

## Knowledge Development and Data Mining Privacy

First, I would like to say that the Wahlstrom and Roddick (2001) article is very well written. I found that as I made notes with questions and thoughts in the first column, they were promptly answered and addressed by the second column.

My first thought came on the heels of the statement, "individual society member develop independent and unique perceptions of their own privacy" (p.2). This caused me to question: how can we determine what privacy is and if it is protected if the concept differs from person-to-person? Wahlstrom and Roddick address this stating "as privacy is contextual and a product of individual perception, an infallible and universal solution to this dilemma is not possible" (p.2). Great, question answered, right? Well actually, while my inquiry was addressed, the solution basically does not exist! I have trouble with this as someone who always believes there is at least some solution to a problem. My brain headed straight to app permissions. If we approach data collection like an app, (in theory we could even create an app for this purpose!), we would require the user accept or reject permissions. For instance, when in the app store, if a game, product, or service wants my location data, and I see no reason why they would need it, I either deny the permission or, if they claim it is necessary, I simply don't download the app. If we use a similar standard or process for data collection, individuals can then determine what their level of acceptable privacy is. Wahlstrom and Roddick propose a similar solution stating, "One proposed compromise with sufficient contextual sensitivity is the empowerment of individuals to dictate the amount and type of personal data they consider appropriate for an organization to use for KD" (p.2). So, it seems like this is a viable solution.

Next, after reading, "Privacy can be violated when information concerning an individual is obtained, used or disseminated, especially if this occurs without their knowledge or consent" (p.2) a huge red flag was raised! This seems like something we could universally agree on as bad practice. And it immediately connected to a recent experience I had. With my tutoring company finally having a website, I wanted to ensure that it could be easily located on the web. After conducting a simple Google search of my name, I found an immense amount of information that I had no clue was connected to me! One of the top results was for a local restaurant my mom visited and had posted on Facebook. I replied to the post asking why she didn't bring me any key lime pie. I had no idea that this post would be so public and especially not a top find for someone Googling me. (Now the world knows my favorite dessert!) Although all my posts are set to private and only friends can see, this is apparently not the case for comments. But this was the least of the problems. The very top link was whitepages.com, and it included my approximate age as well as my physical address. This concerned me. What if a crazy key lime pie hater was offended by my post? They could literally come to my home where my children reside. That is scary. Thankfully, it was easy to remove, just sent a request to the site. But my husband informed me he had already done this a few years ago, and they had reposted the information! As Wahlstrom and Roddick write, "Commonly an individual must adopt a proactive and assertive attitude in order to maintain their privacy" (p.2). This is a problem. While I believe that data is a powerful tool and using it, particularly in Knowledge Development, to find new patterns and trends is amazing, we should not have to take on this role. Especially as I had already asked the site not to publish this information, I suspect this request has a time limit allowing for the ability

to repost after a certain number of months/years. But overall, in the right hands I believe KD can be both a powerful tool and respect the privacy of others.

This brings me to my next point. The authors state, “KD has the potential to violate these [OECD] principles” (p.2). I immediately thought, why not make data anonymous? For instance, when working on the Excel assignment, I took my grade data from a course I assisted and promptly removed all names. Then, with no connection to a particular person, they have no issues with privacy. Similarly, FERPA laws prohibits how we use and disseminate data in a university setting, allowing for clear guidelines to protect student privacy. However, Wahlstrom and Roddick go on to state, “One solution is the anonymization of personal data” (p.2). (They stole my idea!) But they promptly flip and determine that “an abolitionist policy is, we contend inappropriate” (p.2). I don’t feel like the authors have established this point well. While I understand that should any trends arise from the data, not knowing who the data connects to does not allow for solutions to be applied to them; I think that something along the lines of keeping the data anonymous but having some tag on the information would mitigate this issue. Thus, if a solution or pattern is found, the individuals can be notified. The data that connects an anonymous number to an actual individual should be held under lock and key, and I would argue it either not be digital at all or only on an airgap computer to be kept safe from any hacking or data breach. Hence, I am not sure that this idea is completely impossible.

Overall, there is clearly a lot of thought and attention that needs to be applied to KD and DM in consideration of privacy. As a strong proponent of both, I do believe that a fair solution can be obtained.

## References

Wahlstrom, K., & Roddick, J. (2001). On the impact of knowledge discovery and data mining. Presented at 2nd Australian Institute of Computer Ethics Conference (AICE2000), Canberra, Australia.