

CSCI 5070 Project Proposal

Intelligent Cross Domain SNS Router

Group No.?

HU Pili, hupili@ie.cuhk.edu.hk

Motivation

There are so many Social Network Services nowadays. Instead of being trapped in some platforms, people had better focus on information itself. Information shows its value in the process of spreading. Information spreading in a single domain is a well-developed research topic. However, cross-domain behavior is a seldom touched area. Some motivating examples:

- You subscribe to a blog RSS and read a gossip on it. The gossip is interesting so you forward it to Renren. Renren is an OSN, which is typical SNS in our traditional understanding. Blog, from the subscriber's point of view, is also a kind of SNS, but is read-only.
- You receive Google Scholar updates on a certain topic by email. Once there are some eye-catching new papers, you forward to your colleagues by email. Email is also SNS, where the link is formed dynamically and message is targeted instead of being posted on a wall.
- Now iPhone 5 is a big event and you want to get latest news about it. Traditionally, you follow some guy on Sina Weibo or Twitter. This guy always keeps an eye on big tech news so you are informed of the latest news. Here's an alternative: <http://www.baidu.com/s?ie=utf-8&wd=iphone+5>. Instead of following "inbridge" at "twitter", you can follow "iphone+5" at "baidu". You can think of "baidu" as an SNS, where there are many people who usually give you the updated messages on a certain topic. Of course, after getting the news, you are very likely to forward it to Sina Weibo. (Congratulations! You stand at the frontier now!!)

Note the roles people play in the above examples. We know Internet routers connect different IP subnet. Analogously, those people are routers across different SNSs. The routing job is done manually at preset. How about making it automated? Can we learn one's forwarding behaviors and ask a robot to do it?

Background

We observe many such cross-domain forwarding behaviors everyday! However, we lack abstraction of those platforms and differently formatted messages. Users don't have a unified way to perform easy forwarding. Researchers don't have cross-domain data for study. We take an initial step towards this direction. The open-source SNSAPI project [1] develops a middle ware to unify different SNS (OSN, RSS, email, etc). Different SNSs are abstracted in the form of channel and they support primitives like **auth()**, **read()**, **write()**. The SNSAPI project is at its infancy. We'll spend some time contributing to SNSAPI but do not count it in the SNSRouter[2] project. The SNSRouter project lies on SNSAPI (App layer in SNSAPI terminology). We'll focus on prototyping an easy way to route cross-domain messages. On this prototype system, we'll be able to collect trace data. Using this cross-domain routing data, we can train

algorithms to intelligently make forwarding decisions. We plan to leverage **topology** (e.g. come from whom), **content** (e.g. the topic) and **context** (e.g. time of day / location of user) features. It's worth to note that SNSAPI framework enables very flexible usage. One can configure an archive channel (e.g. email) and write all messages he like there for future retrieval. In this case, the routing (originally can be multiple-input-multiple output) degrades to a multiple-in-single-out personal recommender system. To make things tractable, we'll start with the recommender algorithm using all three kinds of features.

Input and Output

- Input: nothing.
- Output:
 - system prototype (web UI to enable easy cross-domain routing for human)
 - algorithm (topology/content/context based filtering)
 - trace data of cross-domain forwarding (expected to be longer than one-month)
 - theoretical paper on the routing algorithm (depending on time)

Objective

- Build up and strengthen SNSAPI community.
- Develop web UI of SNS Router.
- Collect cross-domain (manual) routing data.
- Implement baseline algorithms (e.g. looking at from whom, term frequency, etc)
- Try to build advanced algorithms (e.g. forwarding sequence analysis, using wordnet to mine related terms/concepts/topics, context analysis, etc)
- Depending on the trials of advanced algorithms, possibly tackle with some research issues. e.g. mining the concepts behind similar messages (above word level).

(This is a maximal listing; subject to change at middle review)

Timeline

- Week4-Week6. System develop (email channel on SNSAPI; web UI of sns-router; etc).
- Week7-Week11. Data collection. At least my own trace; hope to attract more users.
- Week4-Week11. Paper reading.
- Week7. Forwarding algorithm baseline.
- Week8-Week11. Advanced algorithm trial and error.
- Week12. Wrap up; final training and testing; report.

Reference

- [1] <https://snsapi.ie.cuhk.edu.hk/>
- [2] <https://github.com/hupili/sns-router>