

A decision support system for chemical R&D analytics and portfolio optimization

Nikolaos Ploskas, Tong Zhang, Nick Sahinidis
Carnegie Mellon University

Flor Castillo and Krishnan Sankaranarayanan
SABIC Americas



Carnegie Mellon



Overview

- Problem statement
- Web based decision support system
- Patent analysis and patent evaluation
 - Patent search and retrieval
 - Evaluation and ranking of patents
 - Case studies
- Portfolio optimization
 - Integrate impact of social media using text mining methods
- Conclusions

Problem statement

- Develop a web-based decision support system for portfolio optimization
 - Gather most relevant patents for each project
 - Evaluate and rank patents
 - Estimate resource allocation
 - Optimize portfolio of projects

Decision support system

- The DSS has been implemented using PHP, MySQL, Ajax and jQuery



Patent analysis and mining

- Patents contain a wealth of information about technological progress and market trends
- Patents have a well-defined format but they are lengthy and include many technical terms
- Patent analysis and mining
 - Finding, processing and analyzing patents
 - Patent search and retrieval
 - Patent visualization
 - Patent evaluation

Proposed methodology

- Define criteria for patent evaluation
- Define weights for the criteria
- Use NLP methods to search and retrieve relevant patents for each project (Quid)
- Apply an MCDM method to evaluate and rank patents for each project (custom web-based DSS)
- Optimize portfolio of projects (custom web-based DSS, GAMS)

Criteria selection

- Criteria related to a patent's connections
 - Degree
 - Betweenness centrality
 - Inter-cluster fraction
- Criteria related to a patent's neighbors
 - Flow
 - Pagerank
 - Triangles
- Criteria related to a patent's citations
 - Forward citations
 - Backward citations

Linear optimization model

- Find weights by linear programming model

$$\text{Min : } \sum_i s_i$$

$$\text{s.t : } w_D + w_B + w_I - w_F - w_P - w_T + s_1 \geq 0$$

$$w_D + w_B + w_I - w_{FC} - w_{BC} + s_2 \geq 0$$

$$w_F + w_P + w_T - w_{FC} - w_{BC} + s_3 \geq \varepsilon$$

$$w_{FC} + w_{BC} + s_4 \leq 0.25$$

$$w_D - w_B - w_I + s_5 \geq 0$$

$$w_B - w_I + s_6 \geq 0$$

$$w_P - w_F - w_T + s_7 \geq \varepsilon$$

$$w_F - w_T + s_8 \geq \varepsilon$$

$$w_{FC} - 2w_{BC} + s_9 \geq 0$$

$$0.05 \leq w_k + s_k \leq 0.33, \quad k \in \{D, B, I, F, P, T, FC, BC\}$$

$$w_D + w_B + w_I + w_F + w_P + w_T + w_{FC} + w_{BC} = 1$$

$$w_D, w_B, w_I, w_F, w_P, w_T, w_{FC}, w_{BC} \geq 0$$

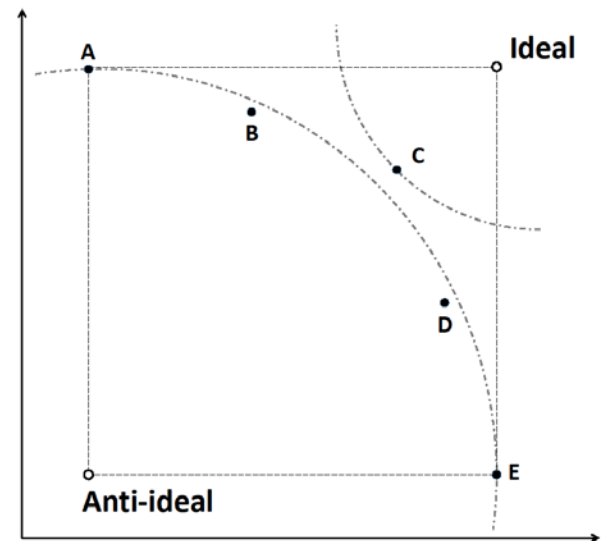
$$s_i \geq 0, \quad \text{for all } i$$

Optimal weights

Variable	Criterion	Weight	
		Solution A	Solution B
w_D	Degree	0.3300	0.3300
w_B	Betweenness centrality	0.0500	0.0500
w_I	Inter-cluster fraction	0.0500	0.0500
w_F	Flow	0.0825	0.0600
w_P	Pagerank	0.1650	0.3100
w_T	Triangles	0.0725	0.0500
w_{FC}	Forward citations	0.1667	0.1000
w_{BC}	Backward citations	0.0833	0.0500

TOPSIS method

- Step 1. Calculation of the weighted normalized decision matrix
- Step 2. Determination of the ideal and anti-ideal solutions
- Step 3: Calculation of the distance from the ideal and anti-ideal solutions
- Step 4: Calculation of the relative closeness to the ideal solution
- Step 5: Ranking the alternatives



Case studies

- Three case studies were conducted
 - Ammonia process synthesis
 - Olefin synthesis
 - Polyethylene synthesis
- Three different sets of weights used
 - The two solutions from the linear optimization model
 - A set of equal weights (naive weights)

Ammonia process synthesis

Set of weights	Top five patents				
	First	Second	Third	Fourth	Fifth
A	5	100	137	48	96
B	5	100	137	48	96
C	137	5	20	100	48

- 166 patents exported from Quid
- Most highly-ranked patents are very important
- Patents 5, 100 and 137 are dealing with generating syngas used in ammonia process
- Patent rankings do not change drastically with different weights

Olefin synthesis

Set of weights	Top five patents				
	First	Second	Third	Fourth	Fifth
A	75	76	22	45	55
B	75	22	76	45	17
C	75	45	22	17	76

- 102 patents exported from Quid
- Most highly-ranked patents are very important
- Patent 75 claims a new catalyst for producing olefin-based polymer
- Patents 17 and 45 are also related with catalyst composition
- Patent rankings do not change drastically with different weights

Polyethylene synthesis

Set of weights	Top five patents				
	First	Second	Third	Fourth	Fifth
A	36	32	37	38	93
B	36	32	37	38	33
C	93	36	67	38	32

- 116 patents exported from Quid
- Most highly-ranked patents are very important
- Patents 32 and 36 describe processes that produce high molecular weight polyethylene
- Patent 38 is related with the production of polyethylene powder
- Patent rankings do not change drastically with different weights

Portfolio optimization

- Create a portfolio of projects
- Find the most promising patents for all projects
- Enter resource allocation information
- Integrate impact of social media
- Export the result report for further processing

Patents selection

- Find the most promising patents for all projects
 - Patents are selected by ranking results
 - Patents cover diverse topics
 - Patents selection is limited by available resources
- Patents found from news trends
 - Utilize news trends using text mining methods
 - ngrams: most frequent words/phrases in a text
 - Find ngrams from top-5 news for all projects
 - Find patents with largest semantic similarity

Portfolio optimization

- Develop a mixed integer linear programming optimization model for portfolio optimization

Max : Risk-adjusted net present value

s.t : Budget limit

Full time employee cost

Project selection

Topic covering

Patents selection by ranking

Patents selection by news trends

Patents selection limit

Portfolio optimization example

- Budget: \$10,000,000
- Patent cost: \$ 1,500,000
- Patents from news trends (20%): 1
- Patents from the portfolio optimization: 5
- Maximum number of patents that can be found: 6

Portfolios

General information Projects

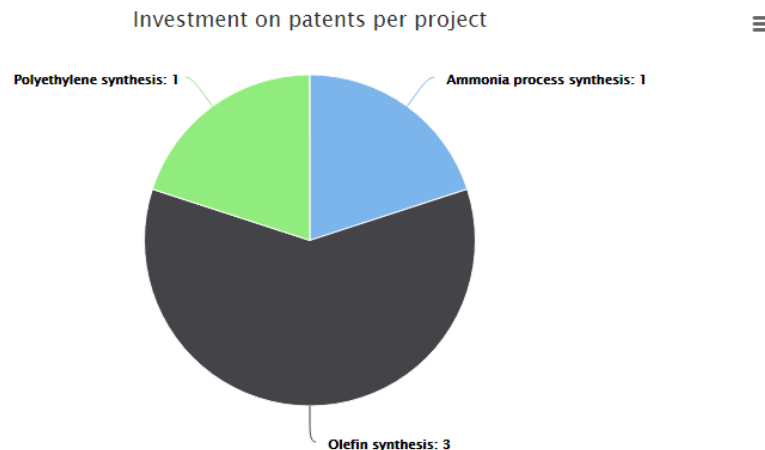
Field	Value	Field	Value
Name:	<input type="text" value="Portfolio 1"/>	Budget:	<input type="text" value="9000000"/>
Description:	<input type="text" value="portfolio including the following projects: ammonia, olefin, polyethylene"/>		
Employee per patent:	<input type="text" value="6"/>	Cost per employee:	<input type="text" value="250000"/>
Percentage of patents that are closely related with news trends:	<input type="text" value="20%"/>		

Save

Portfolio optimization result

- Export a report with the results for further processing

Investment on patents per project			
Project	Patent	Cost	FTEs
Ammonia process synthesis	Removal of carbon dioxide from ammonia synthesis reforming gas by countercurrent contacting ammonia synthesis reforming gas with absorbing liquor, producing large centrifugal force, and contacting absorbing liquor with gas	\$1,500,000	6
Olefin synthesis	New solid insoluble hydrocarbon catalyst component containing magnesium, titanium, halogen, and internal electron donor having carbonic acid dimethyl ester structure useful in polymerization of olefins for producing olefin-based polymer	\$1,500,000	6
	Manufacturing turbine and/or diesel fuels, comprises dehydrating alcohol with catalyst to give olefin mixture, oligomerizing mixture, separating unreacted olefins, converting unreacted olefins, and hydrogenating and distilling oligomers	\$1,500,000	6
	Process for cracking heavy crude oil in steam cracker e.g. pyrolysis furnace, to produce olefin e.g. ethylene, involves separating hydro-processed product into gaseous and liquid products, and passing liquid product to vaporization unit	\$1,500,000	6
Polyethylene synthesis	Producing high molecular weight polyethylene comprises contacting ethylene under polymerization conditions with a slurry of a catalyst composition comprising a Group 4 metal complex of a phenolate ether ligand	\$1,500,000	6
Total	5 patents	\$7,500,000	30



Conclusions

- Web-based DSS automates our methodology
- A novel methodology to rank patents
 - Multiple criteria
 - Intuitive linear optimization formulation determines weights
- Three case studies in chemical R&D analytics
 - Highly-ranked patents represent core processes in their field
 - Rankings are relatively consistent for our two optimal sets of weights
- Portfolio optimization
 - Find the most promising patents for all projects
 - Integrate impact of social media by text mining