International Journal of Web Based Communities

In Search of Disruptive Ideas: Outlier Detection Techniques in Crowdsourcing Innovation Platforms

Adam Westerski*

Institute for Infocomm Research, A*STAR, Singapore
E-mail: adam-westerski@i2r.a-star.edu.sg
*Corresponding author

Rajaraman Kanagasabai

Institute for Infocomm Research, A*STAR, Singapore
E-mail: adam-westerski@i2r.a-star.edu.sg
E-mail: kanagasa@i2r.a-star.edu.sg

Abstract: The key challenge for data science in open innovation web systems is to find best ideas among thousands of community submissions. To date, this has been done with metrics reflecting enterprise needs or community preferences. This article proposes to look in a different direction: inspired by theoretical studies on disruptive innovation, we frame the problem of valuable ideas as those rarely taken up by masses or organisations yet having potential to change industries. Our aim is to find technological means for automatic detection of such innovations to aid decision making. Following past findings from business sciences on nature of disruptive innovations, the article presents a comparative study of multiple outlier detection algorithms applied to two real-world datasets containing textual descriptions of ideas for different industries. Obtained results demonstrate capability of outlier detection and show k-NN algorithm with TF-IDF and cosine distance to be the best candidate for the task.

Keywords: idea management; disruptive innovation; outlier detection; data analytics; innovation management; open innovation systems; crowdsourcing

1 Introduction

The care for innovation as an organised process has been present for long time in business and economical sciences, with initial contributions dating back to 1930s (Fagerberg and Verspagen, 2009). A multitude of theories and proven practices show how to innovate successfully for commercial (Becheikh et al., 2006; Tornjanski et al., 2015; Sanchez et al., 2011), public sector (Osborne and Brown, 2013; Windrum and Koch, 2008) and social (Mulgan et al., 2007) purposes. Different methodologies have been applied depending on what is the subject of innovation, e.g. product, process, organization, or marketing (European Commission, 2005). All of those approaches, formalised as innovation models and categorised under various topologies, have evolved over time by adjusting to changing reality of markets or societies (Popadiuka and Choo, 2006; Chuang et al., 2010; Garcia and Calantone, 2002). Regardless of innovation type and area however, some elemental problems have always been discussed, such as: "why to innovate?"; "how to innovate?";

"what should be the process of innovation?"; and finally about the outcome "what makes the best innovation?"

In the last years, along with the introduction of computer decision support systems into innovation processes, this final question of idea quality and successfulness has become of key significance. That is because one of the more established technologies, Idea Management Systems (IMS), for years attempted to deliver a toolkit for selecting best ideas from innovation proposals submitted by communities relevant to the organization (Westerski et al., 2011). One of the standing problems in the area, has been information overflow, ie. large amount of trivial or poor quality submissions vastly outnumbering quality innovation proposals (Merz, 2018). There have been a number of case studies in the literature, which show how companies such as IBM (Bjelland and Wood, 2008), Cisco (Jouret, 2009) or Google (Google, 2009) struggled with assessment of their community open innovation campaigns due to human resources needed for the task. Therefore, ever since the establishment of this IMS software family, scientists and practitioners have proposed a variety of different methods for improvement of screening and selection of those best ideas: basic community voting (Hrastinski et al., 2010; Gangi and Wasko, 2009), internal expert reviews (Hrastinski et al., 2010; Gangi and Wasko, 2009), prediction markets (Bothos et al., 2008), sentiment analysis (Lee et al., 2018) and multiple other data analytic solutions (Martinez-Torres, 2015; Yoo et al., 2015). However, regardless of toolkit used, in all aforementioned approaches, the perception of what is a good/bad idea has been dominated by a core concept of community needs intertwined with expert opinions involving enterprise strategy, market knowledge and technological feasibility.

In this article, we argue that while such approach seems intuitively correct and is indeed grounded with some past innovation management studies, it is also not the only possibility (Dean et al., 2006). To push the state of the art in Idea Management Systems, we propose to rethink those very basic assumptions. Indeed, in business studies such argument has been already taken up by Bower and Christensen in 1995 (Bower and Christensen, 1995), and refined in years to come (Christensen, 1997) to formulate the theory of disruptive innovation. In it, authors point that the breakthrough progress that makes or destroys market leaders is achieved not by listening to well established customer base but new revolutionary ideas that tend to get little support (community or enterprise) and are unlike anything typically proposed.

Taking inspiration from this theory of disruptive innovation, we connect it with the practical implications of data present for analysis in Idea Management Systems and propose to treat idea selection problem as a variation of outlier detection problem: search for data samples not fitting typical characteristics of global population. The algorithms applied for solving this problem have been long established and used in multiple domains in the past such as: fraud detection, intrusion detection or new event monitoring (Aggarwal, 2013). We hypothesise the same techniques could be applied for IMS but to support idea selection in a way which would adhere to what Bower and Christensen described as discovery of disruptive innovations. Furthermore, we view such outlier detection tools, not as a replacement for traditional idea selection methods but as a supplement that could bring new insights to decision makers of IMS and do so in an efficient way further offloading human referees.

The article is structured as follows: firstly we describe our hypotheses in more detail and their theoretical grounding (see Sec. 2). Next, we present the methodology to address the defined hypotheses in a scientific process (see Sec. 3), followed by a report on results of implementation of that methodology in form of evaluation exercises (see Sec. 4). Finally, we compare our approach to alternatives in the same domain as well as related areas (see

Sec. 5). We conclude the article with final recommendations in light of experimental results; and comments on possible extensions in the future (see Sec. 6).

2 Hypothesis and Theoretical Grounding

Idea Management Systems (IMS) that we analyse in this article are a practical approach to implementing open innovation concept in an organization (Brem and Voigt, 2007; Hossaina and Islam, 2015a,b). Typical platforms of that kind are aimed to collect textual idea descriptions from many different people and publish this content in the open to allow collaboration in a Social Web fashion, including idea rating and discussions. As a consequence of such approach, typically IMS deployments produce huge amounts of submissions. While that might seem like a good outcome, the reality is that just like on the Social Web the quality of content is often underwhelming (Agarwal and Yiliyasi, 2010) and due to size it is difficult to separate valuable input from duplicate, obvious ideas or non-ideas (Christensen et al., 2017). Furthermore, deciding on idea quality and its degree of alignment with interests of the organization is also a hard and time consuming task.

To address those problems and extract valuable submissions out of thousands of contributions, we take inspiration from disruptive innovation theories by Bower and Christensen (Bower and Christensen, 1995). More precisely, according to later work of Christensen: "disruptors deliver innovations for overlooked market segments, while market leaders address their most demanding customers via incremental innovation" (Christensen et al., 2015). Connecting this to Idea Management practice, we hypothesise that the huge amount of trivial and duplicating ideas which create the bulk of IMS content are incremental small innovations requested by loyal customers (also suggested by some prior studies (Westerski et al., 2013)). Therefore, creating metrics based on either majority of community preferences or internal enterprise urge to satisfy those typical customer needs will overlook ideas that do not get popular support but still hold the value of what Christensen labels as disruptive innovation.

Furthermore, it is also notable that key criteria for evaluation of many past metrics and studies developed in IMS research are based on correlation with idea successfulness typically reflected by ratio of past implemented ideas (Gangi and Wasko, 2009; Hossaina and Islam, 2015a). Some solutions even go further utilising those past implemented ideas, as input for supervised machine learning to predict the next big innovation (Yoo et al., 2015). In relation to that approach however, Christensen observes that established incumbent businesses tend to implement only those ideas which "improve their products for most demanding and profitable customers yet omitting needs of others" (Christensen et al., 2015). That in turn creates a gap that disruptive innovators utilise to gain market foothold and eventually challenge for market dominance. Connecting those theories to IMS practice, a possible conclusion could be that trying to find ideas solely based on past implementation criteria will not capture potentially valuable future disruptive innovations.

Building on top of those observations and characteristics of disruptive innovation pointed out by Christensen, we hypothesise that the tool adequate for finding disruptive innovation in text corpora of Idea Management Systems are outlier detection algorithms that enable to find unusual and rare text extracts from a pool of otherwise similar contributions. In context of IMS that could be textual descriptions of potential disruptive innovations lost and neglected among thousands of similar incremental ideas as described in previous paragraphs. In this paper we structure those assumptions regarding IMS content into two hypotheses that we evaluate in a systematic way:

- (1) the outlier rank reflects degree to which idea is a disruptive innovation
- (2) the outlier rank brings a new metric to IMS for rating ideas that is not expressed by any other metric used so far

With the first hypothesis we check if our idea about connecting Christensen theories with reality of Idea Management Systems via outlier detection is indeed a correct direction. The second hypothesis aim is to see if applying outlier detection to Idea Management data brings any new value whatsoever in relationship to prior metrics (ie. if it does not duplicate what is already available). We describe the process of testing both of those hypotheses in the next section.

3 Proposed Approach

Outlier detection algorithms have been established for a long time and since then studied and applied in many different domains. In this article, to attain the vision described in the introduction and thoroughly evaluate hypotheses highlighted in previous section we survey the available outlier detection algorithms and test their performance for the needs of Idea Management Systems. We also describe the customisations made in order to achieve the best performance for tested algorithms. Our goal is to pick the best solution for Idea Management use case out of the most representative and established algorithms as typically presented in the state of the art. Therefore, our approach can be summarised in the following steps:

1. **Analyse** state of the art in outlier detection algorithms
2. **Pick** the most representative candidates based on previous applications and recommendations from past studies and articles that attempt to categorise and classify available algorithms
3. **Apply** selected algorithms to Idea Management problem
4. **Evaluate** algorithms for Idea Management problem using two different public datasets to obtain domain independent results
5. **Compare** results across different algorithms/ algorithm types/ algorithm configurations as well as different industry application areas of Idea Management
6. **Recommend** the best approach across all tested variations and verify the hypotheses

The application of this methodology is possible because majority of algorithms that are typically classified under outlier detection domain work based on similar inputs and delivering results in a similar output form. The inputs being unorganised list of concepts characterised by a certain set of features - in our case list of ideas with features obtained using various text modelling approaches. While the outputs being a ranked list of those input concepts with each associated a score determining degree to which a concept is an outlier - in our case ranked list of ideas with scores that we intend to compare with earlier manually assigned disruptive innovation scores. The best choice for algorithm delivering those outputs is the unknown which we set out to identify through experiments as described in detail in next section.

4 Experiments Setup and Results

The outlier detection domain being quite mature has had multiple approaches to categorise various algorithms and systematically describe the landscape of the state of the art (Knorr et al., 2000; Chandola et al., 2009; Singh and Upadhyaya, 2012; Zhang, 2013; Aggarwal, 2013; Pimentel et al., 2014; Campos et al., 2016). While each of those works bring slightly different point of view on the domain, some common perception of the main types of algorithms and most prominent examples can be extracted (Chandola et al., 2009). For the needs of our study we narrowed down on such broadly established algorithm categories, which are well described and have past examples of being used specifically with text outlier detection, namely: (1) distance based; (2) probabilistic; (3) density based; and (4) clustering based algorithms. For each of those categories we picked one representative algorithm that would be used during our Idea Management evaluation, respectively: k-nearest neighbors (k-NN) (Ramaswamy et al., 2000), latent Dirichlet allocation (LDA) (Aggarwal, 2013; Lu et al., 2011), Local Outlier Factor (LOF) (Breunig et al., 2000), k-Means/k-Medoids (Zhang, 2013). This choice was made based on availability of scalable implementations and reports on prior evaluations with textual data. Each of the algorithms was additionally tested with variety of configuration settings to look for an optimal solution to our problem.

Furthermore, many outlier algorithms were not created specifically with text outlier analysis in mind but originate from more broad and generic approaches. Therefore, outside of outlier detection solution, frequently there is also a need to pick a method in which text would be represented as a numerical vector to facilitate the algorithm. To make our study more comprehensive, for each of the previously mentioned algorithms, if applicable, we tested several different such approaches: (1) term frequency-inverse document frequency (TF-IDF) (Salton, 1983) ; (2) word2vec (Mikolov et al., 2013a); (3) LDA with variational expectation-maximisation (VEM) algorithm (Blei et al., 2003); and (4) LDA with Gibbs Sampling (Phan et al., 2008).

Thirdly, multiple categories of outlier detection algorithms rely on the concept of measuring distance between samples. This can be also performed in multiple ways, for the needs of our study we experimented with: (1) Cosine; (2) Manhattan; and (3) Euclidean distance measures.

Finally, depending on algorithm there can be a number of parameters that affect final performance, where applicable we approached this individually, trying to find the most suitable configuration but also analysing to what degree differences in such settings would affect the final result in the case study of Idea Management Systems. Thus, for k-NN algorithm we considered different k settings, measuring distance from 1st, 2nd, 3rd, 4th, 10, 100, 200, 300, 1000, 2000, 3000, 5000 nearest neighbour as well as for same brackets but taking mean of all distances in a given radius (ie. mean of closest 2 neighbours, closest 3 etc.). For LDA algorithm we experimented with different settings for number of iterations, passes, alpha and eta parameters, optimising for best perplexity metric as well as targeting for modelling 3, 42, 100, 200, 300 or 400 topics. With LOF, which is dependent on perception of neighbourhood like distance based algorithms, we analysed similar setting ranges as with k-NN. Finally, for the clustering approaches, we picked target cluster counts like earlier with topic modelling (3, 42, 100, 200, 300 or 400). All the aforementioned parameter values were chosen to experiment with a possibly wide spectrum of settings to get the idea where algorithm performed well and where not. In case of aforementioned tested neighbourhood distances as the point of reference we used the total count of samples in the datasets, while to pick topic/cluster counts, we used as a guideline the count of predefined idea categories in the datasets. The full list of all tested algorithms and combinations can be seen in Table 1.

Table 1 List of evaluated algorithms and underlying settings.

Category	**Algorithm**	**Feature Vector Generation**	**Distance Measures**	**Tested Parameters**	**Algorithm Outline**
Distance based	k-NN	TF-IDF; word2vec; LDA/VEM; LDA/Gibbs	Cosine; Manhattan; Euclidean	k = {1, 2, 3, 4, 10, 100, 200, 300, 1000, 2000, 3000, 5000}; k = {single / mean dist}	Outliers are determined based on sample distances from its nearest neighbours.
Probabilistic	LDA	-	-	#iterations, #pases, alpha, eta; topics ={3, 42, 100, 200, 300, 400}	Samples are assigned into topic groups. Outliers are samples with weakest links to any topic.
Density Based	LOF	TF-IDF	Cosine; Manhattan; Euclidean	*same as k-NN*	Outliers are picked by smallest density of samples around them.
Clustering	k-Means/ k-Medoids	TF-IDF	Cosine; Manhattan; Euclidean	#cluster = {3, 42, 100, 200, 300, 400}	Samples are formed into groups with one most representative, outliers are those samples furthest from the representative.

4.1 Datasets

All of the aforementioned algorithms and combinations of configurations were evaluated twice - using two different datasets coming from public deployments of Idea Management Systems: Dell IdeaStorm and My Starbucks Ideas. Both systems were created using the same Idea Management technology from SalesForce and therefore presented similar user interface and interaction workflow: published in form of web portals opened to the public, allowing users to register, post new ideas, comment on ideas of others, vote up/down existing idea posts. In both cases ideas were organised in thematic categories and attached a status indicating response of organization to the idea (e.g. under review, implemented etc.). In case of Dell system additionally special focused sessions were organised to collect ideas of particular interest for the company. The datasets were obtained by HTML scraping web pages for all the aforementioned information. The data coming from those IMS has been widely used in other studies and the exact similarities and differences between available fields and semantics of those can be seen in previous publications (Westerski et al., 2010, 2013). For the needs of this article, we only used the basic textual fields being idea title and idea description. Remaining metedata was only used for comparative purposes to generate metrics typically present in Idea Management Systems, this is further described in next sections.

In terms of idea content and meaning for the organization, the main differentiator between the two datasets is related to business areas and characteristics of target audience. Both instances were related to companies with very big customer base and worldwide reach. The first, Dell IdeaStorm, was run by a big international computer technology company, primary occupied with manufacturing and sales of personal computers, server equipment and related peripherals. The goal of Dell instance was to collect any ideas related to its business from customers and other interested parties, often resulting in contributions related to creating new products or services; or improving existing ones. The second instance, My Starbucks Ideas, was operated by a large international coffee company and coffeehouse chain offering brewages, snacks and some related seasonal and locality based products. The audience of this Idea Management System was, similar to Dell, customers but also franchise and other business partners. Both instances collected ideas for a long period resulting in a fairly big and varied database: Dell system was opened to public starting February 2007, while Starbucks March 2008. In both cases we collected all data from instance opening until February 2011. The extraction ending date was limited for several reasons: 1) in following years both instances underwent some changes in UI and submission rules, which might have impacted the collected data introducing some needless complexity to our analysis; 2) we compare our research to state of the art using several prior articles that analyse Dell and Starbucks instances in similar time frame, therefore we found it more accurate to operate on same data; 3) some of the outlier detection algorithms we tested are computationally demanding, therefore to run multiple experiments with exhaustive set of parameter combinations we found it more suitable to limit the data scope. Further statistics of the datasets can be observed in Table 2.

Table 2 Summary of datasets used during experiments.

System name	Description	#Ideas	#Comments	#Users
Dell IdeaStorm	Computers, telecommunication devices and related services.	9.741 (207 implemented)	65.222	8.589
myStarbucks Ideas	Coffee and related products sold in a coffeehouse chain.	10.921 (1069 implemented)	21.870	12.745

For the purpose of evaluation, idea title and description were treated equally and merged into a single block of text. Before proceeding with applying any algorithms both datasets were pre-processed to remove unnecessary noise and clean the textual content from symbols and phrases unrelated to idea meaning. Firstly, we filtered out all text related to web environment and coding language, i.e. all HTML tags and HTML special symbols that sometimes showed in text due to formatting errors by users, as well as all hyperlinks. Further, we performed processing typical for most text mining approaches: removed stop words, numbers and punctuation. Finally, we selected only ideas with text length between 250 and 1250 characters. This choice was made as a consensus between quality of results and fraction of idea database that would be filtered out. During some preliminary experiments with outlier algorithms, we discovered that very short ideas picked up by some algorithms were either accidental submissions with just few random letters or incomprehensible extracts of few words without any real meaning. On the other side of the spectrum, very long submissions that sometimes got elevated rankings were spam not related to ideation process at all thus turning out as outliers due to large amount of very distinct words used. Given both of those observations after applying the aforementioned constraints we ended up using 5861 (60%) ideas for Dell, and 5470 (50%) for Starbucks instances. In both cases vast majority of removed submissions were below 250 characters (see Fig. 1).

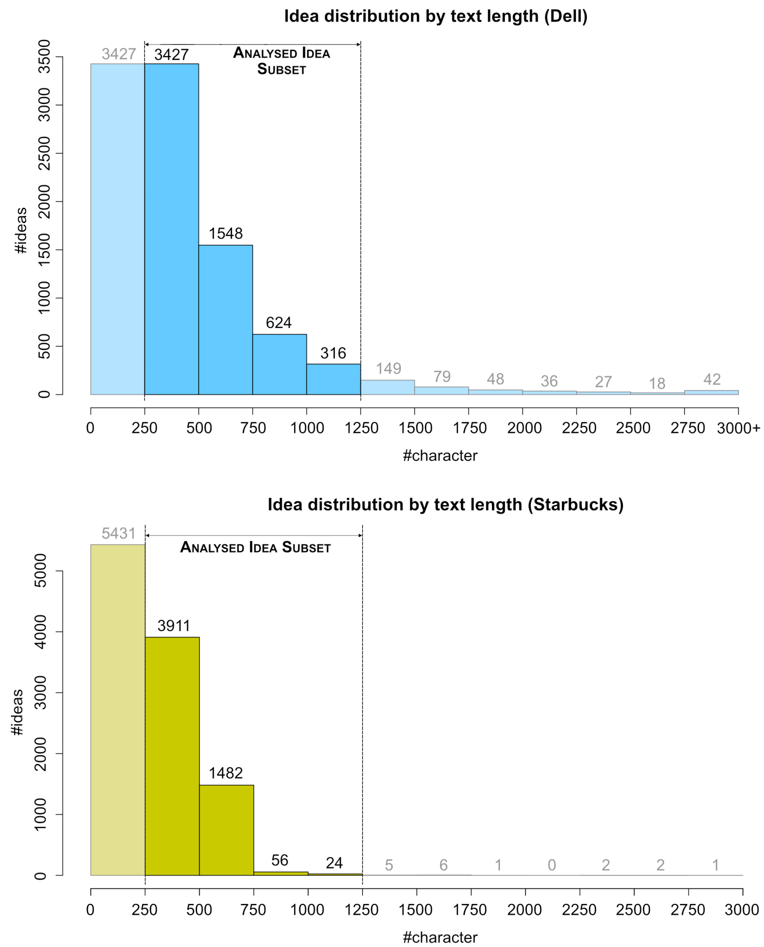


Figure 1 Distribution of idea length in IMS datasets and subsets used during evaluation.

4.1.1 Dataset annotation

In order to evaluate the hypotheses and the degree in which application of outlier detection would be an effective tool we needed a reference assessment for the quality of ideas gathered in the dataset. In prior research on Innovation Management and even Idea Management in specific, there are a number of approaches to this without a single consensus (Ozer, 2002; Girotra et al., 2010; Schulze et al., 2012; Oman et al., 2013). Contrary to the evaluation process proposed by Montoya-Weiss and O'Driscoll (Montoya-Weiss and O'Driscoll, 2000) in Nortel study, we decided to limit the amount of metrics to minimum and simplify them. As shown in past studies: too many metrics structured in a complex and overly formalised taxonomy makes non-expert annotators confused, challenging to reach an agreement regarding the rating and in some cases difficult even to repeat own rating judgements (Westerski et al., 2013). Taken into account our goals and those prior experiences, we performed a manual annotation exercise, where a non-expert annotator having seen only idea title and description would rate it just with one overall rating indicating the *Breakthrough* potential of idea related to idea disruptiveness. The rating was given using 1-10 Likert scale (Likert, 1932).

To verify the capability to objectively rate disruptive innovations we conducted two tests: 1) using a small subset of 10 ideas, we asked 20 annotators to provide ratings and measured inter-rater agreement; 2) we asked the same annotator in time distance of 3 months to rate the same set of ideas twice and checked for agreement as well. In the first experiment with multiple annotators we used an online crowdsourcing platform Mechanical Turk that enables to publish surveys and hire participants to provide answers. We did not put any demographic or knowledge constrains for annotators except of 95% acceptance ratio for their prior annotations done in other surveys. As a result of this experiment, we obtained 0.76 Krippendorf alpha inter-rater agreement equivalent to 84% agreement ratio for Dell ideas, and 0.61 Krippendorf alpha inter-rater agreement equivalent to 78% agreement ratio for Starbucks ideas. According to Taylor and Watkinson (Taylor and Watkinson, 2007), the results obtained are equivalent to excellent agreement for Dell and substantial agreement for Starbucks. In case of second experiment, single annotator repeating his task given 3 month time interval obtained agreement of 100% for Dell and 90% for Starbucks. Given satisfactory results for both tests and having seen such methodology previously applied successfully in similar IMS annotation studies (Westerski et al., 2013), we proceeded with annotation exercise using a single annotator from second test to provide all annotations for each of the datasets.

For applying the breakthrough metric to both of our datasets, we needed a sampling scenario to make the annotation process manageable. In our case, we took two directions to decide which ideas to annotate: (1) based on legacy Idea Management metrics; (2) based on indications from the outlier algorithms used during evaluation. In the first case of legacy metrics, we took top 10, middle 10 and bottom 10 ideas as rated by *idea vote count* and another set as rated by *idea comment count*. That was supplemented by 10 random *implemented* and 10 *unimplemented ideas*. Total of 80 ideas per dataset. Looking from a different angle, outlier algorithm indicators, we took top 10 outlying ideas as pointed by every tested algorithm / configuration combination. In total, this resulted in approximately 1000 ideas annotated per each dataset.

Analysing the outcome of this annotation exercise for all annotations within dataset scope (Fig. 2), we can see from the breakthrough rating distribution there is only a small fraction of very innovative ideas (high breakthrough value). That would indicate that some of the algorithms or legacy metrics did not work very well, in the next section we analyse this in detail and reveal which algorithms managed to correctly indicate the desired small fraction of most valuable ideas with high annotator ratings.

4.2 Evaluation Results

Having both the reference annotations and results of outlier algorithms, the evaluation of the earlier described hypotheses can be seen as a recommender system evaluation problem. In such case, we evaluate to what degree the ranking of ideas by outlier algorithm follows the order established by the reference annotation, corresponding to testing hypothesis (1). Shani and Gunawardana (Shani and Gunawardana, 2011) as well as Schroder et al. (Schroder et al., 2011) point out multiple metrics for such evaluation and present comparative study concluding that across the state of the art the choice of metric is often dependant on particular application, domain and available data. Out of the metrics they highlight, in our study, we

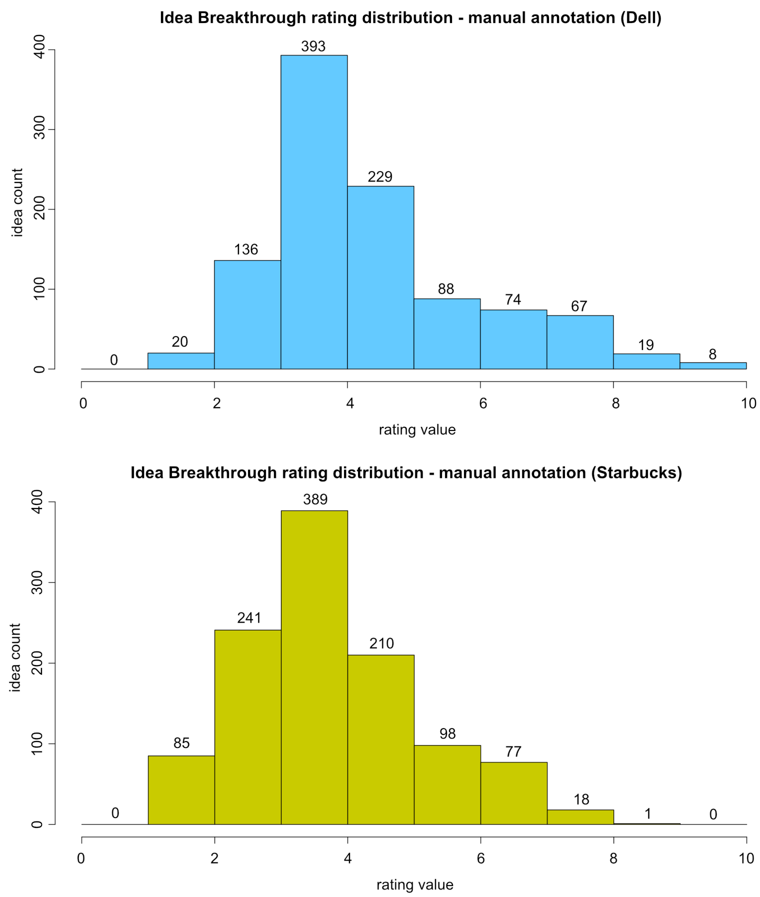


Figure 2 Distribution of breakthrough ratings after annotation exercise.

pick *correlation* as it has been already used in some past Idea Management studies (Gangi and Wasko, 2009; Westerski et al., 2013), as well as *precision@10* which is frequently mentioned by both Shani and Schroder as utilised across multiple past publications. The motivation for using correlation is to know if the overall ordering reflects the expected one across the entire spectrum of idea database, therefore if the value of outlier score is related at all to degree in which idea is disruptive. While precision@10 is used to see if the relationship between those variables is particularly valid for the top outliers, which is what typically reviewers in Idea Management pay attention to.

After applying all algorithms as discussed earlier and calculating correlation metrics, the best performance was achieved for distance based algorithms, regardless of the dataset (see Fig. 3). In particular, in both cases *k-NN* with *TF-IDF* for feature vector generation and *cosine distance* for measuring neighbourhood gave best result - medium correlation with manual annotation scores (using Cohen scale (Cohen, 1988) to interpret 0.28 Pearson correlation for Dell, and 0.32 for Starbucks). On the other side of the spectrum, the worst results were also produced by probabilistic algorithms. Such outcome could be partially attributed to fact that we ran into many difficulties tuning the LDA algorithm given many

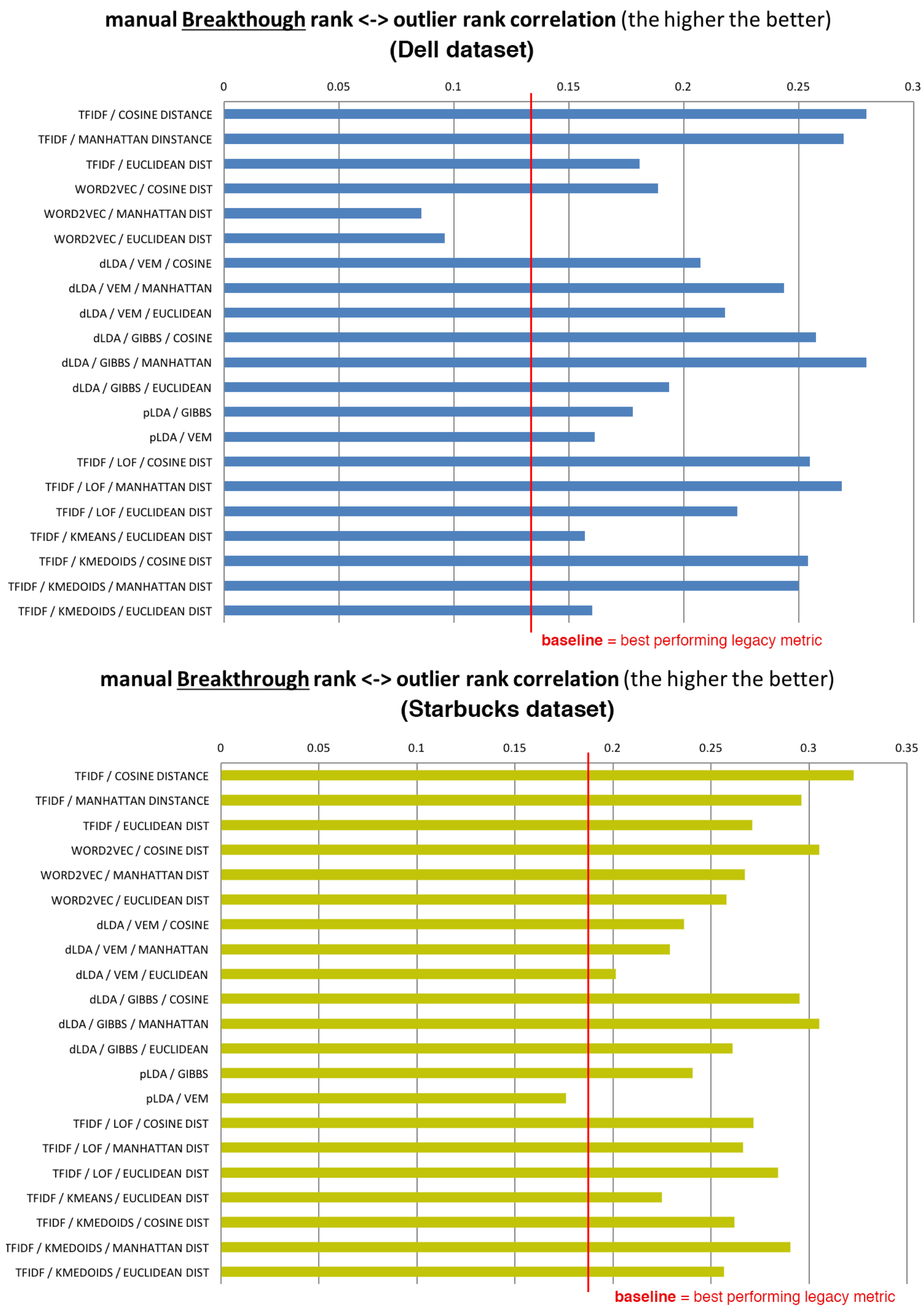


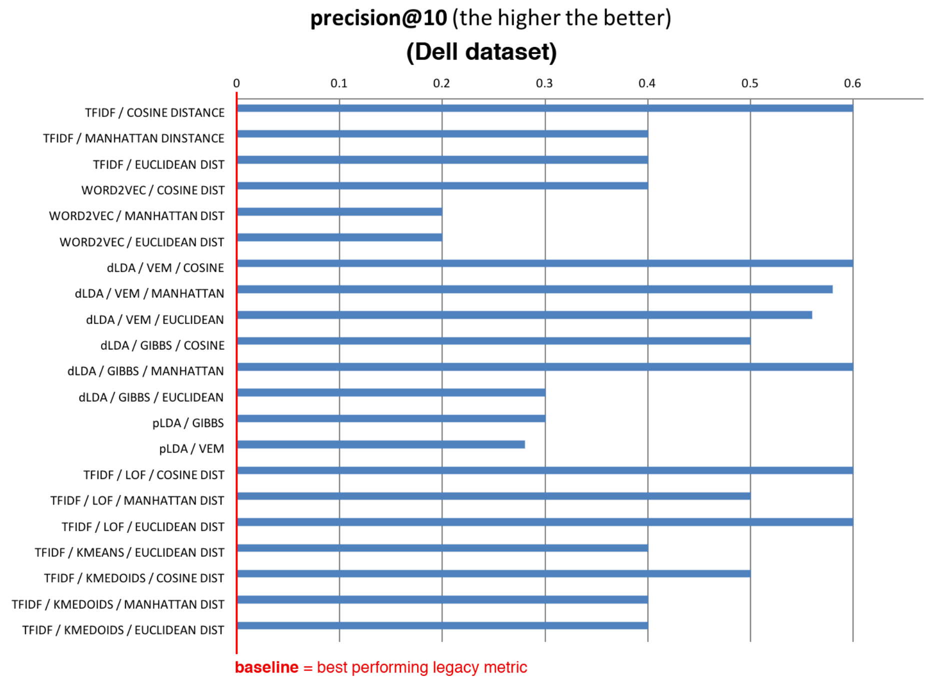
Figure 3 Correlation between outlier ranking and manually rated disruptive ideas: comparison between different outlier detection algorithms.

parameters and the difficulty to assess their best combination that impacts the final result. Further, comparing our results to what could be achieved earlier with legacy metrics, out of all considered at most *vote count* got weak correlation with breakthrough rank (respectively 0.12 and 0.17 for Dell and Starbucks).

Those results got also confirmed by performance of algorithms according to our second test using *precision@10* metric (see Fig. 4). Best score in Dell dataset being 0.6, and 0.3 for Starbucks dataset, in comparison to 0 for all legacy metrics in both datasets. In both cases, same as previously, k-NN being among the top scoring algorithms. The new insight however is that when looking specifically at a very few most innovative breakthrough ideas, many of the outlier detection algorithms seem to be less predictable in terms of performance when applying to different datasets. With correlation results it can be observed that differences were less radical and Dell/Starbucks charts were more similar to each other than in comparison to precision@10 chart. In case of Dell, we can see that density algorithms start to be more competitive when considering only tops of the ranking, while for Starbucks that does not appear to be the case. Such differences could be perhaps explained looking back at the distribution of manual annotations chart (see Fig. 2). In comparison to Dell it seems that Starbucks annotators had a lot more trouble pointing to clear-cut breakthrough innovations (ie. Dell has 94 ideas above rating 7, while Starbucks only 19, making it about 80% difference, while in the lower ranks the quantitive difference is only about 18%). Regardless of this observation however, the conclusions within distance based algorithms (ie. for k-NN and LDA) remain the same for both datasets in both tests. Given such confirmation we might consider both hypotheses as confirmed.

Given a wide scope of our tests with various settings for all algorithms we found some interesting insights revealing particularities of outlier detection in Idea Management context. Specifically, looking at the best performing algorithm (k-NN), contrary to our expectation we got surprisingly bad performance of word2vec when replacing TF-IDF for feature vector generation. Multiple recent publications often show word2vec as improving upon the results of other text modelling approaches (Campr and Jezek, 2015; Baroni et al., 2014). Contrary to those reports, word2vec gave us rather bad results regardless of IMS dataset and test performed. For this reason we investigated it more thoroughly, eventually attributing such outcome to three following reasons: (1) relatively small text size for individual idea; (2) overall small IMS corpus size; and (3) amount of domain dependant words in idea text that could be significant for overall meaning of proposed innovation. Initial results for word2vec as presented on the charts were obtained by training the model only using the related Idea Management text corpus (either Dell or Starbucks depending on experiment). When replacing the training set with more popularly used Google News or Wikipedia corpuses we got slight improvement but still underperforming. Investigating further related literature we can find confirmation in some publications that word2vec does not perform well with those generic training sets when text is highly related to specific domain (Suarez-Paniagua et al., 2015; Yao and Li, 2016; Minarro-Gimenez et al., 2015); when the training corpus is small (Mikolov et al., 2013b) or samples of very short text length (Boom et al., 2015). Those conditions seem to be applicable to typical Idea Management Systems therefore possible explanation for giving edge to other solutions.

Outside of text modelling approach, we also looked more in depth into how distance interpretation would affect final k-NN performance. Comparing in detail different "k" settings, it turned out that fitting it for best performance was to great extent dataset independent further confirming our earlier observation that this algorithm would be quite good for the task regardless of Idea Management dataset used. Comparing our Dell and



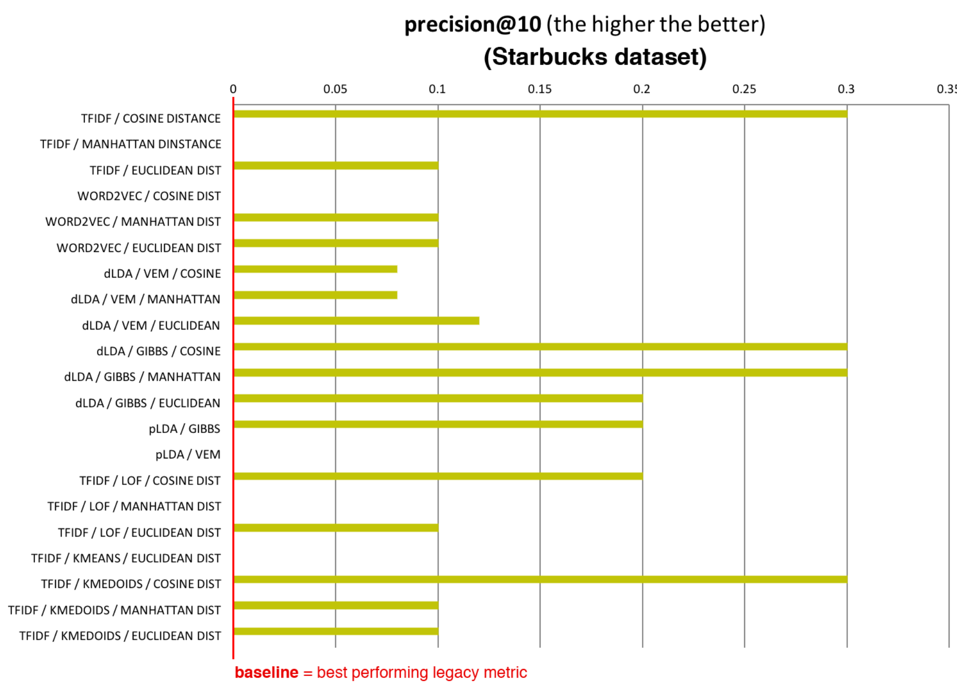


Figure 4 Precision@10 for outlier rankings against manual rated disruptive ideas: comparison between different outlier detection algorithms.

Starbucks results (see Fig. 5) it can be observed that overall behaviour of algorithms in comparison to each other remains the same. Furthermore giving always best results for k being set around 100. Finally comparing to other outlier algorithms that rely on distance measurements (density and clustering), we noticed that for k-NN all three tested variables (text modelling approach, distance metric and neighbourhood setting) were equally important, whereas for clustering, the selection of correct cluster size had more significance than any other parameter.

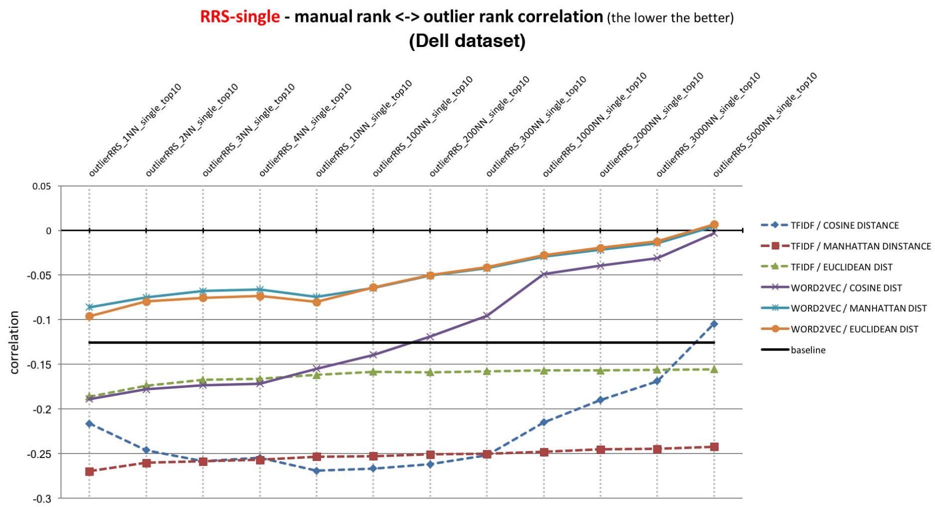
4.2.1 Comparative Analysis of Top Ranking Idea Texts

Apart of a holistic analytical approach using recommender system performance metrics, we also manually investigated the text of top idea picks by outlier algorithms and compared the correctly classified disruptive innovations to those that outlier detection brought to attention incorrectly. This was done to identify the potential problems and shed some light as to why the algorithms did not perform any better.

For ideas ranking high by outlier rank and low by manually obtained breakthrough rank, we noticed the most typical reasons for misclassification were: (1) usage of very informal vocabulary (e.g. slang) or user inventing completely new names therefore making the idea textually very unique but not necessarily innovative; (2) very abstract or not fully serious ideas (e.g. in Dell proposals for making business choices based on superstition, references to witchcraft, wizardry, some ancient practices); (3) simple and obvious ideas but explained in more elaborate ways and in greater detail than by other ideation platform participants (e.g. posts explaining importance of marketing for Dell through psychology theory and references to multiple academic resources etc.); (4) explaining trivial ideas using a lot of domain specific terminology that is rare in other posts (e.g. ideas for Dell to modify its products via comparison to multiple niche products and mentioning many technical parameters with very detailed measures and units rarely mentioned by others). Taken into account all those observed problems, we can see that majority of issues are not related to outlier detection working poor but correctly spotted textual outliers not being interesting innovations. With respect to those observations, the difference between individual outlier algorithms seemed to be related to being more sensitive to the aforementioned textual differences.

To confirm this, moving toward, we also looked at the opposite situation: ideas marked by annotators as highly disruptive but not scoring particularly high in the outlier rankings. The frequently occurring common reasons for those were: (1) very innovative ideas being complex but described with very simple vocabulary frequently used in many other ideas; or (2) some innovative ideas being quite simple in their nature and therefore description also being short and simple (e.g. proposal for Dell to fully shutdown its business, pay off money to shareholders that would give bigger benefit to society).

Finally, to complete the picture, we also checked the positive cases, where ideas scored high in both rankings (manual annotation and automatic outlier detection). Our analysis showed that ideas highlighted by outlier detection correctly were multiple interesting cases that otherwise would be missed just looking at other legacy metrics. This was due to ideas having low community support, not being commented much or getting submitted in a short timeframe together with many other ideas. Those same ideas got spotted by outlier detection because containing words related to unpopular and fresh topics that would bring major shift for businesses of either Dell or Starbucks (e.g. going into politics and taking very specific stances on certain public interest topics; involving cutting edge technologies related to university research; or controversial ideas for radical shift from current business models).



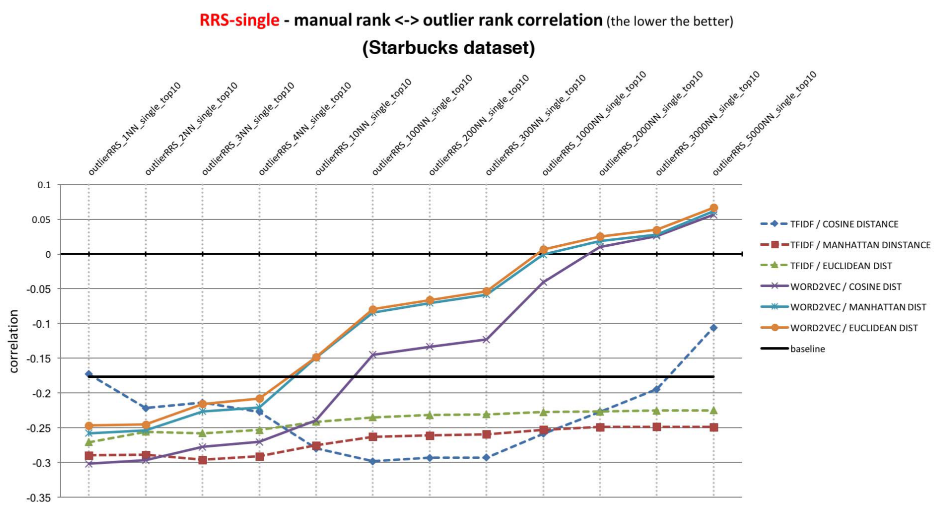


Figure 5 Distance based algorithms (k-NN-RRS) - comparison of different configuration vs. performance.

Overall, comparing the positive cases with the misclassified ones as presented in previous paragraphs, it seems like reasons for either are quite similar - both result in correctly working outlier detection but on sometimes imperfect textual representation. This could lead to a conclusion that perhaps instead of improving on the outlier detection algorithm a better underlying text model that expresses semantics and degree of connection to IMS key business domain might have more impact on overall performance. Evaluation for this type of in-depth understanding of text is outside of the scope of this paper, however we discuss such and other related work in the next section.

5 Related Work

Idea Management Systems have been present as a software solution similar to its current crowdsourcing shape since the late 90s (Rozwell et al., 2002). Being used in practice of enterprise innovation efforts, this technology has had multiple success stories (Jouret, 2009; Bjelland and Wood, 2008; Bailey and Horvitz, 2010) but also multiple problems have surfaced (Klein and Convertino, 2015). Outside of the best idea selection problem described in this paper, some other investigated areas are idea generation (Blohm et al., 2011; Morgan and Wang, 2010), idea refinement and co-creation (Klein and Iandoli, 2008; Shah et al., 2001; Vivacqua et al., 2010). However, the problem taken up by us is also the most predominant in scientific literature with multiple solutions inspired by a variety of side domains. Similar to us, some scientists have observed that typical community metrics are not sufficient and therefore a need to invent a new idea rating approach. Among those, Bothos et al. (Bothos et al., 2008) followed a very different direction to ours, by increasing the community engagement and relying on popular opinion even more via prediction markets. In another extreme, multiple other authors advocate that crowdsourcing data gathered in IMS is insufficient for accurate rating and tap into metrics that originate from different enterprise systems (Ning et al., 2006; Westerski and Iglesias, 2011) or manual annotation methodologies (Conn et al., 2009). In comparison to those research directions, we see our proposal in between the two extremes: still relying on the minimum of community created information (idea text analysis), yet not connecting to behavioural metrics such as Bothos; at the same time, we try to avoid complicated connections to other parts of enterprise where software and metadata integration problems are a frequent weaknesses. Also, in contrast to both of those aforementioned approaches, our proposal does not rely on creating any new information but uses only what is already present in any, even most simplistic IMS. In that regard, there are some even more conservative researchers that avoid inventing any new idea rating methods at all but instead focus on improving the existing metrics or analysing in more detail the origin of problems and nature of IMS with current metrics (Hrastinski et al., 2010; Gangi and Wasko, 2009).

While to our knowledge non of this prior art in IMS research connects to the contexts of disruptive innovation or application of outlier detection, some authors have turned into other forms of data analytics algorithms for expanding Idea Management Systems capabilities or as a tool to help analyse case studies. Helander et al. (Helander et al., 2007) present an analysis of IBM open innovation exercises and use clustering with underlying TF-IDF to connect two consecutive phases of ideation to backtrack on origins of winning ideas. Same authors also turn to supervised learning and seek to analyse key influencing features of winning ideas to arrive at similar conclusions to us confirming that currently available metadata in IMS is not sufficient for making such decisions. Yoo et al. (Yoo et al., 2015) also use both supervised and unsupervised learning but for end user purposes rather than case study analysis. Their conclusions although reached in a different manner, also suggest that for further improvement more semantic relationships between ideas and context would be needed. Such ontological solutions have been proposed before in IMS research (Riedl et al., 2009; Westerski et al., 2010), but non fully successful in delivering an analytical solution leveraging the complex models. Furthermore, many of the authors proposing such schemas for semantic description of idea content and relationships note the difficulties in efficient annotation by either manual or automatic means of existing idea databases (Westerski et al., 2010).

A common point for those semantic models and our approach is the frequent reference to Innovation Management studies as core source of inspiration. While Idea Management Systems research described here predominantly is done by computer scientists, Innovation Management area is more focused on business and economic approach to innovation and its impact on organisation management (Hidalgo and Albors, 2008). Still, in spite of those differences, the problems of finding best innovations given certain organization or market context have been equally if not more explored in Innovation Management (Afuah, 1998). Concepts mentioned by us frequently in this article such as open innovation (Chesbrough, 2003), ability to increment on other ideas or by radical changes (Schumpeter, 1942; Kirzner, 1973) before making their way into IMS were originally proposed in those business studies. Similarly, the corner stone for research described in this paper - notion of disruptive innovation, was originally proposed by Bower and Christensen (Bower and Christensen, 1995) and adopted by us in a way earlier explained in Section 2.

To bring those innovation management theories to the practice of Idea Management Systems we tapped into more technical domain of data analytics. Thus the relationship to outlier detection techniques. Technology that has been very well established in past years in application to variety of datasets and domains (Aggarwal, 2013). According to our exploration not used before in particular context of Idea Management, however multiple times utilised for analysis of other textual content. The bulk of those efforts have been related to Topic Detection and Tracking (TDT) series of events organised between 1998 and 2004 (Allen, 2002). The majority of algorithms evaluated in this paper are direct references to the achievements of that TDT community. This paper however advances the domain by adding one more layer of evaluation related to application area (idea quality) rather than stopping at measuring the quality of outliers. Interestingly, some of our discoveries mentioned in evaluation section are in line with the final conclusions from TDT workshops and benchmarking competitions showing that best results in textual outlier detection are achieved by combination of TF-IDF with outlier detection algorithm tailored for particular problem domain. Such steps of adopting or modifying TDT techniques have been also proposed by other scientists. While in this paper we excluded very short text contributions, Petrovic et al. (Petrovic et al., 2010) focused exactly on this type of content analysing novelties in Twitter posts; Smith (Smith, 2009) went also in a different direction extending the analysed media types beyond text to video and voice in order to label different sections of movies. Since such types of content are of limited quantity in most Idea Management Systems and analysing them would bring more complexity to the problem, we considered the task in terms of future work as described in next section.

6 Conclusions and future work

In this article we proposed a new idea rating scheme based on analysis of ideas disruptive innovation capacity. We have used a novel set of analytic tools, outlier detection algorithms, which to our knowledge were previously not evaluated in the context of Idea Management domain. A series of experiments have concluded that certain algorithms are more suitable for this task. Specifically, within distance based outlier detection: k-NN algorithm has shown best performance and capability to maintain it regardless of dataset used. In contrast, the probabilistic outlier detection methods have proven worst and very difficult to tune. Regardless of the individual algorithm performance, the outlier detection in all our tests has shown superior results with regard to disruptive innovation detection than any prior legacy metrics of Idea Management Systems. This confirmed our hypotheses that metrics typically used in modern day Idea Management software have little to do with discovering disruptive innovations due to their strong relation to idea popularity and need to satisfy established customer base desires.

We supplemented this study with experiments on multiple different outlier detection algorithm configurations, amongst others manipulating algorithm input with different idea text representations and further testing multiple idea similarity metrics. With that regard, our results showing TF-IDF with cosine distance as best choice, substantiate findings of the text outlier detection community and prove that such techniques can applied to Idea Management domain with similar success.

Investigating closer the output of best performing algorithms, we looked into the reasons why certain innovations were recommended incorrectly while other omitted. This is because our end goal was not just mere outlier detection but more ambitious disruptive innovation detection. The conclusions of this study were that the outlier detection algorithms worked correctly to the capacity of delivered input data, however the interpretation of disruptive innovation did not always coincide with the fact of idea being an outlier. The reason for this was algorithms operating on simple textual differences and lacking a deeper understanding for idea meaning in its business context.

Based on the results achieved, we see the lack of such semantic text representation as the biggest limitation of our study. Therefore, we propose modifications of text representation a good direction for future work rather than further manipulation with outlier detection techniques.

References

Afuah, A. (1998). *Innovation management: Strategies, implementation, and profits.* Oxford University Press, New York.

Agarwal, N., Yiliyasi, Y. (2010). “Information quality challenges in social media”, in *Proceedings of the 15th International Conference on Information Quality*, Curran Associates, Inc., Red Hook, NY. pp. 234 – 249.

Aggarwal, C.C. (2013). *Outlier Analysis*. Springer, New York.

Allen, J. (2002). *Topic Detection and Tracking: Event-based Information Organization*. Kluwer Academic Publishers Norwell, MA, New York.

Bailey, B.P., Horvitz, E. (2010). “What’s your idea? a case study of a grassroots innovation pipeline within a large software company”, in *CHI ’10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York. pp. 2065 – 2074.

Baroni, M., Dinu, G., Kruszewski, G. (2014). “Don’t count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors”, in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, ACL, Baltimore, Maryland. pp. 238 – 247.

Becheikh, N., Landry, R., Amara, N. (2006). “Lessons from innovation empirical studies in the manufacturing sector: A systematic review of the literature from 1993-2003”. Technovation, Vol. 26 No. 5–6, pp. 644 – 664.

Bjelland, O.M., Wood, R.C. (2008). "An inside view of ibm's innovation jam". MIT Sloan Management Review, Vol. 50 No. 1, pp. 32 – 40.

Blei, D.M., Ng, A.Y., Jordan, M.I. (2003). "Latent dirichlet allocation". Journal of Machine Learning Research, Vol. 3, pp. 993–1022.

Blohm, I., Bretschneider, U., Leimeister, J.M., Krcmar, H. (2011). "Does collaboration among participants lead to better ideas in it-based idea competitions? an empirical investigation". International Journal of Networking and Virtual Organisations, Vol. 4 No. 2, pp. 106 – 122.

Boom, C.D., Canneyt, S.V., Bohez, S., Demeester, T., Dhoedt, B. (2015). "Learning semantic similarity for very short texts", in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, IEEE Computer Society, Washington. pp. 1229 – 1234.

Bothos, E., Apostolou, D., Mentzas, G. (2008). "Idem: A prediction market for idea management. designing e-business systems. markets, services, and networks", in *7th Workshop on E-Business, WeB 2008*, Springer, Paris, France. pp. 1 – 13.

Bower, J.L., Christensen, C.M. (1995). "Disruptive technologies: Catching the wave". Harvard Business Review, Vol. January–February No. 1, pp. 303 – 312.

Brem, A., Voigt, K.I. (2007). "Innovation management in emerging technology ventures - concept of an integrated idea management". International Journal of Technology, Policy and Management, Vol. 7 No. 3, pp. 304 – 321.

Breunig, M., Kriegel, H.P., Ng, R.T., Sander, J. (2000). "Lof: Identifying density-based local outliers", in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, ACM, New York. pp. 93 – 104.

Campos, G.O., Zimek, A., Sander, J., Campello, R.J.G.B., Micenkov, B., Schubert, E., Assent, I., Houle, M.E. (2016). "On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study". Data Mining and Knowledge Discovery, Vol. 30 No. 4, pp. 891 – 927.

Campr, M., Jezek, K. (2015). "Comparing semantic models for evaluating automatic document summarization", in *International Conference on Text, Speech, and Dialogue. TSD 2015*, Springer, Cham. pp. 252 – 260.

Chandola, V., Banerjee, A., Kumar, V. (2009). "Anomaly detection: A survey". ACM Computing Surveys, Vol. 41 No. 3, pp. 15:1 – 15:58.

Chesbrough, H.W. (2003). *Open Innovation: The new imperative for creating and profiting from technology*. Harvard Business Press, USA.

Christensen, C.M. (1997). *The Innovator's Dilemma*. Harvard Business School Press, USA.

Christensen, C.M., Raynor, M.E., McDonald, R. (2015). "What is disruptive innovation?". Harvard Business Review, Vol. 93 No. 12, pp. 44 – 53.

Christensen, K., Norskov, S., Frederiksen, L., Scholderer, J. (2017). "In search of new product ideas: Identifying ideas in online communities by machine learning and text mining". Creativity and Innovation Management, Vol. 26 No. 1, pp. 17 – 30.

Chuang, L.M., Liu, C.C., Tsai, W.C., Huang, C.M. (2010). "Towards an analytical framework of organizational innovation in the service industry". African Journal of Business Management, Vol. 4 No. 5, pp. 790–799.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum.

Conn, S.J., Torkkeli, M.T., Bitran, I. (2009). "Assessing the management of innovation with software tools: an application of innovation enterprizer". International Journal of Technology Management, Vol. 45 No. 3/4, pp. 323–336.

Dean, D.L., Hender, J.M., Rodgers, T.L., Santanen, E.L. (2006). "Identifying quality, novel, and creative ideas: Constructs and scales for idea evaluation". Journal of the Association for Information Systems, Vol. 7 No. 10, pp. 646–699.

European Commission (2005). *Oslo Manual. Guidelines for Collecting and Interpreting Innovation Data*. OECD Publishing.

Fagerberg, J., Verspagen, B. (2009). "Innovation studies - the emerging structure of a new scientific field". Research Policy, Vol. 38 No. 2, pp. 218–233.

Gangi, P.M.D., Wasko, M. (2009). "Steal my idea! organizational adoption of user innovations from a user innovation community: A case study of dell ideastorm". Decision Support Systems, Vol. 48 No. 1, pp. 303–312.

Garcia, R., Calantone, R. (2002). "A critical look at technological innovation typology and innovativeness terminology: a literature review". The Journal of Product Innovation Management, Vol. 19 No. 2, pp. 110–132.

Girotra, K., Terwiesch, C., Ulrich, K.T. (2010). "Idea generation and the quality of the best idea". Management Science, Vol. 56 No. 4, pp. 591–605.

Google (2009). "Announcing project 10 100 idea themes. google official blog". available at: `https://googleblog.blogspot.com/2009/09/announcing-project-10100-idea-themes.html` (accessed 20 July 2018).

Helander, M., Lawrence, R., Liu, Y., Perlich, C., Reddy, C., Rosset, S. (2007). "Looking for great ideas: Analyzing the innovation jam", in *9th International Workshop on Knowledge Discovery on the Web, WebKDD 2007*, Springer, Berlin. pp. 21–39.

Hidalgo, A., Albors, J. (2008). "Innovation management techniques and tools: a review from theory and practice". R&D Management, Vol. 38 No. 2, pp. 113–127.

Hossaina, M., Islam, K.M.Z. (2015a). "Generating ideas on online platforms: A case study of my starbucks ideaŁž". Arab Economic and Business Journal, Vol. 10 No. 3, pp. 102–111.

Hossaina, M., Islam, K.M.Z. (2015b). "Ideation through online open innovation platform: Dell ideastorm". Journal of the Knowledge Economy, Vol. 6 No. 3, pp. 611–624.

Hrastinski, S., Kviselius, N.Z., Edenius, M. (2010). "A review of technologies for open innovation: Characteristics and future trends", in *Proceedings of the 43rd Hawaii International Conference on System Sciences*, IEEE, USA. pp. 1–10.

Jouret, G. (2009). "Inside cisco's search for the next big idea". Harvard Business Review, Vol. 87 No. 9, pp. 43–45.

Kirzner, I.M. (1973). *Competition and Entrepreneurship*. University of Chicago Press, Chicago.

Klein, M., Convertino, G. (2015). "A roadmap for open innovation systems". Journal of Social Media, Vol. 2 No. 1, pp. 1–16.

Klein, M., Iandoli, L. (2008). "Supporting collaborative deliberation using a large-scale argumentation system: The mit collaboratorium", in *Directions and Implications of Advanced Computing; Conference on Online Deliberation (DIAC-2008/OD2008)*, University of California, Berkeley. pp. 5–12.

Knorr, E.M., Ng, R.T., Tucakov, V. (2000). "Distance-based outliers: algorithms and applications". VLDB Journal, Vol. 8 No. 3–4, pp. 237–253.

Lee, H., Choi, K., Yoo, D., Suh, Y., Lee, S., He, G. (2018). "Recommending valuable ideas in an open innovation community: A text mining approach to information overload problem". Industrial Management & Data Systems, Vol. 118 No. 4, pp. 683–699.

Likert, R. (1932). "A technique for the measurement of attitudes". Archives of Psychology, Vol. 22 No. 140, pp. 5–55.

Lu, Y., , Mei, Q., Zhai, C. (2011). "Investigating task performance of probabilistic topic models: an empirical study of plsa and lda". Information Retrieval, Vol. 14 No. 2, pp. 178–203.

Martinez-Torres, M.R. (2015). "Content analysis of open innovation communities using latent semantic indexing". Technology Analysis & Strategic Management, Vol. 27 No. 7, pp. 303–312.

Merz, A.B. (2018). "Mechanisms to select ideas in crowdsourced innovation contests - a systematic literature review and research agenda", in *26th European Conference on Information Systems (ECIS 2018)*, UK.

Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013a). "Efficient estimation of word representations in vector space". ArXiv e-prints, `arXiv:1301.3781`.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J. (2013b). "Distributed representations of words and phrases and their compositionality", in *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, Curran Associates Inc., USA. pp. 3111–3119.

Minarro-Gimenez, J., Marian-Alonso, O., Samwald, M. (2015). "Applying deep learning techniques on medical corpora from the world wide web: a prototypical system and evaluation". ArXiv e-prints, `arXiv:1502.03682`.

Montoya-Weiss, M.M., O'Driscoll, T.M. (2000). "From experience: Applying performance support technology in the fuzzy front end". Product Innovation Management, Vol. 17 No. 2, pp. 143–161.

Morgan, J., Wang, R. (2010). "Tournaments for ideas". California Management Review, Vol. 52 No. 2, pp. 77–97.

Mulgan, G., Tucker, S., Ali, R., Sanders, B. (2007). *Social Innovation: what it is, why it matters and how it can be accelerated*. Basingstoke Press, UK.

Ning, K., O'Sullivan, D., Zhu, Q., Decker, S. (2006). "Semantic innovation management across the extended enterprise". International Journal on Industrial and Systems Engineering, Vol. 1 No. 1/2, pp. 109–128.

Oman, S.K., Tumer, I.Y., Wood, K., Seepersad, C. (2013). "A comparison of creativity and innovation metrics and sample validation through in-class design projects". Research in Engineering Design, Vol. 24 No. 1, pp. 65–92.

Osborne, S.P., Brown, L. (2013). *Handbook of Innovation in Public Services*. Edward Elgar, UK.

Ozer, M. (2002). *What do we know about new product idea selection*. working paper. City University of Hong Kong.

Petrovic, S., Osborne, M., Lavrenko, V. (2010). "Streaming first story detection with application to twitter", in *HLT '10 Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, ACL, USA. pp. 181–189.

Phan, X., Nguyen, L., Horiguchi, S. (2008). "Learning to classify short and sparse text & web with hidden topics from large-scale data collections", in *Proceedings of the 17th international conference on World Wide Web*, ACM, USA. pp. 91–100.

Pimentel, M.A., Clifton, D.A., Clifton, L., Tarassenko, L. (2014). "A review of novelty detection". Signal Processing, Vol. 99, pp. 215–249.

Popadiuka, S., Choo, C.W. (2006). "Innovation and knowledge creation:how are these concepts related?". International Journal of Information Management, Vol. 26 No. 4, pp. 302– 312.

Ramaswamy, S., Rastogi, R., Shim, K. (2000). "Efficient algorithms for mining outliers from large data sets", in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, ACM, USA. pp. 427–438.

Riedl, C., May, N., Finzen, J., Stathel, S., Kaufman, V., Krcmar, H. (2009). "An idea ontology for innovation management". International Journal on Semantic Web and Information Systems, Vol. 5 No. 4, pp. 1–18.

Rozwell, C., Harris, K., Caldwell, F. (2002). *Survey of Innovation Management Technology*. working paper. Gartner.

Salton, G. (1983). *Introduction to Modern Information Retrieval*. Mcgraw-Hill College, USA.

Sanchez, A., Lago, A., Ferras, X., Ribera, J. (2011). "Innovation management practices, strategic adaptation, and business results: Evidence from the electronics industry". Journal of Technology Management & Innovation, Vol. 6 No. 2, pp. 14–39.

Schroder, G., Thiele, M., Lehner, W. (2011). "Setting goals and choosing metrics for recommender system evaluations", in *UCERSTI 2 Workshop, at the 5th ACM Conference on Recommender Systems*, ACM, USA. pp. 78–85.

Schulze, T., Indulska, M., Geiger, D., Korthaus, A. (2012). “Idea assessment in open innovation: A state of practice”, in *Proceedings of the 20th European Conference on Information Systems*, Spain.

Schumpeter, J. (1942). *Capitalism, Socialism, and Democracy*. Harper, New York.

Shah, J.J., Vargas-Hernandez, N., Summers, J.D., Kulkarni, S. (2001). “Collaborative sketching (c-sketch) - and idea generation technique for engineering design”. California Management Review, Vol. 35 No. 3, pp. 168–198.

Shani, G., Gunawardana, A. (2011). *Evaluating Recommendation Systems*. Springer, USA.

Singh, K., Upadhyaya, S. (2012). “Outlier detection: Applications and techniques”. IJCSI International Journal of Computer Science Issues, Vol. 9 No. 1, pp. 307–323.

Smith, G. (2009). *Video scene detection using closed caption text*. Virginia Commonwealth University, USA.

Suarez-Paniagua, V., Segura-Bedmar, I., Martinez, P. (2015). “Word embedding clustering for disease named entity recognition”, in *Proceedings of the Fifth BioCreative Challenge Evaluation Workshop*, pp. 299–304.

Taylor, J., Watkinson, D. (2007). “Indexing reliability for condition survey data”. The Conservator, Vol. 30 No. 1, pp. 49–62.

Tornjanski, V., Cudanov, M., Savoiu, G. (2015). “A holistic approach to innovation management in nanking: A review”. Econophysics, Sociophysics & Other Multidisciplinary Sciences Journal (ESMSJ), Vol. 5 No. 2, pp. 8–15.

Vivacqua, A.S., Expedito, C., Galuzzo, F. (2010). “Moving from ideas to proposals”, in *CSCW10 Workshop on Collective Intelligence In Organizations: Toward a Research Agenda*, ACM, USA.

Westerski, A., Dalamagas, T., Iglesias, C.A. (2013). “Classifying and comparing community innovation in idea management systems”. Decision Support Systems, Vol. 54 No. 3, pp. 1316–1326.

Westerski, A., Iglesias, C.A. (2011). “Exploiting structured linked data in enterprise knowledge management systems: An idea management case study”, in *Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2011 15th IEEE International*, IEEE, USA. pp. 395–403.

Westerski, A., Iglesias, C.A., Nagle, T. (2011). “The road from community ideas to organisational innovation: A life cycle survey of idea management systems”. A Special Issue of the Journal Web-Based Communities. Community-based Innovation: Designing Shared Spaces for Collaborative Creativity., Vol. 7 No. 4, pp. 493–506.

Westerski, A., Iglesias, C.A., Rico, F.T. (2010). “A model for integration and interlinking of idea management”, in *Metadata and Semantic Research: 4th International Conference, MTSR 2010*, Springer, Germany. pp. 183–194.

Windrum, P., Koch, P. (2008). *Innovation in Public Sector Services: Entrepreneurship, Creativity and Management*. Edward Elgar, UK.

Yao, Y., Li, G. (2016). “Context-aware sentiment word identification: sentiword2vec”. ArXiv e-prints, `arXiv:1612.03769`.

Yoo, D., Choi, K., Lee, H., Suh, Y. (2015). “An ontology-based co-creation enhancing system for idea recommendation in an online community”. The DATA BASE for Advances in Information Systems, Vol. 46 No. 3, pp. 303– 312.

Zhang, J. (2013). “Advancements of outlier detection: a survey”. ICST Transactions on Scalable Information Systems, Vol. 13 No. 01–03, pp. 1–26.