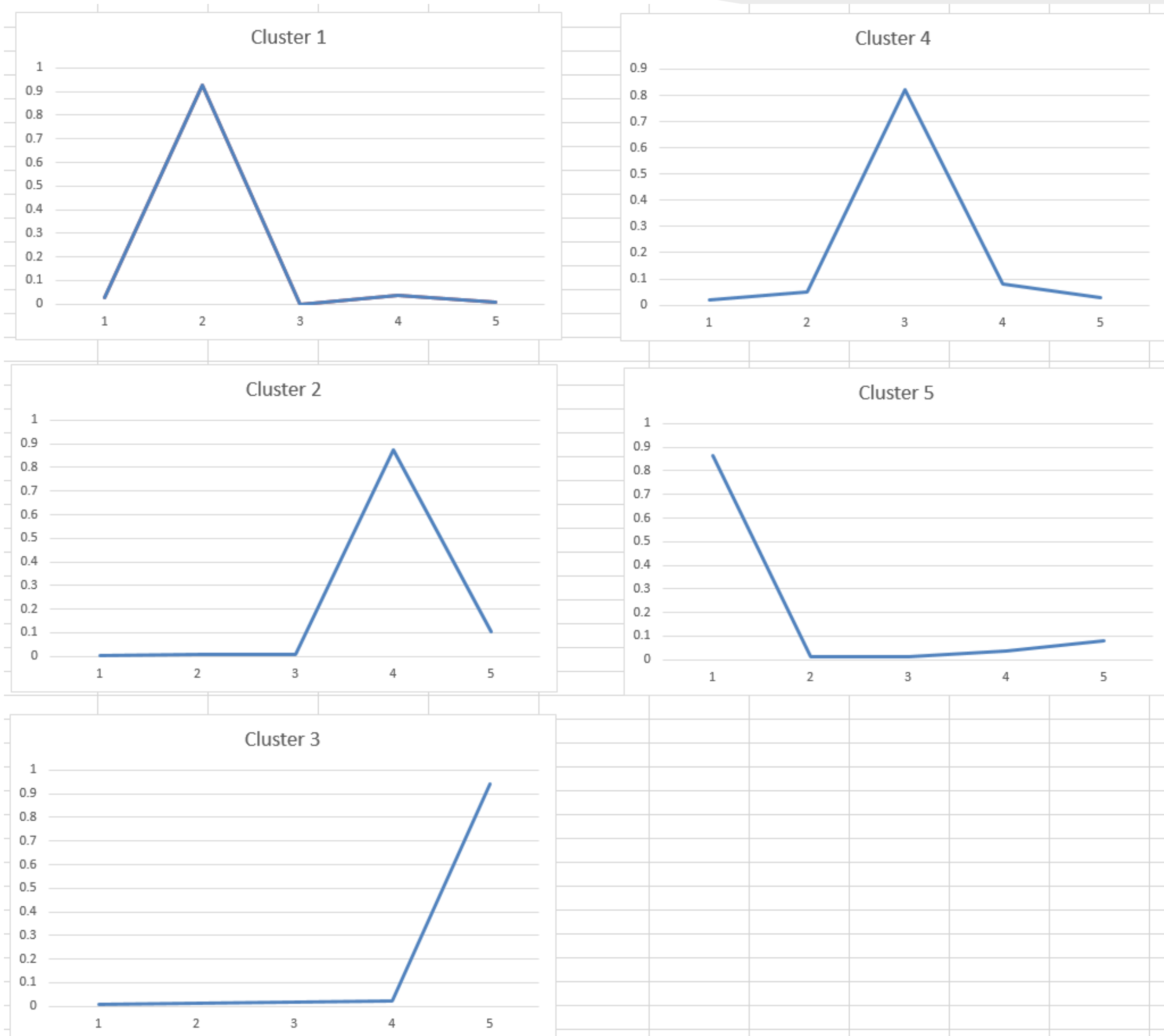
- N-grams
  - unigrams, bigrams, trigrams, 4-grams, and 5-grams
  - top discriminating n-grams
- Author profile
  - Previous rating behaviors
- Stylistic features
  - Review length, negation, readability
- Miscellaneous
  - product type/genre path

Author Rating Pattern Clustering

- Each author represented by a 5-dimensional vector.
- Hierarchical clustering from 10000 author samples.
- Cosine distance between author vectors



Five Clusters



Ten Clusters

Evaluation

- Strict accuracy is not that informative.
- Credit should be given to a close guess.
- Wildly inaccurate guesses should be penalized more harshly.

- Solution: Mean Squared Error

Using Five-Cluster Author Type as Feature

AllBigrams

	1	2	3	4	5	Total Squared Error	Instances	MSE
1	39647	2613	2715	2834	48005	807059	95814	8.423184503
2	11912	4569	7976	6798	31807	333343	63062	5.285956678
3	5881	3132	14731	21955	55344	269987	101043	2.672001029
4	3828	1201	8532	44456	173848	221636	231865	0.955883812
5	5831	857	3372	25533	631164	140030	666757	0.210016543
						1772055	1158541	
							Overal MSE	**1.529557435**
							Normalized MSE	3.509408513

AllBigrams and 5-cluster Author-Type

	1	2	3	4	5	Total Squared Error	Instances	MSE
1	40280	2850	3975	3688	45021	772278	95814	8.0601791
2	11663	3925	8943	7862	30669	328075	63062	5.20241984
3	6018	2533	14914	23721	53857	265754	101043	2.63010797
4	4367	1133	9221	47582	169562	222618	231865	0.96011903
5	7520	1007	4663	29703	623864	177738	666757	0.26657088
						1766463	1158541	
							Overal MSE	**1.52473067**
							Normalized MSE	3.42387937

It helped *a little bit*…

Our best results so far

AllCaseInsensitiveBigramsBalanced

	1	2	3	4	5	Total Squared Error	Instances	MSE
1	67172	16111	4549	2255	5727	146234	95814	1.5262279
2	18318	23840	12458	4144	4302	86070	63062	1.364847293
3	12514	20282	37062	20061	11124	134895	101043	1.335025682
4	16291	13824	42706	85784	73260	317881	231865	1.370974489
5	51675	16602	32257	111473	454750	1216719	666757	1.824831235
						1901799	1158541	
							Overall MSE	**1.641546566**
							Normalized MSE	1.48438132

- Rebalanced training data by down-sampling
- Using case-insensitive bigrams results in error reduction
- Incorporating author-profile actually resulted in performance degradation.
- We tried trigrams, tetragrams, and fivegrams. Nothing beat good ol' bigrams.
- A disproportionate number of 5s got classified as 1s. Perhaps some negation resolution could help here.

Human Performance

- We set up a website showing ten reviews to viewers and asked them to guess the ratings.
- Accuracy of 57.78%
- Mean Squared Error of **0.7889**
- Humans haveHuman much better MSE.
- MaxEnt had better accuracy on unbalanced training data, simply because it guessed 5-star more often.
- MaxEnt has similar accuracy as human when trained on balanced data.

What influences author-type?

We found more than 50% of the data are 5-star reviews.

Most authors also only give 5-star reviews.

Could that be influenced by things like location, time, day of week, etc?

For example, do Americans generally give more positive reviews than people in the UK?

In Summary…

Nothing beats balanced case-insensitive bigrams (so far), but we're still investigating certain style features (negation, length, readability).

We could explore giving author-type features more weight instead of just throwing everything into MaxEnt