A Machine Learning Approach for Evaluating Creative Artifacts

Disha Shrivastava, Saneem Ahmed CG, Anirban Laha, Karthik Sankaranarayanan
{dishriva, saneem.cg, anirlaha, kartsank}@in.ibm.com **(IBM Research, India)**

KDD2017

IBM

Problem Statement

- System which **learns to evaluate creativity** of creative artifacts.

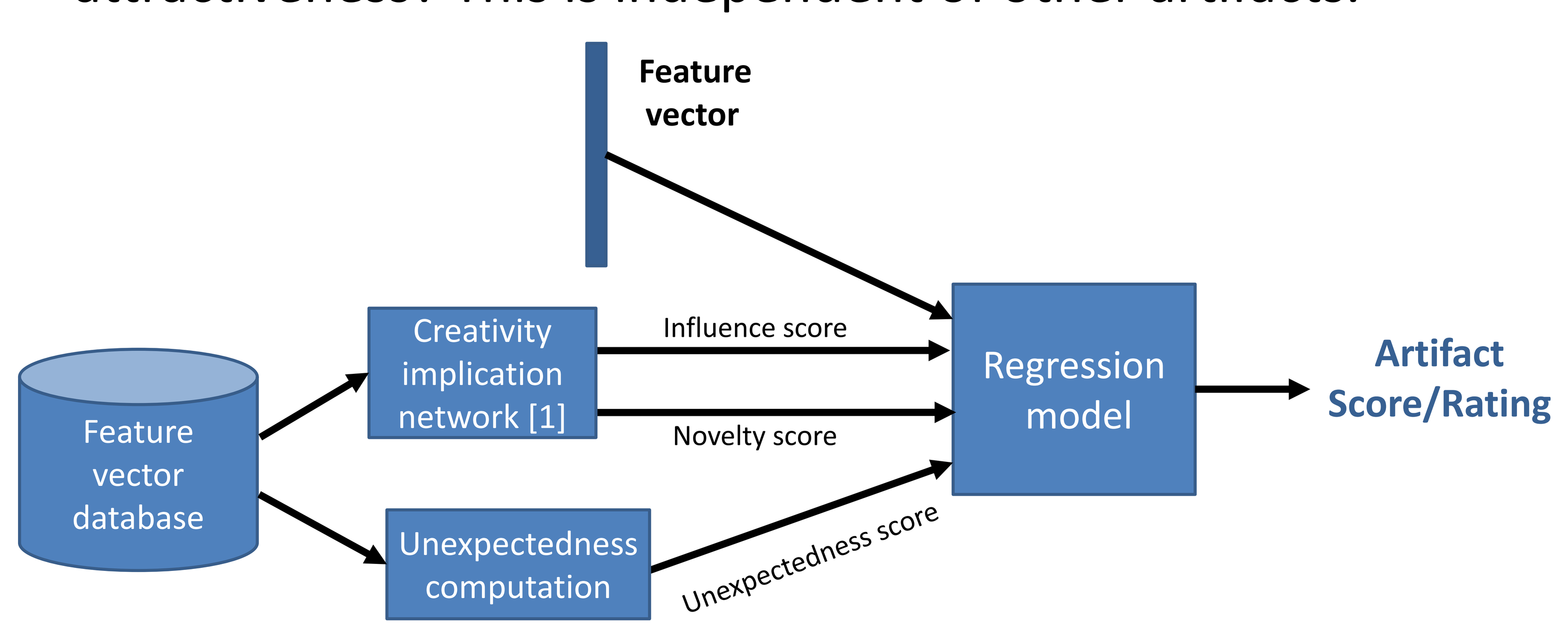
Creativity Evaluator → **Creativity Score**

- First attempt to build a **domain-independent** prediction model which is inspired from philosophies of creativity.
- This system can assist creative agents which could generate creative products.
- Uses audience/critic rating for the artifact as a proxy ground truth for creativity score.

Proposed Approach

- Creative scores are computed not just looking at the artifact in isolation but also with other artifacts which were created before and after it.

- **Creativity Criteria**:
- **Novelty:** or originality is about how different the artifact is from prior works.
- **Influence:** How much impactful or inspiring it has been for artifacts occurring later in time?
- **Unexpectedness:** Not expected at that point of time.
- **Value:** How the artifact is good in utility, performance or attractiveness? This is independent of other artifacts.

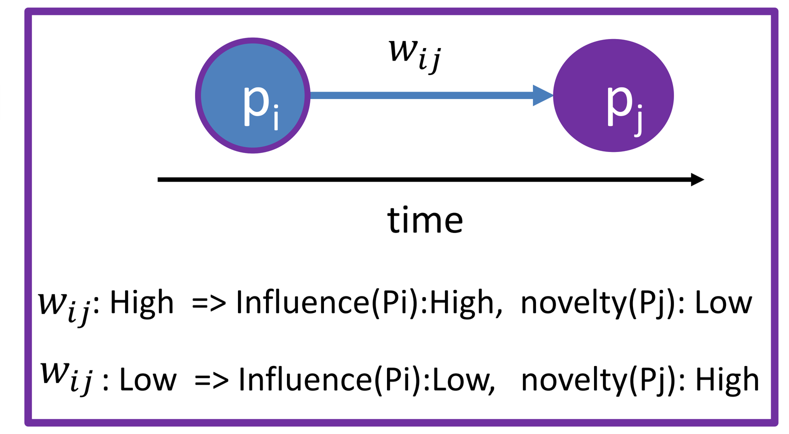


- Regression model to predict **value score** and to combine novelty, influence, value and unexpectedness scores to get creativity score.

Novelty and Influence Scores

- **Construction of Markov chain**
- First construct a graph with edges directed in chronological order with edge weights as similarities (w_{ij}) between them.
- We create a graph with edge weights $\tilde{w}_{ij}$ got from following equation.

$$B(w_{ij}) = w_{ij} - \theta$$
$$B(w_{ij}) > 0 \;\rightarrow\; \tilde{w}_{ij} = B(w_{ij})$$
$$B(w_{ij}) < 0 \;\rightarrow\; \tilde{w}_{ji} = -B(w_{ij})$$



- This network reduces the problem of computing novelty+influence scores as a **traditional centrality problem**.
- Markov chain update rule for node probability $C(p_i)$ is given as.

$$C(p_i) = \frac{(1-\alpha)}{N} + \alpha \sum_j \tilde{w}_{ij} \frac{C(p_j)}{N(p_j)}$$

- Stationary distribution of this network gives **novelty+influence scores.**

Unexpectedness Score

- Computed as **negative mean similarity** between the artifact and all the artifacts created in **K year** window preceding it.

Experiments

1. **Dataset Collection and Curation:**
- **Kaggle IMDb 5000 Movie Dataset[2]**
- **Meta-data scraped from Rotten Tomatoes(RT) website**
- **Wikipedia plots**

➢ *Preprocessing:* Removed movies with missing year and features which are redundant/heavily correlated to output labels.

5021 movies with 21 unique attributes

2. **Attribute Representation and Computing Similarity**

Attribute Type	Example	Representation	Similarity
String with semantic meaning	Genre	Word2Vec	Cosine
String with no Semantic meaning	RT Studio	One-Hot Encoded Vector	Cosine
Paragraphs	Wikipedia Plot	Skip-Thought Vectors	Cosine
Numerical	Director Facebook likes	Real number	Linear, Exponential

➢ PCA on word2vec and skip-thought vectors to reduce dimensionality ***407 features in total (Baseline)***

3. **Computing Creativity Measures:**

➢ Similarity between two movies used as an ***edge*** for forming ***attribute-wise graph*** *(α=0.95, β=0.2, 0.5)*

4. **Prediction Models:** **SVR, Random Forest, KNN, Ridge and Bayesian Regression** for different feature combinations

Results

Feature Combination	IMDB Rating		RT Critic Score	
Prediction model	KNN	Random Forest	KNN	Random Forest
Baseline	0.11224	0.09042	0.26919	0.23472
PUN	0.10777	0.08908	0.26782	**0.23040**
PUI	0.10877	0.08922	0.26776	0.23042
PUNI	0.10835	0.08936	0.26952	0.23136
PUNIA	**0.1071**	**0.08891**	**0.26539**	0.23127
%Improvement	**4.58040**	1.67013	1.41486	**1.8394**

Fig1: Comparison of RMSE for different feature, model and output label combinations

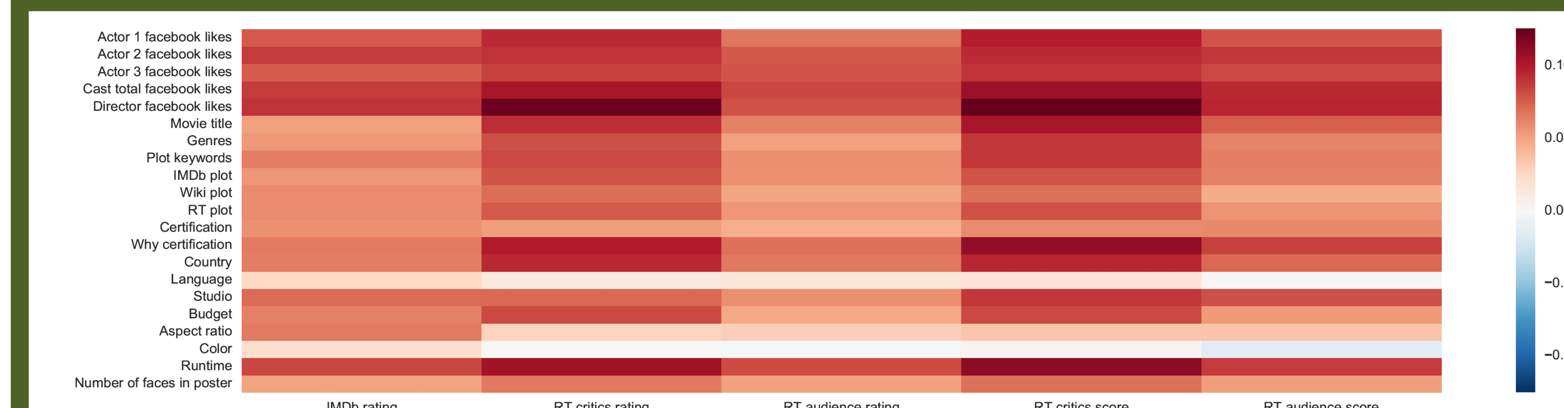


Fig2: Pearson correlation of aggregate of novelty and influence scores of different attributes with labels.

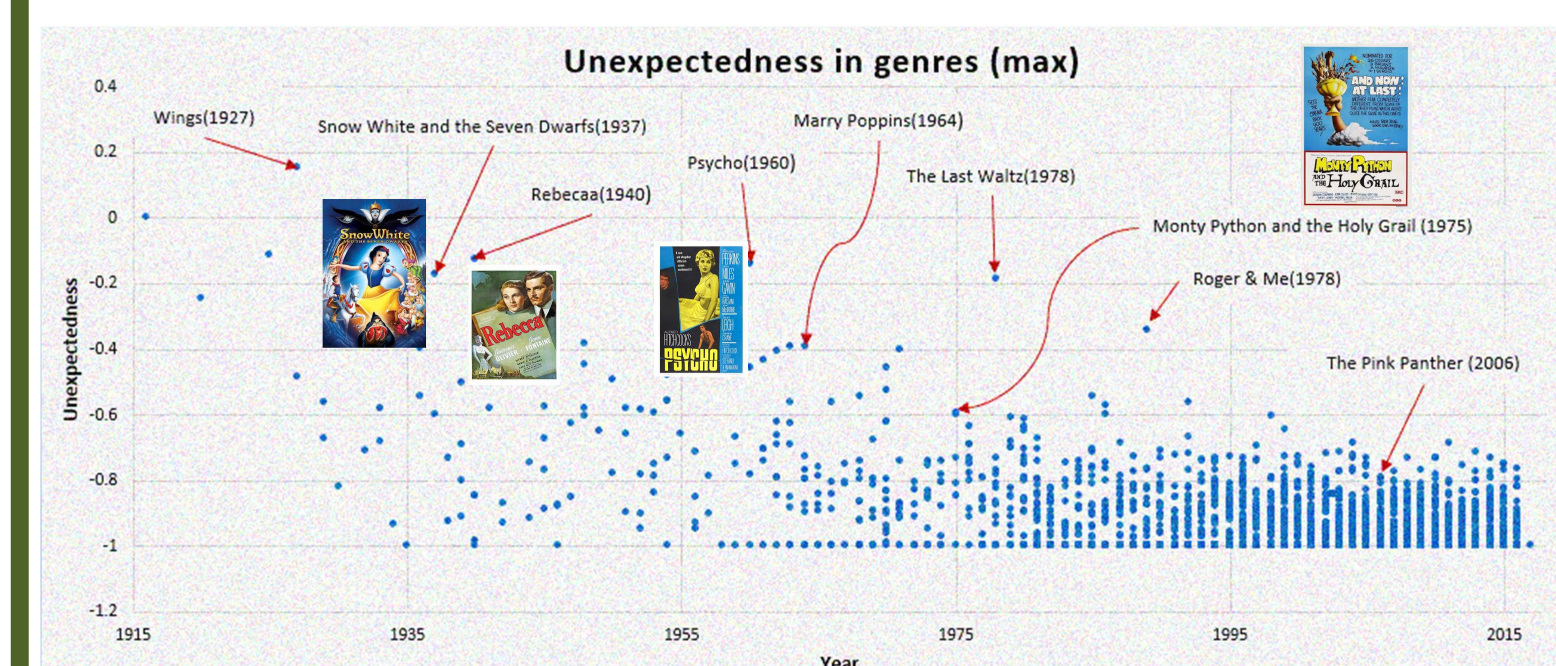


Fig3: Visualization of unexpectedness scores of movies

➢ Inclusion of creativity measures **improves prediction performance** for all models and output labels

➢ Creativity measures are **positively correlated** with all output labels.

➢ **Stronger correlation to critic scores as compared to audience scores** suggest critics consider creativity measures more as opposed to audience who may be biased by other factors.

References

[1] Ahmed Elgammal and Babak Saleh. 2015. Quantifying Creativity in Art Networks. (ICCC 15)

[2] IMDB 5000 Movie Dataset.Kaggle(2016). https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset