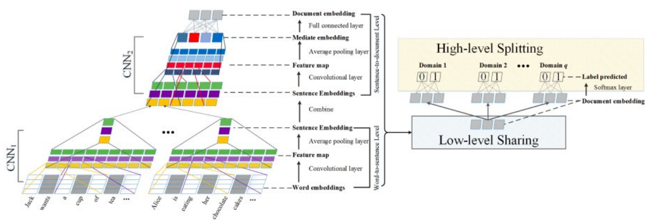
A neural network to extract knowledgeable snippets and documents

5 September 2018, by Ingrid Fadelli



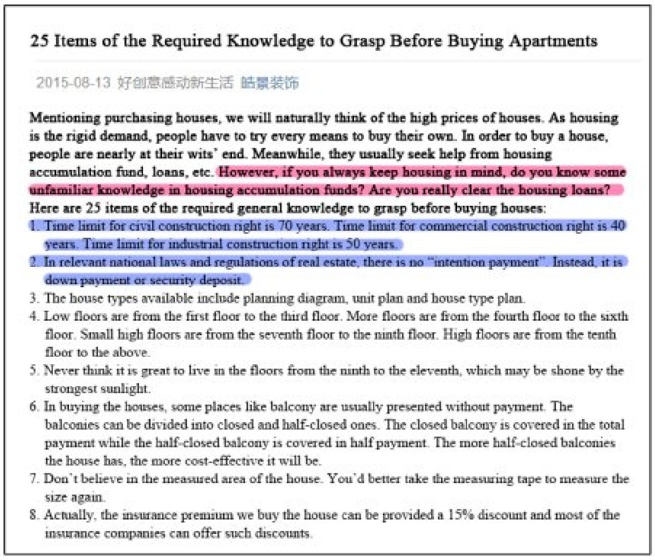
Structure of proposed model. Credit: Zhou et al.

Every day, millions of articles are published on social media and other platforms, receiving a vast amounts of clicks and shares from users navigating the web. Many of these articles contain useful information that, if extracted, could be used to compile knowledge databases or to deliver knowledge retrieval and question answering services.

Researchers at the Chinese Academy of Sciences (CAS) have developed a convolutional neural network (CNN)-based model to extract knowledgeable snippets and annotate documents. Their method, outlined on a paper pre-published on arXiv, was found to perform better than existing tools, despite being trained for shorter periods of time.

In their paper, the researchers define the term "knowledgeable document" as "a document containing multiple knowledgeable snippets, which describe concepts, properties of entities, or the relations among entities." So far, most knowledge bases, such as YAGO or DBpedia, extract knowledge based on Wikipedia, WordNet, GeoNames, and other online resources. However, compared to social media platforms, these resources often contain limited and inflexible information.

"Another recent knowledge base, Probase, with 2.7 million concepts, was automatically harnessed from the so-far largest corpus, consisting of 326 million knowledgeable sentences extracted from 1.68 billion web pages," the researchers wrote in their paper. "However, these sentences are extracted only by the Hearst patterns. For extracting more knowledgeable snippets to construct more comprehensive knowledge bases, semantic-based methods are needed to complement the previous pattern-based ones."



25 Items of the Required Knowledge to Grasp Before Buying Apartments

2015-08-13 好创意感动新生活 皓景装饰

Mentioning purchasing houses, we will naturally think of the high prices of houses. As housing is the rigid demand, people have to try every means to buy their own. In order to buy a house, people are nearly at their wits' end. Meanwhile, they usually seek help from housing accumulation fund, loans, etc. However, if you always keep housing in mind, do you know some unfamiliar knowledge in housing accumulation funds? Are you really clear the housing loans?
Here are 25 items of the required general knowledge to grasp before buying houses:

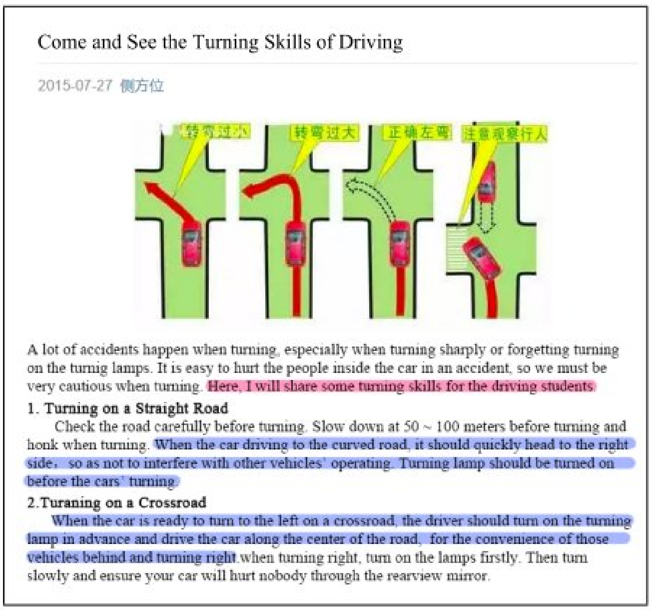
1. Time limit for civil construction right is 70 years. Time limit for commercial construction right is 40 years. Time limit for industrial construction right is 50 years.
2. In relevant national laws and regulations of real estate, there is no "intention payment". Instead, it is down payment or security deposit.
3. The house types available include planning diagram, unit plan and house type plan.
4. Low floors are from the first floor to the third floor. More floors are from the fourth floor to the sixth floor. Small high floors are from the seventh floor to the ninth floor. High floors are from the tenth floor to the above.
5. Never think it is great to live in the floors from the ninth to the eleventh, which may be shone by the strongest sunlight.
6. In buying the houses, some places like balcony are usually presented without payment. The balconies can be divided into closed and half-closed ones. The closed balcony is covered in the total payment while the half-closed balcony is covered in half payment. The more half-closed balconies the house has, the more cost-effective it will be.
7. Don't believe in the measured area of the house. You'd better take the measuring tape to measure the size again.
8. Actually, the insurance premium we buy the house can be provided a 15% discount and most of the insurance companies can offer such discounts.

Example of knowledgeable document. The blue and red sentences are knowledgeable and unknowledgeable snippets respectively. The document introduces the 25 pieces of tips for purchasing real estates. Credit: Zhou et al.

Knowledgeable snippets and articles could also be used to develop knowledge retrieval and question answering services. These services would, for instance, answer questions raised by users who are looking for help with a particular problem. With these applications in mind, the researchers at CAS set out to develop a CNN based model that can analyze the semantics of a document, determine whether it is knowledgeable or not, and extract knowledgeable snippets of information from it.

"Specifically, we propose SSNN, a joint CNN-based model, to understand the abstract concept of documents in different domains collaboratively and judge whether a document is knowledgeable or not," the researchers explain in their paper. "In more detail, the network structure of SSNN is 'low-level Sharing, high-level Splitting," in which the low-level layers are shared for different domains while the high-level layers beyond the CNN are trained separately to perceive the differences of different domains."

The model devised by the researchers offers an end-to-end solution to annotate documents that does not entail extensive and time-consuming feature engineering. They also developed manual features and trained a SVM classifier model to complete the task.


Come and See the Turning Skills of Driving

2015-07-27 侧方位

转弯过小 转弯过大 正确左弯 注意观察行人

A lot of accidents happen when turning, especially when turning sharply or forgetting turning on the turnig lamps. It is easy to hurt the people inside the car in an accident, so we must be very cautious when turning. Here, I will share some turning skills for the driving students.

1. Turning on a Straight Road

Check the road carefully before turning. Slow down at 50 ~ 100 meters before turning and honk when turning. When the car driving to the curved road, it should quickly head to the right side, so as not to interfere with other vehicles' operating. Turning lamp should be turned on before the cars' turning.

2.Turaning on a Crossroad

When the car is ready to turn to the left on a crossroad, the driver should turn on the turning lamp in advance and drive the car along the center of the road, for the convenience of those vehicles behind and turning right.when turning right, turn on the lamps firstly. Then turn slowly and ensure your car will hurt nobody through the rearview mirror.

Example of knowledgeable document. The blue and red sentences are knowledgeable and unknowledgeable snippets respectively. The document introduces the turning skills of driving. Credit: Zhou et al.

The researchers evaluated the effectiveness of their model on a dataset of real documents from three content domains on WeChat, a Chinese messaging, social media and mobile payment platform developed by Tencent. Their findings were very promising, with the SSNN performing consistently better than other CNN models, while saving time and memory consumption thanks to shorter and more efficient training processes.

"Compared with building multiple domain-specific CNNs, this joint model not only critically saves training time, but also improves the prediction accuracy visibly," the researchers wrote in their paper. "The superiority of the proposed model is demonstrated in a real dataset from Wechat public platforms."

In future, the SSNN model proposed in this study could be used to build more comprehensive knowledge databases. It could also aid the development of innovative services that answer user queries both quickly and exhaustively in real-time.

More information: Hierarchical Neural Network for Extracting Knowledgeable Snippets and Documents. arXiv:1808.07228v1 [cs.CL]. arxiv.org/abs/1808.07228

Abstract
In this study, we focus on extracting knowledgeable snippets and annotating knowledgeable documents from Web corpus, consisting of the documents from social media and We-media. Informally, knowledgeable snippets refer to the text describing concepts, properties of entities, or relations among entities, while knowledgeable documents are the ones with enough knowledgeable snippets. These knowledgeable snippets and documents could be helpful in multiple applications, such as knowledge base construction and knowledge-oriented service. Previous studies extracted the knowledgeable snippets using the pattern-based method. Here, we propose the semantic-based method for this task. Specifically, a CNN based model is developed to extract knowledgeable snippets and annotate knowledgeable documents simultaneously. Additionally, a "low-level sharing, high-level splitting" structure of CNN is designed to handle the

documents from different content domains. Compared with building multiple domain-specific CNNs, this joint model not only critically saves the training time, but also improves the prediction accuracy visibly. The superiority of the proposed method is demonstrated in a real dataset from Wechat public platform.

© 2018 Tech Xplore

APA citation: A neural network to extract knowledgeable snippets and documents (2018, September 5) retrieved 28 June 2022 from https://techxplore.com/news/2018-09-neural-network-knowledgeable-snippets-documents.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.