Extracting Common-Sense Knowledge About Change

Benno Kruit, Centrum Wiskunde & Informatica

Motivation. Automatically constructed Knowledge Bases (KBs) have proven to be useful for a broad variety of tasks, including question answering, web search and data integration. These KBs contain rich factual knowledge about entities in semantic classes and their relations. Unfortunately, most KBs only cover static information, without keeping track of a historical record of facts. When KBs do contain temporal data, it is typically only for a small number of relations. This means that KBs have no way to know whether a fact is out of date and that knowledge about change is often absent.

However, many inferences that people make in daily life depend on knowledge of change, which we generally attribute to common sense. Some relations, such as a birth dates, geographic area or names don't change, but many relations do have a temporal scope: population and GDP for countries, employment and net worth for people, or sales numbers and chart positions for music albums. The behaviour of numeric relations can also be attributed to common sense. For instance: awards and sales numbers increase monotonically, the population of a country never doubles in a year, and music albums generally don't top the charts a decade after their release.

Consider the example in Figure 1. The first table expresses a single relation with a temporal scope. It is common sense that a city's population changes in time. The second table expresses two relations, whose values are independent of time. When mapping extracting the data from these tables into a KB, it is mandatory to include the temporal scope in the first case, but not in the second. If the rightmost column was missing from the first table we would have needed to find a timestamp in the table context, while the second table does not need additional information for integration with a KB.

Common-sense knowledge about temporal relations like this has many applications. When performing data integration, it is useful to know when a timestamp is mandatory and when it isn't. In question answering and web search, it is useful to know whether an relation value might be out of date. For any temporal reasoning task you need to know what changes and what stays the same. Additionally, temporal change can cast light on the way relations are related to each other through correlations and events.

Problem. Common-sense knowledge about how relations change through time is not available in KBs, and is challenging to annotate or construct automatically for a number of reasons.

City	Population	Year
Amsterdam	851573	2014
Mogadishu	2425000	2017

Mountain	Height	First Ascent
Mount Everest	8848	1953
Mont Blanc	4810	1786

Figure 1: These tables both have a column of years, but the first table expresses a single relation anchored in time while the second table expresses two time-independent relations.

First of all, temporal scopes can be missing from facts in a KB, either because of incorrect extraction or oversight of contributers, or because the fact should not be temporally scoped at all. However, determining why a temporal scope is missing is non-trivial and cannot be determined from the data in the KB alone.

Secondly, the degree and manner in which information changes over time depends on many different factors. Broad-coverage temporal reasoning and data integration require a large set of relations and classes, but any of these can influence the stability of relation values. Therefore, common-sense knowledge about relations needs to take into account the regularities that exist for groups of entities that behave similarly through time. It is necessary to discover patterns in the data that capture these regularities in a consistent way.

Finally, in order to make this knowledge useful, these structures should allow queries that support common-sense temporal inference and data integration.

Goals. Therefore, the goals of this research are as follows:

- Evaluate several existing strategies for representing temporally scoped facts in KBs with regard to their applicability to extracting common-sense temporal knowledge
- Propose a method to construct weighted rules about relation stability from text and tables, and apply it to several KBs
- Leverage the constructed rules to extract high-precision temporal scopes for existing facts from textual and tabular data sources