2019 KPMG

Ideation Challenge

KG Implementation in Health Insurance

Hall of Fame

王绍兴 软件工程
谭智洋 金融学
邱泽鸿 计算机科学与技术
黄宸 统计学
林诗音 会计学

Agenda

Part I : Business Analysis

- Industry
- Knowledge graph
- Pain points
- Clients

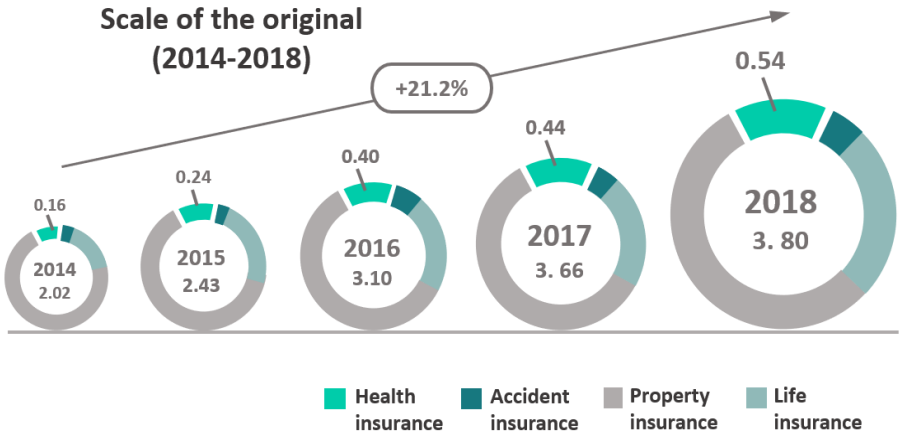
Part II : Construction

Part III : Business Model

Part IV : Prospect

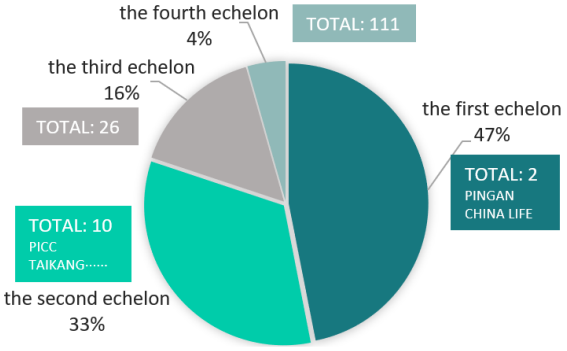
Health insurance industry shows great potential in the long run

1 Current development



Conclusion: Health insurance industry has a promising potential for developing.

2 Market Share



- Conclusion: Health insurance premium market is highly concentrated, with 8% of the companies occupying 80% of the business scale.

3 Background

- Increasingly personal health expenditure
- Intensifying aging population
- Expanding middle income group
- Increasing number of chronic patients
- Boosting people's health awareness

4 Policies supporting health insurance industry

- Building a multi-level medical security system.
- National strategy of "healthy China 2030".

Conclusion: Health insurance can serve as a supplement to social security.

5 Challenge

- High compensation rate
- Increasing competition

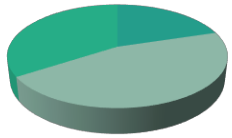
Source : CIRC , CEIC

Health insurance companies: face a status quo with coming internal and external challenges

1 How companies deal with complex data

Actuary analysis

Company internal database

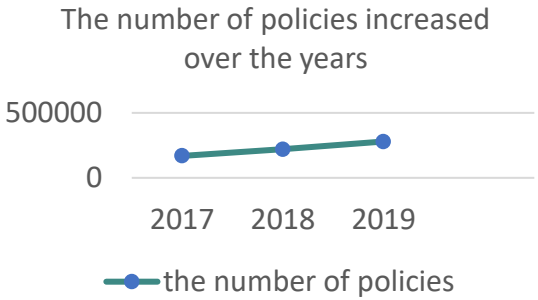


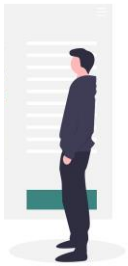
cause

Bottleneck:

High labor cost and low efficiency.
Information screening is difficult and redundant.
Hard to establish internal relations.

2 Trend for increasing data





3 Numerous internal departments

- Marketing department
- Actuarial analysis department
- Product department
- Compensation department
- Operation Department
- Planning department

The cross degree between departments is small, and the efficiency of handling things together is low.

4 Complicated connection with other organizations

Medical institutions
Nursing institutions
Health management institutions
Other insurance companies

It is difficult to exchange information within the industry and integrate due to the scattered connection with other institutions.

Conclusion:

Data: Weak health data system.
System: Lack of professionalism, expansibility and flexibility.
Risk control: Low level of intelligence.
Product: Serious homogenization.
Operation: High operation cost.

Source : CIRC , CEIC

Health insurance knowledge graph is in the early stage

1 Development status and trend

Depending on the number of existing health insurance company and the conversion rate.

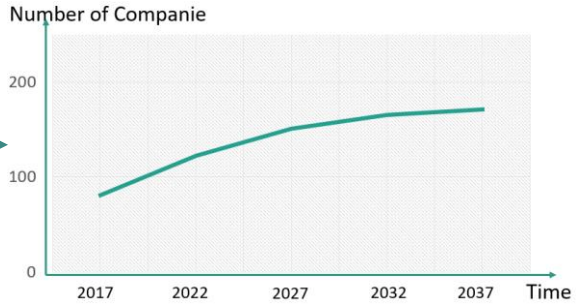
	Market scale	Number	Estimated conversion rate	Number of convertible customers
Large	Ten billion level	12	40%	5
Mid	Billion level	26	80%	21
Small	/	111	50%	55
Sum up				81

Number of convertible customers = number * estimated conversion rate

Conclusion: Measuring the purchasing power and loss ratio, we determine SME's as our main clients.

Market Size of Health insurance industry

2017-2037



Conclusion: the market has relatively high value to explore.

2 Patent protection

No guarantee to prevent potential **threatens.**

3 Potential competitors

- Disease insurance knowledge graph for policyholders.
- 2016 knowledge graph of life insurance.
- Ample data bases.

4 Entry barriers

Technology barrier:

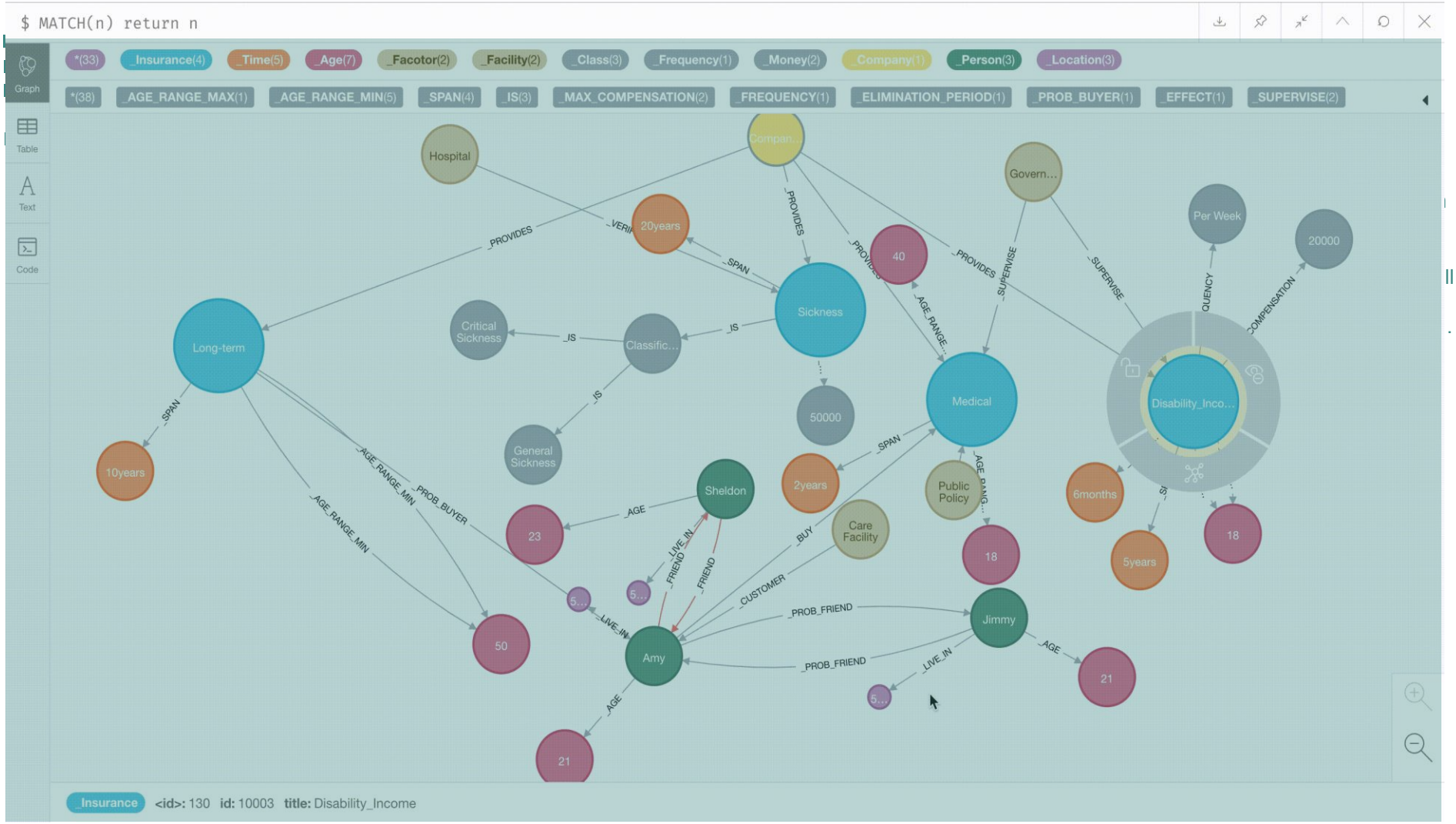
1.Lack of the required skill sets and training.

2.There are multiple initiatives across the organization that are not streamlined or optimized for the enterprise.

Government policy: constraint of collecting secret information.

Source : CIRC , CEIC

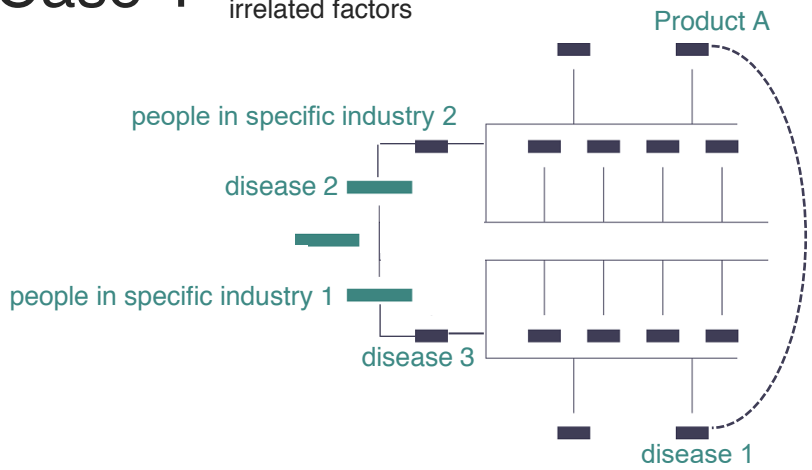
The effect of knowledge graph on health insurance industry



Examples of knowledge graph application in health insurance industry

Case 1

Knowledge graph between products and seemingly irrelated factors



show hidden relationships visually

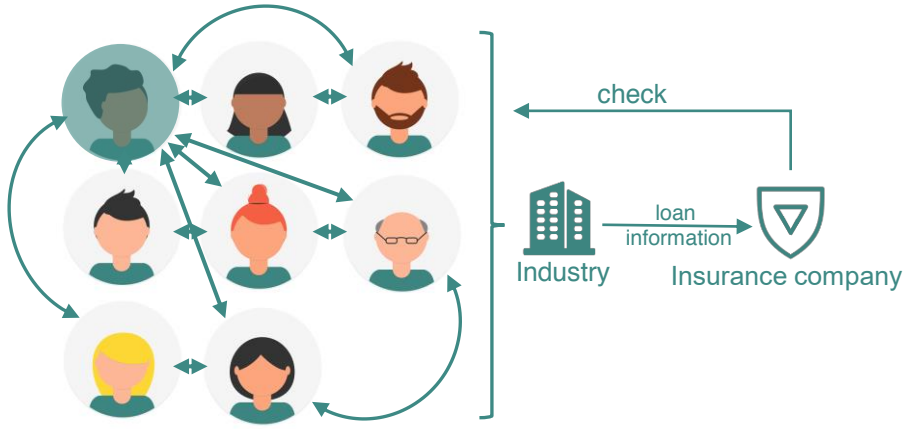
- Based on knowledge graph, we can also realize the function of multi-dimensional visualization of information.

Knowledge graph can tell you if something you care about will change when several seemingly unrelated things change.

Disease1 seems to have nothing to do with product A, but with knowledge map we can directly see disease3 associated with disease1, commonly affects people in industry1, then, the incidence of disease2 will change, ultimately affects people in industry2, who are the major agents of product A, and that can tell the company to look again the product pricing.

Case 2

Knowledge graph in related industries.



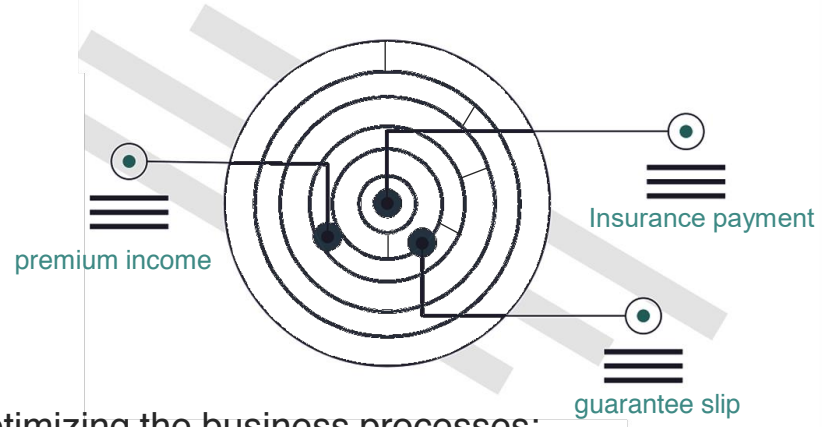
Risk control : Predicting potential risks

- The knowledge graph, based on multi-dimensional data, the connection between clients, enterprises and certain industries is established, and so from the perspective of industry associations we could predict the risks faced by industries or enterprises.
- For instance, by establishing a relational model between industries based on loan information; if there are several overdue loans in a certain region, (probably a high-risk event occurs in the industry,) with the analysis result of the knowledge graph, the insurance company can make predictions in a timely manner and avoid risks as early as possible.

Examples of knowledge graph application in health insurance industry

Case 3

Knowledge graph in the whole market



Optimizing the business processes: thinking with large data

- Knowledge graph can optimize the business processes.

By extracting knowledge and establishing rules, knowledge graph contains a large amount of structured data about health insurance products, which can assist the product designers in design process.

For instance : in 2019

premium income = 55036983

guarantee slip = 3867663

insurance payment = 10387

Case 4

Knowledge graph among agents

Marketing : Find potential agents

- Knowledge graph can better reflect the relationship among users, events, behavior and other factors.

As shown in the figure, using mathematical model, we can mine the potential users with the help of the knowledge graph.

$$\ln\tilde{y}_{re} = \tilde{u}_{re} + \beta_{in}\ln\tilde{x}_{in} + \beta_{ag}\ln\tilde{x}_{ag} + \beta_{fr}\ln\tilde{x}_{fr} + \beta_{wp}\ln\tilde{x}_{wp}$$

$\tilde{y}_{re}$:potential premium income per user bring.

$\tilde{x}_{fr}$:the number of friends in graph.

$\tilde{x}_{in}$:Income. $\tilde{x}_{wp}$:purchase insurance density. $\tilde{x}_{ag}$:age of purchaser.

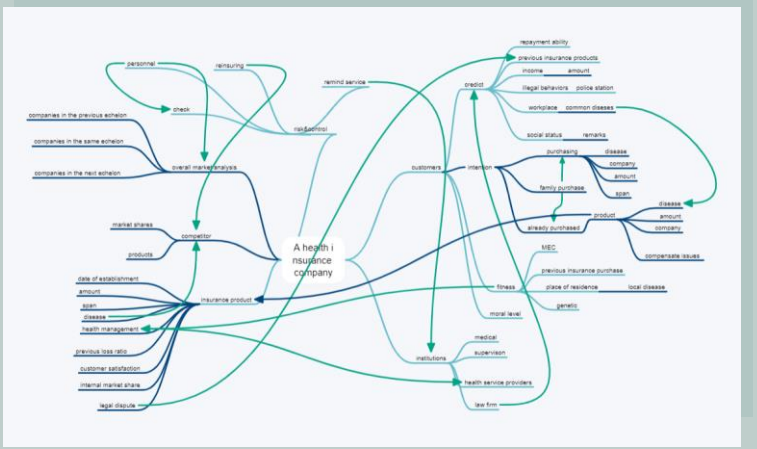
Unit :hundred million , ten thousand

Machine Learning Strategies for constructing knowledge graph

Basic Steps to Construct a Knowledge Graph

Knowledge Modeling → Knowledge Extraction → Knowledge Processing → Knowledge Representation

To construct the knowledge graph, we need the following data:



Data source

cooperation

Work with related industries.

such as getting data from hospitals on the incidence of disease in the medical sector.

internet

Crawlers crawl through legitimate data.
such as a company's insurance policies and product information

guarantee slip

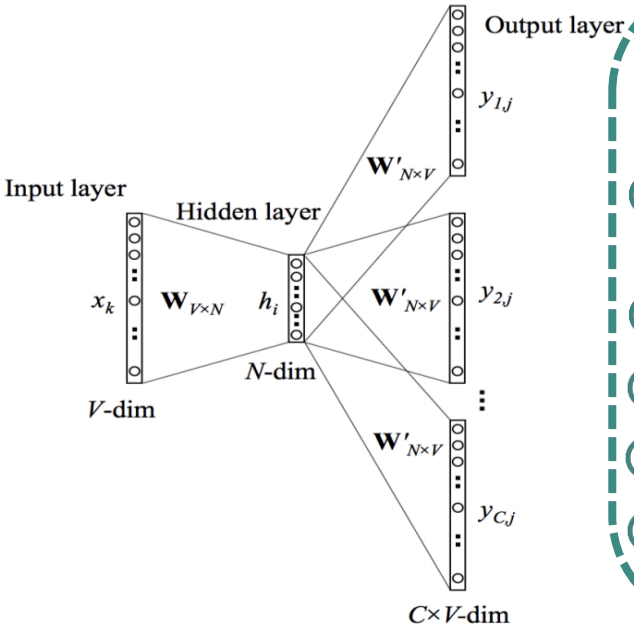
Through the policy we can get almost all the information about the consumer

market

Connections are made by investigating various data inputs into the knowledge map, such as industry condition

Source : Refer to appendix

Data source and data structure of knowledge graph



NLP Using Deep Learning

1 Word Embedding(Skip-Gram Model)

- Principle: Words that are nearby in the text should have similar representations.
- Assign a vector to each word.
- Slide a ‘window’ over the text .
- Word vectors are nudged around to minimize surprise
- Keep iterating until ‘good enough.

Mathematical Clarification of Propagation Functions

Standardize the probability model with Cross-entropy cost function and Derive the propagation function

Model

$$p(w_{c,j} = w_{O,c} | w_I) = y_{c,j} = \frac{exp(u_{c,j})}{\sum_{j'=1}^{V} exp(u_{j'})}$$

$$\begin{aligned} E &= -\log p(w_{O,1}, w_{O,2}, \ldots, w_{0,C} | w_I) \\ &= -\log \prod_{c=1}^{C} \frac{exp(u_{c,j_c^*})}{\sum_{j'=1}^{V} exp(u_{j'})} \\ &= -\sum_{c=1}^{C} u_{c,j_c^*} + C \cdot \log \sum_{j'=1}^{V} exp(u_{j'}) \end{aligned}$$

$$\frac{\partial E}{\partial u_{c,j}} = y_{c,j} - t_{c,j} := e_{c,j}$$

$$\mathrm{EI}_j = \sum_{c=1}^{C} e_{c,j}$$

$$\frac{\partial E}{\partial w'_{ij}} = \sum_{c=1}^{C} \frac{\partial E}{\partial u_{cj}} \cdot \frac{\partial u_{cj}}{\partial w'_{ij}} = \mathrm{EI}_j \cdot h_i$$

$$w'^{(\mathrm{new})}_{ij} = w'^{(\mathrm{old})}_{ij} - \eta \cdot \mathrm{EI}_j \cdot h_i$$

Formula

$$\mathbf{v}'^{(\mathrm{new})}_{wj} = \mathbf{v}'^{(\mathrm{old})}_{wj} - \eta \cdot \mathrm{EI}_j \cdot \mathbf{h} \qquad j = 1, 2, \ldots, V$$

Subproblem: Natural Language Processing

2 Relation Extraction Using Classifiers

Pre-define the collection of the possible relations

Select collection of relevant entities

Label the data for machine learning

--- select the typical corpus
--- label the named entity
--- specify the relationships by hands
--- divides the processed data into 3 parts: training, developing, testing

Design the characteristics of the entities

Word Features: For Chinese, it's been proved that the character pattern outperformed others. Position embeddings, Word embeddings, knowledge embeddings

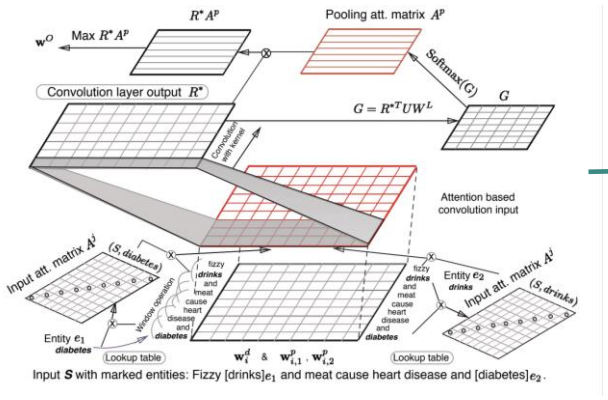
Select and train the classifiers

Evaluate the results

Subproblem: Natural Language Processing

3 Deep learning model — CNNs + attention

With new distance function and cost function introduced by Linlin Wang, Zhu Cao, Gerard de Melo and Zhiyuan Liu in their paper “Relation Classification via Multi-Level Attention CNNs



$$\delta_\theta(S, y) = \left\| \frac{\mathbf{w}^{\mathrm{O}}}{|\mathbf{w}^{\mathrm{O}}|} - W_y^{\mathrm{L}} \right\|$$

$$\mathcal{L} = \left[\delta_\theta(S, y) + (1 - \delta_\theta(S, \hat{y}^-))\right] + \beta\|\theta\|^2$$

$$= \left[1 + \left\| \frac{\mathbf{w}^{\mathrm{O}}}{|\mathbf{w}^{\mathrm{O}}|} - W_y^{L} \right\| - \left\| \frac{\mathbf{w}^{\mathrm{O}}}{|\mathbf{w}^{\mathrm{O}}|} - W_{\hat{y}^-}^{\mathrm{L}} \right\|\right] + \beta\|\theta\|^2,$$

4 Knowledge Assessment

Most commonly used factors:

$$P = \frac{\#\ of\ correctly\ extracted\ relations}{Total\ \#\ of\ extracted\ relations} \rightarrow R = \frac{\#\ of\ correctly\ extracted\ relations}{Total\ \#\ of\ gold\ relations} \rightarrow F_1 = \frac{2PR}{P + R}$$

Source : CIRC , CEIC

Neo4j for Knowledge Representation and PRA for Knowledge Inference

5 Knowledge Representation

We use the open-source graph database management system Neo4j to store the processed data and triples.
However, we have presumed that the materials for training our model rely on both the limited formal records provided by the companies and the open materials found on the Internet, it is an essential step to better the imperfect knowledge. The Path Ranking Algorithm(PRA, introduced by Lao and Cohen(2010b)) is a remarkable method to perfect the balance between effectiveness and reliability.

For each instance, the cross-entropy cost function is implemented, i.e.

$$o_i(\theta) = w_i[r_i \ln p_i + (1 - y_i) \ln(1 - p_i)]$$

where p_i is the predicted relevance defined as $p_i = p(r_i = 1|x_i;\Theta) = \frac{exp(\Theta^T x_i)}{1 + exp(\Theta^T x_i)}$

Parameter Θ is estimated by maximizing the following regularized objective function $O(\theta) = \sum_i o_i(\theta) - \lambda_1|\theta|_1 - \lambda_2|\theta|_2$

Description

For any path in a knowledge graph

$$P: T_0 \xrightarrow{R_1} T_1 \xrightarrow{R_2} \cdots \cdots \xrightarrow{R_l} T_l$$

For some seed node $s \in domain(P)$

We define a distribution $h_{s,P}(e)$ where

When P is empty: $h_{s,P}(e) = 1, if\ e = s$

$h_{s,P}(e) = 0, if\ e \neq s$

When P is not emp $h_{s,P}(e) = \sum_{e' \in range(P')} h_{s,P'}(e') \cdot P(e|e'; R_\ell)$.

Where $P(e|e'; R_\ell) = \frac{R_\ell(e',e)}{|R_\ell(e',\cdot)|}$ is the probability of reaching node *e* from node *e'* a one step random work. . The $h_{s,P}(e)$ is called path feature.

And the we rank nodes in the graph by a linear combination of weighted path features $\theta_1 h_{s,P_1}(e) + \theta_2 h_{s,P_2}(e) + \ldots \theta_n h_{s,P_n}(e)$

The Θ_i is the parameters to be trained, the scoring function $score(e;s) = \sum_{P \in \mathbf{P}_\ell} h_{s,P}(e)\theta_P$ will be used to train the model, where P_l is the set of relation paths with length $\leq l$.

Separate sales model for high and low-end customers

Market share

Premium market

customization

Strategic position compare

- Excellent programmers
- Effective and accurate services
- Customer-orientation

Customer preference

- Optimal reliability
- More specific to the operating characteristics Higher accuracy

Enhance competitiveness

- Cooperate with other technology companies.
- Develop in R & D.

Low-end market

Purpose: expanding SME market; focus on standardization

1 Overall sale

- Sell the knowledge graph as a whole.
- Customers can use all the functions.

- 8-year ownership of our knowledge graph after paying the fee
- Need to pay the maintenance costs in quarter.
- Updated versions are available and customers can decide to buy.

2 Partly sale

- Sell the whole knowledge graph but partly functions.
- Need to pay the maintenance cost in quarter.
- Updated versions are available for customers to buy or not

3 Bundle sale

- Bundling package is available and more profitable.
- Bundling the maintenance service and knowledge graph, the 8-year fault can be maintained free of charge.

4 Subscription

- Customer obtain the usage rights in a certain period instead of the ownership, depending on the usage cycle they choose. For instance, in months or quarters.
- Free from the maintenance costs and updated costs, for they have already been calculated in the subscription fee.

Feasibility

Customer are not willing to pay for the whole knowledge graph due to the constraint of cash flow or other specially internal operation.

Subscription service helps customers to form dependence on products.

Source : Wikipedia: Enterprise Resource Planning

Adopt innovative profit model and cooperation model

Profit Model

Income and Cost mode

Premium (Customization)

Proportion of customers 5%

Low-end market (Standardization)

Overall sale	20%
Bundle sale	25%
Partly sale	25%
Subscription	25%

Revenue = Customized R & D revenue + Maintenance revenue + Updated revenue

Costs = Development costs + Maintenance costs

Profit =∑((Company pay per day -Daily maintenance costs in average)*days – First entry fee) *Company size

Gross profit = ∑ Net profits * Number of companies * Proportion

Enterprises that have purchased knowledge graphs are available to sacrifice some internal information in order to pay less service fee.

Cooperation Model

Set up cooperation network with insurance company.

Strengthen their position as suppliers and cooperate in depth.

Partnership is diversified, seller, lender or Service provider.

Barriers model

Additional costs required to break through industry **technology** and **data source restrictions.**

Major customers : Health insurance company; Companies planning to develop health insurance businesses

Other customers : Consulting companies; Insurance agency

Source : CIRC , CEIC

Establish good cooperation with customers

1 The need to contact customers

- Model: **Changes in the number of users**

$$dN=a\times[\overline{N}-N(t)]\times dt+\sigma(t)\times dz(t) \quad a<0$$

N : the threshold of the number of users
N(t) : the current number of users
Dz : standard wiener process
a : the speed of the change in the number of users

- During the dynamic evolution of the number of users, only if the initial number of users reach the threshold(N), users will increase exponentially in a short time. If the initial level of the number of users falls below the threshold for a long time, the knowledge map will decline due to negative feedback.

2 Fit company's internal business

Marketing
Prospect for convertible customers
Expand marketing target
Precision marketing

Product design
Need extensive data as support
Dynamic data as a design breakthrough
Coordination of product portfolio

Pricing
Build connections between items
Market and competitor pricing
Product related properties

Risk & Control
Real-time data update
Factor that could grow into a big risk
Predicting potential risk

2 Way to contact customers

Get in touch: contact with Purchasing Department of insurance company; Provide one month probation period

Way: make customer keep the aware of significant position; Customers rely on the use of knowledge map

Maintain: stay in touch

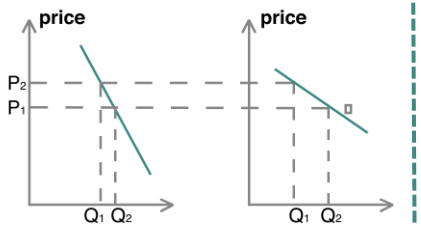
4 Pricing

✕ **Competitive analysis**

Since there is not any existing competitor of our product.

✔ **Price-based costing**

price that customers are willing to pay
Value comparison with manual handling costs
Supply and demand

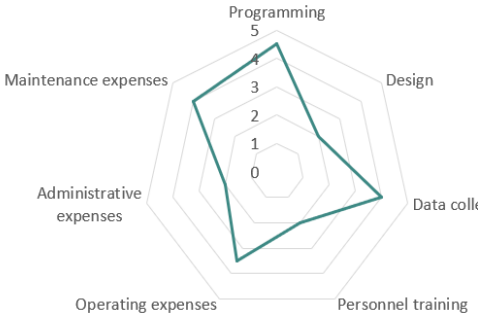


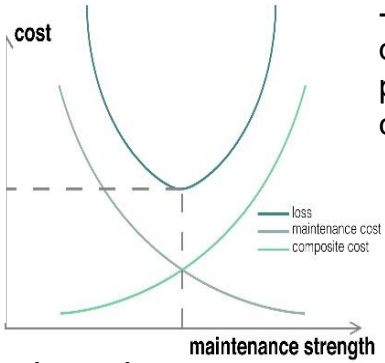
Conclusion:
If our knowledge graph enters the health insurance market, it will affect their resilience to data integration products. Pricing appropriately, the market demand will be satisfying.

Source : CIRC , CEIC

To earn certain profit and expand market share, knowledge graph should be priced reasonably

Cost structure





Conclusion: Determining a certain maintenance intensity is beneficial for acquiring minimized costs.

Other factors in pricing

In addition to cost structure, demand-orientation and competition-orientation are two significant models to be based on.

- Expected profit
- S&D situation
- Overall price level of existed products

750000 × (1+15 %) = 862500 Overall sale

Ways to price the subscription fee(SF)

——take maintenance costs for example

-Include maintenance costs as part of the subscription fee in the diversified profit model. When the maintenance cost targeting at premium enterprises is equal to the subscription fee for SMEs, it can be successfully shifted to clients.

P_1:the possibility of error （increasing rate r_1）
C_1:Maintenance cost （increasing rate r_2）
t : Assumed time point （discount rate r_3）
C_2:Normal fee
C_3:Growing fee (increasing rate r_4)

One-time charges:
-Fee> NPV_1: Equal to NPV_1, the SF will make no profit for company.

$$NPV_1 = \frac{P_1 C_1 \left\{1 - \left[\frac{(1+r_1)(1+r_2)}{1+r_3}\right]^t\right\}}{r_3 - (1+r_1)(1+r_2)}$$

Multiple charges:
-Fee > C_2: If covering maintenance cost by charging for the same amount, the fee is suggested to surpass C_2.

$$NPV_1 = C2\frac{\left[1 - \left(\frac{1}{1+r_3}\right)^t\right]}{r_3}$$

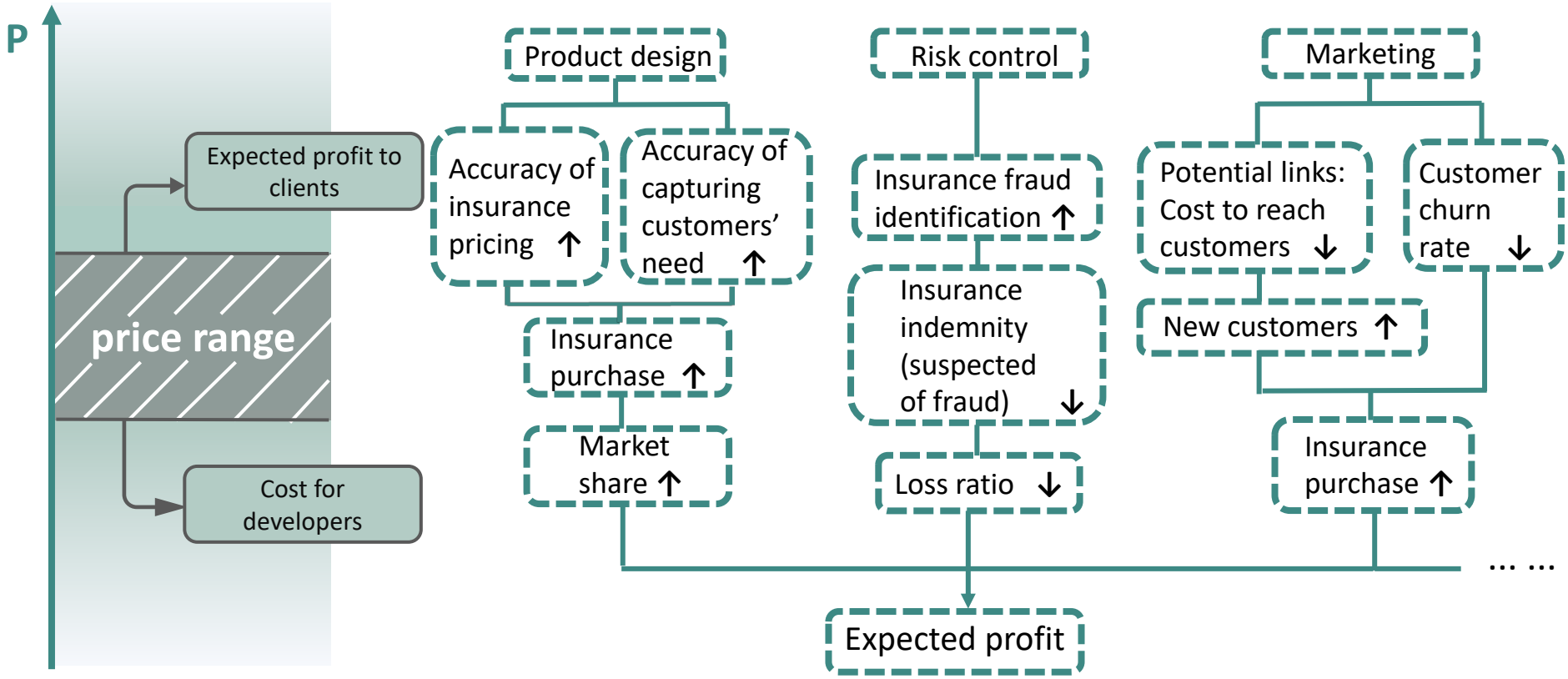
-Fee > C_3: The amount of fee is increasing at the rate of r_4, and the SF in the first period is suggested to be greater than C_3.

$$NPV_1 = C_3\frac{\left[1 - \left(\frac{1+r_4}{1+r_3}\right)^t\right]}{r_3 - r_4}$$

Conclusion: Taking all costs into account in this way, the price floor can be set.

Source : CCGP

To earn certain profit and expand market share, knowledge graph should be priced reasonably



Conclusion:

Expected profit from product design, risk control, marketing and other aspects represents the price ceiling.

The price floor and price ceiling consist of the appropriate price range.

Source : Essay: Software pricing strategy and model design

Expansion: feasibility analysis of expanding overseas market

1 Why Expand?

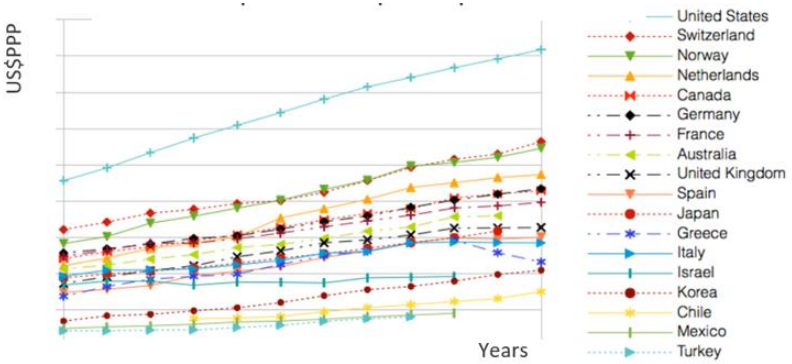
- Commercial health insurance has become the main way of health maintenance and disease.
- Prevention in many countries.
- Great mobility of knowledge gragh.

- Expand our influence.
  Global Internet platform.

Time for expansion: when the domestic market is nearly saturated

2 A promising market

The trend of commercial health insurance in the world



Conclusion: the knowledge graph industry for health insurance has a good market prospect

3 Potential overseas markets

Focused position:
South Africa; Asia

- Considerations:
- Relatively large market share;
- Geographical position;
- Friendly and relationship

4 Strengths

- Regional economic level is high.
- Countries are concerned about health insurance.
- Knowledge graph's strong service level.

5 Barriers

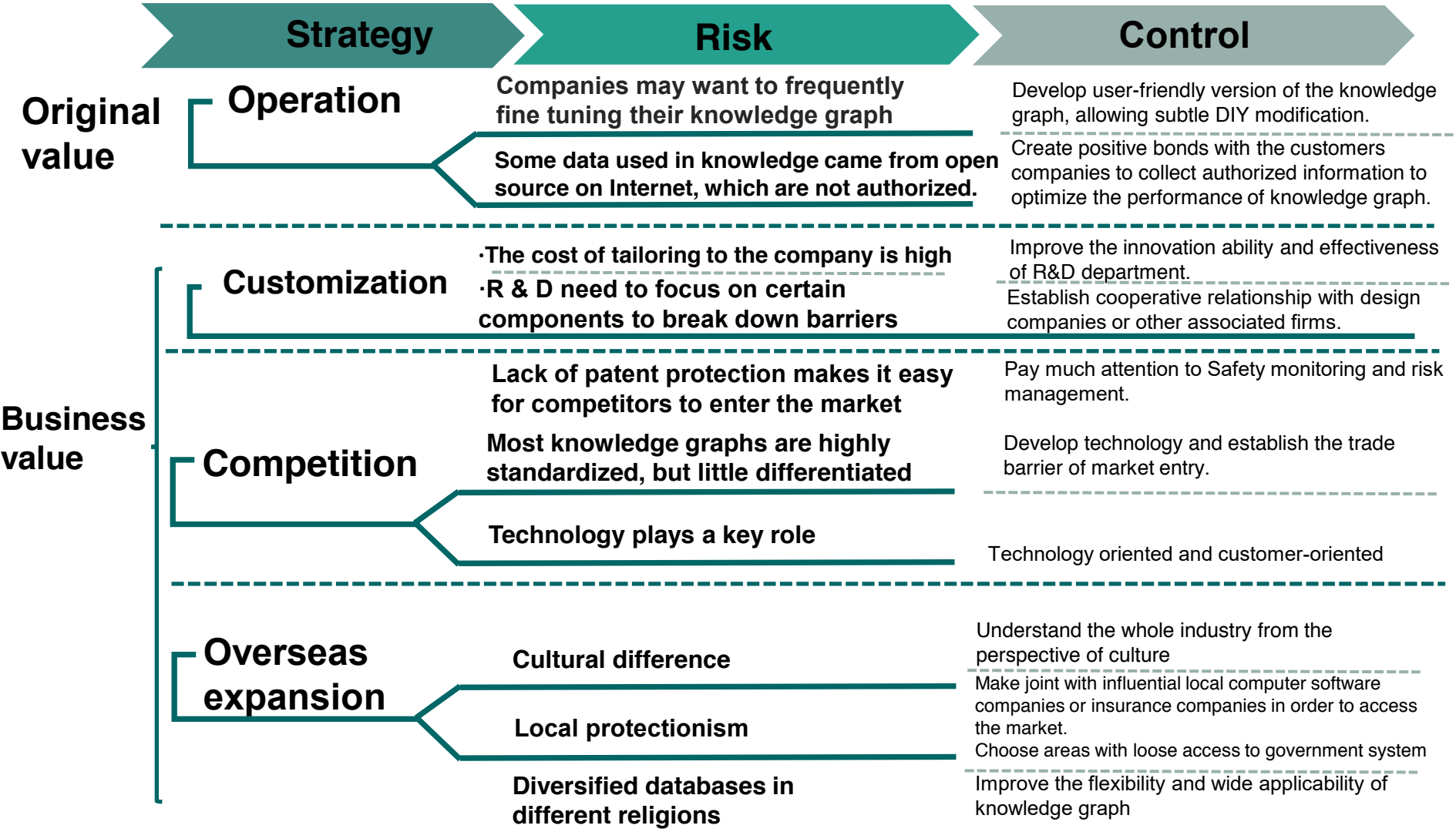
- More secret for privacy than domestic market.
- Great difference in legal system and insurance system.
- Cultural difference.
- Limited data source channel.

6 Strategy to expand

- Get local industry information in advance
- Establish diverse market information base for different countries.
- Information can be collected through agents, local embassies or professional organizations.
  Establish partnerships with influential local computer software companies.

Source : CIRC , CEIC

Prospect: the original value and business value of knowledge graph realize the continuous growth



Source : CIRC , CEIC