Towards Computational Assessment of Idea Novelty

Kai Wang[1] Boxiang Dong[2] Junjie Ma[1]

[1]School of Management and Marketing
Kean University
Union NJ

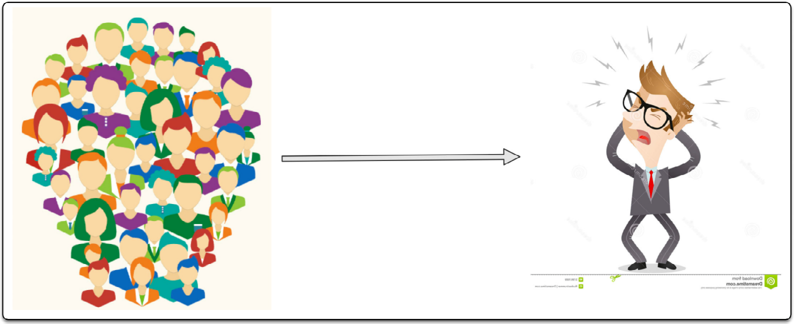
[2]Department of Computer Science
Montclair State University
Montclair, NJ

Jan 11, 2019

Idea Collection

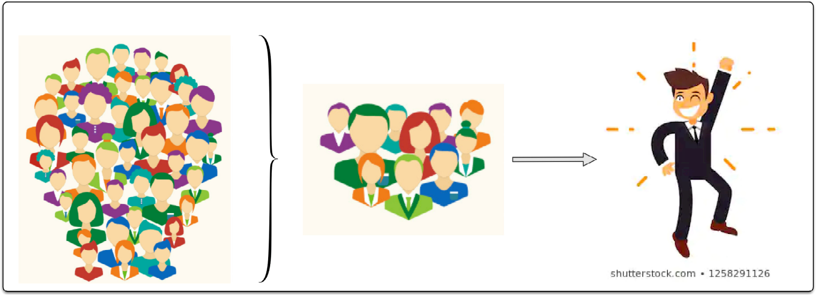
- Companies collect ideas from a large number of people to improve existing offerings [AT12, WN17].

Idea Novelty Assessment



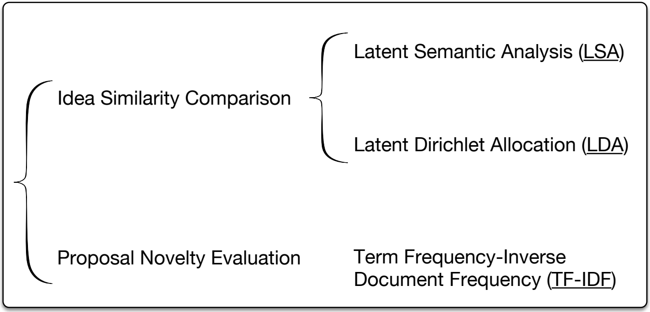
- Manually selecting the most innovative ideas from a large pool is not effective.

Idea Novelty Assessment



- Manually selecting the most innovative ideas from a large pool is not effective.
- It would be very helpful to automate the evaluation of creative ideas.

Idea Novelty Assessment



However, none of these approaches have been validated through the comparison with human judgment.

Our Contribution

- Three computational idea novelty evaluation approaches
    - LSA
    - LDA
    - TF-IDF
- Three sets of ideas
- Comparison with human expert evaluation

Outline

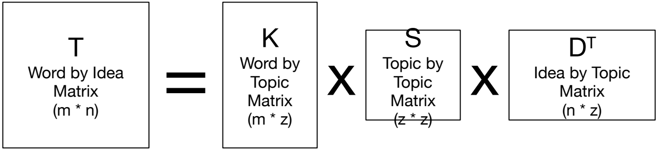
1. Introduction
2. **Background**
3. Methods
4. Results
5. Conclusion

Background - LSA [CS15, TN16]

Input Idea by word matrix

Output Idea by topic matrix

Key Idea Apply *Singular Value Decomposition (SVD)* on the input matrix.

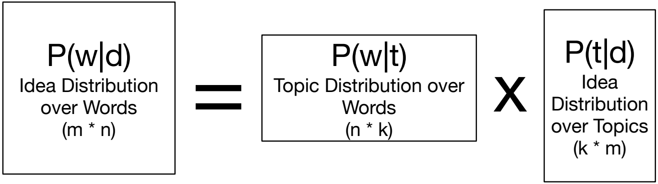


Background - LDA [WNS13, Has17]

Input Idea by word matrix

Output Idea by topic matrix

Key Idea
- Each idea is represented as a mixture of latent topics.
- Each topic is characterized as a distribution over words.



Background - TF-IDF [WB13]

Input Idea by word matrix

Output Idea by word tf-idfs

Key Idea Determine how important a word is to an idea.

$$tf\text{-}idf(w_i, d_j) = tf(w_i, d_j) \times \log\left(\frac{n}{df(w_i)}\right)$$

$tf(w_i, d_j)$: # of times that w_i appears in d_j

$df(w_i)$: # of ideas that include w_i

n: # of ideas

Methods - Data Collection

We use Amazon Mechanical Turk (www.mturk.com) to employ crowd workers to collect three set of ideas.

Alarm Ideas about a mobile app of an alarm clock.

Fitness Ideas to improve physical fitness.

Advertising Ideas to promote TV advertising.

Dataset	# of Ideas	Avg. # of Characters
Alarm	200	555
Fitness	240	586
Advertising	300	307

Methods - Human Expert Evaluation

We hire a group of human experts to evaluate the collected ideas.

- Each idea is evaluated by at least two human experts.
- Novelty is defined by using a Likert scale of 1 to 7 (1 being not novel at all, 7 being highly novel).
- Human experts demonstrate reasonable level of agreement in the ratings (Intraclass correlation coefficient is higher than 0.7).
- We take the average of human ratings as the ground truth of idea novelty.

Methods - Computational Novelty Evaluation

LSA Cosine distance to average

LDA
- Use Gibbs sampling with 2,000 iterations
- Cosine distance to average

TF-IDF Sum of all tf-idfs in an idea

Experiments

We compare the following methods with the ground truth.

LSA

LDA

TF-IDF

Crowd We hire 20 crowd workers to manually evaluate the idea novelty, and take their average.

Experiments

Table 1. Correlation between different measures and expert evaluation.

Ideation Tasks	Alarm Clock App (n=200)	Fitness App (n=240)	TV Advertising (n=300)
Expert Evaluated Novelty (Mean±SD)	4.55±1.01	4.64±1.38	3.97±1.20
Correlation of LSA measure with expert evaluation	0.114 (p=0.107)	0.230 (p<0.001)	0.230 (p<0.001)
Correlation of LDA measure (k=10) with expert evaluation	0.184 (p=0.009)	0.231(p<0.001)	0.198 (p<0.001)
Correlation of LDA measure (k=20) with expert evaluation	0.253 (p<0.001)	0.190 (p=0.003)	0.229 (p<0.001)
Correlation of LDA measure (k=30) with expert evaluation	0.226 (p=0.001)	0.219 (p<0.001)	0.235 (p<0.001)
Correlation of TF-IDF measure with expert evaluation	0.340 (p<0.001)	0.319 (p<0.001)	0.307 (p<0.001)
Correlation of crowd evaluation and expert evaluation	0.748(p<0.001)	0.501 (p<0.001)	0.648 (p<0.001)

- LSA correlates well with the ground truth on the Fitness and TV Advertising datasets.
- LDA and TF-IDF performs well on all three datasets.
- Crowd evaluation correlates with expert evaluation better than all the three computational methods.

Experiments

Table 2. Top ten novel ideas according to different measures.

Ideation Tasks	Alarm Clock App	Fitness App	TV Advertising
True novelty of the top ten ideas by LSA measure (Mean±SD)	4.38±0.82	5.45±1.01	5.1±1.07
Number of correctly identified top ten ideas by LSA measure	0	1	0
Point-biserial correlation comparing top 10 novel ideas by LSA measure and the remaining ideas	-0.039 (p=0.587)	0.122 (p=0.059)	0.176 (p=0.002)
True novelty of the top ten ideas in LDA measure (k=20) (Mean±SD)	5.13±0.82	5.30±0.90	4.55±1.21
Number of correctly identified top ten ideas by LDA (k=20) measure	1	0	1
Point-biserial correlation comparing top 10 novel ideas by LDA (k=20) measure and the remaining ideas	0.132 (p=0.062)	0.099 (p=0.125)	0.09 (p=0.119)
True novelty of the top ten ideas in TF-IDF measure (Mean±SD)	4.98±0.97	5.40±0.49	5.35±0.90
Number of correctly identified top ten ideas by TF-IDF measure	1	0	3
Point-biserial correlation comparing top 10 novel ideas by TF-IDF measure and the remaining ideas	0.098 (p=0.168)	0.130 (p=0.045)	0.215 (p<0.001)
True novelty of the top ten ideas in crowd evaluation (Mean±SD)	5.73±0.66	5.9±0.44	5.95±0.61
Number of correctly identified top ten ideas by crowd evaluation	3	1	1
Point-biserial correlation comparing top 10 novel ideas by crowd evaluation and the remaining ideas	0.269 (p<0.001)	0.190 (p=0.003)	0.300 (p<0.001)

- Crowd evaluation identifies more top-10 novel ideas than all computational approaches.
- Crowd evaluation resulted in significant point-biserial correlation for all three ideation tasks

Conclusion

We experimentally compare three computational novelty evaluation approaches with ground truth.

- TF-IDF outperforms LSA and LDA in matching expert evaluation.
- All three computational approaches fall far behind crowd evaluation.
- Much more research is needed to automate the evaluation of creative ideas.

References I

[AT12] Allan Afuah and Christopher L Tucci.
Crowdsourcing as a solution to distant search.
Academy of Management Review, 37(3):355–375, 2012.

[CS15] Joel Chan and Christian D Schunn.
The importance of iteration in creative conceptual combination.
Cognition, 145:104–115, 2015.

[Has17] Richard W Hass.
Tracking the dynamics of divergent thinking via semantic distance: Analytic methods and theoretical implications.
Memory & cognition, 45(2):233–244, 2017.

[TN16] Olivier Toubia and Oded Netzer.
Idea generation, creativity, and prototypicality.
Marketing science, 36(1):1–20, 2016.

[WB13] Thomas P Walter and Andrea Back.
A text mining approach to evaluate submissions to crowdsourcing contests.
In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pages 3109–3118. IEEE, 2013.

[WN17] Kai Wang and Jeffrey V Nickerson.
A literature review on individual creativity support systems.
Computers in Human Behavior, 74:139–151, 2017.

[WNS13] Kai Wang, Jeffrey V Nickerson, and Yasuaki Sakamoto.
Crowdsourced idea generation: the effect of exposure to an original idea.
2013.

Q & A

Thank you!

Questions?