Towards Learning a Knowledge Base of Actions from Experiential Microblogs

Emre Kıcıman
Microsoft Research
emrek@microsoft.com

Abstract

While today's structured knowledge bases (e.g., Freebase) contain a sizable collection of information about entities, from celebrities and locations to concepts and common objects, there is a class of knowledge that has minimal coverage: *actions*. A large-scale knowledge base of actions would provide an opportunity for computing devices to aid and support people's reasoning about their own actions and outcomes, leading to improved decision-making and goal achievement. In this short paper, we describe our first efforts towards building a distributional representation of actions and their outcomes, as learned from the timelines of individuals posting experiential microblogs.

Introduction

While today's structured knowledge bases (e.g., Freebase) contain a sizable collection of information about entities, from celebrities and locations to concepts and common objects, there is a class of knowledge that has minimal coverage: *actions*. Simple information about common actions, such as the effect of eating pasta before running a marathon, or likely outcomes after adopting a puppy, are missing. While some of this information may be found within the free text of Wikipedia articles, the lack of a structured or semi-structured representation make it largely unavailable for computational usage. With computing devices continuing to become more embedded in our everyday lives, and mediating an increasing degree of our interactions with both the digital and physical world, knowledge bases that can enable our computing devices to represent and evaluate actions and their possible consequences have the potential to aid individuals in making better decisions about their actions, and thus being more likely to achieve their individual goals.

Representing and reasoning about actions—knowledge about how an actor can intervene to change the state of the world—has been widely studied in classical symbolic planning systems, such as STRIPS and PDDL (Fikes and Nilsson 1971) (McDermott et al. 1998), though applied to restricted domains. In the context of these planning systems, actions consist of pre-conditions that must be satisfied before an action can be performed, and a set of post-conditions that hold true afterwards. By chaining actions together, planning systems produce intricate step-by-step plans to achieve some goal state. In contrast, the goal of our knowledge base of actions is not necessarily to enable generation of multi-step plans, but simply to enable better analysis and selection of a single action from many possibilities.

In this short paper, we discuss our first efforts in the *Quantified-All* project to build such a knowledge base of actions and outcomes, based on the published experiences of the hundreds of millions of people posting every day on social media about the actions they take and what happens in their lives afterwards. While there are many data sources (including web documents, search queries, and a variety of wearable sensors) that potentially capture the relationship between actions and their consequences, our initial focus is on social media data for several reasons. First, experiential social media data naturally captures the temporal occurrences of events experienced by individuals, allowing our analysis to exploit temporal relationships among actions and outcomes. Secondly, explicit statements in social media messages capture both the actions that people take as well as the outcomes across a wide variety of domains. Finally, social media messages are annotated with persistent user identifiers that allow us to condition our learned semantic relationships on user demographics, past actions and other relevant information.

Our basic approach is conceptually straightforward: for any given action, we first find all social media users who have reported taking the action. Then, we analyze the post-action social media timelines of these users and, comparing

Submitted to AAAI Spring Symposium on Knowledge Representation and Reasoning.

them to the timelines of users who did not take the action, identify the events that are likely to occur after taking some action. Our first implementation reports simple correlations between actions and consequent events, and we are now exploring more sophisticated propensity score analyses and temporal prediction algorithms (Rosenbaum and Rubin 1983) (Gunawardana, Meek and Xu 2011).

Approach

In the Quantified-All project, we are building a knowledge base of actions through analysis of the experiential reports made by individuals in social media to infer semantic relationships between actions, outcomes, and higher-level goals. Our use of social media for this purpose has implications on our representation of actions and their consequences. First and foremost, whereas classical planning systems make a clear distinction between actions performed and the state of the world; we do not. As all of our data comes from textual observations—both of actions and properties of the world—we simply represent all of these observations as *events* that occur on a user's timeline. Whether an event happens to be an action that can be performed by an actor is left for interpretation outside of the core data representation and analysis. Thus our analysis mechanics are tasked simply with extracting relationships between some event occurrence (an event which might be an action) and other events observed to occur afterwards and in correlation with it. There are several other implications of our use of social media and the incomplete and biased view it represents, that we touch briefly upon in the Discussion section.

The first stage of our analysis identifies experiential social media messages and extracts a per-user timeline of events. While many kinds of conversations occur over social media, from news discussions to celebrity gossip, we are specifically interested in experiential messages where the author is speaking directly about their own recent, first-person experiences. We identify these experiential messages with a simple trained classifier. Once we have identified experiential messages, we encode their content into per-user timelines. Currently, we are using a simple phrase segmentation algorithm (Jin et al. 2014), actively exploring various vector space mapping algorithms to represent the semantic meaning of phrases (Mikolov et al. 2013) (Pennington, Socher and Manning 2014), and considering alternatives involving more sophisticated parsings of the text (Bordes et al. 2012) (Banko et al. 2007). Note that at least in this approach, every phrase of every message is a potential event of some kind, and thus we avoid distinguishing between events which are actions or outcomes (or neither) and building an ontology of actions.

Conceptually, the second stage of analysis begins with a query for information about a particular target action (event). We identify all user timelines which include this event and analyze them in aggregate to identify subsequent events that are likely to be outcomes the target action. Conversely, given a target outcome (event), we can identify all user timelines which include that event and analyze the aggregated timelines to identify earlier events that may lead up to this outcome.

Correlated Phrase	**Sample of Original Tweet**
`Love_my_new_dog`	I love my new dog :)
`Chilling_with_my_cat`	Good morning, chilling with my cat and drinking some tea
`Watching_birds`	She fell asleep like this after watching birds out the window.
`Flea_medicine`	cuz I put flea medicine on her
`Doggie_day_care`	Dog had a great time at doggie day care. At 9:15pm, he's passed out.
`Won_t_stop_biting`	My dog won't stop biting me :((
`Dog_just_s***`	The dog just s*** on my son bed
`Cat_smells`	my cat smells like she's just had *** and I don't know why

Table 1—Sample of top correlated phrases that follow tweets that indicate new pet ownership, selected to demonstrate diversity of detected consequences.

The third and final stage of our approach consists of adapting the models to specific application scenarios. For example, the usage of these models in re-ranking restaurant recommendations; the usage of these models for explaining the potential consequences of an action; or the usage of these models to predict higher-level goal achievement all imply different requirements for the precision and recall of relationships, as well as the need to differentiate among classes of events.

Example Analysis #1: Getting a New Pet

As a demonstration, we consider the mundane action of a person considering the action of buying a pet, such as a dog or cat. For this analysis, we have identified all English-language Twitter users who mentioned getting a dog, cat, puppy or kitten during the period August 1-15, 2013, and collected their complete Twitter timelines for the period August 1-September 15, 2013. This dataset included 6232 Twitter users, posting 4.6M tweets in this 6 week period.

Given the phrases extracted from the experiential tweets of new pet owners, as a first analysis we calculate the relative likelihood of phrases in this data set as compared to their likelihoods in a large, random sample of Twitter users. More formally, let $N(e_i)$ be the number of occurrences of an event e_i after the target event (pet ownership). We estimate a Laplace smoothed probability of occurrence for e_i as:

$$\widehat{\theta_i} = \frac{N(e_i) + p_i m}{|N| + m}$$

Where $|N|$ is the count of all event occurrences after the target event, and p_i is a prior on the probability of occurrence of e_i, estimated as the larger of the maximum-likelihood estimates of $P(e_i)$ before the target event's occurrence or in timelines where the target event is not observed at all. Finally, we calculate a ranking score for e_i as its relative likelihood: $S_i = \widehat{\theta_i}/p_i$

While this simplistic analysis does not yet include more sophisticated propensity scoring, or attempt to account for temporal correlations unrelated to pet ownership, it does give us a sense of the kinds of relationships we might expect to extract from social media data. Table 1 shows a selection of these extracted phrases. We find that this data set exposes a broad selection of consequences of pet ownership, from initial social interactions with friends announcing and naming the new pet, enjoying spending time with a pet (`love_my_new_dog`, `chilling_with_my_cat`), activities the pet does (`watching_birds`), to pet care issues (`flea medicine`, `doggy_day_care`) to decidedly less pleasant consequences (biting, defecating on floors and pet odors). Common phrases not related to pet ownership (not shown in Table 1), such as Justin Bieber related phrases, mentions of Starbucks, etc., show up with much lower correlations.

Example Analysis #2: Marathon Training

In a second analysis, we consider the effect of selected actions on a specific, declared goal. In particular, we choose to look at the relative importance of various marathon training actions on the eventual outcome of a marathon race. We gather the timelines of over 600 Twitter users who are running a marathon during the month of March 2014, as well as their timelines for the four weeks preceding the date of their marathon run. The resulting data set consists of approximately 430k tweets.

While there are certainly several ways that individuals might determine the success of their own marathon, we use a simple definition here: whether the individual declared that they achieved a personal record (PR) after running the marathon. Against this, we measure the correlation between a person tweeting about taking a specific training-action (whether they chose to "taper", trained with "long runs", ate carbs before the race) and reporting that they achieved a personal record. Table 2 shows the results. Overall, we found that reporting the actions of long runs and tapering (reducing exercise before the marathon) were most correlated with later reporting a personal record. Reporting eating carbohydrates (carbs) before the race had a minor effect as well.

To better understand some of the potential hazards in inferring and interpreting actions and results from social media reports, we further analyzed the varieties of actions related to eating carbohydrates ("pasta", "spaghetti", "carboloading" and "carbloading"). Interestingly we found that reporting "carboloading" was strongly correlated with reporting a personal record, whereas the variant "carbloading" had no correlation. This discrepancy has several possible explanations, all rooted in confounding factors, such as the existence of distinct communities with different propensities to achieve or report personal records due to other, correlated, habits; or possibly different implementations of the "carbloading" and "carboloading" actions. More sophisticated analyses such as propensity score analyses given high-dimensional user history vectors may help separate some of these confounds.

A similarly confounding case occurs when we examine the effects of the action of eating "pasta" versus eating "spaghetti". Here we see that pasta has a small but positive correlation with reporting a personal record, but spaghetti has a strong negative correlation with reporting a personal record. Upon inspecting the original messages, we found that while "pasta" referred to the Italian dish of that name, the usages of the word "spaghetti" were actually referring to the vegetable "spaghetti squash". This demonstrates a case where semantic ambiguities must be carefully resolved.

Action	Increase/Decrease in PR likelihood
Taper	+45%
Long run	+27%
Carbs	+9%
Carboload	+172%
Carbload	0%
Pasta	+13%
Spaghetti	-40%

Table 2—Actions reported by marathon runners on Twitter and the relative increase or decrease in reporting a personal record.

The semantic interpretation of these results, of course, immediately leads one to wonder whether people who are more likely to be on a rigorous training regime, more likely to tweet about their training, or simply more aware of the technical terms, such as "carboload," are predisposed to tweet about their personal records.

Discussion

There are of course, several challenges that our presentation above has so far elided. For example, relying on experiential social media data to learn outcomes can introduce significant bias due to population biases as well as self-reporting biases (Mislove et al. 2011) (Diaz et al. 2014) (Kıcıman 2012). Significantly, the absence of an event in our social media timeline does not necessarily mean that an event did not occur. Understanding the implications of previous empirical studies for our inference processes, as well as the implications for how such biases circumscribe our ability to learn parts of the semantic space of relationships is important future work, as is incorporating additional data sources to augment textual social media data.

Another significant challenge is that a true model of actions and consequences is essentially a model of causal relationships. There is a rich literature on the inference of causal relationships from purely observational data (Spirtes and Glymour 1991) (Pearl 2000) though there is debate about the reliability of causal inference in the absence of randomized, active intervention (Robins and Wasserman 1999). Luckily, at least for some initial applications of these models, inference of the true causal seems unnecessary and simpler analyses such as temporal prediction and propensity scored relationships may be sufficient for the extracted results to be useful. Finally, understanding how a knowledge base of actions will be applied in different application scenarios and determining appropriate evaluation methodologies is critical future work.

Summary

As computing devices continue to become more embedded in our everyday lives, they are mediating an increasing number of our interactions with the world around us. From helping people search for the best product to buy among many, to recommending a restaurant we are likely to enjoy, computing services enable users to evaluate options and take action with “one click”. While such services model many facets of the options they present, they do not model the higher-level implications and trade-offs inherent in deciding to take one action instead of another. For example, a restaurant recommender service will not know that suggesting a carb-heavy Italian restaurant the evening before a person is going to run a marathon might improve their race outcomes. Today, people reason (or don’t reason) about these trade-offs based on their own past experiences and learnings, combined with their own “gut instinct”. People with relevant learnings and experiences and good “gut instinct” do well; but many others do not. By aggregating the combined experiences of hundreds of millions of people into a knowledge base of actions and their consequences, we believe that our computing devices may provide significant assistance to human decision-making abilities.

In this short paper, we have presented our first investigations into the feasibility of automatically extracting such a knowledge base of actions from the experiential microblogs available via social networking services such as Facebook and Twitter. As demonstrated in our initial results and despite the relatively limited data we are working with in these preliminary analyses (just a few weeks of data, instead of the many years of social data potentially available for analysis), we are able to extract relevant consequences of mundane actions such as becoming a pet owner. While many challenges remain to be addressed, these initial results gives us some hope that we will be able extract broad, useful models of the consequences of actions from such data.

References

Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. 2007. Open information extraction for the web. In IJCAI (Vol. 7, pp. 2670-2676).

Bordes, Antoine, et al. 2012. Joint learning of words and meaning representations for open-text semantic parsing. International Conference on Artificial Intelligence and Statistics.

Diaz, F., Gamon, M., Hofman, J., Kıcıman, E., Rothschild, D. 2014. Online and social media data as a flawed continuous panel survey. Working paper. http://research.microsoft.com/flawedsurvey/

Fikes, R. and Nilsson, N. 1971. STRIPS: a new approach to the application of theorem proving to problem solving. Artificial Intelligence, 2:189-208.

Gunawardana, A., Meek, C., and Xu, P. 2011. A model for temporal dependencies in event streams. In NIPS.

Jin, Yuzhe, Emre Kıcıman, Kuansan Wang, and Ricky Loynd. 2014. Entity linking at the tail: sparse signals, unknown entities, and phrase models. In Proceedings of WSDM. ACM.

Kıcıman, Emre. 2012. OMG, I have to tweet that! A study of factors that influence tweet rates. In Proceedings of ICWSM.

McDermott, D., Ghallab, M., Howe, A., Knoblock, C., Ram, A., Veloso, M., Weld, D., and Wilkins, D. (1998). PDDL-the planning domain definition language. Technical Report CVC TR98003/DCS TR-1165.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of NIPS.

Mislove, A., Lehmann, S., Ahn, Y. Y., Onnela, J. P., & Rosenquist, J. N. 2011. Understanding the Demographics of Twitter Users. In Proceedings of ICWSM.

Pearl, J. 2000. Causality: models, reasoning and inference (Vol. 29). Cambridge: MIT press.

Pennington, J., Socher, R., Manning, C. D. 2014. Glove: global vectors for word representation. To appear in EMNLP.

Robins, J. M., & Wasserman, L. (1999). On the impossibility of inferring causation from association without background knowledge. Computation, causation, and discovery, 305-321.

Rosenbaum, P. R., and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Spirtes, P., and Glymour, C. 1991. An algorithm for fast recovery of sparse causal graphs. Social Science Computer Review, 9(1), 62-72.